Dark Trading at the Midpoint:
Does SEC Enforcement Policy Encourage Stale Quote Arbitrage?

Robert P. Bartlett, III
University of California, Berkeley

Justin McCrary
University of California, Berkeley, NBER

October 26, 2016

JEL codes: G10, G15, G18, G23, G28, K22
Keywords: latency arbitrage, high-frequency trading; dark pools; tick sizes; market structure

***Statement of Financial Disclosure and Conflict of Interest*: Neither author has any financial interest or affiliation (including research funding) with any commercial organization that has a financial interest in the findings of this paper. The authors are grateful to the University of California, Berkeley School of Law, for providing general faculty research support.**

Dark Trading at the Midpoint:
Does SEC Enforcement Policy Encourage Stale Quote Arbitrage?

Abstract:

Prevailing research posits that liquidity providers bypass queue lines on exchanges by offering liquidity in dark venues with *de minimis* sub-penny price improvement, thus exploiting an exception to the penny quote rule. We show that (a) the SEC enforces the quote rule to prevent sub-penny queue-jumping in dark pools unless trades are "pegged" to the NBBO midpoint and (b) the documented increase in dark trading due to investor queue-jumping stems from increased midpoint trading. Although encouraging pegged orders can subject traders to stale quote arbitrage, we show it could have affected no more than 5% of our sample midpoint trades.

1. Introduction

Recent theoretical and empirical work has shown that the formal rules regulating the minimum price variation (MPV) for quoting equity securities create a competitive advantage for non-exchange "dark" venues over conventional stock exchanges. Specifically, when traders compete to provide liquidity on exchanges, the current MPV of $0.01 means the minimum spread between the national best bid and the national best offer (NBBO) will often collapse to a penny, producing long queue-lines for traders seeking to post orders on exchanges at the most competitive prices. However, because the MPV rule regulates quotes but not *trades*, liquidity providers can "queue jump" exchanges by offering to trade in dark venues at sub-penny prices (Buti, Consonni, Rindi, Wen and Werner, 2015; Buti, Rindi, Wen and Werner, 2013; Kwan, Masulis & McInish, 2014). For instance, if the NBBO for a security is $10.00 x $10.01, a literal reading of securities laws might suggest that a trader could post an offer to buy at $10.001 in a non-exchange venue, thus incentivizing a liquidity-taking seller to look first to dark pools for optimal pricing.[1] In this fashion, liquidity providers in dark pools could intercept marketable orders that would otherwise be filled on exchanges, potentially harming the incentive to post displayed orders without any meaningful price improvement for liquidity takers.

In this study we evaluate these claims in light of a largely unexamined enforcement policy of the Securities Exchange Commission (SEC) regarding the MPV rule—law on the ground as opposed to law on the books.[2] Under this policy, the SEC selectively enforces the MPV rule against dark pools to prevent their customers from queue-jumping exchanges using sub-penny orders unless their orders are "pegged" to the midpoint of the NBBO. This enforcement policy substantially reduces the possibility that traders can bypass long queue-lines at the NBBO by placing sub-penny orders in dark pools offering just *de minimis* improvement over the NBBO (a practice which, in keeping with prior literature, we refer to as "stepping ahead"), because doing so would expose the trading venue to substantial legal liability.

[1] Importantly, the liquidity-providing trader would be prohibited from posting a passive buy order on an exchange at $10.01 (one tick better than $10.00) because such an order would lock the NBBO at $10.01 x $10.01. Under Rule 610 of Regulation National Market System (Reg. NMS), all exchanges must establish and enforce rules that prohibit their members from displaying orders that lock or cross the NBBO. Combined with the MPV rule, Rule 610 thus constrains an aggressive buyer to either crossing the spread and buying at $10.01, or posting a passive bid at $10.00. However, because exchanges generally observe price-time priority rules, the trader's passive order at $10.00 will stand behind other passive bids that arrived earlier and which are also priced at the national best bid of $10.00.

[2] The one manuscript in the literature of which we are aware that acknowledges the SEC enforcement actions we discuss below is Buti, Consonni, Rindi, Wen, and Werner (2015).

The legal basis for this SEC enforcement policy arises from the agency's authority to interpret and enforce Rule 612 of Reg. NMS. Rule 612 sets forth the formal MPV rule by prohibiting the "display, rank or accept[ance]" of an order for any NMS stock priced in an increment smaller than $0.01 for orders priced at $1.00 per share or more. In adopting the rule in 2005, however, the SEC emphasized the rule was purposefully silent on the treatment of sub-penny trades to permit their use so long as the trade did not result from an impermissible sub-penny order or quotation. The SEC gave as examples of permissible sub-penny trades the common practice among broker-dealers of providing sub-penny price improvement to their customer orders and sub-penny trades arranged to be completed at the midpoint of the NBBO. As a formal rule, any sub-penny trade arising from an order not explicitly priced in sub-pennies would therefore appear to be permissible.

Prominent enforcement actions in 2015 and 2016, however, reveal that the SEC does not agree with such a literal interpretation of its regulations. Instead, the SEC interprets a "priced" order restrictively with respect to how dark pool customers can submit orders designed to execute in sub-pennies. In the 2015 action, the SEC documented how UBS—then the operator of the largest U.S. dark pool—permitted its subscribers to submit orders that were "pegged" to the NBBO plus or minus a specified percentage of the NBBO spread. Although the SEC noted a broker-dealer operating a dark pool "does not violate Rule 612 by accepting and ranking an order *pegged* to the midpoint of the NBBO (even if the midpoint is a sub-penny price)," these other forms of pegged orders "facilitated the very result that Rule 612 was designed to prevent: it allowed one subscriber to gain execution priority over another in the order queue by offering to pay an economically insignificant sub-penny more per share."[3] As such, UBS was fined for using this order type between 2008 and 2011 when the order type was discovered in an SEC examination. In the 2016 action, the SEC pursued Credit Suisse's dark pool for using a similar type of pegged order type during the same time period. Together, these two actions illustrate that while a broker-dealer can offer *de minimis* sub-penny price improvement to its customers directly, broker-dealers are prohibited from offering these opportunities to their dark pool clients—unless their clients are willing to trade at the midpoint of the NBBO.

The existence of this enforcement policy has significant implications for the theoretical and empirical literature examining how the MPV rule affects trading in dark venues relative to public

[3] In the Matter of UBS Securities LLC, Exchange Act Release 34-74060 (Jan. 15, 2015).

exchanges. For instance, Buti, Rindi, Wen, and Werner (2013) model how the current MPV rule can cause trades to occur in dark venues. Their model highlights the incentive of traders in dark venues to use limit orders with smaller tick sizes to undercut the displayed price in exchanges' limit order books that must display prices using wider ticks. Buti, Consonni, Rindi, Wen, and Werner (2015) (BCRWW) empirically test the model and find that sub-penny trading in non-exchange venues is increasing in a stock's displayed depth and decreasing in quoted spread. They interpret this finding as consistent with the model's predictions that traders in dark venues use sub-penny orders to undercut traders posting liquidity on exchanges. Kwan, Masulis & McInish (2014) (KMM) similarly document that a wider MPV, by increasing the queue-line for limit orders on exchanges, results in a greater share of trades occurring in dark pools, which they also attribute to the ability of dark pool traders to queue-jump exchanges by posting sub-penny orders that provide little or no price improvement.

These papers assess the formal MPV rule in the same manner that UBS did through 2011, insofar that Rule 612 would appear to allow clients of dark pools to execute trades at any price point within the NBBO.[4] However, the existence of this SEC enforcement policy constrains the dimension on which customers of dark pools can queue-jump exchanges: To the extent a wider MPV incentivizes traders to use dark pools to gain execution priority over exchanges, they will be limited to posting passive orders pegged to the midpoint of the NBBO, thereby providing price improvement in the form of the quoted half spread.

At the same time, while this enforcement policy nominally diminishes the risk that liquidity providers in dark pools "step ahead" of displayed liquidity on exchanges, it potentially enhances the level of adverse selection liquidity providers face within a dark pool. By definition, orders that are pegged to the NBBO obligate dark venues to calculate the NBBO when filled by a marketable order. Because the NBBO represents the best bid and best offer available across approximately one dozen exchanges, a dark pool can either accomplish this by calculating the NBBO directly through subscribing to direct data feeds from each venue or by relying on the slower, centralized Securities Information Processors (SIPs) to which all exchanges are required

[4] In their important paper, BCRWW acknowledge the SEC enforcement action against UBS for violating Rule 612. However, because this action indicates that sub-penny orders have been used in dark pools and because Rule 612 permits sub-penny price improvement by internalizers, they nevertheless model for tractability the ability of traders to post sub-penny limit orders in dark venues. This modeling assumption also appears grounded in an understanding that dark pools "do not have to comply with Rule 612 of Reg. NMS, which only refers to quoted prices." (BCRWW, p. 20). As we show, Rule 612 applies to all trading venues—whether lit or dark—which forms the basis for the SEC enforcement policy against dark pools.

to report updates to their best bids and offers. In either case, any latency in a dark venue's determination of the NBBO creates the possibility that traders with a comparative advantage in processing exchanges' data feeds (e.g., because of faster access to exchanges' quotations through direct feeds and co-location) can exploit midpoint orders priced using a dark venue's stale NBBO (Katsuyama, 2014). Indeed, it is precisely this type of "stale quote arbitrage" that is alleged to harm institutional investors in the widely-followed book *Flash Boys*.[5]

In this fashion, the SEC's policy of encouraging pegged midpoint orders might inadvertently have contributed to the controversial practice of stale quote arbitrage, thereby undermining the SEC's very aim of permitting investors to receive price improvement in non-exchange venues. At the same time, the rents from stale quote arbitrage can contribute to a socially wasteful arms race for trading speed (Budish, Cramton & Shim, 2015). The extent to which dark midpoint trades are "picked off" through stale quote arbitrage, however, remains contested within the industry (Tabb, 2014) and empirically unexamined.[6]

To understand better the role of the SEC's policy of encouraging midpoint orders in dark venues, we conduct two series of empirical tests using trade and quote data from the NYSE Trade and Quote (TAQ) files from 2011 through 2014. Because the TAQ data record all quote updates that exchanges make to the SIPs as well as all trades reported to them (whether made on or off exchanges), the TAQ data permit a comprehensive analysis of how a policy allowing midpoint peg orders affects off-exchange trading in light of the prevailing NBBO.

[5] As an illustration of this behavior, consider the following example given in Fox, Glosten & Rauterberg (2015). In it, an institutional investor posts to a dark venue a midpoint buy order for a security when the NBBO is $161.11 x $161.15 so that an incoming market order to sell would result in this order being filled at $161.13. However, if the exchange holding the best ask subsequently decreases its displayed quote from $161.15 to $161.12 while the midpoint order rests in the dark pool, a fast trader co-located at this exchange can detect the new NBBO before the dark venue, providing it a momentary opportunity to send an immediate-or-cancel sell order to the dark venue that will execute at the stale midpoint of $161.13. Upon receiving confirmation, the fast trader can cover the resulting short position by sending a marketable buy order to an exchange to execute at the new national best bid of $161.12, producing a penny of risk-free profit. In the meantime, the institutional investor—rather than buying at $161.115, the actual midpoint—buys at $161.13.

[6] Wah (2016) estimates the prevalence of latency arbitrage opportunities created by market fragmentation when two or more exchanges create a crossed market (i.e., when the best bid on one exchange creates a national best bid that is greater than the national best offer). Accordingly, her data excludes stale quote arbitrage arising from dark midpoint orders. More relevant to the question of stale quote arbitrage is Ding, Hann & Hendershott (2014). They study the latency between NBBO updates provided by the publicly-available SIP and NBBO updates calculated using direct data feeds to exchanges. Using a single trading day for Apple, Inc., they find that price dislocations average 3.4 cents and last on average 1.5 milliseconds. Based on these averages, they estimate that a fast trader could theoretically earn up to $32,000 over the course of the trading day by trading against stale midpoint orders in dark pools based on the volume of off-exchange trades. This estimate, however, assumes each off-exchange trade is made during a period of price dislocation. Using new timestamp data from the SIPs, Bartlett & McCrary (2016) study trades that occur during a period of price dislocation, but they find little evidence that fast traders initiate liquidity-taking orders to pick-off stale quotes priced at the SIP NBBO. Because their analysis focuses on trades at the national best bid and national best offer, however, they do not analyze the incidence of stale quote arbitrage surrounding midpoint peg orders.

In our first series of tests, we examine how the SEC enforcement policy interacts with the penny MPV to draw trading interest to non-exchange venues. To test whether the penny MPV induces queue-jumping in dark pools, we follow KMM and exploit the fact that Rule 612 requires all quotations priced at or above $1.00 per share to be priced in penny increments, but quotations below $1.00 per share may be priced in sub-penny increments as fine as a hundredth of a penny. Given that a sub-penny MPV should no longer constrain the bid-ask spread, Rule 612 accordingly permits a regression discontinuity (RD) design to examine the effect of the discontinuous increase in the incentive to queue-jump at the $1.00 cut-off.

