Sub-Penny and Queue-Jumping

Sabrina Buti
Rotman School of Management, University of Toronto

Francesco Consonni
Bocconi University

Barbara Rindi
Bocconi University and IGIER

Yuanji Wen
Deakin University

Ingrid M. Werner
Fisher College of Business, Ohio State University

First Version: September, 2013
Current Version: June, 2014

We thank Carole Comerton-Forde and Ariadna Dimitrescu for very useful comments, as well as participants at the NBER Market Microstructure Meeting, December 2013 and EFMA 2014 conference. Financial support from GRI is gratefully acknowledged. The theoretical model presented in Section 3 previously circulated as an extension included in the paper titled "Tick Size Regulation and Sub-Penny Trading" (Buti, Rindi, Wen and Werner, 2013). This was an initial theoretical manuscript on both the regulation of the minimum price increment and sub-penny trading. We have now separated the two topics into two different contributions, one on tick size titled "Tick Size: Theory and Evidence" (Buti, Consonni, Rindi, Wen and Werner, 2014), and the paper presented here focused on sub-penny trading.

ABSTRACT

Sub-Penny Trading (SPT) is a form of dark trading that allows traders to undercut displayed liquidity. We distinguish between SPT that is queue jumping (QJ) and mid-crossing (MID) and find that QJ is higher for NASDAQ than NYSE stocks. Consistently with a model in which a Limit Order Book (LOB) competes with a Sub-Penny Venue (SPV), QJ is positively related to depth and negatively related to stock price. We also find that QJ is associated with improved lit market quality, especially for large capitalization stocks. Sub-penny quotes are allowed for stocks priced below $1.00, and we use this fact to show that the spread improves but depth deteriorates as the price of a stock crosses from above to below ($1.00). Finally we show that our proxy of High Frequency Trading (HFT) is negatively related to QJ.

1 Introduction

The microstructure of financial markets during the last decade has been characterized by a growing importance of inter-market competition between lit and dark venues, and a significant increase in High Frequency Trading (HFT). These two elements have fostered the growth of a trading practice called sub-penny trading (SPT) which now involves approximately 10% of the U.S. consolidated equity volume and which is the objective of the investigation in this paper.

SPT takes place when traders undercut orders at the top of public limit order books (PLBs) by a fraction of the tick size, which is the minimum price increment allowed by regulators on financial trading platforms. SPT may either take place "in house" when broker-dealers match orders internally, against other clients' orders or against their own inventory.[1] Or it may take place on dark pools, trading platforms that do not publicly display price quotes.[2] Broker-dealers may use dark pools to internalize customers' orders at sub-penny increments, but they may also use dark pools to post limit orders at sub-penny prices. Also institutional traders with access to dark pools may use them to post orders that jump ahead of the orders sitting at the National Best Bid Offer (NBBO) on the lit market. This queue-jumping feature of SPT in dark pools is particularly important for traders pursuing algorithmic and HFT strategies. Fast trading includes smart order routers (SORs) which are programs largely used by both the buy and the sell side to search for the best quotes in regular exchanges as well as in dark pools. Therefore, the use of SORs guarantees that orders posted at sub-penny increments are actually executed if they offer the best price. This is why the rise of dark markets combined with the development of fast trading facilities has paved the way for the expansion of SPT.

Market participants can trade in sub-penny increments because of an exception established by Rule 612 of Regulation National Market System (Reg NMS). Rule 612 prohibits sub-penny quoting by banning traders from accepting, ranking or displaying orders or quotations in price increments smaller than a penny. At the same time, however,

[1] Broker-dealers' internalization accounts for a substantial share of the U.S. consolidated equity volume according to the Securities and Exchange Commission's Concept Release on Equity Market Structure (SEC, 2010), being equal to 17.5% in 2009. According to the SEC Concept Release, in 2009 around 10 exchanges, 5 electronic communication networks, 32 dark pools, and over 200 broker-dealers were active in the U.S..

[2] More precisely, SPT can be observed in specific dark markets called Bank-Brokers or Internalization Pools. According to Rosenblatt Securities Inc.'s estimates for January 2013, Bank-Brokers Pools represent 7.09% of U.S. consolidated equity volume, and approximately 50% of total dark volume.

Rule 612 does not prohibit SPT. Specifically, SPT is allowed provided that it does not result from executions of visible quotations in sub-penny increments. This can occur for two reasons. First, Reg NMS exempts Alternative Trading Systems (ATSs) from pre-trade transparency as long as they execute less than 5% of the average aggregate daily volume in a particular stock. This implies that a broker may operate an ATS with no pre-trade transparency (a dark pool) where invisible quotations can be posted in sub-penny increments. Second, Rule 612 allows broker-dealers to internally execute non-displayed orders (typically retail orders) provided that this is done in compliance with their duty of best execution and so that orders are filled at prices that are better than the NBBO.

Rule 612 was introduced to limit the negative effects of the decimalization that took place in the U.S. in 2001.[3] The rationale for the smaller tick size was to lower trading costs for investors. The unanticipated consequences were the possible effects on PLBs' depth and on brokers' profits. The U.S. Securities and Exchange Commission (SEC) recognized that very small tick sizes could have detrimental effects on the liquidity of PLBs. In particular, the regulator realized that if traders could undercut limit orders sitting on the book by an economically insignificant amount, it would potentially reduce the incentive for traders to post limit orders at the top of the PLB, and therefore could have a detrimental effect on inside depth. The SEC also realized that a smaller tick size and hence a lower inside spread would lead to lower profit opportunities for brokers. Therefore, the SEC introduced an exception to the Rule so that broker-dealers could internalize small orders and at the same time offer their clients best execution by trading in sub-penny price increments inside the NBBO. As will be clarified later in this paper, another aspect of Rule 612 is relevant for our analysis of SPT. Rule 612 established a two-tiered market by setting a minimum price increment of \$0.01 (1 penny) only for stocks priced over \$1, but a minimum price increment of \$0.0001 for stocks priced below \$1.

For all the reasons discussed above SPT is one of the main concerns of the SEC and in this paper we aim at studying SPT in NASDAQ and NYSE stocks with the objective to understand the frequency of SPT, the factors which induce traders to undertake SPT and the effects that SPT may have on market quality. We then tackle the regulatory issue of how to mitigate the possible negative effects of SPT.

[3] In the U.S. from 2001 the minimum price improvement was gradually reduced to 1 cent for stocks above \$1 and 0.01 cent for stocks below \$1.

Our empirical analysis is based on a sample of Trade and Quote (TAQ) data for U.S. equities that includes all trades executed in sub-penny increments during 42 trading days in 2010. These trades may come from broker-dealers' internalization or from dark pools' executions. We build a stratified sample of 90 NASDAQ and 90 NYSE listed stocks and use Fama-MacBeth regressions to investigate which factors are associated with more SPT, i.e., under which conditions broker-dealers trade more intensively at sub-penny increments. We consider spread, depth, stock price, share volume, volatility and order imbalance. We also study in which type of stocks, small or large cap, low or high priced, there is more SPT as a fraction of consolidated volume. We then use a simultaneous equations model that takes into account endogeneity issues to investigate the association between SPT and market quality, measured by relative quoted spread and inside depth. Finally, we build a proxy for HFT based on the ratio between quote updates on lit markets and total number of trades, and investigate a possible relation between HFT and SPT.

Before summarizing the main results of the paper, it is important to clarify how we classify SPT. We generally classify an execution as SPT when the price improvement associated to the execution price is smaller than \$0.01. However, note that there are two main reasons why we observe trades executed on ATSs at fractions of a penny in the data. First, traders may aim to undercut orders posted at the top of transparent PLBs. Second, the execution system of some dark pools follows a derivative pricing rule according to which all trades in the pool execute at the midpoint of the primary market's inside spread.[4] To differentiate between these two categories, we treat executions at half a cent price improvement as a separate type of SPT and name it Mid-Crossing (MID), as it may originate from dark pools with midpoint derivative execution systems.[5] We group the remaining SPT with a positive sub-penny price improvement different from half-cent into a category that we label Queue-Jumping (QJ). This way we adopt a parsimonious classification of QJ, and we are sure that we do not mix data which derive from substantially different trading strategies, and therefore could have different driving forces and different effects on the quality of the lit markets. However, by doing so we miss the executions originating from undercutting at exactly half a penny.[6] In

[4]To mention but a few, see ITG Posit, Liquidnet and BATS.

[5]While not included in our dataset, mid-quote executions can also occur within exchange operated dark pools.

[6]We also miss instances of QJ in stock at the whole penny increment, which may occur in high-priced and wide-spread stocks.

our study, we will focus on QJ that is the main concern of regulators, and use MID primarily to highlight the differences between executions in dark pools that rely on a derivative pricing rule with QJ.

To develop hypotheses about the factors affecting QJ and about the effects of QJ on market quality, we rely on a framework which models competition between a PLB and a Sub-Penny Venue (SPV). The model predicts that SPT takes place mainly for liquid low-priced stocks and that the effects of SPT on market quality are positive for liquid stocks (especially when the price is high), and negative for illiquid stocks (especially when the price is low).

Our empirical results are as follows. We find that for our sample QJ is on average 7.16% of the consolidated volume for NASDAQ stocks and 6.03% for NYSE stocks. Consistently with the model's predictions, QJ is positively related to the liquidity of the stock, increasing with depth and decreasing with inside spread. QJ is also negatively related to the price of the stock; it is somewhat positively related to volatility and, for the group of large capitalization stocks, to the relative order imbalance. The opposite holds for small capitalization stocks for which QJ decreases with relative order imbalance. Still in line with the model's predictions, we find that for large capitalization stocks QJ positively affects depth and inside quoted spread, measured both in cents and in basis points. By contrast, we do not find significant results for the group of small stocks. Our results show that MID executions increase with share volume and order imbalance, but have no significant effects on market quality, measured either by depth or inside quoted spread.

Having investigated the overall magnitude and effects of QJ and MID, we extend the empirical analysis to address the regulatory issue. So far, two opposite proposals have been put forward from regulators and exchange officials that relate to SPT. In 2010, the SEC in its Concept Release on Equity Market Structure stressed that the larger percentage spread that characterizes low-priced stocks may lead to greater internalization by Over-The-Counter (OTC) market makers or more trading volume in dark pools, and proposed a reduction in the tick size for lower priced stocks.

While major U.S. exchanges, e.g., NYSE, NASDAQ and BATS, in their comment letters to the SEC concept release responded positively to the suggestion of reducing the tick size, there was at the same time a widespread sentiment among practitioners that decimalization curtailed brokers' profits and therefore incentives to supply liquidity for less liquid stocks and for initial public offerings (IPOs). This led to the 2012 Jumpstart

Our Business Startups Act (Section 106(b) of the "JOBS Act") which instructed the SEC to investigate the possible effects of raising the minimum price increment, i.e., the tick size, for stocks of high growth companies and authorized the SEC to set a pilot program to raise the tick size for small and medium capitalization stocks to $0.10. Our empirical work aims at assessing whether an increase or a decrease in the tick size is the optimal regulatory response to the widespread use of SPT.

The SEC concept release and the ensuing comment letters from market players stress that it is the tick-to-price ratio that matters for policy. It is the size of the minimum price increment relative to the price of the stock that is the relevant policy instrument to investigate when dealing with QJ.[7]

Our results from the Fama-MacBeth regressions allow us to draw a first conclusion on the effect of a change in the relative tick size on QJ. We find a strong negative relationship between QJ and the price of the stock. Because the sample of NASDAQ and NYSE stocks described above includes only stocks priced above $1, the tick size is constant for that sample and therefore when the price increases the tick-to-price ratio decreases. Consequently, we find a positive empirical relationship between QJ and the tick-to-price ratio. Hence a reduction in the tick-to-price ratio is associated with a reduction in QJ. If the aim of regulators is to reduce QJ, then the right policy action would therefore appear to be to decrease rather than to increase the tick size.

However, in related work, Buti, Consonni, Rindi, Wen and Werner (BCRWW, 2014) find that a reduction in relative tick size is associated with a deterioration of market quality (wider spread and less depth). This suggests that while a lower tick size reduces the incentives to QJ, the drawback is that a lower relative tick size encourages traders to switch from limit orders to market orders. In equilibrium, this decrease in liquidity provision results in wider spreads and less depth.

An important caveat is that BCRWW (2014) show that a change in the relative tick size may not affect market quality in the same way as a change in the absolute tick size. Intuitively, a reduction in the absolute value of the tick size increases the

[7]The SEC (2010) concept release (page 72) reads: *There may be greater incentives for broker-dealer internalization in low-priced stocks than in higher priced stocks. In low-priced stocks, the minimum one cent per share pricing increment of Rule 612 of Regulation NMS is much larger on a percentage basis than it is in higher-priced stocks. For example, a one cent spread in a $20 stock is 5 basis points, while a one cent spread in a $2 stock is 50 basis points – 10 times as wide on a percentage basis. Does the larger percentage spread in low-price stocks lead to greater internalization by OTC market makers or more trading volume in dark pools? If so, why? Should the Commission consider reducing the minimum pricing increment in Rule 612 for lower priced stocks?*

number of price levels that investors can use to post liquidity on the limit order book, and therefore, when there is at least one order already at the top of the book, it allows traders to compete for the provision of liquidity by undecutting the queue thus reducing the spread and distributing the liquidity supply over more price levels. The empirical evidence from European liquid stocks reported in BCRWW (2014) confirms the model's prediction that a reduction in the absolute tick size reduces spread and depth at the best bid and offer (BBO).

In order to investigate the effect of a reduction in the absolute value of the tick size for U.S. stocks, we exploit the fact that U.S. stocks priced below \$1 trade at a smaller tick size (\$0.0001) than the stocks priced above \$1 (\$0.01). By comparing market quality for stocks crossing the \$1 threshold from above, we can study the effects of a reduction in absolute tick size. We build a new sample of NASDAQ stocks whose price is between \$0.8 and \$1.2 at September 30, 2010, and which cross, at least once, the \$1 threshold during the sample period. We then study the effects of a reduction in the tick size on spread and depth around the \$1 threshold by means of a discontinuity regression analysis. We find that when the stock price falls below \$1, spread decreases and depth deteriorates. We therefore conclude that a reduction of the absolute tick size may have mixed effects on market quality and potentially magnify some of the negative effects of QJ. Hence, a reduction of the tick size might not lead to an improvement of market quality.

There are two more limitations of our analysis that we would like to mention. First, Rule 612 does not restrict SPT on lit markets for stocks priced below \$1, and therefore a natural question is how a reduction in the absolute value of the tick size affects QJ, this time at \$0.00001 increments or smaller. Unfortunately the existing data collected by the Financial Industry Regulatory Authority (FINRA) are recorded only up to four decimal digits thus not allowing researchers to study trades at such tiny increments.[8] Second, U.S. stocks priced around \$1 could potentially differ from stocks with higher prices in other ways that are important for QJ.