Building on KMM, we begin with an RD analysis of the incentive to queue-jump at the $1.00 cut-off, using a sample of 796 securities that traded in the vicinity of the $1.00 cut-off from 2011 through 2014. By interleaving the quote and trade data, however, our analysis permits a direct analysis of whether the increase in market share of dark venues at the $1.00 cut-off is the result of midpoint trading compared to stepping ahead. Consistent with KMM, we first find a discontinuous increase in non-exchange trading when the MPV increases to $0.01. Yet we also find this result is driven by a discontinuous increase of approximately 12 percentage points at the $1.00 cut-off in the frequency of midpoint trading within non-exchange venues. In contrast, we find a marginal, discontinuous *decrease* in the percentage of non-exchange trades offering just *de minimis* price improvement over the NBBO. In combination, these findings are consistent with longer-queue lines encouraging traders to turn to dark pools, but only if they are willing to offer price improvement in the form of the quoted half-spread. These findings show that the prior narrative of *de minimis* price improvement fails to capture the nuance of actual market behavior in the current trading environment.

Turning to the question of whether a policy that encourages midpoint peg orders facilitates stale quote arbitrage, we similarly rely on the interleaved TAQ data to identify potential instances where a dark midpoint order might have been subjected to this form of latency arbitrage. In our empirical design, we leverage the fact that due to the SEC enforcement policy, any trade in our sample priced above $1 per share and that ends in five one-thousandths of a dollar (i.e, the “trailing 5s” of $0.005, $0.015, and so on) is highly likely to be the result of a midpoint peg order. Our analysis is further aided by the fact that the SIPs report non-exchange trades slower than they report quote updates comprising the NBBO. For instance, Bartlett & McCrary (2016) find that the time between when a trade occurs in a non-exchange venue and the time it is reported by a SIP has a mean (median) of approximately 100 (7) milliseconds. In contrast, the delay between a quote report occurring on an exchange and its dissemination on the SIP has a mean (median) of approximately 1.15 (0.5) milliseconds (Bartlett & McCrary, 2016).

Given these reporting delays, comparing the price at which a dark midpoint trade appears (or "prints") in the interleaved TAQ data to the midpoint of the TAQ-generated NBBO at the time of the trade report can provide insight into whether a midpoint trade was executed at a stale NBBO price. Specifically, a midpoint trade that prints at a price equal to the midpoint of the NBBO at the time of the trade report indicates a trade that did not occur when the NBBO had already transitioned to a new price, as even the SIP NBBO would have changed prior to the trade printing. In contrast, a mismatch between a midpoint trade's price and the midpoint of the prevailing SIP-generated NBBO presents a *potential* instance of stale quote arbitrage. To be sure, evidence of such trades could also reflect a simple delay in trade reporting. For example, the trade itself could induce a change in the NBBO that gets reported to the SIP prior to the trade. However, the incidence of these "mismatched" trades provides an outer estimate of the frequency with which midpoint trades are at risk of stale quote arbitrage strategies.

Applying this approach to all off-exchange midpoint trades in our sample, we find that for over 95% of all midpoint trades, the SIPs report the trade at a price equal to the midpoint of the SIP-reported NBBO disseminated at the time of the trade report. Given the considerably longer reporting latency for non-exchange trades relative to quote updates, the fact that the SIPs print nearly all dark midpoint trades in our sample while reporting an NBBO having the same midpoint price highlights how these midpoint trades typically execute during periods of price stability and, accordingly, create no opportunity for stale quote arbitrage. Moreover, midpoint trades that print at a price equal to the midpoint of the prior two NBBOs account for less than 2.7% of all non-exchange midpoint trades, even if we assume up to a 10 second delay in trade reporting. Overall, these results indicate that while an SEC policy favoring pegged midpoint orders can facilitate stale quote arbitrage, this form of latency arbitrage would appear to be far less prevalent than depicted in contemporary debates about market structure.

These findings emphasize the important role of SEC enforcement policy in shaping U.S. market microstructure, as well as the challenge it poses for market microstructure research. As this study illustrates, the SEC retains considerable discretion in interpreting and enforcing the complex array of rules that govern equity trading, highlighting the importance for empirical research of understanding law on the ground. Unlike formal rule-making, however, the development of SEC enforcement policies occurs in ways that are considerably less transparent and which can evade public analysis and commentary for potentially extended periods of time. Indeed, the fact that prominent papers examining stepping ahead in dark pools emerged after the SEC first sanctioned UBS in 2011 but before the agency published its 2015 enforcement action represents a telling example of the challenge this discretion poses for analyzing the *de facto* regulatory regime governing U.S. equity markets.

Our results also have implications for several policy debates concerning U.S. market microstructure. First, our finding that traders use the current MPV rule to queue-jump exchanges by means of providing midpoint liquidity highlights an important limitation of the SEC's current pilot to increase the MPV for certain securities. In October 2016, the SEC commenced a two-year pilot tick size program that widens the MPV from a penny to a nickel for select companies with the goal of encouraging market-making in the securities of these firms. Due to concerns that the wider MPV will induce queue-jumping in dark pools, a component of the pilot includes a controversial "trade-at" rule, which prohibits a venue from filling an incoming order unless the venue is displaying the NBBO. In so doing, the trade-at rule strives to provide a window into how queue-jumping affects the provision of displayed liquidity; however, it exempts from its application any trades executed at the midpoint of the NBBO. Our finding that queue-jumping occurs primarily by means of midpoint trading suggests the proposed trade-at rule is unlikely to stem the flow of trading from exchanges to dark venues and represents a missed opportunity for examining how dark midpoint trading affects the incentive to provide displayed liquidity. This is an example of how observational work can and should inform experimental design.

Finally, our findings have immediate implications for resolving some of the controversy emanating from the 2014 publication of *Flash Boys* regarding whether markets are "rigged" in favor of high speed traders. Notwithstanding its widely followed claims regarding the prevalence of stale quote arbitrage, our findings indicate that, at least with respect to traders placing midpoint orders in dark pools, the probability of being subjected to stale quote arbitrage is likely to be far lower than depicted in Lewis' narrative and in debates about market structure that have followed in its wake. At the same time, by selectively enforcing the MPV rule to prohibit stepping-ahead in dark pools and not in other dark venues (e.g., internalizers), the SEC enforcement policy we document highlights how the SEC itself can operate to favor certain market participants over others.

The remainder of this paper is organized as follows. Section 2 provides institutional details regarding Rule 612 and the SEC's enforcement policy concerning sub-penny trading in dark pools. Section 3 analyzes whether the increase in non-exchange trading for orders priced at more than $1.00 per share reflects midpoint trading or stepping ahead. Section 4 tests the claim that midpoint orders place traders at a significant risk of being subjected to stale quote arbitrage. Section 5 concludes.

2. Institutional Details

Promulgated as part of Reg. NMS, Rule 612 reflects the SEC's desire in 2005 to rectify what it perceived to be several adverse consequences arising from the decimalization of stock prices in 2001. Because the original decimalization order did not specify a precise decimal increment, trading venues differed after decimalization in how they permitted market participants to submit decimalized orders. For instance, while the major exchanges and Nasdaq required orders to be priced in penny increments, several electronic communication networks (ECNs) permitted quotations in sub-penny increments, with many of these quotations resulting in trades at the $0.001 and $0.009 price points.

For the SEC, the frequency of these sub-penny orders was troubling. As it noted in proposing Reg. NMS, the prevalence of these orders suggested "that much of the trading that currently takes place in sub-pennies is the result of market participants attempting to step ahead of penny-priced limit orders for the smallest economic increment possible."[7] In adopting Reg. NMS, it further noted that "[i]f investors' limit orders lose execution priority for a nominal amount, investors may over time decline to use them, thus depriving the markets of liquidity."[8] Accordingly, by prohibiting the "display, rank or accept[ance]" of an order for any NMS stock priced in an increment smaller than $.01 for orders priced at $1.00 per share or more, the SEC concluded that "Rule 612 will deter the practice of stepping ahead of exposed trading interest by an economically insignificant amount."

[7] See Securities Exchange Act Release No. 49325 (Feb. 26, 2004), 69 FR 11125 (Mar. 9, 2004).
[8] See Securities Exchange Act Release No. 51808 (June 9, 2005), 70 FR 37495 (June 29, 2005) (the "Reg NMS Adopting Release"). The Reg NMS Adopting Release is the source of all subsequent quotations in this paragraph.

In keeping with its separate commitment to minimize trading costs, however, the Commission also declined to prohibit sub-penny *trades*. By 2005, many exchanges and non-exchange venues had systems in place for facilitating trades at the midpoint of the NBBO. Moreover, brokers also routinely internalized customer orders, whereby broker-dealers fill incoming market orders from retail investors either as an agent matching their customers' buy and sell orders or as a principal taking the other side of those orders.[9] For example, if the market for a security stands at $10.00 x $10.05, a broker receiving from a customer a market order to buy the security could do one of two things. First, she could route the order to the venue offering to sell at $10.05. Second, the broker could simply sell the customer the same stock at a price that is $10.05 or lower (e.g., $10.04). Indeed, a broker choosing to internalize such an order will often provide such price improvement over the NBBO to comply with her best execution obligations, providing a common justification for the practice.[10]

In permitting sub-penny trades, the SEC noted that "every commenter that addressed [sub-penny trades] agreed that Rule 612 should permit sub-penny trades that result from midpoint and average-price algorithms."[11] The agency further noted that "sub-penny executions due to price improvement are generally beneficial to retail investors." Accordingly, it emphasized that, in light of Rule 612's silence regarding the price of executed trades, "Rule 612 will not prohibit a sub-penny execution resulting from a midpoint or volume-weighted algorithm or from price improvement, so long as the execution did not result from an impermissible sub-penny order or quotation."

---

[9] To be sure, internalization is potentially problematic for a broker-dealer seeking to internalize an order if that broker also holds a customer limit order on the same side of the market for the same security, because in that case the broker would be competing with her customer. Under Finra Rule 5320 (the "Manning Rule"), a broker-dealer holding such a limit order "is prohibited from trading that security on the same side of the market for its own account at a price that would satisfy the customer order." To avoid this conflict of interest, brokers often sell market orders (but not limit orders) to dedicated broker-dealer internalizers under payment-for-order flow arrangements.

[10] As discussed in Ferrell (2001), the pressure to provide price improvement over the NBBO arose in large part due to the Third Circuit's decision in *Newton v. Merrill, Lynch, Pierce, Fenner & Smith, Inc.*, 135 F.3d 266 (3d Cir. 1998), where the Third Circuit found that a broker-dealer that automatically executed customer trades at the NBBO may not be in compliance with its best execution obligations. Additionally, the manner in which Reg. NMS discussed the desirability of brokers' providing price improvement for their customers has also created a perception within the industry that best execution may require a broker to seek out opportunities for customer price improvement. In a comment letter to the SEC outlining how internalizers can often be subject to significant market risk when trading with their customers, TD Ameritrade (2010) articulated this perception: "One could certainly suggest that the [market-maker] simply avoid the price improvement opportunity and that the market maker or broker should have simply sent the order to fill at the NBBO. In such case, however, the broker would run the risk of being accused of violating its best execution obligation, as Regulation NMS elevated price improvement above all else." Finally, the incentive for offering price improvement over the NBBO is also encouraged by Rule 605 of Reg. NMS, which requires that broker-dealers publicly disclose their rate of price improvement over the NBBO as a core measure of execution quality.

[11] See Reg NMS Adopting Release. The Reg NMS Adopting Release is the source of all subsequent quotations in this paragraph.

Moreover, following the promulgation of Rule 612, the SEC also released several "Frequently Asked Questions Concerning Rule 612" which further clarified that the test for a permissible order was whether the order was "explicitly priced" in sub-pennies, leaving open the possibility for orders pegged to the NBBO that can result in sub-penny trades.[12] Indeed, the agency subsequently approved in 2006 a "midpoint match" order for the International Stock Exchange on the basis that it was an "unpriced" order to trade at the midpoint of the NBBO. It similarly noted in granting a limited exemption from Rule 612 to Liquidnet, Inc. in 2006 that "[i]n a typical midpoint trade by a dealer" the trade occurs at the midpoint of the NBBO "without there being an order or quotation priced in sub-pennies."[13]

This emphasis on Rule 612's prohibition on orders that are "explicitly priced" in sub-pennies undoubtedly explains in part the subsequent use through 2011 of the impermissible pegged orders at dark pools operated by UBS and Credit Suisse. Consistent with Rule 612, both venues had protocols that rejected orders explicitly priced in sub-pennies, relying instead on orders whose price would be determined from the NBBO. For instance, CrossFinder, the dark pool operated by Credit Suisse, permitted customers to enter an order with an execution price specified as an adjustment to the national best bid (NBB) or national best offer (NBO) that was either a penny or 10% of the spread of the NBBO. In its enforcement action against the dark pool, the SEC concluded that, because of this order type, "Crossfinder accepted and ranked orders priced in increments smaller than one-cent ('sub-penny orders') in violation of Rule 612."[14]

UBS similarly permitted customers to enter orders using a "PrimaryPegPlus" (PPP) feature in which the execution price was pegged to the NBB or NBO, "plus or minus a subscriber-entered percentage of the 'spread.'"[15] As in its case against CrossFinder, the SEC found such orders to violate Rule 612 "[b]ecause the second component of the formula determining the price of a PPP order … nearly always yielded a sub-penny amount." Remarkably, the SEC noted elsewhere in the same action that a dark pool "does not violate Rule 612 by accepting and ranking an order *pegged* to the midpoint of the NBBO (even if the midpoint is a sub-penny price)" (emphasis in

[12] See Division of Market Regulation: Responses to Frequently Asked Questions Concerning Rule 612 (Minimum Pricing Increment) of Regulation NMS, available at https://www.sec.gov/divisions/marketreg/subpenny612faq.htm.

[13] See Order Granting Exemption to Liquidnet, Inc. from Certain Provisions of Rule 612 of Regulation NMS under the Securities Exchange Act of 1934, SEC Release No. 34-53193 (Jan. 30, 2006).

[14] See In the Matter of Credit Suisse Securities (USA) LLC, Exchange Act Release 34-77002 (Jan. 31, 2016).

[15] See In the Matter of UBS Securities LLC, Exchange Act Release 34-74060 (Jan. 15, 2015). This document is the source of all quotations in this and the subsequent paragraph.

original), revealing an obvious inconsistency in declaring PrimaryPegPlus orders "priced" in prohibited sub-penny increments.

In both enforcement actions, the SEC emphasized the broader goal of deterring stepping ahead. As it stated in the UBS order, "the PPP order type facilitated the very result that Rule 612 was designed to prevent: it allowed one subscriber to gain execution priority over another in the order queue by offering to pay an economically insignificant sub-penny more per share." Yet even this justification only highlighted how the contours of Rule 612 have become subject to the SEC's interpretive discretion. Had either UBS or Credit Suisse simply chosen to fill its in-bound marketable orders with *de minimis* price improvement (rather than offer such trades to the customers of their affiliated dark pools), such behavior would presumably have been justified as permissible internalization.

Regardless of whether this policy represents a consistent approach to "stepping ahead," the end result for dark pools has been to limit their customers to using midpoint peg orders when they wish to trade in sub-pennies, potentially explaining the important role of midpoint trading for these venues. According to Tabb Group (2015), for instance, prominent dark pools such as Barclays DirectEx, and BIDs report more than seventy percent of their trades are done at the NBBO midpoint. Moreover, for purposes of understanding the competition between dark pools and exchanges, an enforcement policy that encourages midpoint orders can also help explain why liquidity takers might first look to dark pools before routing to exchanges: While an order to a dark a venue raises execution risks compared to the certainty of accessing an exchanges' displayed liquidity, the greater likelihood of receiving price improvement in the form of the quoted half-spread can provide a potentially offsetting benefit for trading in these venues (Buti, Rindi and Werner, 2016).

3. Dark Trading and the MPV

3.1 Sample Construction

To analyze the effect of the MPV on queue-jumping and the use of midpoint orders, we use the NYSE's daily TAQ files. The TAQ data is obtained directly from the two SIPs responsible for collecting and disseminating real-time trade and quote information from all exchanges and FINRA members pursuant to Section 11A of the Securities Exchange Act of 1934. As such, the dataset reflects the two distinct reporting systems that apply to trades and quotes for all exchanged-listed securities.

Pursuant to the Consolidated Tape Association (CTA) Plan and the Unlisted Trading Privileges (UTP) Plan, all U.S. exchanges and FINRA are obligated to collect and report to the appropriate SIP last sale data in all exchange-listed securities. For transactions in securities listed on the NYSE, the NYSE Mkt (formerly the American Stock Exchange) and any regional exchanges, the appropriate SIP is the NYSE; for securities listed on Nasdaq, the appropriate SIP is Nasdaq. Trade reports recorded by the SIPs are recorded in the TAQ daily Consolidated Trade File. Although the SIP data does not directly record the identity of non-exchange participants reporting a trade, the SEC has required since March 2007 that all off-exchange transactions be reported to a formal FINRA-managed Trade Reporting Facility (TRF) established at certain stock exchanges which report directly to the appropriate SIP. As described by O'Hara and Ye (2011), this requirement means that off-exchange trades made through a broker-dealer internalizer or in a dark pool (both of which were historically reported to an exchange and then consolidated with the exchanges' own trades when reported) are now effectively segregated and reported to the appropriate SIP as having been executed at a FINRA TRF.

In addition to the Consolidated Trade File, TAQ also includes a daily Consolidated Quote File that records historical quotation data reported separately to the SIPs. As with their trade reporting obligations, all exchanges and FINRA are required to report to the appropriate SIP any change in the best bid and best offer (including aggregate quotation sizes) currently available on each trading venue.[16] The Consolidated Quote File thus provides for any moment of the trading day a snapshot of the total, consolidated trading interest at the best bid and offer ("Consolidated BBO") available at each exchange and through a FINRA member. We use TAQ's Consolidated Quote File to calculate the NBBO over the course of each trading day for every security in our sample.

In light of the research questions posed in this paper, the specific TAQ files we use often vary. Some of our analyses rely exclusively on the Consolidated Trade File, others rely exclusively on the Consolidated Quote File, and yet others require that we interleave the two files (i.e., align them in chronological order for the same security). For some of our analyses, it

[16] FINRA operates an Alternative Display Facility, through which venues (such as an ECN) might choose to disseminate quotations from their subscribers. At present, no venue disseminates any quotations through the FINRA ADF.

is additionally necessary to classify trades as having been buy- or sell-side initiated. We follow much of the literature and use the Lee and Ready (1991) algorithm to do so.[17]

Because we are interested in how the change in MPV at $1.00 affects dark trading, we limit our sample to trades and quotes that are priced at less than $4.50 per share during the four-year period spanning 2011-2014.[18] To ensure that all quotes and trades occur during the trading day after the opening cross and before the closing auction, we also filter the TAQ data to exclude quotes and trades occurring before 9:45:00.000 and after 15:35:00.000. Since the only identifier for securities in the TAQ data is ticker symbol, which does not uniquely identify securities due to the retirement and recycling of ticker symbols, and also because TAQ contains some ticker symbols that do not correspond to actual securities, we further limit our sample to those ticker symbols that could be matched to a CRSP record on a day-by-day basis.[19]

Finally, given that many of these securities trade at spreads wider than the penny MPV, we focus our analysis on those securities where the penny MPV is most likely to be a binding constraint. Visual inspection of the data indicated that roughly the third most liquid securities traded at penny spreads; therefore, we restrict the sample to those securities that fell within the top tercile of traded securities based on average daily trading volume.[20]

With these restrictions imposed on the TAQ data, the final sample contains 793 securities that, over 2011-2014, were associated with roughly 776 million trades and 8 trillion updates to venues' Consolidated BBO.[21] Average quoted (midpoint) spreads were approximately $0.011

---

[17] A challenge with the analyses involving interleaving, including analyses using the Lee-Ready algorithm, is that the timestamps in the two files are not perfectly synchronized, as is widely recognized in the literature. Since our analyses are from recent years, we follow the recent literature in assuming that trades occur contemporaneously with quotes (e.g., Bessembinder and Venkataraman 2010).

[18] To efficiently access those records, and in particular to take advantage of the index structure of the TAQ files as stored on Wharton Research Data Services (WRDS) where we performed the bulk of our computations, we accomplish this subsetting in a multi-step procedure. We first identify the subset of CRSP universe securities that had a closing price of below $4.50 at some point during 2011-2014, using the CRSP dsf file. We then used the CRSP dsenames file to identify the corresponding ticker symbols on a day-by-day basis. We then constructed time series plots of each identified security over time, and verified that the CRSP data on closing price was in tight agreement with TAQ end-of-day prices (or bid-ask midpoint when trade prices were missing, consistent with CRSP measurement protocols). We then pulled extracts of all trades and quotes for those securities from the full TAQ data, taking advantage of the index structure using key merging. Finally, we restrict our attention to prices and NBBO values that are below $4.50. In addition to being computationally efficient, this ensures that all of our securities are true securities, as opposed for example to the test securities that are present in the TAQ data and about which documentation is uneven.

[19] When calculating the NBBO, we additionally restrict our analysis to those quotations that are eligible to establish an exchanges' consolidated BBO (i.e., quotation updates having a condition of A, B, H, O, R, or W).

[20] Securities having an average daily trading volume of 336,000 qualified to be in the sample.

[21] The 8 trillion BBO quote updates correspond to just over 6 trillion NBBO updates.

($0.006), indicating that the penny MPV generally represents a binding constraint for traders seeking to display liquidity for these securities.[22]

3.2 Regression Discontinuity Design

We follow KMM in adopting an RD research design to examine the effects of the MPV on non-exchange trading, though our methodology differs in several respects. As noted by Hahn, Todd and van der Klaauw (2001, p. 1), the "regression discontinuity data design is a quasi-experimental data design with the defining characteristic that the probability of receiving treatment changes discontinuously as a function of one or more underlying variables" (p. 1). The MPV rule fits nicely within this data design on account of the sharp regulatory distinction involving the MPV created by Rule 612. Under this rule, the MPV regulation that applies to any given trading order varies sharply: an order is allowed to be posted in below penny increments if and only if the order is priced less than $1.00.[23]

Using this discontinuous treatment of MPV regime, we develop the following baseline model to evaluate the effect of changing the MPV on the trading environment by measuring directly the conditional expectations of market measures given two-decimal prices, or $E[\textit{Market Measure}_i \mid \textit{Price}_i]$, where $\textit{Market Measure}_i$ is an outcome for security-time i and $\textit{Price}_i$ is the running variable, or price truncated to two decimals (e.g., $0.98, $.0.99, $1.00, $1.01, etc.). The flexibility of this baseline model allows us to apply the RD design to the actual quotation data to which the MPV rule pertains, as opposed to trade data as used by KMM. For example, in our core results we define "security-time" as one of our sample securities during a given second and "price" as the NBB for that security as of the beginning of the second.[24] We specify below in our empirical results the particular definition of "security-time" and "price" being adopted for each analysis.[25]

[22] That is, a trader seeking to submit a competitive buy (sell) order will be required to submit the order at the national best bid (ask) as any price that is more aggressive than the national best bid will lock the market, causing the order to be rejected or converted into a marketable buy (sell) order.
[23] In unreported results, compliance with this regulatory rule is complete—all quotes above $1.00 are made in penny increments, whereas it is common for quotes below $1.00 to be made in hundredths of a penny.
[24] We use the NBB in our empirical analyses for ease of exposition. Our results, however, are robust to using the NBO.
[25] An important point to note in analyses using the Consolidated Quote Data is that reported data include only updates to exchanges' Consolidated BBO. As such, randomly selecting reported data from this file would not yield a randomly drawn quotation for a security at a moment in time, and rather would yield an oversample of the most liquid securities. To ensure that our analyses of daily data and data relying on the intra-day quotation data (e.g., the NBBO) use the same sampling scheme, we utilize all intra-day quotation records, but generate a weighting scheme such that our intra-day analyses of the quoting environment can be thought of as reflecting randomly sampled seconds or milliseconds for our sample securities.

Finally, while the sharp cut-off in the MPV at $1.00 per share makes Rule 612 a candidate for an RD design, two issues relating to the trading environment around the $1.00 cut-off could potentially bias our results. First, all major U.S. exchanges impose a $1.00 minimum bid price requirement for continued listing. Second, both trades and quotes are known to cluster at price increments of five and ten cents (nickels and dimes) as described in Ikenberry & Weston (2008). By distinguishing the $1.00 price point from other prices, both issues potentially violate the core underlying assumption of the RD design of "smoothness," or continuity of potential confounders given the running variable. We formally evaluate these issues in Appendix A and conclude that neither invalidates the use of the RD design and that smoothness is likely to be satisfied.

3.3 Empirical Results

We present our empirical results in two stages. First, we replicate KMM's primary finding in our data, which cover more securities for a longer period of time. Our analysis confirms that increasing the MPV to a penny at the $1.00 cut-off enhances the incentive for traders to engage in so-called "queue-jumping" in dark venues. Second, moving beyond KMM's findings, we examine whether traders queue-jump in dark pools using midpoint orders or through stepping ahead.

3.3.1 MPV and the Incentive to Queue-Jump in Dark Venues

As noted previously, a wider tick size can affect queue jumping by forcing traders who post limit orders to join long queue-lines at the NBBO. For example, a trader looking to post a bid at the displayed NBB when a stock trades at $0.994 x $0.9999 will face a shorter queue of limit orders than if the same stock had a penny MPV. The reason arises from the simple fact that a penny MPV would require the NBBO to be priced at $0.99 x $1.00, pushing any traders willing to buy at prices between $0.99 to $0.994 to post bids at $0.99. For an investor looking to avoid paying the spread, this heightened risk of non-execution when posting a limit order at the NBBO can increase the attractiveness of midpoint trading within a dark venue.

To examine whether this in fact occurs, we first analyze whether a wider tick size results in longer queue lines at the NBBO. Figure 1 plots for the sample securities the average log quoted depth at the end of a millisecond as a function of the two-decimal NBB for that millisecond.[26] This figure contains a number of distinctive features that are specific to the RD context and

[26] We utilize log quoted depth because of the long right-hand tail of the quoted depth distribution.

deserve explanation. First, quoted depths for each two-decimal NBB are presented as open circles of different colors. Dollars are displayed in red, quarters are displayed in blue, nickels and dimes are displayed in light gray, and all other price points are displayed in dark gray. This color scheme relates to the potential distinctiveness of these price points due to the phenomenon of price clustering discussed previously. Second, a black vertical line is superimposed at the actual cutoff of $1.00, and green vertical lines are superimposed at "placebo" cutoffs of $2.00, $3.00, and $4.00. These vertical lines can be used to assess visually whether the evidence of a discontinuity at $1.00 stands in contrast to the pattern around other dollar points, where no policy changes discontinuously.