Our empirical analysis shows that a reduction of both the absolute and the relative tick size reduces QJ. However, the side effects on market quality are that depth at the BBO worsens and spread only improves following a reduction in the absolute tick size. Considering that we observe a positive rather than a negative effect of QJ on

[8] Informal conversations with U.S. traders informed us that banks are not even aware of the possibility to trade at more than 4 decimals, suggesting that perhaps QJ below \$1 does not even exist.

market quality for liquid stocks and that QJ concentrates in liquid stocks, we argue that regulators' concern about QJ may be exaggerated.

Finally, we find that our proxy of HFT activity on lit markets is inversely related to QJ, suggesting a possible substitution effect between transparent and opaque venues.

The paper is organized as follows. In Section 2 we discuss the related literature and in Section 3 we lay out the model and the resulting testable hypotheses. In Section 4 we describe the dataset; in Section 5 we investigate the factors that affect SPT; in Section 6 we show the effects of SPT on market quality and in Section 7 we present the results from our discontinuity regression analysis. We conclude in Section 8.

2 Related Literature

As SPT takes place on dark venues, our paper is related to the empirical literature on dark pools. Ready (2013) investigates the determinants of dark trading by considering monthly volume by stock for the period June 2005 to September 2007 in two dark pools, Liquidnet and ITG POSIT, that executed approximately 1% of total market consolidated volume. He finds that dark pools execute most of their volume in stocks with low spread and high share volume. Considering that both Liquidnet and ITG POSIT execute at the mid-quote of the primary market spread, Ready's results are comparable (and consistent) with our findings for mid-crossing which show that the effect of depth on MID is positive and that of spread is negative.

Our results are also consistent with Buti, Rindi and Werner (BRW, 2011) who examine a unique dataset for the calendar year 2009 on dark pool activity for a large cross section of U.S. securities. BRW find that liquid stocks are those characterized by more intense dark pool activity. They also find that dark pool volumes increase for stocks with narrow quoted spreads and high inside bid depths, suggesting that a higher degree of competition in the PLB enhances dark pool activity. BRW also investigate the effect of dark trading on market quality and show that increased dark pool activity improves spreads, depth, and short-term volatility. Degryse, de Jong and van Kervel (2014) consider a sample of 52 Dutch stocks and analyze both internalized trades and trades sent to dark pools. They find that when these two sources of dark liquidity are combined, the overall effect on global liquidity is detrimental. Hatheway, Kwan and Zheng (HKZ, 2013) too study the effects of dark trading on market quality, even though they only focus on effective spread. HKZ look at a sample of NASDAQ and NYSE stocks during the period January-March 2011 and find a positive relation between dark pool market share and effective spread. They explain this result with the effects of SPT that should drive uninformed traders away from the lit markets into dark pools thus increasing adverse selection costs on lit markets. We cannot directly compare HKZ' results with ours because they consider dark pool activity as a whole and do not distinguish between dark trading and SPT when looking at the effects on market quality. However, by controlling for endogeneity we find that SPT does not harm spread and depth at the top of the lit market which does not seem consistent with the conjectured cream skimming effect. Our paper is also related to Kwan, Masulis and McInish (2013) who study the effects of competition for order flow on the fragmentation of U.S. equity markets and point out the relevance of SPT, and in particular of QJ, in the distribution of market shares across lit and dark venues. More precisely, they consider NASDAQ and NYSE stocks and show that when the price of a stock crosses the $1 threshold, volume in dark venues increases while volume in traditional exchanges decreases.

Finally, Comerton-Forde and Putnins (2013) analyze the effect of dark trading on price discovery by considering the 500 largest stocks listed at the Australian Securities Exchange (ASX). They observe that informational efficiency and price discovery are negatively related to the share of volume executed on dark venues. However, they do not find any evidence that block trades harm price discovery. Still on price discovery, Nimalendran and Ray (2014) study detailed data from one dark pool and find evidence suggesting that price discovery may take place in the dark venue, particularly for less liquid stocks.

On the theory side, our paper is related to the models that study how limit order books work.[9] As SPT crucially depends on the tick size and the state of the book -measured by depth and spread- to deliver testable empirical hypotheses the model must have discrete prices. Moreover, the model's equilibrium prices cannot be derived under the assumption of steady state which would inevitably imply a constant state of the PLB.[10] To satisfy these requirements and explicitly analyze QJ, we therefore extend Parlour (1998) and draw our main empirical hypotheses from the resulting framework.

[9]To mention but a few, see Buti and Rindi (2013), Goettler, Parlour and Rajan (2005), and Parlour (1998).

[10]See, e.g., Foucault (1999), Foucault, Kadan and Kandel (2005), and Rosu (2009).

3 Model and Empirical Hypotheses

We draw our main empirical hypotheses from a theoretical framework that extends Parlour (1998) to model competition between a PLB and a SPV. In what follows we briefly describe the design of both the PLB and the SPV. We first consider the PLB (our benchmark model) in which traders can only post orders to a standard limit order book. We then compare this model with the extended framework in which traders are allowed either to trade on the PLB, or to undercut limit orders posted at the top of the PLB by trading in sub-penny on the SPV. We will draw our empirical hypotheses by investigating how the state of the PLB affects traders' decisions to opt for the SPV and by studying the effects that the SPV trading option has on the quality of the PLB.

3.1 Public Limit Order Book

The PLB works like a double auction market that enforces time and price priority and runs over 3 periods of time, $t = \{t_1, t_2, t_3\}$. At each period t a risk neutral trader arrives and can submit an order of size 1 to buy or to sell an asset with fundamental value equal to v. Each trader comes to the market with a private value, β, drawn from the uniform distribution with support $U(0, 2)$, which indicates the trader's degree of aggressiveness. A non aggressive trader has a β close to 1 while an eager one has values of β close to either 0 or 2.

Upon arrival in period t, the trader observes the state of the book, which is characterized by the number of shares available at each level of the price grid (Table 1), and includes the following possible prices: $A_1 = v + \frac{\tau}{2}, A_2 = v + \frac{3}{2}\tau$ and $B_1 = v - \frac{\tau}{2}$, $B_2 = v - \frac{3}{2}\tau$. The difference between two adjacent prices -the minimum price increment- is the tick size which we set equal to a constant $\tau_{PLB} = \tau$ and which also corresponds to the minimum inside spread. We define as $S_t = [Q_t^{A_1}, Q_t^{A_2}, Q_t^{B_1}, Q_t^{B_2}]$ the state of the book that specifies the number of shares Q_t available at each price level. A trading crowd provides liquidity at the highest levels of PLB, and traders are allowed to submit limit orders queuing in front of it. Upon arrival each trader can choose between limit orders $(+1)$ and market orders (-1), that cannot be modified thereafter.

[Insert Table 1 about here]

The trader's strategy space at time t is therefore $H_t = \left\{-1^{A_{k'}}, +1^{A_k}, 0, +1^{B_k}, -1^{B_{k'}}\right\}$, where $+1^{A_k}(-1^{A_{k'}})$ indicates a limit (market) order to sell submitted at the k^{th} level of the book, with $k = 1, 2$, and k' the best available ask or bid price; analogous strategies are available to the trader on the buy side and 0 indicates that the trader decided not to trade.

The state of the PLB evolves according to $S_t = S_{t-1} + h_t$, where S_{t-1} and S_t are respectively the states of the book before and after the trader's order submission and h_t is the change in the PLB induced by the trader's strategy H_t. We assume that when a trader arrives at t_1, he observes the initial state of the book denoted by S_0, which is exogenous. To proxy stocks with different degrees of liquidity, we consider two different starting books: an empty book for illiquid stocks, and a book with one unit on the first level of the price grid for liquid stocks. Without loss of generality, in our numerical simulations we assume that the tick size (τ) is equal to 0.1 and that the price of the asset can take four values, i.e., $v = \{1, 5, 10, 50\}$.[11] By considering different prices of the asset, we can study the effect of a change in the tick-to-price ratio (or relative tick size), which is a crucial determinant of the order submission decisions of real traders.

Figure I in the Online Appendix shows the extensive form of the game and indicates a seller's possible order submission strategies. The trader arriving at the market at each trading round maximizes his payoff and faces a trade-off. If he chooses a market order, which guarantees immediate execution, he minimizes non-execution costs; whereas if he chooses a limit order, which enables him to obtain a better price, he reduces the price opportunity costs.[12]

To choose an optimal order submission strategy, the trader observes the state of the PLB and computes the execution probability of his limit orders. Because he knows that at the last period of the trading game, t_3, no trader will submit limit orders but rather only market orders, he can compute by backward induction the execution probabilities of his limit orders both at t_2 and at the previous periods. To economize space we report in the Online Appendix traders' equilibrium strategies and a discussion on how to solve the model.

From the equilibrium order submission strategies of the model we compute indicators of market quality like volume, spread and depth. We then use these metrics to investigate how the state of the PLB changes when traders are allowed to undertake QJ.

[11] For brevity we only present results for v=1 and v=10.

[12] Table I in the Online Appendix reports the trader's payoffs from the different possible strategies.

3.2 Sub-Penny Venue

In order to include QJ among the trading strategies of market participants, we extend the benchmark PLB model to include competition from a SPV. All else equal, the SPV has a smaller tick size, $\tau_{SPV} = \frac{\tau}{3}$, and a finer price grid with five levels on both the ask and the bid side, a_l and b_l ($l = 1, .., 5$), which allows market participants to trade at sub-penny increments (Table 1). Furthermore, in order to approximate the nature of real SPVs, we assume that the SPV does not have a trading crowd sitting at a_5 and at b_5.

To add further realism to this new framework, which we indicate as PLB&SPV, we assume that the market is now populated by two group of traders, namely regular traders (RTs) and broker-dealers (BDs), who arrive in each period with probabilty $1 - \alpha$ and α respectively, and that only BDs are allowed to supply liquidity to the SPV.[13] Yet, while only BDs can post limit orders to the SPV, all traders can take advantage of the liquidity offered by both trading platforms and demand liquidity on both the PLB and the SPV. This assumption is consistent with a fast market in which a SOR technology allows all investors to search the best quotes on the consolidated limit order book (PLB&SPV). So, as long as the standing orders in the SPV offer better prices than the PLB, all traders, RTs included, can execute against them. Orders posted to the SPV will offer a better price than the PLB when BDs engage in QJ and therefore compete on price against the limit orders posted at the top of the PLB.

To sum up, we consider two protocols: the benchmark (PLB), where only one trading platform is available to all traders, and the PLB&SPV framework, where a SPV competes with the PLB.[14] We further assume that at t_1 the SPV opens either empty or with one unit on the second level of the book, with equal probability.[15]

[13] In real markets retail traders cannot access SPVs. Technically, we assume that when the BDs' payoffs from trading across the two markets are the same, they submit orders to both markets with equal probability.

[14] The latter case further differentiates into a transparent and an opaque setting, where the lack of transparency refers to the visibility of the SPV by RTs. To economize space we discuss the opaque case in the Online Appendix.

[15] To compare markets characterized by different price grids we need to start from the same initial state of the book. This implies that in the SPV depth at the price levels not in common with the PLB must be zero.

3.3 Empirical Hypotheses

We now use this model to draw our main empirical hypotheses, first on the factors driving QJ and, second, on the effects of QJ on market quality. More precisely, we draw from this model our hypotheses number 1, 2 and 4. We derive hypothesis 3 from previous empirical results and hypothesis 5 from BCRWW (2014).

By comparing traders' equilibrium strategies for liquid and illiquid stocks (Figure 1), we observe that BDs engage more intensively in QJ when competition for the provision of liquidity is strong on the PLB, i.e., when the PLB is deep and the spread is small. With greater inside depth the queue at the BBO is longer, so that the incentive for BDs to gain price priority by stepping ahead of limit orders posted at the top of the PLB and submitting limit orders on the SPV is greater. With smaller spread there is less room to post limit orders within the inside spread on the PLB and therefore there is more incentive for traders to undercut existing liquidity via QJ on dark venues.[16]

[Insert Figure 1 about here]

This leads to our first hypothesis on the factors driving QJ.

Hypothesis #1 QJ increases when the book is deep and when the inside spread is narrow.

Our results also show that the effect of the PLB liquidity on the propensity of traders to undercut via QJ is inversely related to the price of the stock. When the stock price is low, the price-to-tick ratio (or relative tick size) is high, and therefore traders' propensity to supply rather than demand liquidity (to use limit rather than market orders) is high. It follows that competition for the provision of liquidity is higher for low-priced stocks, and so is the incentive to undertake QJ in dark markets. Moreover, our model's results on the order submission strategies of traders show that a smaller tick size has analogous effects on the incentive to undertake QJ (Figure 2).

[Insert Figure 2 about here]

This leads to our second hypothesis.

[16]For example, when the inside spread is equal to the tick size (liquid stocks), traders cannot gain price priority on lit markets. This effect is stronger when the proportion α of BDs increases from 10% to 20% and the sub-penny activity builds up in the SPV (Figure III in the Online Appendix).

Hypothesis #2 QJ is higher for stocks with low-price (high relative tick size), and high tick size.[17]

QJ can take place in dark venues that allow executions in sub-penny increments, and trading in the dark can be affected by volatility and order imbalance. BRW (2011) show that for a given stock dark pool activity is higher on days with low intraday volatility and low order imbalances relative to share volume. In general traders approaching dark markets are worried about the uncertainty of executions and this uncertainty may increase with volatility, thus suggesting a negative relation between dark trading and both volatility and order imbalance. However, when trading in the dark entails undercutting existing liquidity on the lit market, an increase in the order book imbalance might signal a long competitive queue on one side of the market and have a positive rather than a negative effect on QJ.

Hypothesis #3 QJ increases when volatility decreases, and may be positively or negatively related to order imbalance.

We now move to the hypotheses on the effects of QJ on the quality of regular exchanges. When traders post limit orders on the SPV to undercut existing liquidity on the PLB, liquidity provision and hence depth and spread on the PLB may deteriorate. However, liquidity demand on the PLB may also decrease. The reason is that, due to the existence of SORs, the better prices available in the SPV intercept market orders away from the PLB. The reduced liquidity demand preserves depth and spread on the PLB. Therefore, the effects of QJ on the quality of regular exchanges depend on the relative proportion of limit and market orders that migrate from lit to dark venues. If the reduction in liquidity supply prevails, market quality -measured by spread and depth- on lit markets decreases; if instead it is the reduction in liquidity demand that prevails, liquidity improves as the drop in market orders preserves liquidity on regular exchanges. The net effect of the reduced supply and demand of liquidity on the quality of the PLB depends on the characteristics of the stock considered; specifically it depends on the liquidity as well as on the price of the stock. It also depends on the absolute value of the tick size.