[Insert Figure 1]

As the figure shows, quoted depth at the NBB increases discontinuously above the $1.00 cut-off. Table 1 provides point estimates of this discontinuity. Because this is our first reported result, and because the structure of Table 1 applies to our other core results discussed below, we pause here to describe in some detail our methodological approach and choices before returning to a substantive discussion of these results.

The first column in Table 1 presents a basic local linear regression estimate of the discontinuous change in quoted depth at the $1.00 cut-off. We use the triangle kernel since that is known to be boundary optimal (for references to this idea, see McCrary 2008) and a bandwidth of fifty cents. Column 2 presents an alternative approach that eliminates dollar, quarter, dime, and nickel price increments to address price clustering. The models in Columns 3 and 4 are the same as those in Column 1, but assess robustness to alternative bandwidths.[27] For all four models, we report HC3 standard errors in parentheses below point estimates, as is appropriate for the RD context (cf. Angrist and Pischke, 2009, Chapters 3 and 8; MacKinnon and White, 1985; and Hausman and Palmer 2011). Additionally, because even the conservative approach

[27] Column 3 uses a narrow bandwidth of twenty-five cents, and Column 4 uses the data-driven bandwidth of Imbens and Kalyanaraman (2012). The tension between the models in Columns 1 and 3 is the same as the familiar tension between bias and variance: a narrower bandwidth is closer to the regression discontinuity ideal of "in the limit" and potentially is less biased, yet a narrower bandwidth yields a more noisy estimate. A further wrinkle is that the accuracy of standard errors becomes more suspect with bandwidths that are too small. There is also tension between the models in Columns 1 and 4. In our data, fifty cents is a generally robust choice of bandwidth and has the important result of being comparable across outcomes and transparent. However, this transparency is not without costs. We examine curves of a variety of shapes and variances, and for some of these a bandwidth of fifty cents is probably too wide and for others it is probably too narrow. Column 4 seeks to replace transparency with a reliable data-driven method. While this method generally performs well, it also occasionally yields bandwidth choices and point estimates which human judgment squarely rejects. Overall, we have a mild preference for the model in Column 1, but we think it is important to consider what each of the four models tells us about the data.

recommended by Angrist and Pischke (2009) may lead to spurious rejection, we complement the standard reporting of standard errors with the reporting of a randomization inference p-value.[28] This latter value is reported in brackets below the standard error.[29] All four specifications confirm the visual impression from Figure 1: Quoted depth increases significantly and discontinuously at the $1.00 cut-off. More specifically, considering that average quoted bid depth just below $1.00 was approximately 7 round lots (i.e., e^2), estimates from Column 1 (our preferred bandwidth) indicate that quoted depth at the NBB increases by 87 lots—a nearly twelve-fold increase—when priced at $1.00 relative to when the two-decimal NBB is priced at $0.99.

[Insert Table 1]

Given these longer queue lines in the penny quoting regime, we next analyze whether the penny MPV is also associated with a greater incidence of non-exchange trading as measured by trade executions reported to a FINRA TRF (exchange code of "D"). This analysis requires us to examine the venue of a trade execution (tracked in the Consolidated Trade File) as a function of the quoting regime (tracked in the Consolidated Quote File); therefore, we use the interleaved trade and quote data. A practical challenge in utilizing the interleaved data, however, is the need to assign trades to the penny or sub-penny quoting regime. We address this challenge by assuming that the NBB in effect for a security at the beginning of a trading second (a "security-

[28] In Appendix A, we demonstrate empirically how HC3 standard errors can falsely reject smoothness across the $1.00 cut-off. Such spurious rejection can arise in an RD setting given that even with large datasets such as ours, the focus of the econometric modeling is fitting a curve to perhaps a few dozen precisely estimated averages that do not necessarily follow a linear trend. As such, model misspecification dominates as the sample size grows, rendering the sample effectively small despite the abundance of microdata. As shown in a pair of important recent papers, Cattaneo, Frandsen, and Titiunik (2015, "CFT") and Sales and Hansen (2015, "SH"), drawing on randomization inference approaches rather than standard Wald approaches places inference in the RD context on a stronger footing. Randomization inference avoids asymptotic approximations by contemplating a distribution of the test statistic of interest under the null hypothesis obtained in a particular way: the outcomes are held fixed at their observed values, treatment indicators for the sample are chosen at random holding fixed the number of treated, and the test statistic of interest is calculated for each such shuffling of the treatment indicators. CFT and SH both extend these notions to the RD context, but differ somewhat in their implementation. CFT assumes that the indicator for being to the right of the cutoff can be treated as if assigned randomly in a neighborhood of the cutoff, leading to an assumption that that indicator is independent of counterfactual outcomes. The approach taken in SH weakens the assumption of unconditional independence to conditional independence given the running variable. We follow the approach of SH.

[29] These p-values are computed by engaging in 1000 permutations of the treatment indicator and calculating the fraction of F-statistics obtained from those permutations that are more extreme than the F-statistic for the actual data. A p-value below 0.05 is interpreted as evidence of a discontinuous function, whereas a p-value above 0.05 is consistent with the function being smooth.

second") represents a good estimate of the NBB in effect for the duration of that security-second.[30]

Using this approach, we calculate for each security-second the number of trades reported to a FINRA TRF as a function of the two-decimal NBB prevailing at the beginning of that security-second. Figure 2A provides a scatterplot of the results. As the figure shows, TRF-reported trades occurring when the NBB was at or above $1.00 per share reveal a sharp increase relative to TRF-reported trades below this price per share. To examine how this phenomenon affects the market share of trading between exchanges and non-exchange venues, Figure 2B plots the fraction of all trades in a second that are reported to a FINRA TRF as a function of the two-decimal NBB prevailing at the beginning of that security-second. Similar to Figure 2A, the fraction of trades that occur off-exchange reveals a large, discontinuous increase at the $1.00 cut-off. Table 1 provides point estimates of both discontinuities. Overall, these results are consistent with the findings of KMM concerning the effect of a wider tick size on the incidence of dark trading.[31]

[Insert Figure 2]

3.3.2 The Vehicle for Queue-Jumping: Midpoint Trades or Stepping Ahead?

While Figures 1 and 2 are consistent with claims that wider tick sizes enhance the incentive to use Rule 612's trade exclusion to trade in dark venues, neither speaks to whether the primary mechanism for this result is due to stepping ahead, as has generally been assumed. To examine this question, we analyze the incidence of three types of trade executions that occurred away from exchanges during the sample period.

First, for non-exchange trades we estimate the incidence of pricing exactly at the midpoint of the NBBO. We assume that such trades reflect the execution of midpoint peg orders that are designed to be filled at the midpoint of the NBBO.

Second, for non-exchange trades we estimate the incidence of pricing with exactly $0.0001 of price improvement over the NBBO, using the Lee and Ready (1991) algorithm to classify

[30] To examine the reasonableness of this proxy, we regress the NBB for a security-*millisecond* against the NBB as of the beginning of the associated security-second. The regression results (unreported) yield a constant of 0.0010, a slope coefficient of 0.9993, and an R-squared of 0.9957. These results confirm that the NBB as of the beginning of a security-second was virtually a perfect predictor of any randomly drawn NBB for the duration of that second. Accordingly, we use the NBB as of the beginning of a security-second as our proxy for the NBB that applied to all trades executed over the course of that second.

[31] These results are also consistent with the findings of BCRWW that dark trading is increasing in quoted depth and decreasing in quoted spreads.

orders as buy or sell initiated.[32] As noted previously, we refer to this type of trade as "stepping ahead" given that the trade was filled by a liquidity provider in a dark venue offering *de minimis* price improvement over the NBBO. We focus on this level of price improvement for two reasons. First, this fixed dollar amount of price improvement represents roughly the same economic value of a trade for transactions priced at $1.00 per share as those priced at $0.99, making the measure robust to changes in the tick size at $1.00. Second, commentators that are critical of queue-jumping in dark venues have commonly focused on the use of off-exchange trades that offer just $0.0001 of improvement over the NBBO as evidence that queue-jumping provides few pricing benefits to liquidity takers (see, e.g., Buti, Rindi, Wen and Werner (2013); Dick (2010)).[33]

Finally, we estimate for non-exchange trades the incidence of trades priced at exactly the NBBO (again, based on the direction of the trade). These trades offer no price improvement to liquidity takers. On the contrary, these trades reflect the execution of marketable orders in dark venues at the NBBO that would otherwise have been executed on exchanges at the NBBO. Therefore, they simply redistribute trading gains from providers of displayed liquidity on exchanges to providers of non-displayed liquidity in dark venues. Similar to the prior category of trades, a jump in this form of trading at the $1.00 cut-off would also indicate that liquidity providers in dark pools were intercepting marketable orders that would otherwise go to exchanges without providing any form of price improvement. Given well known limitations of the Lee and Ready (1991) algorithm with regard to mismatching trades to the appropriate NBBO (Holden and Jacobsen, 2011), we limit our analysis in all three cases to trades that occurred at or within the prevailing NBBO and to trades where the NBBO was neither locked nor crossed.

Figure 3 plots the frequency of each form of off-exchange trading as a function of the two-decimal NBB that applied to each trade. Figure 3A presents our core finding regarding the frequency of midpoint trading at the $1.00 cut-off. As the figure reveals, the percentage of non-exchange trades that were midpoint trades demonstrates a sharp, discontinuous increase as the NBB crosses above the $1.00 threshold, highlighting a dramatic change in the incidence of midpoint trading between the sub-penny and penny quoting environments. Table 1 reports point estimates of this change at the $1.00 cut-off; these indicate a discontinuous increase of

[32] Where quoted spreads were $0.0002, we classified a trade with $0.0001 of price improvement as a midpoint trade.
[33] Our results are also robust to using any amount of price improvement up to $0.005 per share (or one-half of a penny spread).

approximately 12 percentage points in the frequency of midpoint trading among non-exchange trades. In combination with the sharp change in quoted depth presented in Figure 1, Figure 3A is consistent with traders in the penny quoting environment opting to seek midpoint executions in dark venues rather than posting orders to exchanges given the longer queue lines at the NBBO for orders priced at or above $1.00 per share.

[Insert Figure 3]

Figures 3B and 3C provide similar analyses regarding the percentage of non-exchange trades that provide little or no price improvement. In contrast to Figure 3A, both figures reveal the opposite result with respect to the frequency of these trades among all non-exchange trades in our sample. Indeed, point estimates in Table 1 reveal a discontinuous *drop* just above the $1.00 cut-off in the percentage of non-exchange trades reflecting each form of trading behavior, particularly with respect to trades offering just $0.0001 of price improvement.

Given the discontinuities in Figures 3B and 3C, it is worth emphasizing that Figure 3 presents the *relative* frequencies of all three forms of dark trading within our sample, as opposed to absolute frequencies. This is important, because as discussed above, Figure 2 also demonstrates an increases in FINRA trades above the $1.00 cutoff. In absolute terms, combining these results with our point estimate showing a discontinuous increase of 0.0132 FINRA trades per second at the $1.00 cut-off suggests a discontinuous increase at the $1.00 cut-off of 172%, 32% and 42% trades per second for midpoint trades, trades that step ahead and trades at the NBBO, respectively. In this regard, our findings are consistent with those of BCRWW who similarly find a positive association between a stock's quoted depth and the incidence of sub-penny trades that do not represent midpoint executions.

Our results differ, however, insofar that by juxtaposing the relative frequencies of all three forms of dark trading, we show that the previously documented increase in queue-jumping at the $1.00 cut-off is driven primarily by the increase in midpoint trading rather than stepping ahead. Moreover, the existence of the SEC enforcement policy would also suggest that our evidence of stepping ahead reflects price improvement offered by internalization pools and market making desks rather than sub-penny limit orders submitted by a dark pool's customers.[34] In short and in contrast to the conventional wisdom, these results demonstrate that dark pool customers seeking to bypass long queue lines associated with a wider MPV do so through turning to midpoint trading rather than through stepping ahead.

3.3.3 Robustness Check: Maker-Taker Fees

As a further check on this finding, we also examine how our results might be affected by rebates paid by exchanges for providing displayed liquidity and to fees charged to traders for taking this liquidity. To draw liquidity providers to a venue, exchanges commonly pay liquidity providers a small rebate for every share they sell on an exchange and assess liquidity takers a slightly larger "access fee" for taking it. A small number of exchanges (e.g., BATS Y, Nasdaq BX) also utilize so-called "inverted pricing" whereby the reverse payment structure applies. In these cases, liquidity takers receive a rebate for every share purchased on the venue, and exchanges charge a slightly larger fee to the liquidity provider who provided the liquidity. While the size of rebates paid to liquidity providers is unregulated, Reg. NMS limits access fees for transactions priced above $1.00 per share to 30 cents per hundred shares while limiting access fees for all other transactions to 0.3% of the transaction value. As such, exchanges during the sample period routinely used different pricing regimes depending on whether a transaction was priced above or below $1.00 per share.

To examine whether this differential treatment of fees and rebates might affect our results, we collected data on all exchanges' fee and rebate programs from 2011-2012. During this time period, exchanges competed vigorously for trading volume on the basis of their maker-taker pricing, commonly changing their fees and rebates on a weekly basis. This competition was also characterized by considerable experimentation, with some exchanges flipping from standard to inverted pricing structures either entirely or for select securities. We exploit these time-varying changes in maker-taker fees to explore the sensitivity of our results to these pricing programs. Our analysis, which we set forth in Appendix B, demonstrates that even after controlling for fees and rebates, our results remain consistent with those obtained without regard to maker-taker pricing.

[34] BCRWW likewise note the possibility that sub-penny trades can reflect price improvement provided by internalizers. However, as their empirical study relies on data from 2010 when stepping ahead was documented to have occurred at UBS, they posit their findings may reflect the use of sub-penny limit orders, consistent with their formal model.

4. Midpoint Trading and Stale Quote Arbitrage

Given the critical role of midpoint orders in non-exchange trading, an important question is whether these orders facilitate adverse selection in the form of stale quote arbitrage and, with it, an associated arms-race for trading speed. Indeed, the possibility that a delay in re-pricing a "pegged" order can allow fast traders to profit from stale midpoint orders has played a central role in stirring public concern over HFT.[35] An important episode in *Flash Boys* (Lewis, 2014, pp.113-118), for instance, surrounded the discovery by a portfolio manager that when he posted an order to an exchange that improved the NBBO in a stock, his resting midpoint orders for that stock in a dark pool would often be filled at the prior, stale midpoint price while another trade filled his exchange order. Similar concerns that this new form of front-running is "happening all over" (Lewis, 2014, p. 116) led to the emergence of the Investors Exchange (IEX) as a primary trading venue. To reduce the chance that pegged orders posted to IEX will be subject to stale quote arbitrage, IEX imposes a 350 microsecond delay on accessing resting orders so that IEX can process any changes to the NBBO for pegged orders. As a result, if an HFT firm submits an order to IEX upon seeing a change in the NBBO, all pegged orders at IEX will reflect the new price by the time the order arrives at the IEX matching engine. IEX's widely-followed application to become a stock exchange produced nearly five hundred comment letters—an unprecedented number for an exchange application.

To understand better whether a policy that encourages midpoint peg orders does in fact facilitate stale quote arbitrage, we empirically examine our sample of securities for evidence of this form of trading. While the nature of our data makes it impossible to estimate the precise incidence of stale quote arbitrage in our sample, our empirical strategy nevertheless allows us to estimate a maximum incidence rate. As such, our findings can be viewed as a conservative estimate of the probability that any given midpoint order in our sample might have been affected by this form of adverse selection.

4.1 Empirical Strategy

To identify midpoint trades that might have been subject to stale quote arbitrage, we first identify those trades within our sample that are likely to reflect the execution of a midpoint peg

[35] While our focus is on midpoint peg orders, pegged orders are also used at trading venues to allow traders to peg orders to the NBB or NBO, often with a user-defined offset to improve on the NBBO where doing so does not lock the market. See, e.g., https://www.nasdaqtrader.com/content/ProductsServices/Trading/OrderTypesG.pdf.

order in an non-exchange venue. Because our sample securities generally have a quoted spread of 1 penny when trading above $1.00 per share, we assume that any non-exchange trade priced above $1.00 per share and that ends in five one-thousandths of a dollar (i.e, the "trailing 5s" of $0.005, $0.015, and so on) was the result of a midpoint peg order. We also exclude any trades with a trade condition of "F" (reflecting a fill by an Inter-market Sweep Order) given that a trader pursuing stale quote arbitrage will not be trading through any protected quotations. To be sure, the possibility exists that the resulting trades may also reflect internalized retail orders given that trades listed in TAQ with an exchange code of "D" include both dark pools and broker-dealer internalizers. However, because of the oft-noted practice among internalizers of providing *de minimis* amounts of price improvement, the frequency of midpoint trades among internalizers should be minimal.[36]

For each midpoint trade, we next compare the price of the trade with the midpoint of the NBBO at the time the trade is reported by the SIP using the interleaved TAQ data. Where a midpoint trade appears in the TAQ data shortly after a change in the NBBO and reflects the midpoint of the *prior* NBBO, we posit that the potential exists for this midpoint trade to have been the product of stale quote arbitrage.

However, we emphasize that such trades can reflect only the potential for stale quote arbitrage in light of known delays in the reporting of FINRA trades. In particular, FINRA trade reporting rules require trades to be reported to a trade reporting facility as soon as practicable, but no later than 10 seconds, following trade execution.[37] As such, when a SIP reports a quote update that changes the NBBO followed by a midpoint trade report from a FINRA TRF, a mismatch between the new NBBO midpoint and the reported trade price could reflect one of two possibilities. First, the mismatch could indicate a trade occurred *after* the NBBO changed because a dark venue "picked off" a midpoint peg order priced at the stale NBBO. Second, it could indicate a trade that occurred *before* the NBBO changed and is simply recorded by the SIP with a delay. Notwithstanding this challenge, our analysis is aided by the fact that over 90% of non-exchange trades are reported within 1 second of execution, and over 97% are reported

[36] This conclusion is further supported by the monthly Rule 605 reports of prominent internalizers. For instance, Knight Securities disclosed share-weighted average price improvement for CSCO and MSFT—two securities that traded with penny spreads during the sample period—of $0.0012 and $0.0014 per share for December 2012 (a randomly selected month within our study period).

[37] See FINRA Regulatory Notice 14-46 (November 2014), available at https://www.finra.org/sites/default/files/notice_doc_file_ref/Notice_Regulatory_14-46.pdf.

within 2 seconds of execution (FINRA, 2014; Bartlett & McCrary, 2016). Moreover, regardless of trade reporting delays, any midpoint trade that reflects the midpoint of the NBBO at the time of the trade report indicates the trade was not priced using a stale NBBO. As such, we posit these trades were not subject to stale quote arbitrage.

Figure 4 provides a graphical representation of our empirical strategy using a hypothetical example of an HFT firm engaged in stale quote arbitrage of midpoint orders. As shown in the figure, stale quote arbitrage is made possible by the fact that our hypothetical HFT firm has a view of the NBBO derived synthetically from exchanges' direct data feeds. In contrast, the dark pool relies on the slower SIP-generated NBBO to price resting midpoint peg orders. This allows the HFT firm to see the change in the NBBO from 4.00 x 4.01 to 4.01 x 4.02 occurring at 12:00:00.010 three milliseconds before the dark pool. This 3 millisecond delay is in keeping with the findings of Ding, Hanna, and Hendershott (2014) concerning the latency of the SIP-generated NBBO relative to one generated with direct data fees. Upon seeing the NBBO change, the HFT firm submits a marketable buy order to the pool at 12:00:00.012 to fill a resting order to sell at the midpoint. Because the SIP-generated NBBO remains at $4.00 x $4.01, the HFT firm buys at 4.005, putting it in a position to earn risk-free profits of $0.005 per share by sending a marketable sell order to the exchange holding the new NBB of $4.01.

The bottom portion of Figure 4 depicts how the latency of trade reporting by non-exchange venues affects the reporting of the dark trade within the interleaved TAQ data. Evidence of the trade priced at 4.005 suggests the filling of a midpoint peg order, while the fact that it equals the midpoint of the prior, recently updated NBBO makes it potentially the product of stale quote arbitrage. Moreover, this should be the case regardless of the latency of the SIP-generated NBBO (and therefore, the TAQ-generated NBBO) so long as this latency is less than or equal to the latency with which trades are recorded by the SIPs (and therefore, recorded in the TAQ trade data). Using recently released SIP data that includes new time stamps for the occurrence of a quote update or a trade execution, Bartlett & McCrary (2016) provide data supporting this assumption. While quote updates to the SIP have a mean latency in 2015 of just 1.13 milliseconds, trade reports by non-exchange venues have a mean reporting latency over 100 milliseconds. In short, FINRA trade reporting to the SIPs is sufficiently slow so that trades generally follow NBBO updates. But it is also sufficiently fast that the reported price of a midpoint order will reflect the midpoint of the interleaved NBBO where the NBBO is stable.

4.2 Empirical Results

Table 2 presents our results. Given the growing attention to stale quote arbitrage over our sample period, the table presents results separately for all 16 calendar quarters comprising our sample period. In column 1, we present the size-weighted percentage of midpoint trades whose reported price matched the prevailing NBBO in the interleaved TAQ data. Overall, the rate fluctuated from a low of approximately 94.8% in the third quarter of 2011 to a high of approximately 97% in the second quarter of 2011 and in the middle two quarters of 2014. The consistently high percentage of matching trades throughout the sample period underscores how midpoint trades in our sample were overwhelmingly filled during periods of price stability. Because stale quote arbitrage of midpoint orders requires a moving NBBO, this finding highlights the overall low probability that any given midpoint trade in the sample was the product of stale quote arbitrage.

The remaining columns of Table 2 examine the extent to which the unaccounted, "mismatched" midpoint trades can be matched to the midpoint of a *prior* NBBO. Given that over 97% of FINRA trades are reported within 2 seconds of execution, we begin by imposing a two second "look-back" window to minimize the chance that a mismatched trade paired with a prior NBBO is purely the result of an extended trade reporting delay. In columns 2 and 3, we present the frequency that mispriced trades can be matched to the immediately preceding NBBO (*Lag1*) as well as to the second-to-last NBBO (*Lag2*), respectively.[38] As shown in the final row of the table, *Lag1* matches represent approximately 1.2% of midpoint trades per quarter, while *Lag2* matches represent 1.3%. Columns 4 and 5 repeat the analysis using a look-back window equal to 10 seconds—the maximum time period permitted for a trade report to be considered timely. As the table indicates, increasing the look-back window to 10 seconds leaves the results largely unchanged. None of the columns in the table reveals any notable time variation.

Even without the ability to disentangle trades that are subject to stale quote arbitrage from simple reporting delays, we find these estimates notable in light of claims concerning the widespread incidence of stale quote arbitrage during our sample period. Were it to be the case that every mismatched trade in our sample was due to stale quote arbitrage, our estimates would still indicate that it affected less than 5% of midpoint trades. Equally important, there are good

[38] In unreported results, we also match to the third-to-last, fourth-to-last, and fifth-to-last NBBO. In the aggregate, these matches account for less than .5% of the mispriced midpoint trades in our sample.

reasons to believe the likely incidence is significantly less than this outer estimate. In particular, given that trading activity can induce a change in the NBBO (e.g., by depleting trading interest at the bid or ask, or by signaling investor sentiment), trades should commonly occur just before a change in the NBBO in normal markets. Accordingly, reporting delays should inevitably cause some trade reports to be recorded by the SIP following a change in the NBBO used to price the trade given that, as noted previously, the mean reporting delay for off-exchange trades is over 100 milliseconds while the latency for quote updates to the SIP of just 1.13 milliseconds (Bartlett & McCrary, 2016). For similar reasons, even if trades randomly occurred, volatility in a stock's NBBO could create mismatched trade reports for midpoint trades that occur just prior to an NBBO change.

5. Conclusion

A significant body of research has emerged examining how the current MPV rule induces liquidity providers to queue-jump stock exchanges by posting liquidity in dark venues. These theories posit that when the NBBO is constrained by the penny MPV, traders use an exception to the MPV rule to execute trades at sub-penny prices in dark pools, potentially harming the incentive to provide displayed liquidity without providing any meaningful benefit to liquidity takers.

In this paper, we evaluate these claims in light of a largely unexamined SEC enforcement policy that constrains traders in dark pools from queue-jumping exchanges in this fashion unless their orders are priced at the midpoint of the NBBO. Consistent with this policy and using trading data from 2011-2014, we show that the documented increase in dark venues' market share of trading associated with a wider MPV (and therefore, longer queue lines) is due primarily to the greater incidence of off-exchange trades priced at the midpoint of the NBBO rather than trades offering *de minimis* price improvement. That is, SEC policy as applied is inconsistent with the literature's prior understanding of the legal rule, and our data show that SEC policy has the desired impact on market behavior. Finally, because a policy favoring dark midpoint orders can favor high frequency traders if dark pools are slow to update the NBBO as argued by Katsuyama (2014), we empirically estimate the incidence of so-called stale quote arbitrage of midpoint orders. We find that, at most, stale quote arbitrage could have occurred in no more than 5% of all midpoint trades in our sample; however, trade reporting delays suggest the actual incidence is likely to be substantially lower than this outer estimate.

These findings have immediate relevance for two distinct policy questions in current debates about U.S. market microstructure. First, with respect to the competitive advantage the MPV rule gives to dark pools, our findings highlight a previously unappreciated trade-off a trader must consider when seeking to use dark pools to queue-jump exchange liquidity providers. When the NBBO is constrained by the MPV, the relevant choice for a trader is not whether to offer liquidity at the NBBO on exchanges with their long queues or to queue-jump in a dark pool by posting orders at (or just better than) the NBBO. Rather, such a trader must assess the value of queue-jumping in light of the need to offer price improvement in the form of the quoted half-spread. As such, the SEC enforcement policy we document goes some way in diminishing widely-held concerns that queue-jumping constitutes a mere wealth transfer from liquidity providers on exchanges to liquidity providers in dark pools.