The model shows that for liquid stocks the main effect of sub-penny trading is to

[17] While we can test the effect of a change in price and therefore in relative tick size, we cannot directly test the effect on QJ of a change in tick size. Because moving from above to below $1 the tick size moves from 0.01 to 0.0001, we could test the effect on QJ for stocks crossing the threshold. However, data on QJ below $1 ($0.00001 and smaller) is not recorded by the Security Information Provider (SIP).

foster price competition. When the SPV platform is introduced, BDs submit limit orders to the SPV at $a_1(b_1)$ to undercut the existing depth at $A_1(B_1)$ (Figure 1), thus intercepting incoming market orders away from the PLB. Therefore, the effect of BDs' undercutting is that both limit and market orders move from the PLB to the SPV thus reducing both liquidity provision and volume on the public venue (Figure 3). The result is that depth increases and spread narrows. The reason for why market quality improves in the PLB despite the reduction of the liquidity provision is that when the book is liquid, traders use more market than limit orders and consequently the positive effect of the reduced liquidity demand - which helps preserve the low spread and high depth - prevails. Moreover, when the price of the stock is higher, the relative tick size is smaller and traders use more market than limit orders; hence, when the SPV is introduced the effect on liquidity demand is magnified and so is the effect on market quality. The result is that volume decreases even more on the PLB and market quality improves.[18] This positive effect is reinforced by the fact that the change in limit orders following the introduction of a SPV in a market for a high-priced stock becomes almost irrelevant, and hence, compared to the $v = 1$ case, the reduction in liquidity supply becomes tiny. The same intuition explains the results shown in Figure 3 for the case with a smaller τ, i.e., a smaller tick size for both the PLB and the SPV. A smaller tick size reduces traders' incentive to post limit orders thus making the change in limit orders following the introduction of the SVP smaller.

[Insert Figure 3 about here]

We now turn to the framework with an empty opening book (Figure 4). Even for illiquid stocks when the SPV is introduced both limit and market orders decrease on the PLB. However, contrary to the previous case, depth and spread worsen. In this scenario, the effect of the reduced liquidity supply outweighs the reduction in market orders resulting from their interception by the SPV. The reason is twofold. First, when RTs perceive the potential competition from BDs, they react by supplying less liquidity to the PLB. This is due to the fact that when the book is illiquid they are afraid of being undercut not only on the SPV but also on the PLB. Hence, the reduction of the liquidity supply in the PLB is stronger than for liquid stocks. Second, while traders

[18]Our results also illustrate whether the effects of the competition from a SPV depend on the degree of transparency of the SPV. According to our model, when the stock is liquid, a decrease in the transparency of the SPV has only a moderate effect on traders' order submission strategies. This moderate effect has a subtle explanation that we discuss in the Online Appendix.

generally use more market than limit orders for liquid stocks, when the stock is illiquid traders prefer limit orders which therefore play a dominant role. These effects become much weaker as the stock price increases and the relative tick size decreases so that traders switch from limit to market orders. This second effect attenuates the negative reduction in the liquidity supply, and at the same time it makes the reduction of the liquidity demand more relevant: both effects have a positive impact on the quality of the PLB thus making the overall negative effect on market quality much smaller.[19] These results are summarized in the following hypothesis.

[Insert Figure 4 about here]

Hypothesis #4 QJ improves (worsens) depth and spread when the stock is liquid (illiquid), and the effect is stronger (weaker) for high-priced (low-priced) stocks, i.e., when the tick-to-price ratio is small (large).

The policy debate on SPT aims at identifying the right policy instrument that regulators can use to influence QJ. Our model shows that when the tick size is smaller in a PLB that competes with a SPV, QJ decreases as broker-dealers have lower incentives and profits from undercutting in the dark.

At the same time, BCRWW (2014) show that depth deteriorates due to the reduced competition in liquidity provision while spread improves due to the pressure from HFT following the movement of the price grid generated by the variation of the tick size. This leads to our fifth empirical hypothesis.

Hypothesis #5 A reduction in the tick size has mixed effects on market quality with depth and spread both decreasing.

It is less straightforward to present hypotheses for MID because when MID is reported OTC, it could be either SPT or trading in dark pools which execute at the midpoint of the spread. Therefore, we focus on the hypotheses for QJ.

[19]When instead the SPV is opaque, these effects become stronger because the uncertainty on the SPV depth and on the actual level of competition makes RTs even more reluctant to post limit orders to the PLB. Hence we can conclude that when the stock is illiquid, a reduction in the transparency of the SPV makes the negative effects on the quality of the PLB even more problematic, especially for low-priced stocks.

4 Data Description

4.1 Data and Sample

We construct a sample of stocks stratified by price and market capitalization for both NASDAQ and NYSE. As of December 31, 2009 we identify all common stocks in CRSP which are NYSE or NASDAQ listed. Then we divide all NYSE listed common stocks (i.e., share code 10 or 11), into terciles by market capitalization and price and form nine mutually exclusive groups with the same dimension. From each group we randomly draw 10 stocks for a total of 90 NYSE stocks. We repeat the procedure for NASDAQ stocks using the NYSE breakpoints. This way we create groups of stocks which are comparable across exchanges by market capitalization and price. Our final sample includes the 180 stocks reported in the Online Appendix and spans from October 1 to November 30, 2010, for a total of 42 trading days.

4.2 Descriptive Statistics and Definitions

The data we use come from the following sources. Data on number of shares outstanding, market capitalization, closing prices, listing exchanges and share codes are from CRSP. Data for S&P 500 Stock Price Index (SP500) and CBOE S&P 500 Volatility Index (VIX) used to capture market-wide activity are from Federal Reserve of Economic Data (FRED). Our main data source is the TAQ database which we describe in more detail below.

TAQ contains intraday transactions data (trades and quotes) for all securities listed on the NYSE and American Stock Exchange (AMEX), as well as NASDAQ. The TAQ database includes a flag indicating the exchange where the trade was executed or the quote displayed. Each exchange is identified by a symbol/capital letter (except for NASDAQ which has two equivalent symbols, T and Q). The letter “D” is used for the Trade Reporting Facility (TRF) and for the NASD Alternative Display Facility (ADF). These facilities record transactions from OTC markets, some non-exchange Electronic Communications Networks (ECNs), broker internalization and other dark pools. The crucial information for our analysis is that dark pools trades are reported as “D”. The reason is that dark pools do not have to publicly report their quotes and so they do not have to comply with Rule 612 of Reg NMS, which only refers to quoted prices. As a result all sub-penny trades are reported with the exchange code “D”.[20]

We consider only trades and quotes which take place between 9:30:00 AM and 4:00:00 PM. For each day and for each stock we derive the NBBO using the Wharton Research Data Services (WRDS) suggested procedure and we compute the time-weighted bid, ask and total depth, the time-weighted quoted spread in cents and in percentage of the mid-quote, and the intraday price range, defined as (high-low)/high, as a measure of intraday volatility.

We remove erroneous and irregular trades; in particular we keep only trades whose correction indicator is either “00” or “01”. We then compute the share volume and (buy) order imbalance defined as the absolute value of (buys-sells)/share volume where buys are classified using a modified Lee and Ready (1991) algorithm.[21] Table 2 shows descriptive statistics for our sample of stocks divided by exchange and capitalization.

[Insert Table 2 here]

4.3 Price Improvement

In order to identify and classify SPT, we construct an auxiliary variable, the price improvement (PI), which is computed as follows. First, prices are rounded up to the closest cent for sell orders and rounded down to the closest cent for buy orders. Second, PI is obtained as the difference between the rounded price and the reported price. It is always restricted to the interval (0.00, 0.01). Therefore PI is equal to:

$$PI = |rounded(price) - price| \tag{1}$$

Table 3 reports an example of how we compute PI. We are now able to properly define SPT as a function of PI. In particular, we have no SPT when PI is equal to zero and SPT when PI is different from zero.

[Insert Table 3 here]

[20] We can also observe sub-penny trades with other exchange codes. In this case, however, the price improvement is exactly equal to half-cent and is the result of a trade in exchange operated dark pools pricing at the NBBO mid-quote.

[21] We first apply a tick-test considering at most two previous trades. We classify a trade as a buy if its price is above the price of the previous trade (or two trades before); otherwise we classify the trade as a sell. If the trade is still unclassified, we classify it as a buy (sell) if the execution price is above (below) the mid-quote at the time of the trade.

SPT can be further classified into MID, when PI is exactly equal to half-cent (0.005), and QJ, when PI is strictly positive but different from half-cent.[22] Hence, our rounding procedure is immaterial to the definition of SPT, MID and QJ.

We can use PI to illustrate how significant SPT is in the U.S. equity markets. Interpreting PI as the gain from SPT and multiplying it by the number of shares traded in SPT, we obtain the dollar volume captured by stepping ahead of the queue. For example, consider Adobe Systems Inc.: in our sample period the daily average of shares traded is 14.8 million. Of these, 1.62 million are traded in SPT divided between QJ (1.25 million) and MID (0.37 million), which correspond respectively to 8.4% and 2.5% of consolidated share volume. Therefore the gross gain from SPT, measured as PI multiplied by the corresponding number of shares traded, is equal on average to $2,500 for QJ and $1,850 for MID on a daily base.

Figures 5 and 6 report the plot of the average QJ and MID across the sample period for each stock as a function of price, for NASDAQ and NYSE. Graphically we see that on both markets QJ generally decreases with price, while MID increases. Overall the percentage of volume traded in SPT increases when the stock value decreases and seems to be consistent with the fact that a one cent tick size is a binding constraint for low-priced stocks.

[Insert Figures 5-6 here]

To study the distribution of PIs, in Figures 7 - 8 we group stocks into 10 bins of size 0.001 for two groups of 30 high-priced NASDAQ stocks and 30 high-priced NYSE stocks, separately. For each bin, we report the associated SPT dollar volume as a percentage of the total traded volume; measuring it in share volume or number of trades yields qualitatively the same results. For example, Figure 7 shows that 3.05% of total NASDAQ traded volume is executed with a PI which lies in the interval (0,0.001], while for NYSE it amounts to 2.44%. The most common type of SPT corresponds to PI equal to 0.005. The distribution of PIs decreases almost monotonically as the PI increases, except for the spike at PI exactly equal to half-cent.

[Insert Figures 7-8 here]

[22]Our measure of MID is a lower bound of executions at the mid-quote. The reason is that we are considering as mid-crossing only those executions which result in a sub-penny price. If the spread turns out to be an even multiple of the tick, we will not classify it as MID with our methodology.

5 Factors Driving Sub-Penny Trading

To study how SPT varies with market characteristics, and test Hypotheses 1 and 2, we start with daily Fama-MacBeth cross-sectional regressions for both QJ and MID. More precisely, the regressors used in the different specifications are the following: dummy variable which is equal to one when the stock is NYSE listed, log of market capitalization, log of share volume, closing price, time-weighted cent quoted spread, time-weighted percent quoted spread, log of time-weighted bid depth, relative order imbalance in percent and intraday price range. The average daily estimated coefficients and t-statistics are reported in Tables 4 - 5. The t-statistics are based on the Newey-West adjusted standard errors with 5 lags.[23]

[Insert Tables 4-5 here]

In specification (1) we control for listing exchange by including a dummy variable for NYSE listing. We also control for the logarithm of market capitalization. The results show that QJ is decreasing in market capitalization and is higher for NASDAQ than for NYSE stocks after controlling for market capitalization. Conversely, MID is increasing in market capitalization and is higher for NYSE than for NASDAQ stocks.

In specification (2) we replace market capitalization with share volume and closing price, and the results show that QJ is decreasing in price, while MID is increasing in price.

We then add the quoted spread in cents and the log of (time-weighted) bid depth in specification (3) to include measures of liquidity and we find that stocks with greater depth have more QJ and less MID. A wider quoted spread is also associated with higher MID, holding listing exchange, share volume, and price constant.

We replace quoted spread in cents and price with quoted spread in basis points in specification (4) and find that this variable is statistically significant only for QJ. Stocks with wider basis point spreads have more QJ, controlling for listing exchange and share volume.

Finally, in specification (5) we drop share volume and include the relative order imbalance in percent of share volume and volatility as measured by the intraday price range. We find that for the whole sample of NASDAQ and NYSE stocks the relation

[23] Results for NYSE and NASDAQ as well as for small and large capitalization stocks are available in the Online Appendix.

between QJ and volatility is not significant. However, by looking at the results for the different groups of stocks (Table 6) we find that for small low-priced stocks and large high-priced stocks the coefficient is positive and significant. We also find that MID increases significantly in relative order imbalance.[24]

[Insert Table 6 here]

In sum the multivariate Fama-MacBeth regression analysis shows that QJ is significantly higher for NASDAQ than NYSE stocks all else equal, while for MID we observe the opposite. In terms of our hypotheses, we find that QJ is positively related to time weighted bid depth and negatively related to price. These results are consistent with both Hypotheses 1 and 2. When the book is deep competition for the provision of liquidity is high and therefore QJ becomes a very attractive option for liquidity providers. Moreover the effect is stronger for low-priced stocks as, when the price is low and the tick-to-price is large, the gain traders make by undercutting orders at the top of regular exchanges by a fraction of the tick size is higher as a percentage of the stock price.

The results for spread do not confirm the negative relation with QJ outlined in Hypothesis 1. The coefficient for quoted spread is not significant and that for percent spread is positive. This could be due to the interaction between spread and price. As discussed above, when the stock price decreases (and hence the relative tick size increases) the costs of non-execution are smaller and the gains from QJ are higher so that traders are more inclined to use limit rather than market orders. Hence submitting aggressive limit orders via QJ increases. However, when the price decreases, the relative spread increases with the result that the relation between relative spread and QJ becomes positive. Moreover, the spread itself can be influenced by QJ and therefore to find a clean relation between QJ and spread we need to control for endogeneity, which we do in the next Section.

Our results on volatility do not confirm Hypothesis 3 as we find a positive relation between QJ and volatility, for small low-priced and large high-priced stocks (Table 6). This may be because a significant fraction of QJ is due to internalization for which some factors driving dark pool trading in separate venues may not be as relevant. Even though higher volatility increases execution uncertainty, if broker-dealers are active on

[24] In addition, Table IV to VII in the Online Appendix show that with this specification the coefficient on bid depth for QJ is statistically significant and positive for NASADQ stocks, whereas for MID it is significant and negative for both NASDAQ and NYSE.

both sides of the market, they might be insured against unexpected changes in the fundamental value of the asset.

The results we obtain for order imbalance only partly confirm Hypothesis 3. Results for the overall sample are not significant, the reason being that, consistent with our hypothesis, QJ is positively related to imbalance for large stocks, yet -and surprisingly- it is negatively related to order imbalance for small stocks (Table 6).

5.1 HFT and QJ

To proxy HFT we derive from TAQ data the ratio between quotes updates on lit markets and consolidated number of trades.[25] This way we should capture HFTs that affect quotes, in other words only the lit market trading by potential HFTs. We cannot exclude that QJ is also HFT, but because it is not captured by our proxy, we can only try to assess the relationship between lit market HFT and QJ. This is an important caveat to keep in mind.

We find that, for the whole sample of stocks, lit market trading by potential HFTs is negatively related to QJ (Table 6) which could be due to HFTs switching from lit to dark venues. By looking at the different groups of stocks, we find that the negative relation is stronger for both small low-priced and large high-priced stocks. In the case of large stocks, in which HFTs are generally more active, a possible explanation for the negative relation is that -depending on the state of the PLB- market participants dealing with HFT strategically switch from market making to QJ. In the case of small stocks -especially low-priced for which QJ is strongly negatively related to price- the negative relation between QJ and HFT could be explained as follows: when the price is low and therefore QJ is particularly profitable, traders might switch from HFT market making to QJ, the opposite holding when the price is high. Therefore, our results suggest that HFT market making and QJ could be strategic substitutes in traders' order submission decisions.

[25]To normalize quote updates, we also use number of shares instead of trades and the new HFT proxy obtained delivers similar results. Share volume is instead not appropriate to normalize quote updates as it is strongly correlated with the asset price, which seems to be one of the determinants of HFT (Yao and Ye, 2013). Our results do not qualitatively change when we use as a proxy for HFT the ratio between quote updates and total number of trades net of those trades market with "D" that should include both internalized and dark pool executions.

6 Sub-Penny Trading and Market Quality

We now move to test our hypotheses on the effects of SPT, and in particular QJ, on the quality of regular exchanges. The issue is that market quality and dark trading are jointly determined as pointed out in BRW (2011), so to establish a causal relationship we have to address the endogeneity issue.

To deal with the inherent endogeneity of SPT and market quality, we need to find good instruments for SPT and market quality, respectively. In a recent paper studying the impact of low latency trading on market quality, Hasbrouck and Saar (2013) propose using low latency trading in other stocks during the same time period as an instrument for low latency trading in a particular stock. We follow their suggestion and use SPT for other stocks (not i) on day t as an instrument for SPT in stock i. Because we have observed that there are systematic differences between exchanges and across market capitalization groups in SPT, we refine their instrument slightly. We require that the other stocks (not i) are listed on the same exchange as stock i and that their market capitalization is in the same market capitalization group as stock i. The market quality measures are the time-weighted percent quoted spread and the logarithm time-weighted bid-depth. We estimate a two-equation simultaneous model for SPT, which can be either QJ or MID, and market quality measures (MQMs) using both traditional 2SLS and a two step generalized method of moments (GMM) procedure. Specifically, we estimate the following two-equation simultaneous model for each MQM:

$$\begin{aligned} MQM_{i,t} &= a_1 SP_{i,t} + a_2 MQM_{NOTi,t} + \varepsilon_{1,t} \\ SP_{i,t} &= b_1 MQM_{i,t} + b_2 SP_{NOTi,t} + \varepsilon_{2,t} \end{aligned} \tag{2}$$

As instruments for $SP_{i,t}$, we use $SP_{NOTi,t}$, which is the average SPT of other stocks listed on the same exchange, in the same market capitalization group. Note that we exclude stock i. Similarly, as an instrument for $MQM_{i,t}$, we use $MQM_{NOTi,t}$, which is the average market quality measure for other stocks listed on the same exchange, in the same market capitalization group. We again exclude stock i. This estimation method is chosen to address the endogeneity of SPT and MQM. So we obtain a consistent estimate of the a_1 coefficient that tells us how SPT affects market quality.

We estimate the above system of equations for all stocks and days in a panel. To control for stock fixed effects, we de-mean all variables by deducting the in-sample average and divide the de-meaned variables by their in-sample standard deviation. As a result, the estimated coefficients can be interpreted as the response to a one standard deviation shock. We do not include any trend in the system nor we detrend our variables since the visual inspection of the data tells us that the variables are stationary in the sample period. We estimate system (2) using a 2SLS procedure; standard errors are double clustered by stock and day.[26] As market quality measures we consider the logarithm of time-weighted bid depth and both the time-weighted relative spread and the time-weighted cent quoted spread.[27]

Tables 7 - 10 report the results from the simultaneous equation model. We have four different specifications of system (2), which arise from the combination of the two different types of SPT (QJ and MID) and the two market quality measures (bid depth and relative spread). We are primarily interested in the a_1 and b_1 coefficients: a_1 measures the effect of SPT on market quality and b_1 measures the effect of market quality on SPT. The coefficients on our instruments, a_2 and b_2, are positive and highly significant. In other words, they appear to be good instruments. We present the results for the whole sample and for small and large capitalization stocks.

Results for the factors driving QJ (b_1) are now consistent with Hypothesis 1 as QJ is positively related with the liquidity of the stock not only for depth, as in the Fama-MacBeth regressions, but also for the spread, measured both in cents and in basis points. Table 7 (Panel A) shows that b_1 is positive and statistically significant for depth, while Table 9 (Panel A) shows that it is negative and statistically significant for spread. For MID, we obtain similar results for depth (Table 8, Panel A) and spread (Table 10, Panel A).

Results for the effects of QJ on market quality show a positive effect on both depth and spread. The coefficient a_1 is positive and statistically significant in the depth system (Table 7, Panel A) and negative and statistically significant in the spread system (Table 9, Panel A). Interpreting market capitalization as a rough proxy of liquidity, we can see that consistently with Hypothesis 4 for liquid stocks (large cap) an increase in QJ determines an improvement of market quality measured by depth (Table 7, Panel A) and spread (Table 9, Panel A). By contrast, we cannot confirm this hypothesis for illiquid stocks as the results are not significant. Moreover, we do not observe any significant effect of MID on market quality (Table 8, Panel A, and Table 10, Panel A).

[26]We repeat the estimation using a GMM approach allowing for heteroskedasticity of unknown form, still double clustering standard errors. The results are nearly identical.

[27]Results for time-weighted cent quoted spread are available upon request.

A word of caution is due, however, as we do not control for make-take fees. In practice, traders use dark markets to trade at sub-penny not only to step ahead of exiting limit orders at the top of regular exchanges but sometimes also to save on take fees. By internalizing orders a broker-dealer avoids paying the fees imposed by lit venues. Therefore, one should take the effects of different possible make-take fee structures into account both to correctly interpret our results, and to appropriately discuss our policy implications.[28]

[Insert Tables 7-10 here]

6.1 Robustness Check

We rerun our simultaneous equation models including additional controls which are exogenous and can affect both market quality and SPT. In particular, following again Hasbrouck and Saar (2013), we include the daily return on SP500 and its volatility (proxied by VIX) to take into account market-wide activity. The same concern has been addressed by Brogaard, Hendershott and Riordan (2013) when studying the relation between HFT and volatility. Specifically, we estimate the following system of equations, which is a modification of system (2):

$$\begin{aligned} MQM_{i,t} &= a_1 SP_{i,t} + a_2 MQM_{NOTi,t} + a_3 SP500_t + a_4 VIX_t + \varepsilon_{1,t} \\ SP_{i,t} &= b_1 MQM_{i,t} + b_2 SP_{NOTi,t} + b_3 SP500_t + b_4 VIX_t + \varepsilon_{2,t} \end{aligned} \quad (3)$$

We present the results in Tables 7 - 10, Panel B. The results are robust to the inclusion of the additional controls.

7 Policy Instruments and Regression Discontinuity Design

In this Section we discuss our last empirical implication which aims at investigating whether a reduction in the tick size is the adequate policy instrument to use in a market in which traders can undercut existing liquidity posted to lit markets.

To this end we exploit an additional feature of Rule 612 of Reg NMS which states

[28] We cannot take into account make-take fees as for trades marked with the code "D" we cannot distinguish among different trading venues.

that the minimum price improvement (tick size) changes from \$0.01 to \$0.0001 for stocks priced less than \$1. Given this cutoff, we employ a regression discontinuity design (RD). According to Cameron and Trivedi (2005), a RD is a quasi-experimental design in which the probability of receiving a treatment is a discontinuous function of one or more underlying variables. Such a design can arise in circumstances where a treatment is triggered by an administrative or organizational rule. This is exactly our case: since regulation imposes a cutoff at \$1, the treatment is having a price greater or equal to \$1 and so we can evaluate the effect of the treatment in the neighborhood of \$1.[29] To sum up, the intuition behind a RD is that observations below and above the cutoff can be compared directly to draw inference on the effect of the treatment.

To run the RD we select a different sample than the one previously used. Our prior objective was to have a random sample representative of all NASDAQ and NYSE stocks. Now, we need stocks around \$1. So we take all NASDAQ listed stocks whose closing price at September 30, 2010 was between \$0.8 and \$1.2. Then, we restrict further our sample by retaining only the stocks whose price crosses the \$1 threshold at least once during the sample period. The reason for applying this restriction is that identification of the treatment effect relies on stocks which cross the threshold (Lee and Lemieux, 2010).

The design is valid if the stock's price can be considered as good as randomly assigned below or above the threshold. We are dealing with the case of treatment assignment which is public knowledge (i.e., Rule 612 of Reg NMS). So, in principle, the limit orders (or the trades, depending on the assignment variable we are using) can crowd on one side of the cutoff. If it is the case that the assignment variable has been manipulated, the RD is not valid. However, in our case, the assumption of no manipulation seems to be legitimate since traders should not be able to move a stock price across the cutoff, so the assignment can be considered as good as randomly assigned. Nevertheless, as an additional check we run the density discontinuity test presented by McCrary (2008). The assignment variables are the bid price, the ask price and the execution price, depending on the specification we consider. We make this classification since also the Rule 612 of Reg NMS does.[30] We run the test for the continuity of the assignment variables and we do not find any evidence of manipulation. We report the graphical representation of the test for the bid price in Figure 9. Results are analogous for ask price and execution

[29] We will specify later what we are referring to with price.

[30] To be precise, Rule 612 of Reg NMS refers only to quotes. We extend the analysis to the price at which the trade has been executed.

price.

[Insert Figure 9 here]

We first study the effect of regulation on market quality (i.e., log of bid depth and relative spread). To build our dataset, for each stock and day we take a snapshot of the NBBO every hour starting from 9:30:00 AM. We estimate the following pooled OLS regression:

$$y_{it} = b_0 + b_1 D_{i,t} + b_2(Price_{i,t} - 1) + b_3(Price_{i,t} - 1)D_{i,t} + \varepsilon_{i,t} \tag{4}$$

where $Price_{i,t}$ is the price rounded down to the closest cent. $Price_{i,t}$ is reduced by one to locate the threshold at zero. $D_{i,t}$ is an indicator that takes the value of 1 if $Price_{i,t}$ is equal to or greater than \$1 and a value of 0 if $Price_{i,t}$ is less than \$1. D_{it} generates a discontinuity in the treatment around the threshold which allows for the estimation of the effect of the treatment. Index i is for stock and index t is for time. We include all the observations which satisfy the following condition: $-h \leq Price_{i,t} - 1 \leq h$, where h is the bandwidth. We select a bandwidth equal to \$0.10.[31] We estimate a pooled OLS regression even though we are dealing with a panel dataset. Indeed, in RD fixed effects are unnecessary for identification; it is sufficient to take into account the within-stock correlation of the errors over time using clustered standard errors (Lee and Lemieux, 2010). We present the results for the logarithm of bid depth and quoted spread percent as outcome variable (y) and the bid price as assignment variable. The results of the estimation are in Tables 11 - 12.

[Insert Tables 11 - 12 here]

The coefficients on the dummy variable $D_{i,t}$ show that for both ask depth and relative spread the discontinuity is statistically significant. Bid depth and relative spread both increase moving from below to above the \$1 cutoff. Therefore in line with Hypothesis 5, the implications for market quality are mixed: the inside spread improves but bid depth deteriorates. Interestingly, these conclusions are in line with the predictions of BCRWW (2014) on the effects of a reduction of the tick size. A reduction in the tick size decreases traders' incentive to post limit orders and consequently it reduces market

[31] We replicate the analysis also using the optimal bandwidth according to Imbens and Kalyanaraman (2012): the results still hold.

depth. It also reduces the inside spread as in today's fast markets traders active at the top of the PLB follow the downwards movement of the tick size and move towards the new narrower inside spread. Our results are robust when using 1.5 and 0.75 times the initial bandwidth h (Tables 11 - 12) and when using the ask price as assignment variable.[32] Note that the conclusion of the RD is valid only in the neighborhood of the cutoff and cannot be generalized to all stocks (this is why we have already studied the relation between SPT and market quality with other methods).

8 Conclusions and Policy Implications

During the last decade financial markets have been characterized by the growth of dark markets and fast trading. These two elements have fostered the development of sub-penny trading (SPT) which is a particular form of dark trading. SPT takes place when traders take advantage of the Sub-Penny Rule (Rule 612) and its exceptions by posting orders in the dark market or internalizing customers' orders at fractions of the tick size. In this way they gain price priority over the orders sitting at the inside quotes of regulated markets. Sub-penny orders are then executed against traders who demand liquidity by using smart order routing programs that allow them to hit the best quotes posted both in lit and in dark markets.

In this paper we show that in the U.S. approximately 10% of share volume executes at sub-penny increments due to queue-jumping (QJ). Approximately 3% executes exactly at the mid-quote, and could be the result either of QJ or of a midpoint cross in an opaque venue. These volumes are rapidly increasing and SPT is a concern to regulators as it can reduce the incentive for liquidity providers to post limit orders on regular exchanges and hence worsen market depth. For this reason in 2010 the SEC proposed the Trade-At Rule to curtail dark trading, and more recently such rule has been introduced in Canada (October 2012) and in Australia (May 2013).[33] In the 2010 concept release

[32]In this case the bid depth is substituted by the ask depth for consistency.

[33]On October 12, 2012 the Investment Industry Regulatory Organization of Canada introduced a Trade-At Rule that now gives lit orders priority over dark orders in the same venue. In particular small dark orders under 5,000 shares or C$100,000 dollars in value must offer at least half a tick in price improvement for stocks that have a one tick spread, and a full tick of price improvement for stocks with higher spreads. Following the implementation of this rule, the Canadian dark share of volume dropped by more than 50% (Rosenblatt Securities Inc., February 2013). Even the Australian Securities and Investments Commission on May 26, 2013 adopted a new regulation for dark venues aimed at containing dark trading for transactions of size smaller than blocks. The key component of the new regulatory regime is the adoption of a minimum size threshold for dark orders.

on market microstructure the SEC also opened the debate on SPT by asking questions on the factors that drive SPT, the effects of SPT on the quality of lit markets and more precisely on the adequate policy instrument to use to influence SPT. In this paper we answer most of these questions and to draw our empirical hypotheses we build a model of competition between a Public Limit Order Book and a Sub-Penny Venue.