Additionally, our findings speak uniquely to the current controversy surrounding the extent to which investors are subject to adverse selection in form of stale quote arbitrage when investors post passive liquidity in dark pools. Because of the ability to identify midpoint orders within the historical record, these orders provide an unusual opportunity to examine the incidence with which the pricing of midpoint trades is consistent with conventional understandings of this arbitrage strategy.

We show that approximately 95% of all midpoint trades in our sample had prices equal to the NBBO midpoint at the time the SIPs ultimately print the trade. That is, the overwhelming majority of our midpoint trades reflect no evidence of stale quote arbitrage. In light of widespread claims concerning the incidence of stale quote arbitrage, this finding itself provides a rare source of data regarding the probability that midpoint peg orders might be subject to this form of adverse selection. Moreover, the likelihood that simple trade reporting delays account for much of these mismatched trades, combined with the fact that so few mismatched trades could be matched to a preceding NBBO midpoint, further suggest that the actual incidence in our sample securities is likely to be substantially lower than this outer maximum of 5%. Together with Bartlett & McCrary (2016), these results accordingly highlight the possibility that to the extent latency arbitrage induces an arms race for trading speed, it is likely to take a form that is different from the strategy of stale quote arbitrage so commonly depicted in policy debates about contemporary market structure.

We leave to future research exploration of the extent to which stale quote arbitrage of midpoint trades occurs more frequently outside of our sample securities, as well as how increased scrutiny of it following 2014 might have affected its overall incidence. In this regard, we note that the SEC's recent efforts to enhance the public's awareness of the latency with which the SIPs process quote and trade data will greatly facilitate our understanding of stale quote arbitrage using the methodology adopted here. Specifically, at the urging of the SEC, the SIPs commenced publishing in August 2015 the actual transaction time for trades in addition to the conventional SIP timestamp (which reflected the time the SIP ultimately processed the trade report) (Bartlett & McCrary, 2016). These new timestamps will permit a direct comparison of the price of a midpoint trade at the moment of execution to the prevailing SIP-generated NBBO. As such, they provide a unique means to estimate more precisely the incidence of stale quote arbitrage, while representing more generally an important new source of data for studying whether and how SIP reporting latencies can contribute to adverse selection in equity market structure.

Appendix A: Delisting and Price Clustering Concerns

This appendix describes how exchange delisting rules and the phenomenon of price clustering might compromise the use of our RD design and examines why the smoothness assumption is nonetheless likely to be satisfied.

A. Delisting Risk

All major U.S. exchanges impose a $1.00 minimum bid price requirement for continued listing. On Nasdaq and the NYSE, for instance, a firm that trades for thirty consecutive trading days with a closing bid below $1.00 per share risks triggering a review of its continued listing eligibility. While the existence of this rule potentially increases the probability of a security's delisting to the extent it trades at less than $1.00 per share, the rule can compromise the use of our RD framework only if this risk changes discontinuously at the cut-off. Two factors suggest this is not the case. First, an exchange's decision of whether to initiate a delisting proceeding is discretionary. Second, even if a proceeding is commenced, firms are entitled to a lengthy compliance period of 180 days to increase their stock price (e.g., through a reverse stock split). Indeed, we observe in the data many securities engaging in reverse stock splits when prices remain below $1.00 per share for many days.

While these institutional rules make it unlikely that delisting could operate in a discontinuous manner, the best way to assess the possibility is to examine it empirically. We examined each of our securities for each millisecond during our time window from 2011 to 2013. For each such security millisecond, we generate an indicator for whether that security was trading 180 calendar days in the future. We use that data to compute the probability that the security would still be trading in 180 days, and Figure A.1 plots this probability as a function of a security's two-decimal NBB.

[Insert Figure A.1]

We present in Table A.1 our formal RD estimates of the discontinuity in the probability of a security trading 180 days later.[39] All four specifications confirm the visual impression from Figure A.1: Delisting risk is substantial at price points below, say, $0.25, but does not appear to vary discontinuously at $1.00 per share. While the estimated t-ratios range from 2.4 to 3, the

[39] The four columns in Table A.1 reflect the same four RD specifications used in Table 1.

more reliable randomization inference p-values range from 0.116 to 0.293, with three of the four values larger than 0.27.

[Insert Table A.1]

To understand the discrepancy between the t-ratios, which suggest smoothness should be rejected, and the randomization inference p-values, which suggest smoothness should be accepted, consider the distribution of placebo estimates of the discontinuity in delisting risk, presented in Figure A.2. We generated this distribution by estimating a discontinuity at each price point from $0.50 to $4.00 using the approach of Column 1 of Table A.1. The estimated discontinuities range widely, from -0.076 to 0.036 and the associated t-ratios range similarly widely, from -6.82 to 6.57. Indeed, in the placebo distribution, a t-ratio at least as extreme as the 2.4 observed at the actual cutoff of $1.00 occurs 117 times out of the 300 placebo estimates. That can be understood as a p-value of 0.39, where the p-value arises from the placebo distribution. It is the evident fragility of the t-ratio approach in this context that originally motivated our consideration of the randomization inference approach.

[Insert Figure A.2]

A. Price Clustering

A second potential challenge for designing an RD analysis based on Rule 612 arises from the phenomenon of price clustering of trades at increments of five and ten cents (nickels and dimes) described in Ikenberry & Weston (2008). As noted by Barreca, Lindo, and Waddell (2011), price clustering or "heaping" at particular price points has the potential to undermine the smoothness assumptions that are at the heart of the RD approach. As noted in the main text, we present all RD plots to facilitate "pulling out" potential heap points from the other averages, and in some of our figures, the phenomenon of heaping is readily apparent. Heaping in the running variable means that the running variable density function is not continuous. One of us has written on how a discontinuous density function may indicate a violation of smoothness. On the other hand, McCrary (2008) also emphasizes that a continuous density function is not necessary for the RD approach to be valid and gives an example of an application in which the running variable density function is discontinuous and yet the RD approach is valid. We argue here that price heaping does not invalidate the use of the RD design in financial markets and that smoothness is likely to be satisfied.

Our first set of arguments is theoretical and rooted in the efficient markets hypothesis. If smoothness were not correct—that is, if a price to the right of $1.00 were discontinuously predictive of the underlying "latent" value of a security—then a profitable trading strategy exists that would take advantage of that discontinuity. A "no arbitrage" assumption thus points to the validity of the RD approach in this context. Moreover, since a security may trade below $1.00 in one millisecond and above $1.00 in another millisecond, any invalidity of the RD approach here must point to within-day mechanisms that are consistent with arbitrage possibilities.

Our second set of arguments is empirical. There are a variety of pre-determined characteristics of securities that we can measure, and we can show that these pre-determined characteristics are in fact smooth functions of price. This is precisely the form of testing recommended by Lee (2008), Imbens and Lemieux (2008), and Lee and Lemieux (2010).

In Figure A.3, we plot as a function of a security's closing price the average implied volatility for all outstanding call option contracts on that security for that trading day.[40] For purposes of this analysis, information concerning an option contract's strike price and expiration date, as well as the closing stock price, was obtained from OptionMetrics. For simplicity, we calculated implied volatility using Black-Scholes with a risk free rate of 0.5%.[41] The plot reveals a smooth, downward slopping curve, reflecting a decline in implied volatility as stock prices increase. Table A.1 provides point estimates, standard errors, and randomization inference p-values using the same four specifications used throughout the paper. None indicate evidence of any discontinuity at the cut-off.

[Insert Figure A.3]

Figure A.4 presents a second smoothness test, this time conducted using cumulative 5-day and 100-day returns based on data from CRSP. For each security on trading day *t*, we compute

---

[40] We focus on call contracts because the low average price of our sample securities has the effect of greatly diminishing the demand for put contracts. As such, there are substantially fewer put contracts outstanding than call contracts, adding noise to empirical analysis of the put contract data. In unreported estimates using put contracts, however, we corroborate the qualitative conclusion from the call contracts that implied volatility is a smooth function of the closing price in the equities market.

[41] While this approach does not necessarily reflect volatility for each option contract given that they are generally American options, it provides a rough estimate of market-based volatility that is consistently calculated across issuers. In any event, regression of this measure of volatility against the volatility measure calculated by OptionMetrics for an issuer's outstanding option contracts yields an R-squared of .99 with a constant (slope) of approximately 0 (1).

5-day (100-day) returns as the conditional expectation, given the security's closing price on day *t*, of the cumulative returns over the past 5 (100) trading days, respectively, which we think of as prior weekly returns and prior long-run returns. Cumulative returns are calculated using CRSP daily closing prices as $-1 + \exp(\sum_{t=1}^{T} \ln(1 + R_{r-t}))$, where T is either 5 or 100 and R_r is the daily return. Panel A plots weekly returns as a function of the intra-day NBB truncated to two decimal places, and Panel B plots long-run returns. Neither shows evidence of any discontinuity at the $1.00 cut-off. Point estimates, standard errors, and randomization inference p-values are provided in Table A.1 and confirm this conclusion.

[Insert Figure A.4]

Finally, following the recommendation of Sales and Hansen (2015), we additionally compute for all bandwidths from 0.05 to 1.50 randomization inference p-values using Hotelling's T^2 statistic, testing jointly for zero discontinuities in delisting, the implied volatility of call options, prior weekly returns, and prior long-run returns. These results, presented in Figure A.5, show that bandwidths larger than 0.65 are squarely rejected by the data, but that those in the range of 0.05 to 0.55 are likely reliable in this context.

[Insert Figure A.5]

Appendix B: Midpoint Trading with Maker-Taker Pricing Controls

Because the structure of exchanges' maker-taker pricing models commonly change around the $1.00 cut-off, this Appendix investigates how changes in maker-taker pricing might affect our finding of a discontinuous increase in midpoint trading above $1.00 per share. Table B.1 illustrates the typical structure of these fees for transactions above and below $1.00 per share by summarizing the fees and rebates across exchanges for June 3, 2011, a randomly selected trading day. Because exchanges often differ in whether a fee or rebate is payable on a per share basis or as a percentage of the trade, we convert all fees and rebates to reflect their percentage of a $10,000 trade.[42] Moreover, because exchanges often calculate fees and rebates based on the volume of trading conducted by a trader, Table B.1 provides the maximum and minimum fee or rebate for each exchange.

[Insert Table B.1]

Table B.1 highlights the difficulty of assessing the effect of maker-taker pricing on trading across the $1.00 cut-off. Although the average maximum fee assessed on liquidity takers for a trade priced below $1.00 per share was generally less than the average maximum fee imposed on a trade priced above the $1.00 cut-off, several trading venues (including Nasdaq) assessed the same maximum take-fee regardless of transaction price. At the same time, two exchanges (BATS Y and Nasdaq BX) applied inverted pricing only to trades above $1.00 per share such that liquidity takers on these venues received a rebate for accessing displayed liquidity for trades priced above this price. Rebates paid to liquidity providers were generally lower for trades priced below $1.00 per share, but situations could similarly arise where the reverse applied. For instance, a liquidity provider for shares priced below $1.00 per share on the National Stock Exchange could receive a rebate as high as 0.25% of the transaction value, while a liquidity provider for shares priced at $1.00 per share on the CBOE would be required to pay a fee of 0.02% of transaction value.

In light of this heterogeneity, we re-examine each of our results in Table 1 controlling for the expected fee and rebate payable on a trade during each security-millisecond. Given the range of fees and rebates payable on a venue, we run the analysis separately using the maximum, the

[42] Converting fees and rebates to a percentage of a $10,000 trade also allows us to analyze the fee and rebate owing on a similarly valued trade regardless of the price paid per share.

minimum, and the midpoint of the fee and rebate payable on each security on each exchange for each security-millisecond from 2011-2012.

Our approach to controlling for fees and rebates takes advantage of the natural variation in fees and rebates. The fee and rebate structure changes on a millisecond by millisecond basis, as the NBBO moves from one exchange to another, changes over time as exchanges alter the fee and rebate structure, and differs across securities as exchanges sometimes offer specialized programs for some securities. We merge the fee and rebate structure onto our TAQ data so that each security-millisecond is characterized by the NBB, by fees, and by rebates and then compute averages for each unique configuration of the NBB, fees, and rebates, as well as the total number of security-milliseconds corresponding to each configuration. This first step is motivated by the econometric result that a regression of an outcome on covariates can be accomplished in a numerically equivalent way by running a weighted regression of *average* outcomes on the set of all possible covariate configurations represented in the original data with weights equal to the number of observations underlying the average outcome. This first step is computationally advantageous because it greatly reduces the scale of the data. The second step regression in question is a fixed effect regression of each outcome on a set of indicators for the NBB, as well as fees and rebates. As in the main text, we use a triangle kernel, and we use the same bandwidths as in our main analyses. The fixed effects for the NBB from this regression, up to scale, are a fee-rebate adjusted version of the local average displayed in the main figures in Table 1.[43] We then perform the same analysis as in Section 3, but using the NBB fixed effects in place of the raw means.