We find that SPT varies significantly by listing exchange, as QJ is significantly higher for NASDAQ than for NYSE listed stocks. Consistently with our model's predictions, we show that QJ is positively related to depth and negatively related to stock price. This means that broker-dealers use dark markets for QJ especially for stocks where competition for the provision of liquidity is high and hence it is difficult to gain price priority on lit platforms. The use of QJ is also intense in low-priced stocks for which the profit from price improvement is higher relative to the asset value.

We find that SPT improves both depth and spread, especially for large capitalization stocks. This is also consistent with our model's predictions. We do not find that SPT harms liquidity even though we do not investigate whether market participants benefit or are harmed by SPT. We leave this interesting topic for future research. We also build a proxy of HFT as the ratio between quotes updates on lit markets and consolidated number of trades and find that QJ is negatively related to lit market HFT, suggesting that they could be strategic substitutes in traders' order submission decisions.

Finally, our analysis allows us to draw conclusions on the efficacy of the tick size as a policy instrument to influence SPT. In the U.S. Rule 612 sets the minimum price improvement of 1 penny only for stocks priced above \$1, while it permits executions at sub-penny increments for stocks priced below \$1. We exploit this feature of the Sub-Penny Rule to conduct a regression discontinuity analysis and study what happens to market quality as the price of a stock moves across the trigger point (\$1.00) below which the minimum price increment is 10 time smaller. We show that the inside spread improves and the depth deteriorates when a stock moves from above to below the \$1 threshold. This means that consistently with BCRWW (2014) a reduction in the tick size would have a detrimental effect on the provision of liquidity: not only depth would decrease, but by reducing spreads a smaller tick size would further curtail broker-dealers' profits and hence potentially decrease rather than increase their incentive to supply liquidity.

References

Brogaard, J., T. Hendershott, and R. Riordan, 2013. High Frequency Trading and Price Discovery. Review of Financial Studies, forthcoming.

Buti, S. and B. Rindi, 2013. Undisclosed Orders and Optimal Submission Strategies in a Limit Order Market. Journal of Financial Economics, 109, 3, 797-812.

Buti, S., B. Rindi, and I. M. Werner, 2011. Diving into Dark Pools. Working paper. Fisher College of Business.

Buti, S., B. Rindi, Y. Wen and I. M. Werner, 2013. Tick Size Regulation and Sub-Penny Trading. Working paper. Fisher College of Business.

Buti, S., F. Consonni, B. Rindi, Y. Wen and I. M. Werner, 2014. Tick Size: Theory and Evidence. Working paper. Fisher College of Business.

Cameron, A. C., and P. K. Trivedi, 2005. Microeconometrics: Methods and Applications, Cambridge University Press.

Comerton-Forde, C. and T. J. Putnins, 2013. Dark Trading and Price Discovery. Woking paper.

Degryse, H., F. de Jong and V. van Kervel, 2014. The impact of dark trading and visible fragmentation on market quality. Review of Finance, forthcoming.

Foucault, T., 1999. Order flow composition and trading costs in a dynamic limit order market. Journal of Financial Markets 2, 99-134.

Foucault, T., O. Kadan, and E. Kandel. 2005. Limit Order Book as A Market for Liquidity. Review of Financial Studies 18:1171-1217.

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

Hasbrouck, J., and G. Saar, 2013. Low Latency Trading. Journal of Financial Markets, 16: 645-678

Hatheway, F., A. Kwan, and H. Zheng, 2013. An Empirical Analysis of Market Segmentation on U.S. Equities Markets. Working Paper, SSRN FEN Library.

Imbens, G., and K. Kalyanaraman, 2012. Optimal Bandwidth Choice for the Regression Discontinuity Estimator. The Review of Economic Studies 79: 933-959.

Kwan, A., R. Masulis, and T. McInish, 2013. Trading rules, competition for order flows and market fragmentation. Working Paper, SSRN FEN Library.

Lee, C., and M. Ready, 1991. Inferring Trade Direction from Intraday Data. Journal of Finance 46: 733-746.

Lee, D. S., and T. Lemieux, 2010. Regression Discontinuity Designs in Economics. Journal of Economic Literature 48(2): 281-355.

McCrary, J., 2008. Manipulation of the Running Variable in the Regression Discontinuity Design: A Density Test. Journal of Econometrics 142: 698–714.

Nimalendran, M. and S. Ray, 2014. Informational Linkages Between Dark and Lit Trading Venues. Journal of Financial Markets 17, 230-261.

Parlour, C. A., 1998. Price Dynamics in Limit Order Markets. Review of Financial Studies 11, 789-816.

Ready, M., 2013. Determinants of Volume in Dark Pools.Working paper. University of Wisconsin.

Rosenblatt Securities Inc., 2013. Trading Talk, Let there be light, February 26.

Rosu, I., 2009. A dynamic model of the limit order book. Review of Financial Studies 22, 4601-4641.

U.S. Securities and Exchange Commission (SEC), 2010. Concept Release on Equity Market Structure No. 34-61358.

U.S. Securities and Exchange Commission (SEC), 2012. Report to Congress on Decimalization.

Yao, C and M. Ye, 2013. Tick Size Constraints, Market Structure, and Liquidity. Working Paper, University of Illinois at Urbana-Champaign.

Figure 1: **Trading Strategies -** This Figure reports - for $t = t_1$ - the equilibrium submission probabilities of limit orders posted on the ask side at different levels of the book, under the two states of the book "Liquid Stock" (Panel A) and "Illiquid Stock" (Panel B), and for 2 different asset values, $v = \{1, 10\}$. The submission probabilities of limit orders for the Public Limit Order Book (PLB) refer to the framework with tick size $\tau_{PLB} = \tau$, and price grid $\{A_1, A_2\}$. The submission probabilities of limit orders are instead indicated as SPV when posted to the Sub-Penny Venue with tick size $\tau_{SPV} = \frac{\tau}{3}$, and price grid $\{a_1, a_2, a_3, a_4, a_5\}$. They are indicated as (PLB&SPV) when posted to the PLB that competes with the SPV. The circle indicates the price at which the order is submitted in equilibrium. Under the state of the book "Liquid stock", both the book of the PLB and the book of the PLB&SPV open at t_1 with one share on the first level (A_1, B_1); while under the regime of "Illiquid stock" both the book of the PLB and the book of the PLB&SPV open empty. In both cases the SPV opens with equal probability either empty or with one share on the second level (a_2, b_2). The broker-dealers' arrival rate is $\alpha = 10\%$ and $\tau = 0.1$.

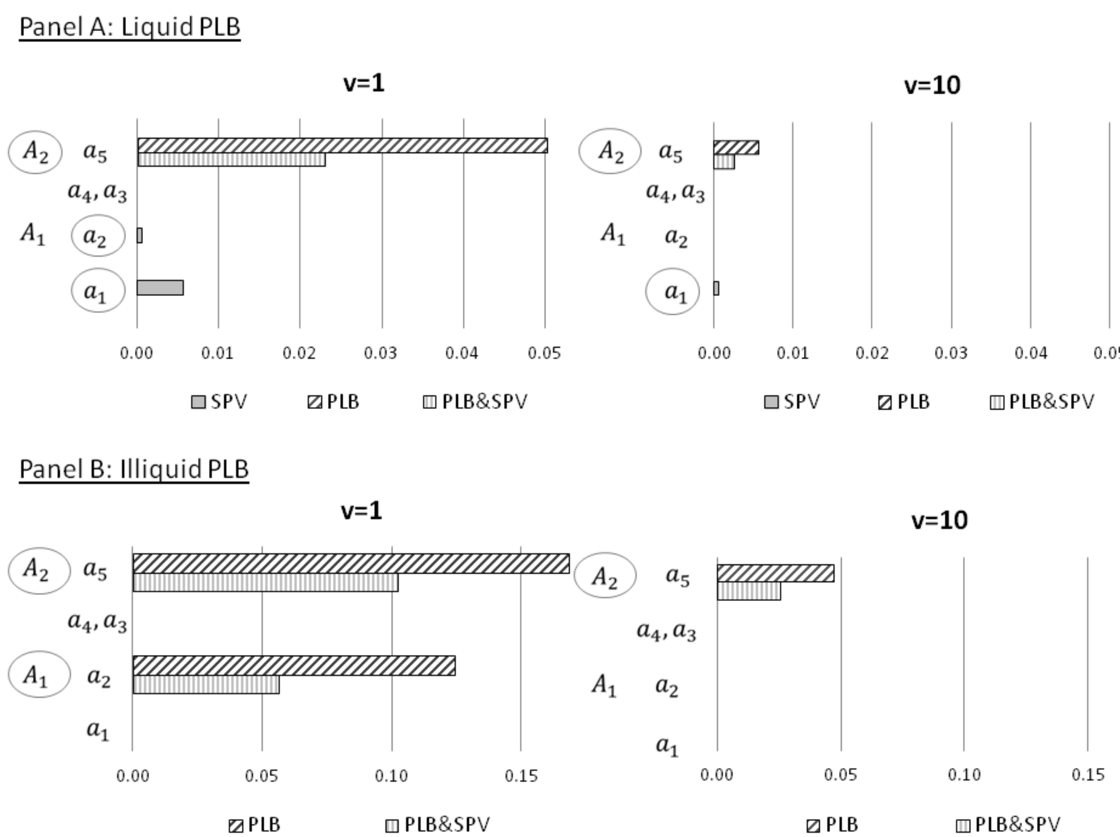

Figure 2: **Trading Strategies - Small Tick Size** - This Figure reports - for $t = t_1$ - the equilibrium submission probabilities of limit orders posted on the ask side at different levels of the book, under the two states of the book "Liquid Stock" (Panel A) and "Illiquid Stock" (Panel B), for 2 different asset values, $v = \{1, 10\}$, and for $\tau = \frac{0.1}{3}$. The submission probabilities of limit orders for the Public Limit Order Book (PLB) refer to the framework with tick size $\tau_{PLB} = \tau$, and price grid $\{A_1, A_2\}$. The submission probabilities of limit orders are instead indicated as SPV when posted to the Sub-Penny Venue with tick size $\tau_{SPV} = \frac{\tau}{3}$, and price grid $\{a_1, a_2, a_3, a_4, a_5\}$. They are indicated as (PLB&SPV) when posted to the PLB that competes with the SPV. The circle indicates the price at which the order is submitted in equilibrium. Under the state of the book "Liquid stock", both the book of the PLB and the book of the PLB&SPV open at t_1 with one share on the first level (A_1, B_1); while under the regime of "Illiquid stock" both the book of the PLB and the book of the PLB&SPV open empty. In both cases the SPV opens with equal probability either empty or with one share on the second level (a_2, b_2). The broker-dealers' arrival rate is $\alpha = 10\%$.

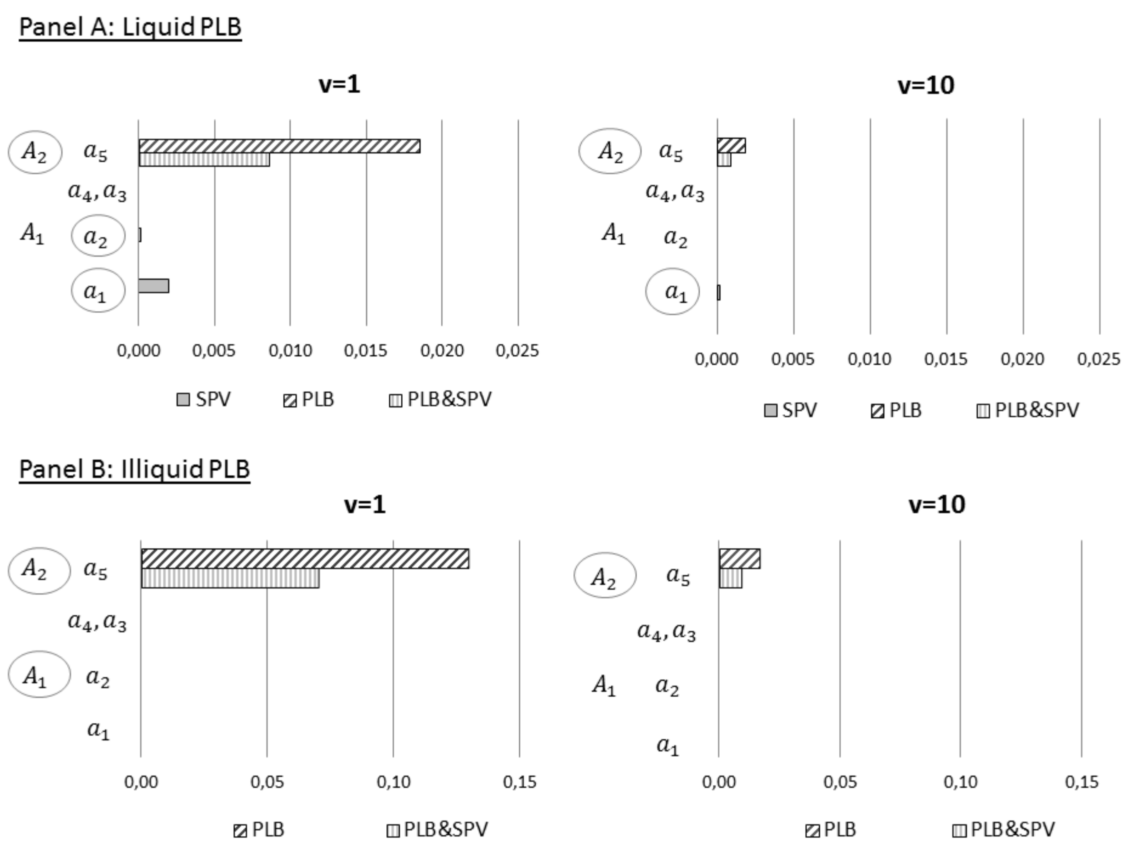

Figure 3: **Order Flows and Market Quality - Liquid stocks -** This Figure reports statistics for order flows and market quality that result from the comparison between the benchmark Public Limit Order Book (PLB) with tick size $\tau_{PLB} = \tau$, and the extended model (PLB&SPV) in which a PLB with the same tick size as the benchmark model competes with a Sub-Penny Venue (SPV) characterized by a smaller tick size, $\tau_{SPV} = \frac{\tau}{3}$. We refer to "Liquid stock" as the regime under which both the benchmark PLB and the PLB that competes with the SPV open at t_1 with one share on the first level of both sides of the book $\{A_1, B_1\}$, and at the same time the SPV opens with equal probability either empty or with one share on the second level of the book (a_2, b_2). We present results for the broker-dealers' arrival rate $\alpha = 10\%$, and for $v = \{1, 10\}$. For order flows this panel reports statistics on liquidity provision and trading volume both for the PLB (LP^{PLB} and VL^{PLB}) and for the SPV (LP^{SPV} and VL^{SPV}). For market quality we report statistics on total depth (DPT^{PLB}), depth at the best bid-offer (DPI^{PLB}) and spread (SP^{PLB}). The statistics are computed as the average difference between the value of the measure in the PLB that competes with the SPV, and the value of the same statistic in the benchmark PLB: $\Delta y = \frac{1}{k}\sum_{t=t_1}^{t_k}[y_t^{PLB\&SPV} - y_t^{PLB}] \times 100$, where $y = \left\{DPI^{PLB}, DPT^{PLB}, SP^{PLB}, VL^{PLB}, LP^{PLB}, VL^{SPV}, LP^{SPV}\right\}$, $k = 2$ for $\left\{DPI^{PLB}, DPT^{PLB}, SP^{PLB}, LP^{PLB}, LP^{SPV}\right\}$, and $k = 3$ for $\left\{VL^{PLB}, VL^{SPV}\right\}$. This difference is computed for both $\tau = 0.1$ and $\tau = \frac{0.1}{3}$.

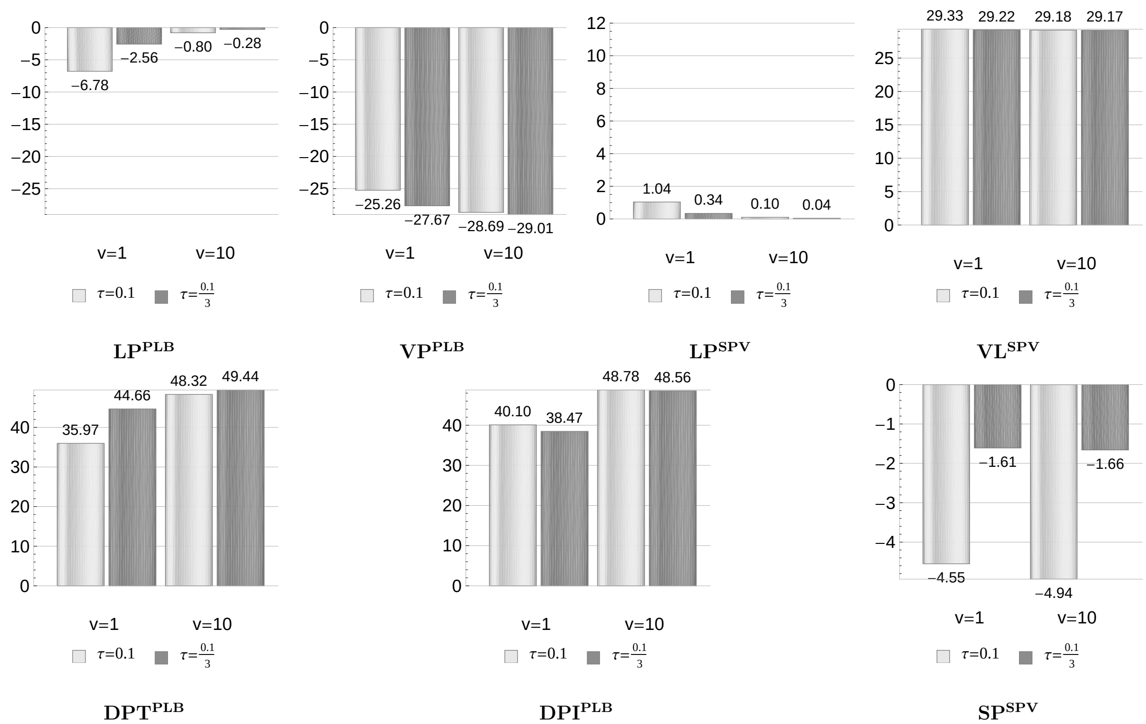

Figure 4: **Order Flows and Market Quality - Illiquid stocks -** This Figure reports statistics for order flows and market quality that result from the comparison between the benchmark Public Limit Order Book (PLB) with tick size $\tau_{PLB} = \tau$, and the extended model (PLB&SPV) in which a PLB with the same tick size as the benchmark model competes with a Sub-Penny Venue (SPV) characterized by a smaller tick size, $\tau_{SPV} = \frac{\tau}{3}$. We refer to "Illiquid stock" as the regime under which both the book of the benchmark PLB and the book of the PLB that competes with the SPV open empty at t_1, and at the same time the SPV opens with equal probability either empty or with one share on the second level of the book (a_2, b_2). We present results for the broker-dealers' arrival rate $\alpha = 10\%$, and for $v = \{1, 10\}$. For order flows this panel reports statistics on liquidity provision and trading volume both for the PLB (LP^{PLB} and VL^{PLB}) and for the SPV (LP^{SPV} and VL^{SPV}). For market quality we report statistics on total depth (DPT^{PLB}), depth at the best bid-offer (DPI^{PLB}) and spread (SP^{PLB}). The statistics are computed as the average difference between the value of the measure in the PLB that competes with the SPV, and the value of the same statistic in the benchmark PLB: $\Delta y = \frac{1}{k}\sum_{t=t_1}^{t_k}[y_t^{PLB\&SPV} - y_t^{PLB}] \times 100$, where $y = \left\{DPI^{PLB}, DPT^{PLB}, SP^{PLB}, VL^{PLB}, LP^{PLB}, VL^{SPV}, LP^{SPV}\right\}$, $k = 2$ for $\left\{DPI^{PLB}, DPT^{PLB}, SP^{PLB}, LP^{PLB}, LP^{SPV}\right\}$, and $k = 3$ for $\left\{VL^{PLB}, VL^{SPV}\right\}$. This difference is computed for both $\tau = 0.1$ and $\tau = \frac{0.1}{3}$. $\tau = 0.1, \alpha = 10\%$.

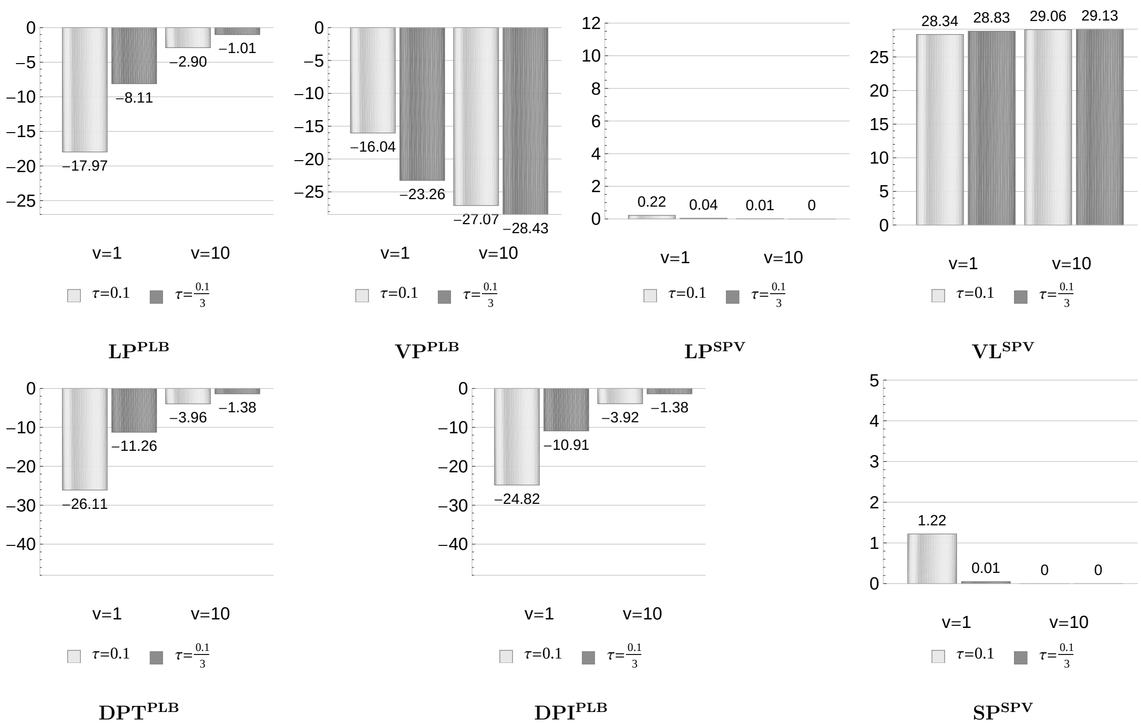

Figure 5: **NASDAQ queue-jumping and mid-crossing versus stock price**
This figure reports the queue-jumping and mid-crossing percentage over the sample period for each NASDAQ stock against the price of the stock.

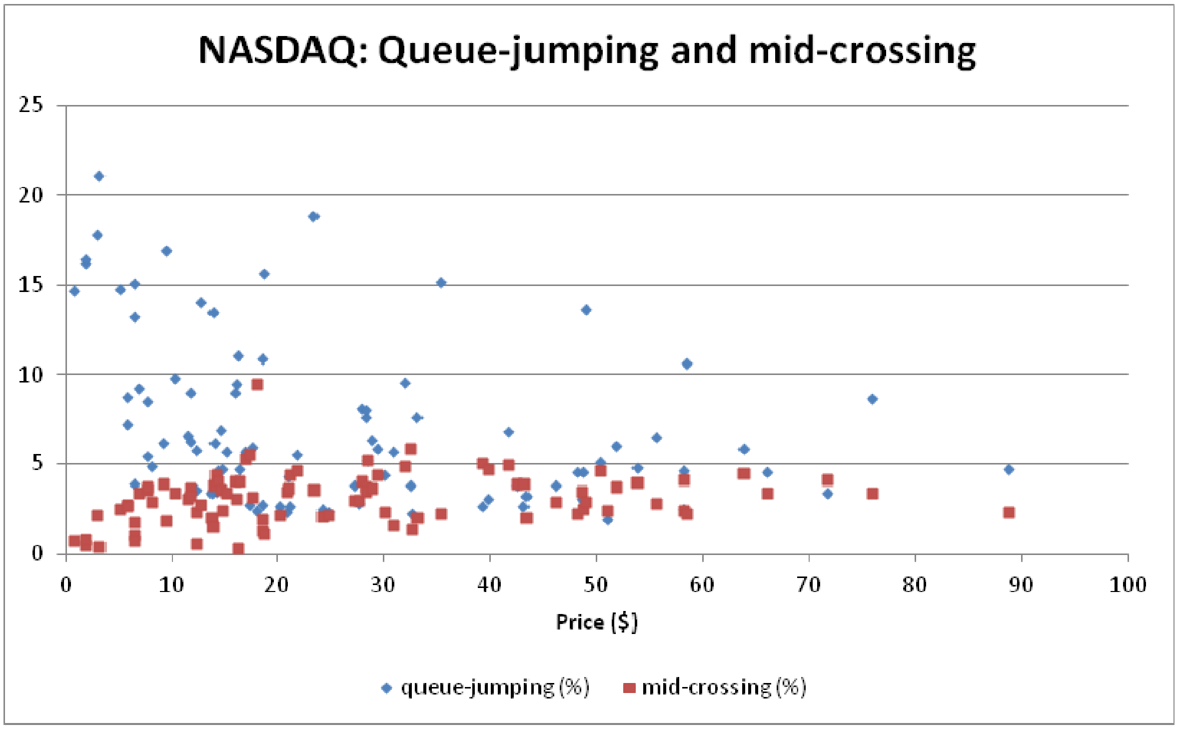

Figure 6: **NYSE queue-jumping and mid-crossing versus stock price**
This figure reports the queue-jumping and mid-crossing percentage over the sample period for each NYSE stock against the price of the stock.

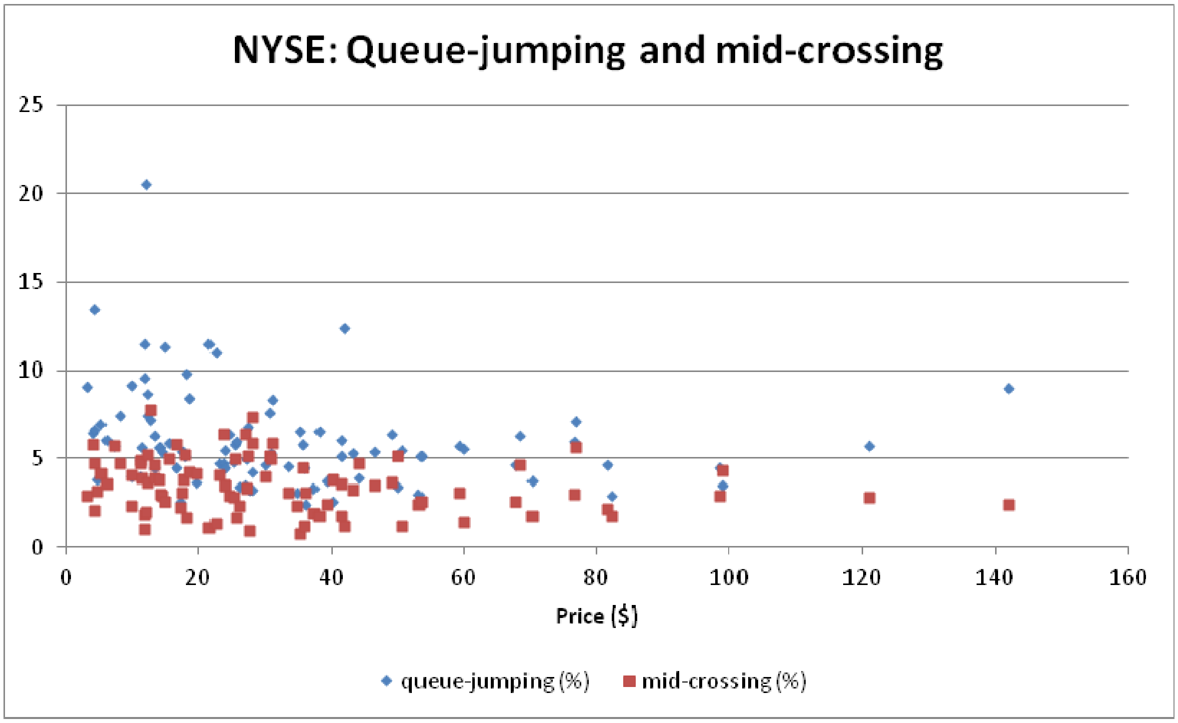

Figure 7: **NASDAQ distribution of Sub-Penny (including mid-crossing)**
This figure reports the Sub-Penny in percentage of consolidated volume for each bin over the sample period and for the group of 30 high-priced NASDAQ stocks.

Figure 8: **NYSE distribution of Sub-Penny (including mid-crossing)**
This figure reports the Sub-Penny in percentage of consolidated volume for each bin over the sample period and for the group of 30 high-priced NYSE stocks.

Figure 9: **Density Representation for bid price**
This figure reports the density of the assignment variable (i.e., bid price) to assess the continuity of the variable itself across the cutoff. The points represent the smoothed histogram while the solid black line represents the smoothed density with confidence bands (solid grey lines). Both histogram and density are smoothed using a triangle kernel.