Table B.2 presents point estimates, standard errors, and randomization inference p-values for the analyses, analogous to the results in Table 1. Regardless of the method for calculating maker-taker fees, our results remain consistent with those obtained without regard to maker-taker pricing. Intuitively, the similarity of results holds because while fees and rebates are sometimes predictive of the outcomes under study, they do not exert a large enough influence to change the results qualitatively.

[Insert Table B.2]

[43] Note that were fees and rebates not to be included in the regression then the fixed effects estimates would be the simple averages reported in the main text.

References

Angrist, Joshua D. and Jörn-Steffen Pischke, 2009, *Mostly Harmless Econometrics: An Empiricist's Companion*, Princeton University Press.

Barreca, Alan I., Jason M. Lindo, and Glen R. Waddell, 2011, Heaping-Induced Bias in Regression-Discontinuity Designs, NBER Working Paper No. 17408.

Bartlett, Robert P., III, and Justin McCrary, 2016, How Rigged Are Stock Markets? Evidence from Microsecond Timestamps, Unpublished working paper, University of California, Berkeley.

Bessembinder, Hendrik, and Kumar Venkataraman, 2010, Bid-Ask Spreads: Measuring Trade Execution Costs in Financial Markets, in "Encyclopedia of Quantitative Finance" (Rama Cont, Ed.), John Wiley & Sons.

Budish, Eric B., Peter Cramton, and John J. Shim, 2015, The High-Frequency Trading Arms Race: Frequent Batch Auctions as a Market Design Response, *Quarterly Journal of Economics* 130, 1547-1621.

Buti, Sabrina, Francesco Consonni, Barbara Rindi, Yuanji Wen and Ingrid Werner, 2015, Sub-Penny and Queue-Jumping, Unpublished working paper. The Ohio State University.

Buti, Sabrina, Barbara Rindi, and Incrid Werner, 2016, Dark Pool Trading Strategies, Market Quality and Welfare, *Journal of Financial Economics* (in press).

Buti, Sabrina, Barbara Rindi, Yuanji Wen and Ingrid Werner, 2013, Tick Size Regulation and Sub-Penny Trading. Unpublished working paper. The Ohio State University.

Cattaneo, Matias D, Brigham R. Frandsen, and Rocío Titiunik, 2015, Randomization Inference in the Regression Discontinuity Design: An Application to Party Advantages in the U.S. Senate, *Journal of Causal Inference*, 3(1), 1—24

Dick, Dennis, 2010, The Hidden Cost of "Sub-Pennying," *CFO Magazine*, Jan-Feb, 8-9.

Ding, Shengwei, John Hanna, and Terrence Hendershott, 2014, How Slow is the NBBO? A Comparison with Direct Exchange Feeds, *Financial Review* 49, 313–332.

Ferrell, Allen, 2001, A proposal for solving the "payment for order flow" problem, *Southern California Law Review* 74, 1029-1088.

Financial Industry Regulatory Authority (FINRA), 2014, Equity Trading Initiatives: OTC Trade Sequencing, Regulatory Notice 14-46, available at https://www.finra.org/sites/default/files/notice_doc_file_ref/Notice_Regulatory_14-46.pdf.

Fox, Merritt, Lawrence R. Glosten, and Gabriel V. Rauterberg, 2015, The New Stock Market: Sense and Nonsense, *Duke Law Journal* 65, 191-277.

Hahn Jinyong, Petra Todd and Wilbert Van Der Klaauw, 2001, Identification and estimation of treatment effects with a regression-discontinuity design, *Econometrica* 69, 201–209.

Hausman, J., and C. Palmer, 2011, Heteroskedasticity-Robust Inference in Finite Samples, NBER Working Paper 17698.

Holden, Craig W., and Stacey Jacobsen, 2011, The Breakdown of Standard Microstructure Techniques and What to Do About It, Unpublished working paper, Indiana University.

Ikenberry, David L. and James P. Weston, 2008, Clustering in US Stock Prices after Decimalisation. *European Financial Management*, 14, 30–54.

Imbens, Guido W., and Karthik Kalyanaraman, 2012, Optimal Bandwidth Choice for the Regression Discontinuity Estimator, *Review of Economic Studies*, xx, 1–27.

Imbens, Guido W., and Thomas Lemieux, 2008, Regression Discontinuity Designs: A Guide to Practice, *Journal of Econometrics*, 142, 615–35.

Katsuyama, Bradley, 2014, Testimony Before the U.S. Senate Permanent Subcommittee on Investigations of the Committee on Homeland Security and Governmental Affairs, available at https://www.hsgac.senate.gov/download/?id=dba2d5e8-ebe2-48ae-af19-e32e0d95a423.

Kwan, Amy, Ronald Masulis, and Thomas McInish, 2014. Trading rules, competition for order flows and market fragmentation. *Journal of Financial Economics* (forthcoming).

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

Lee, David S., 2008, Randomized Experiments from Non-Random Selection in U.S. House Elections, *Journal of Econometrics*, 142, 675-697.

Lee, David S., and Thomas Lemiuex, Regression Discontinuity Designs in Economics, *Journal of Economic Literature*, 48, 281-355.

Lewis, Michael, 2014, *Flash Boys: A Wall Street Revolt*, W.W. Norton & Co.

MacKinnon, J., and H. White, 1985, Some Heteroskedasticity-consistent Covariance Matrix Estimators with Improved Finite Sample Properties, *Journal of Econometrics*, 29, 305-325.

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

O'Hara, Maureen and Mao Ye, 2011, Is market fragmentation harming market quality? *Journal of Financial Economics* 100, 459-474.

Sales, Adam and Ben B. Hansen, 2015, Limitless Regression Discontinuity, Unpublished working paper, University of Michigan.

Tabb, Larry, 2014, No, Michael Lewis, the US Equities Market Is Not Rigged, available at http://tabbforum.com/opinions/no-michael-lewis-the-us-equities-market-is-not-rigged.

Tabb Group, 2015, Equities LiquidityMatrix: December 2014, available at http://tabbforum.com/liquidity-matrix/equities

TD Ameritrade, 2010, Comment Letter on Concept Release on Equity Market Structure, available at http://www.sec.gov/comments/s7-02-10/s70210-124.pdf.

Wah, Elaine, 2016, How Prevalent and Profitable are Latency Arbitrage Opportunities on U.S. Stock Exchanges?, Unpublished working paper, University of Michigan.

Table 1:
Queue-Jumping at the $1.00 Cut-off

Outcome Measure	**Model 1**	**Model 2**	**Model 3**	**Model 4**
Quoted Bid Depth	2.5398 (0.0389) [0.0000]	2.4966 (0.0331) [0.0000]	2.4972 (0.0549) [0.0000]	2.5267 (0.0411) [0.0000]
FINRA Trades in a Second	0.0132 (0.0012) [0.0000]	0.0132 (0.0014) [0.0000]	0.0095 (0.0012) [0.0010]	0.0118 (0.0012) [0.0000]
FINRA Trades as a % of All Trades in a Second	0.0751 (0.0031) [0.0000]	0.0737 (0.0035) [0.0000]	0.0804 (0.0042) [0.0000]	0.0751 (0.0031) [0.0000]
Probability of Midpoint Trade	0.1189 (0.0028) [0.0000]	0.1177 (0.0033) [0.0000]	0.1149 (0.0036) [0.0000]	0.1185 (0.0026) [0.0000]
Probability of De Minimis Price Improvement	-0.0164 (0.0022) [0.0000]	-0.0150 (0.0025) [0.0000]	-0.0120 (0.0034) [0.0030]	-0.0164 (0.0022) [0.0000]
Probability of Trading at the NBBO	-0.0378 (0.0082) [0.0000]	-0.0388 (0.0096) [0.0000]	-0.0457 (0.0152) [0.0000]	-0.0409 (0.0102) [0.0000]

Table 2:
Outer Maximum of Stale Quote Arbitrage for Midpoint Orders in Sample Securities

		(1)	(2)	(3)	(4)	(5)
			Trades Matching Prior NBBO Midpoint (2 second window):		Trades Matching Prior NBBO Midpoint (10 second window):	
Year	Quarter	*Trades Matching Current NBBO*	*Lag1*	*Lag2*	*Lag1*	*Lag2*
2011	1	0.955	0.015	0.016	0.016	0.018
2011	2	0.972	0.010	0.007	0.010	0.008
2011	3	0.948	0.021	0.018	0.022	0.020
2011	4	0.949	0.018	0.019	0.019	0.021
2012	1	0.963	0.011	0.013	0.011	0.014
2012	2	0.967	0.009	0.013	0.009	0.014
2012	3	0.959	0.010	0.014	0.010	0.015
2012	4	0.963	0.010	0.013	0.010	0.014
2013	1	0.969	0.010	0.012	0.010	0.013
2013	2	0.962	0.013	0.014	0.013	0.015
2013	3	0.967	0.012	0.014	0.013	0.014
2013	4	0.962	0.013	0.015	0.013	0.015
2014	1	0.964	0.011	0.010	0.011	0.011
2014	2	0.972	0.010	0.011	0.011	0.011
2014	3	0.972	0.009	0.010	0.010	0.010
2014	4	0.961	0.015	0.014	0.016	0.015
Mean:		0.963	0.012	0.013	0.013	0.014

Table A.1:
Delisting and Smoothness Tests

Outcome Measure	**Model 1**	**Model 2**	**Model 3**	**Model 4**
Probability of Trading in 180 Days	0.0114 (0.0046) [0.2750]	0.0136 (0.0052) [0.2550]	0.0182 (0.0057) [0.0680]	0.0162 (0.0054) [0.2880]
Implied Volatility	-0.0155 (0.0263) [0.9270]	-0.0241 (0.0269) [0.8960]	-0.0463 (0.0397) [0.6870]	-0.0296 (0.0287) [0.8220]
5 Day Prior Returns	-0.0023 (0.0013) [0.1720]	-0.0025 (0.0013) [0.1770]	-0.0045 (0.0015) [0.1940]	-0.0036 (0.0015) [0.5760]
100 Day Prior Returns	-0.0056 (0.0064) [0.2700]	-0.0027 (0.0072) [0.4280]	-0.0100 (0.0086) [0.6110]	-0.0033 (0.0067) [0.8470]

Table B.1:

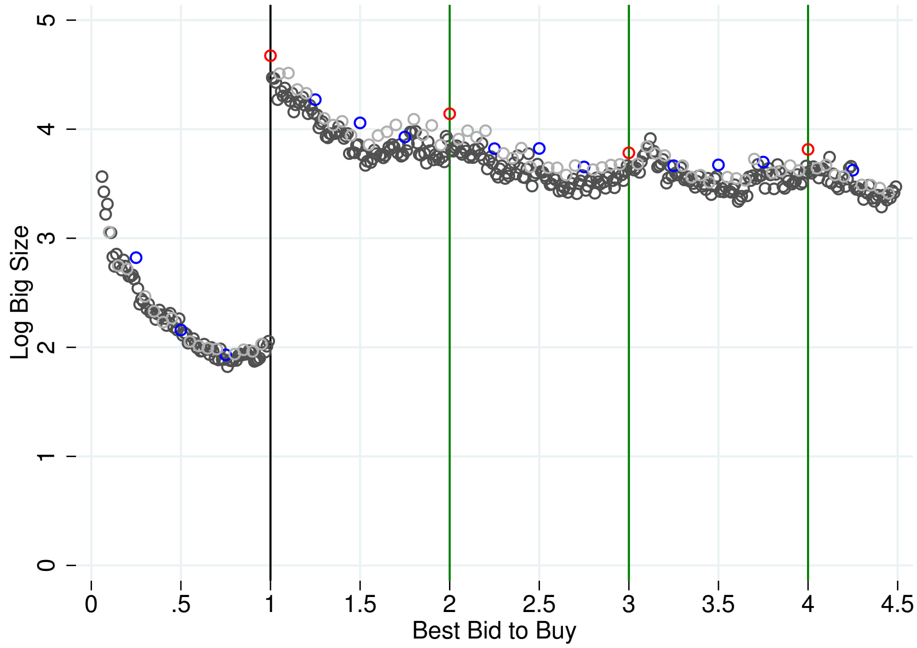
Maker-Taker Fees as of June 3, 2011

	Fee/Rebate for Accessing Liquidity				Rebate/Fee for Providing Liquidity			
	Share Price = $1.00		Share Price = $0.99		Share Price = $1.00		Share Price = $0.99	
Exchange:	Max.	Min.	Max.	Min.	Max.	Min.	Max.	Min.
BATS Z	-0.28%	-0.28%	-0.10%	-0.10%	0.27%	0.27%	0.00%	0.00%
BATS Y	0.03%	0.03%	-0.10%	-0.10%	0.00%	-0.02%	0.00%	0.00%
DirectEdge A	-0.30%	0.02%	-0.20%	0.00%	-0.01%	-0.03%	0.00%	0.00%
DirectEdge X	-0.30%	-0.30%	-0.10%	-0.10%	0.34%	0.23%	0.00%	0.00%
Chicago Board Options Exchange	-0.02%	-0.02%	-0.02%	-0.02%	-0.02%	-0.02%	0.00%	-0.02%
National Stock Exchange	-0.30%	-0.30%	-0.30%	-0.30%	0.30%	0.26%	0.25%	0.00%
Chicago Stock Exchange	-0.30%	-0.30%	-0.30%	-0.30%	0.31%	0.25%	0.01%	0.00%
Nasdaq	-0.30%	-0.30%	-0.30%	-0.30%	0.30%	0.20%	0.01%	0.00%
Nasdaq BX	0.05%	0.14%	-0.10%	-0.10%	-0.15%	-0.18%	0.00%	0.00%
Nasdaq PSX	-0.25%	-0.25%	-0.20%	-0.20%	0.24%	0.22%	0.00%	0.00%
NYSE	-0.23%	-0.23%	-0.23%	-0.23%	0.35%	0.15%	0.05%	0.00%
NYSE Arca	-0.30%	-0.28%	-0.10%	-0.10%	0.32%	0.21%	0.00%	0.00%
NYSE Amex	-0.28%	-0.19%	-0.25%	-0.20%	0.45%	0.16%	0.25%	0.00%
Mean	-0.21%	-0.17%	-0.18%	-0.16%	0.21%	0.13%	0.04%	0.00%
Maximum	-0.30%	-0.30%	-0.30%	-0.30%	-0.15%	-0.18%	0.00%	-0.02%
Minimum	0.05%	0.14%	-0.02%	0.00%	0.45%	0.27%	0.25%	0.00%