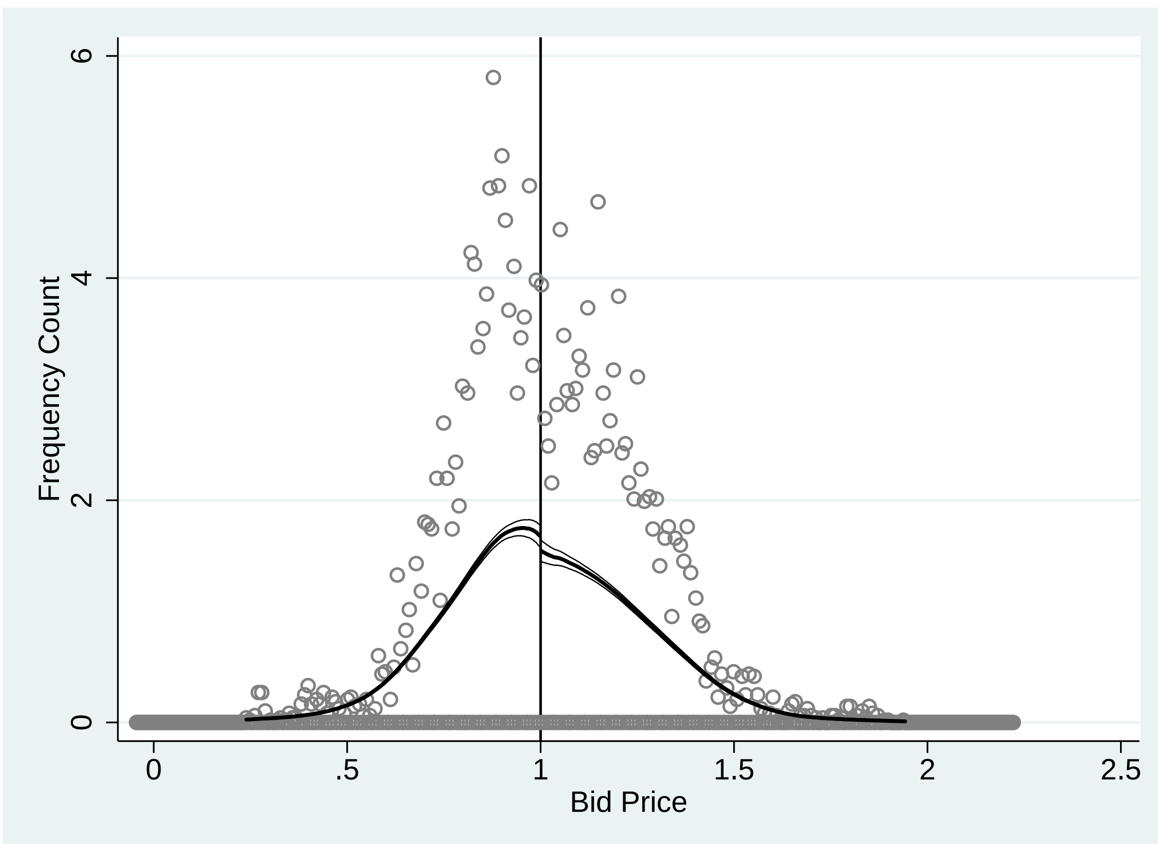

Table 1: **Price Grid**
This Table shows the price grid for both the Public Limit Order Book (PLB) and the Sub-Penny Venue (SPV). v is the asset value and τ is the tick size.

PLB	Price	SPV
A_2	$v + \frac{9}{6}\tau$	a_5
	$v + \frac{7}{6}\tau$	a_4
	$v + \frac{5}{6}\tau$	a_3
A_1	$v + \frac{3}{6}\tau$	a_2
	$v + \frac{1}{6}\tau$	a_1
	$v - \frac{1}{6}\tau$	b_1
B_1	$v - \frac{3}{6}\tau$	b_2
	$v - \frac{5}{6}\tau$	b_3
	$v - \frac{7}{6}\tau$	b_4
B_2	$v - \frac{9}{6}\tau$	b_5

Table 2: **Stocks descriptive statistics**
The sample period is October 1, 2010 - November 30, 2010. All variables are daily (time-weighted) averages. Market cap is the stock's market capitalization in millions from CRSP. Closing price is in dollar from CRSP. Relative spread is the difference between the bid and ask price over the midquote. Bid depth is the bid depth at NBBO in unit of trade (100 shares). Price range is the difference between the maximum and the minimum price over the maximum price. Queue-jumping is the QJ on D computed as the ratio of volume executed in QJ over the total consolidated volume. Mid-crossing is the MID on D computed as the ratio of volume executed in MID over the total consolidated volume. Sub-Penny is the SPT on D computed as the ratio of volume executed in SPT over the total consolidated volume. Statistics are reported for the full sample (NASDAQ and NYSE) and for the NASDAQ and the NYSE separately. Furthermore, for each panel, we report the summary statistics for the three market capitalization groups (SMALL, MEDIUM and LARGE).

	Market Cap (Million $)	Closing Price ($)	Time-weighted relative spread ($^0/_{000}$)	Time-weighted bid depth (UoT)	Price range (%)	Queue-jumping on D (%)	Mid-crossing on D (%)	Sub-Penny on D (%)
NASDAQ+NYSE	**7324.30**	**29.70**	**0.149**	**14398.67**	**0.04**	**6.59**	**3.34**	**9.93**
Small	534.48	26.78	0.272	957.70	0.05	7.68	2.51	10.19
Medium	1740.48	29.10	0.112	2321.72	0.03	5.49	3.40	8.89
Large	19697.94	33.22	0.062	39916.59	0.02	6.61	4.10	10.71
NASDAQ	**5807.27**	**26.58**	**0.153**	**3766.59**	**0.03**	**7.16**	**3.16**	**10.32**
Small	462.41	22.75	0.287	1439.62	0.04	9.11	2.11	11.23
Medium	1648.04	26.92	0.111	2700.43	0.03	5.79	3.51	9.30
Large	15311.36	30.05	0.061	7159.73	0.02	6.57	3.86	10.43
NYSE	**8841.33**	**32.83**	**0.145**	**25030.75**	**0.04**	**6.03**	**3.51**	**9.54**
Small	606.55	30.81	0.257	475.77	0.05	6.25	2.90	9.15
Medium	1832.93	31.28	0.114	1943.02	0.03	5.18	3.29	8.48
Large	24084.53	36.40	0.064	72673.45	0.02	6.65	4.35	11.00

Table 3: **Price improvement calculation**
Hereafter an example of price improvement calculation.

SYMBOL	DATE	TIME	PRICE	TRADE SIGN	PRICE ROUNDED	PRICE IMPROVEMENT
ASEI	20101018	10:13:12	78.7501	BUY	78.75	0.0001
ASEI	20101018	10:25:46	78.7975	SELL	78.8	0.0025

Table 4: **NASDAQ and NYSE sample queue-jumping**

This table reports the results of regressions of queue-jumping on contemporaneous market characteristics based on daily cross-sectional Fama-MacBeth regressions. Queue-jumping is 100 times daily queue-jumping dollar volume on Alternative Display Facility (ADF) divided by daily consolidated dollar volume (volume computed during trading hours 9:30 AM - 4:00 PM ET). We take the log of market capitalization, volume and bid depth. Cent spreads are multiplied by 100. The relative order imbalance is the absolute value of (buys-sells)/share volume. The intraday price range is defined as (high-low)/high. NYSE is a dummy variable which takes the value of 1 for stocks whose primary listing exchange is NYSE. We report the average daily coefficients on top and t-statistics below, computed according to Newey-West with 5 lags.

	(1) queue-jumping	(2) queue-jumping	(3) queue-jumping	(4) queue-jumping	(5) queue-jumping
NYSE	-0.97***	-0.83***	-0.83***	-0.77***	-0.83***
	(-19.61)	(-16.60)	(-17.33)	(-15.15)	(-14.83)
market capitalization (CSRP)	-0.49***				
	(-24.25)				
share volume (TAQ)		0.13***	-0.28***	0.76***	
		(8.00)	(-7.97)	(18.95)	
closing price (CSRP)		-0.042***	-0.024***		
		(-33.68)	(-10.16)		
quoted spread cents (TAQ)			-0.016		
			(-2.00)		
time-weighted bid depth (TAQ)			0.66***	0.013	0.69***
			(14.03)	(0.31)	(40.38)
quoted spread percent (TAQ)				14.9***	10.1***
				(26.88)	(17.04)
relative order imbalance in percent (TAQ)					0.17
					(0.37)
(high-low)/high (TAQ)					4.45
					(1.66)
Constant	17.5***	5.81***	8.42***	-9.02***	0.68**
	(40.50)	(17.71)	(21.54)	(-15.80)	(3.32)
Observations	7560	7560	7560	7560	7560
R^2	0.04	0.07	0.09	0.17	0.16

t statistics in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 5: **NASDAQ and NYSE sample mid-crossing**

This table reports the results of regressions of mid-crossing on contemporaneous market characteristics based on daily cross-sectional Fama-MacBeth regressions. Mid-crossing is 100 times daily mid-crossing dollar volume on Alternative Display Facility (ADF) divided by daily consolidated dollar volume (volume computed during trading hours 9:30 AM - 4:00 PM ET). We take the log of market capitalization, volume and bid depth. Cent spreads are multiplied by 100. The relative order imbalance is the absolute value of (buys-sells)/share volume. The intraday price range is defined as (high-low)/high. NYSE is a dummy variable which takes the value of 1 for stocks whose primary listing exchange is NYSE. We report the average daily coefficients on top and t-statistics below, computed according to Newey-West with 5 lags.

	(1)	(2)	(3)	(4)	(5)
	mid-crossing	mid-crossing	mid-crossing	mid-crossing	mid-crossing
NYSE	0.22**	0.44***	0.44***	0.41***	0.37***
	(2.85)	(5.61)	(5.60)	(5.28)	(4.99)
market capitalization (CSRP)	0.42***				
	(30.03)				
share volume (TAQ)		0.41***	0.88***	0.69***	
		(28.64)	(23.62)	(16.72)	
closing price (CSRP)		0.0054***	-0.016***		
		(4.01)	(-6.29)		
quoted spread cents (TAQ)			0.045***		
			(8.07)		
time-weighted bid depth (TAQ)			-0.69***	-0.46***	0.15***
			(-13.39)	(-11.29)	(6.78)
quoted spread percent (TAQ)				-0.33	-4.55***
				(-0.88)	(-15.99)
relative order imbalance in percent (TAQ)					1.25*
					(2.67)
(high-low)/high (TAQ)					3.00
					(1.60)
Constant	-5.84***	-4.40***	-7.84***	-6.13***	2.48***
	(-17.87)	(-14.49)	(-19.63)	(-10.33)	(16.73)
Observations	7560	7560	7560	7560	7560
R^2	0.05	0.07	0.10	0.09	0.07

t statistics in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 6: **NASDAQ and NYSE sample queue-jumping and High-Frequency Trading (HFT)**

This table reports the results of regressions of queue-jumping on contemporaneous market characteristics based on daily cross-sectional Fama-MacBeth regressions. Queue-jumping is 100 times daily queue-jumping dollar volume on Alternative Display Facility (ADF) divided by daily consolidated dollar volume (volume computed during trading hours 9:30 AM - 4:00 PM ET). HFT proxy is the ratio between quotes updates on lit markets and consolidated number of trades. The relative order imbalance is the absolute value of (buys-sells)/share volume. Log return is the daily logarithmic return and the intraday price range is defined as (high-low)/high. NYSE is a dummy variable which takes the value of 1 for stocks whose primary listing exchange is NYSE. We divide the sample between the small (first and second columns) and the large (third and fourth columns) stocks by market capitalization, within them we further divide between high and low-priced stocks. We report the average daily coefficients on top and t-statistics below, computed according to Newey-West with 5 lags.

	(1) Small - Low	(2) Small - High	(3) Large - Low	(4) Large - High	(5) All
NYSE	-1.800***	-1.536***	0.191	0.614***	-0.722***
	(-9.92)	(-8.65)	(1.11)	(4.50)	(-16.90)
closing price (CSRP)	-0.438***	-0.016*	-0.115***	0.010*	-0.033***
	(-15.76)	(-2.35)	(-4.53)	(2.18)	(-15.99)
HFT proxy	-0.037**	-0.007*	-0.049	-0.104***	-0.017***
	(-2.71)	(-2.05)	(-2.00)	(-5.60)	(-5.43)
relative order imbalance in percent (TAQ)	-8.760***	0.122	1.546	4.987***	-0.772
	(-6.11)	(0.11)	(1.72)	(5.24)	(-1.40)
Log Return	6.044	-7.549	-6.315	0.624	-1.998
	(0.66)	(-0.48)	(-0.74)	(0.05)	(-0.40)
(high-low)/high (TAQ)	28.019***	8.548	-24.995	57.365***	30.965***
	(3.96)	(0.76)	(-1.85)	(4.06)	(5.27)
Constant	15.683***	7.095***	8.992***	4.143***	7.305***
	(29.62)	(14.05)	(21.90)	(13.64)	(30.87)
Observations	1260	1260	1260	1260	7560
R^2	0.33	0.21	0.25	0.28	0.13

t statistics in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 7: **NASDAQ and NYSE queue-jumping and time-weighted bid depth**
This table reports the results from the analysis of the relationship between queue-jumping activity and time-weighted bid depth. We measure queue-jumping activity as QJ, which is defined as 100 times daily queue-jumping dollar volume on Alternative Display Facility (ADF) divided by daily consolidated dollar volume (volume computed during trading hours 9:30 AM - 4:00 PM ET). Due to the potential simultaneity between time-weighted bid depth (log of) and queue-jumping activity, in Panel A we estimate the following two-equation simultaneous model for QJ and the log of time-weighted bid depth (MQM):
$MQM_{i,t} = a_1 QJ_{i,t} + a_2 MQM_{NOTi,t} + \varepsilon_{1,t}$
$QJ_{i,t} = b_1 MQM_{i,t} + b_2 QJ_{NOTi,t} + \varepsilon_{2,t}$
In Panel B, we estimate the following two-equation simultaneous model, including as control, the return on SP500 and the VIX:
$MQM_{i,t} = a_1 QJ_{i,t} + a_2 MQM_{NOTi,t} + a_3 SP500_t + a_4 VIX_t + \varepsilon_{1,t}$
$QJ_{i,t} = b_1 MQM_{i,t} + b_2 QJ_{NOTi,t} + b_3 SP500_t + b_4 VIX_t + \varepsilon_{2,t}$
As an instrument for $QJ_{i,t}$ we use $QJ_{NOTi,t}$, which is the daily average QJ activity of other stocks listed on the same exchange, in the same market capitalization grouping (LARGE, MEDIUM, SMALL; excluding stock i). Similarly, as an instrument for $MQM_{i,t}$ we use $MQM_{NOTi,t}$, which is the average time-weighted bid depth (log of) for other stocks listed on the same exchange, in the same market capitalization grouping (LARGE, MEDIUM, SMALL; excluding stock i). We estimate the simultaneous equation model by pooling observations across all stocks and days in the sample. To make the pooling meaningful, we de-mean all variables by deducting the stock-specific average and scale all variables by dividing by the stock-specific standard deviation to control for stock fixed effects. We report the estimated coefficients on top and p-values below. Estimation is done with 2SLS with two-way clustered standard errors (i.e., stock and day).

	Panel A			Panel B		
	(1)	(2)	(3)	(4)	(5)	(6)
	Full sample	Small	Large	Full sample	Small	Large
a_1	0.34***	0.044	0.28*	0.32***	0.055	0.25*
	(0.000)	(0.802)	(0.022)	(0.000)	(0.738)	(0.032)
a_2	0.52***	0.37***	0.64***	0.52***	0.36***	0.62***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
a_3				0.015	-0.017	0.035
				(0.295)	(0.521)	(0.193)
a_4				0.000025	0.000090	-0.000042
				(0.965)	(0.928)	(0.968)
Observations	7560	2520	2520	7560	2520	2520
b_1	0.28**	0.058	0.25*	0.26***	0.071	0.23*
	(0.001)	(0.818)	(0.035)	(0.001)	(0.776)	(0.024)
b_2	0.42***	0.42***	0.55***	0.41***	0.41***	0.54***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
b_3				0.026*	0.016	0.018
				(0.045)	(0.505)	(0.499)
b_4				-0.00024	-0.00015	-0.00023
				(0.628)	(0.885)	(0.812)
Observations	7560	2520	2520	7560	2520	2520

p-values in parentheses
* p<0.05 ** p<0.01 *** p<0.001

Table 8: **NASDAQ and NYSE mid-crossing and time-weighted bid depth**
This table reports the results from the analysis of the relationship between mid-crossing activity and time-weighted bid depth. We measure mid-crossing activity as MID, which is defined as 100 times daily mid-crossing dollar volume on Alternative Display Facility (ADF) divided by daily consolidated dollar volume (volume computed during trading hours 9:30 AM - 4:00 PM ET). Due to the potential simultaneity between time-weighted bid depth (log of) and mid-crossing activity, in Panel A we estimate the following two-equation simultaneous model for MID and the log of time-weighted bid depth (MQM):
$MQM_{i,t} = a1 * MID_{i,t} + a2 * MQM_{NOTi,t} + \varepsilon_{1,t}$
$MID_{i,t} = b1 * MQM_{i,t} + b2 * MID_{NOTi,t} + \varepsilon_{2,t}$
In Panel B, we estimate the following two-equation simultaneous model, including as control, the return on SP500 and the VIX:
$MQM_{i,t} = a_1 MID_{i,t} + a_2 MQM_{NOTi,t} + a_3 SP500_t + a_4 VIX_t + \varepsilon_{1,t}$
$MID_{i,t} = b_1 MQM_{i,t} + b_2 MID_{NOTi,t} + b_3 SP500_t + b_4 VIX_t + \varepsilon_{2,t}$
As an instrument for $MID_{i,t}$ we use $MID_{NOTi,t}$, which is the daily average MID activity of other stocks listed on the same exchange, in the same market capitalization grouping (LARGE, MEDIUM, SMALL; excluding stock i). Similarly, as an instrument for $MQM_{i,t}$ we use $MQM_{NOTi,t}$, which is the average time-weighted bid depth (log of) for other stocks listed on the same exchange, in the same market capitalization grouping (LARGE, MEDIUM, SMALL; excluding stock i). We estimate the simultaneous equation model by pooling observations across all stocks and days in the sample. To make the pooling meaningful, we de-mean all variables by deducting the stock-specific average and scale all variables by dividing by the stock-specific standard deviation to control for stock fixed effects. We report the estimated coefficients on top and p-values below. Estimation is done with 2SLS with two-way clustered standard errors (i.e., stock and day).

	Panel A			Panel B		
	(1)	(2)	(3)	(4)	(5)	(6)
	Full sample	Small	Large	Full sample	Small	Large
a1	0.20	-0.041	0.25	0.16	-0.046	0.16
	(0.376)	(0.889)	(0.439)	(0.494)	(0.872)	(0.644)
a2	0.60***	0.37***	0.72***	0.59***	0.36***	0.69***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
a3				0.030*	-0.016	0.044
				(0.028)	(0.521)	(0.088)
a4				-0.000094	0.000080	-0.00014
				(0.870)	(0.935)	(0.889)
Observations	7560	2520	2520	7560	2520	2520
b1	0.11*	-0.055	0.15	0.096	-0.060	0.12
	(0.045)	(0.804)	(0.086)	(0.095)	(0.797)	(0.230)
b2	0.25***	0.28***	0.28***	0.24***	0.28***	0.26***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.001)
b3				0.016	-0.0060	0.032
				(0.270)	(0.793)	(0.249)
b4				-0.000081	0.00010	-0.00020
				(0.886)	(0.914)	(0.810)
Observations	7560	2520	2520	7560	2520	2520

p-values in parentheses
* p<0.05 ** p<0.01 *** p<0.001

Table 9: **NASDAQ and NYSE queue-jumping and time-weighted relative spread** This table reports the results from the analysis of the relationship between queue-jumping activity and time-weighted relative spread. We measure queue-jumping activity as QJ, which is defined as 100 times daily queue-jumping dollar volume on Alternative Display Facility (ADF) divided by daily consolidated dollar volume (volume computed during trading hours 9:30 AM - 4:00 PM ET). Due to the potential simultaneity between time-weighted relative spread and queue-jumping activity, in panel A we estimate the following two-equation simultaneous model for QJ and the log of time-weighted relative spread (MQM):
$MQM_{i,t} = a_1 QJ_{i,t} + a_2 MQM_{NOTi,t} + \varepsilon_{1,t}$
$QJ_{i,t} = b_1 MQM_{i,t} + b_2 QJ_{NOTi,t} + \varepsilon_{2,t}$
In Panel B, we estimate the following two-equation simultaneous model, including as control, the return on SP500 and the VIX:
$MQM_{i,t} = a_1 QJ_{i,t} + a_2 MQM_{NOTi,t} + a_3 SP500_t + a_4 VIX_t + \varepsilon_{1,t}$
$QJ_{i,t} = b_1 MQM_{i,t} + b_2 QJ_{NOTi,t} + b_3 SP500_t + b_4 VIX_t + \varepsilon_{2,t}$
As an instrument for $QJ_{i,t}$ we use $QJ_{NOTi,t}$, which is the daily average QJ activity of other stocks listed on the same exchange, in the same market capitalization grouping (LARGE, MEDIUM, SMALL; excluding stock i). Similarly, as an instrument for $MQM_{i,t}$ we use $MQM_{NOTi,t}$, which is the average time-weighted relative spread for other stocks listed on the same exchange, in the same market capitalization grouping (LARGE, MEDIUM, SMALL; excluding stock i). We estimate the simultaneous equation model by pooling observations across all stocks and days in the sample. To make the pooling meaningful, we de-mean all variables by deducting the stock-specific average and scale all variables by dividing by the stock-specific standard deviation to control for stock fixed effects. We report the estimated coefficients on top and p-values below. Estimation is done with 2SLS with two-way clustered standard errors (i.e., stock and day).

	Panel A			Panel B		
	(1)	(2)	(3)	(3)	(4)	(5)
	Full sample	Small	Large	Full sample	Small	Large
a_1	-0.49**	-0.55	-0.29	-0.46**	-0.60	-0.27*
	(0.007)	(0.070)	(0.164)	(0.002)	(0.090)	(0.046)
a_2	0.63***	0.58***	0.68***	0.61***	0.50***	0.67***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
a_3				-0.028	-0.068*	-0.017
				(0.057)	(0.012)	(0.608)
a_4				0.00014	0.00027	0.00017
				(0.781)	(0.774)	(0.846)
Observations	7560	2520	2520	7560	2520	2520
b_1	-0.27***	-0.26*	-0.23*	-0.26***	-0.32**	-0.21*
	(0.000)	(0.024)	(0.015)	(0.000)	(0.008)	(0.011)
b_2	0.35***	0.28**	0.54***	0.35***	0.26*	0.53***
	(0.000)	(0.005)	(0.000)	(0.000)	(0.013)	(0.000)
b_3				0.0086	-0.035	0.026
				(0.525)	(0.230)	(0.291)
b_4				-0.000084	0.000084	-0.00013
				(0.868)	(0.930)	(0.883)
Observations	7560	2520	2520	7560	2520	2520

p-values in parentheses
* p<0.05 ** p<0.01 *** p<0.001

Table 10: **NASDAQ and NYSE mid-crossing and time-weighted relative spread**
This table reports the results from the analysis of the relationship between mid-crossing activity and time-weighted relative spread. We measure mid-crossing activity as MID, which is defined as 100 times daily mid-crossing dollar volume on Alternative Display Facility (ADF) divided by daily consolidated dollar volume (volume computed during trading hours 9:30 AM - 4:00 PM ET). Due to the potential simultaneity between time-weighted relative spread and mid-crossing activity, in Panel A we estimate the following two-equation simultaneous model for MID and the time-weighted relative spread (MQM):
$MQM_{i,t} = a_1 MID_{i,t} + a_2 MQM_{NOTi,t} + \varepsilon_{1,t}$
$MID_{i,t} = b_1 MQM_{i,t} + b_2 MID_{NOTi,t} + \varepsilon_{2,t}$
In Panel B, we estimate the following two-equation simultaneous model, including as control, the return on SP500 and the VIX:
$MQM_{i,t} = a_1 MID_{i,t} + a_2 MQM_{NOTi,t} + a_3 SP500_t + a_4 VIX_t + \varepsilon_{1,t}$
$MID_{i,t} = b_1 MQM_{i,t} + b_2 MID_{NOTi,t} + b_3 SP500_t + b_4 VIX_t + \varepsilon_{2,t}$
As an instrument for $MID_{i,t}$ we use $MID_{NOTi,t}$, which is the daily average MID activity of other stocks listed on the same exchange, in the same market capitalization grouping (LARGE, MEDIUM, SMALL; excluding stock i). Similarly, as an instrument for $MQM_{i,t}$ we use $MQM_{NOTi,t}$, which is the average time-weighted relative spread for other stocks listed on the same exchange, in the same market capitalization grouping (LARGE, MEDIUM, SMALL; excluding stock i). We estimate the simultaneous equation model by pooling observations across all stocks and days in the sample. To make the pooling meaningful, we de-mean all variables by deducting the stock-specific average and scale all variables by dividing by the stock-specific standard deviation to control for stock fixed effects. We report the estimated coefficients on top and p-values below. Estimation is done with 2SLS with two-way clustered standard errors (i.e., stock and day).

	Panel A			Panel B		
	(1)	(2)	(3)	(4)	(5)	(6)
	Full sample	Small	Large	Full sample	Small	Large
a_1	-0.32	-0.076	-0.58	-0.30	-0.13	-0.54
	(0.157)	(0.817)	(0.245)	(0.103)	(0.685)	(0.199)
a_2	0.74***	0.72***	0.67***	0.71***	0.66***	0.66***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
a_3				-0.039***	-0.057*	-0.015
				(0.000)	(0.020)	(0.624)
a_4				0.00023	0.00027	0.00020
				(0.636)	(0.758)	(0.841)
Observations	7560	2520	2520	7560	2520	2520
b_1	-0.13**	-0.045	-0.22*	-0.12**	-0.067	-0.20*
	(0.002)	(0.362)	(0.015)	(0.003)	(0.380)	(0.010)
b_2	0.22***	0.28***	0.20*	0.22***	0.27***	0.19
	(0.000)	(0.000)	(0.033)	(0.000)	(0.000)	(0.057)
b_3				0.0040	-0.016	0.027
				(0.789)	(0.583)	(0.212)
b_4				0.0000055	0.00015	-0.000072
				(0.992)	(0.875)	(0.933)
Observations	7560	2520	2520	7560	2520	2520

p-values in parentheses
* p<0.05 ** p<0.01 *** p<0.001

Table 11: **Regression Discontinuity for bid depth**

This table reports the results of regression discontinuity for the bid depth. We estimate the following pooled OLS regression model:
$MQM_{i,t} = b_0 + b_1 D_{i,t} + b_2(Price_{i,t} - 1) + b_3(Price_{i,t} - 1)D_{i,t} + \varepsilon_{i,t}$
where $Price_{i,t}$ is the price rounded down to the closest cent. $D_{i,t}$ is an indicator that takes the value of 1 if $Price_{i,t}$ is equal to or greater than $1 and a value of 0 if $Price_{i,t}$ is less than $1. $Price_{i,t}$ is reduced by one to have the threshold at zero. $MQM_{i,t}$ is the logarithm of the bid depth. Estimation is done using a pooled OLS regression with standard errors clustered by stock. Estimation is carried out for three different bandwidths (0.10, 0.15 and 0.075).

	(1) Bandwidth = 0.10	(2) Bandwidth = 0.15	(3) Bandwidth = 0.075
Price	-1.46	-2.82	-1.92
	(-0.48)	(-1.54)	(-0.41)
Dummy	1.57***	1.46***	1.55***
	(4.46)	(4.29)	(3.95)
Price*Dummy	-8.53	-2.30	-6.49
	(-1.43)	(-0.48)	(-0.80)
Constant	6.09***	6.05***	6.08***
	(28.16)	(31.87)	(25.60)
Observations	1887	2483	1532
R^2	0.12	0.10	0.13

t statistics in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 12: **Regression Discontinuity for relative spread**

This table reports the results of regression discontinuity for the relative spread. We estimate the following pooled OLS regression model:

$MQM_{i,t} = b_0 + b_1 D_{i,t} + b_2(Price_{i,t} - 1) + b_3(Price_{i,t} - 1)D_{i,t} + \varepsilon_{i,t}$

where $Price_{i,t}$ is the price rounded down to the closest cent. $D_{i,t}$ is an indicator that takes the value of 1 if $Price_{i,t}$ is equal to or greater than \$1 and a value of 0 if $Price_{i,t}$ is less than \$1. $Price_{i,t}$ is reduced by one to have the threshold at zero. $MQM_{i,t}$ is the relative spread. Estimation is done using a pooled OLS regression with standard errors clustered by stock. Estimation is carried out for three different bandwidths (0.10, 0.15 and 0.075).

	(1) Bandwidth = 0.10	(2) Bandwidth = 0.15	(3) Bandwidth = 0.075
Price	-0.11	-0.12*	-0.063
	(-1.42)	(-2.02)	(-0.78)
Dummy	0.012*	0.013*	0.012**
	(2.36)	(2.58)	(2.91)
Price*Dummy	0.11	0.095	0.0092
	(0.68)	(0.96)	(0.05)
Constant	0.019**	0.019**	0.021**
	(3.21)	(3.25)	(3.25)
Observations	1887	2483	1532
R^2	0.01	0.01	0.01

t statistics in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$