Table B.2:
Discontinuities at the $1.00 Cut-Off Controlling for Maker-Taker Fees

Outcome Measure	**Model 1**	**Model 2**	**Model 3**	**Model 4**
Quoted bid depth	2.5132 (0.0393) [0.0000]	2.4694 (0.0332) [0.0000]	2.4715 (0.0558) [0.0000]	2.5309 (0.0372) [0.0000]
FINRA Trades in a Second	0.0138 (0.0012) [0.0000]	0.0138 (0.0014) [0.0000]	0.0102 (0.0012) [0.0360]	0.0143 (0.0014) [0.0000]
FINRA Trades as a % of All Trades in a Second	0.0823 (0.0028) [0.0000]	0.0814 (0.0033) [0.0000]	0.0860 (0.0037) [0.0000]	0.0833 (0.0032) [0.0000]
Probability of Midpoint Trade	0.1224 (0.0028) [0.0000]	0.1212 (0.0033) [0.0000]	0.1184 (0.0036) [0.0000]	0.1198 (0.0026) [0.0000]
Probability of De Minimis Price Improvement	-0.0193 (0.0022) [0.0000]	-0.0180 (0.0025) [0.0000]	-0.0148 (0.0034) [0.0030]	-0.0208 (0.0023) [0.0000]
Probability of Trading at the NBBO	-0.0314 (0.0082) [0.0030]	-0.0323 (0.0096) [0.0020]	-0.0395 (0.0152) [0.7170]	-0.0329 (0.0098) [0.0010]

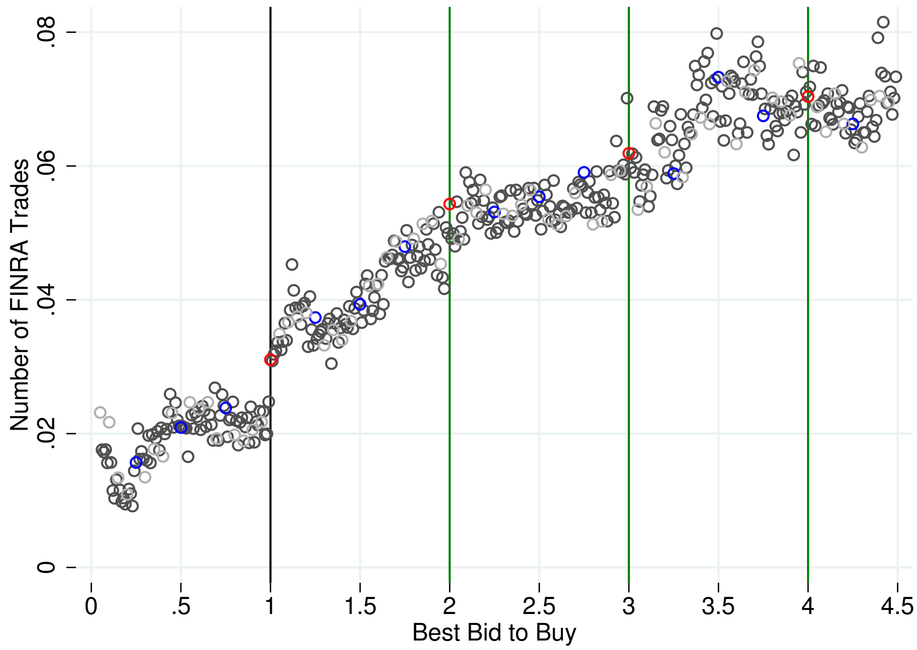
Figure 1. Inside Depth



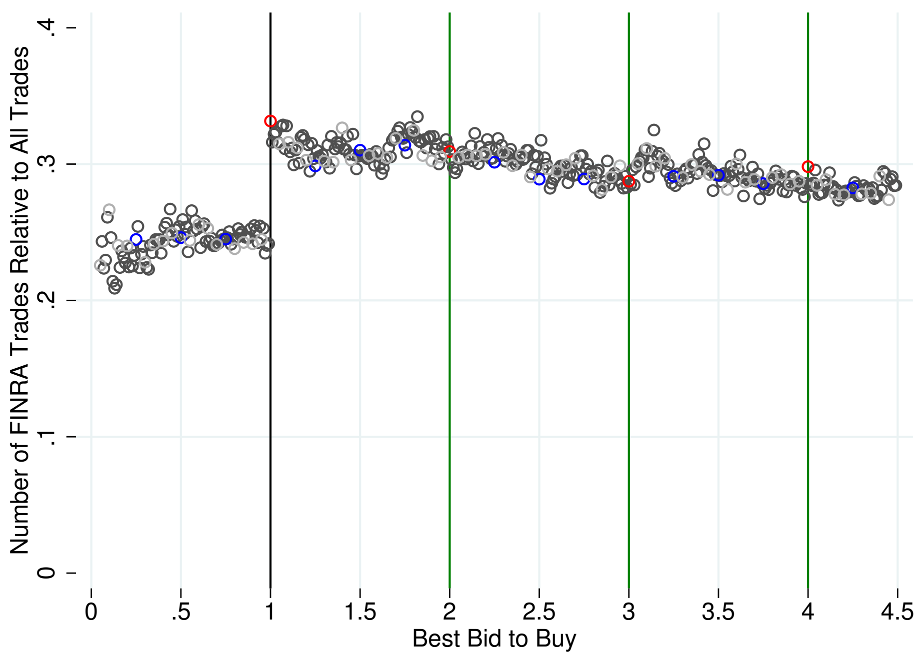
Note: Figure shows the conditional expectation of log bid size given the NBB. Estimates based on TAQ data. See text for details.

Figure 2. Prevalence of FINRA Trades by Price

A. Number of FINRA Trades in a Second



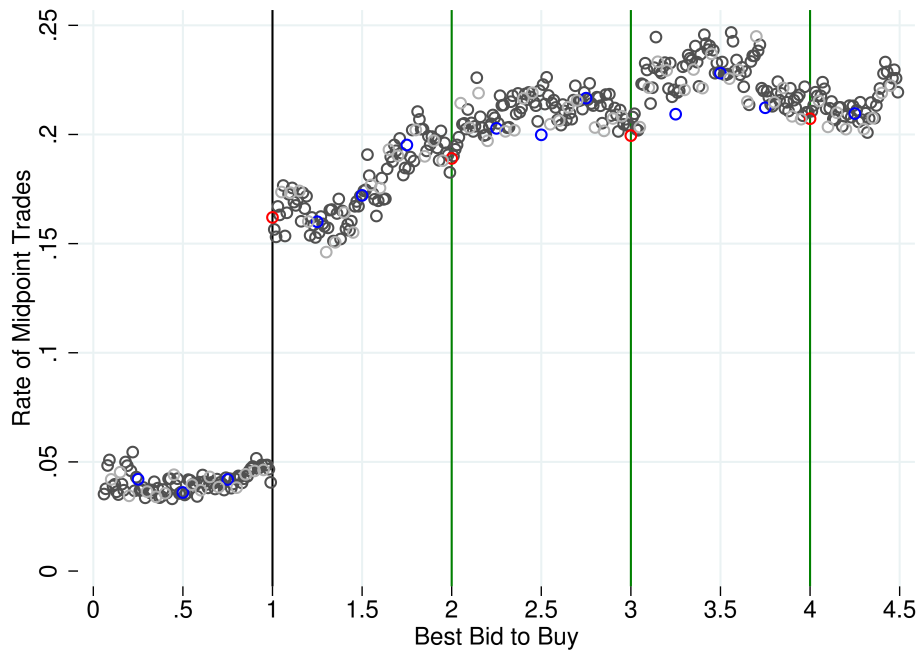
B. Number of FINRA Trades in a Second Relative to Trades in a Second



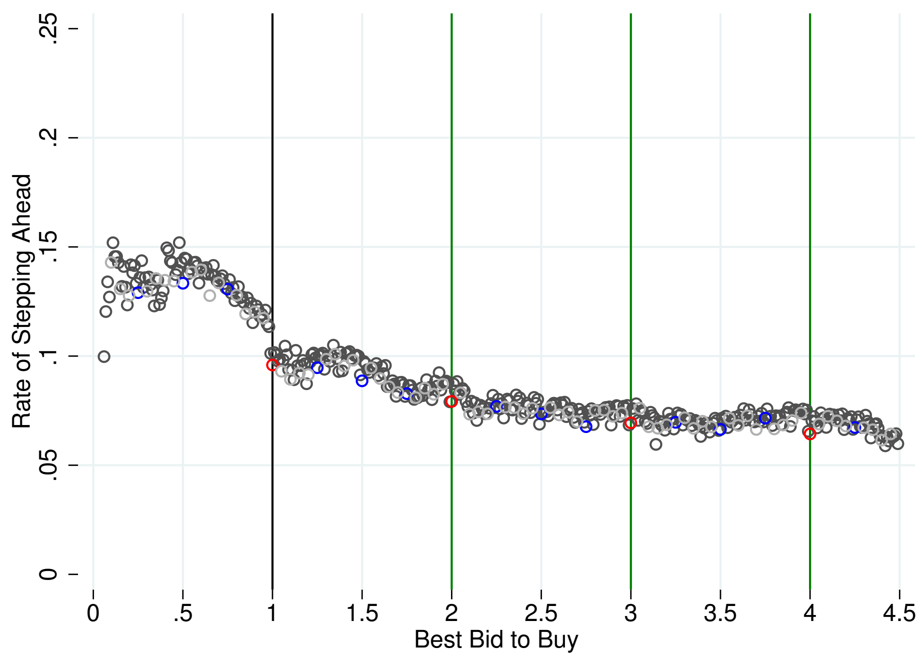
Note: Figure shows the conditional expectation of the number of FINRA trades in a second given the NBB as of the beginning of the second (A) and the fraction of trades in a second that are FINRA trades (B). Estimates based on TAQ data. See text for details.

Figure 3. Price Improvement for FINRA Trades

A. Rate of Midpoint Trades

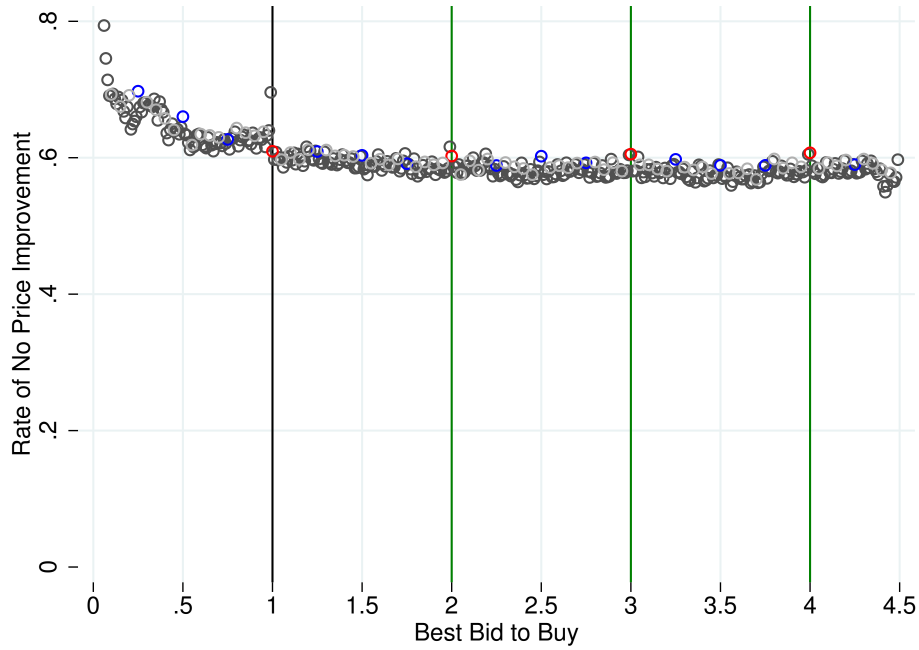


B. Rate of Stepping Ahead



Notes at end of table.

C. Rate of Zero Price Improvement



Note: Figure shows conditional probabilities of functions of price improvement, or the difference between trade price and NBO (for buy orders) or NBB (for sell orders), for off-exchange trades. Orders classified using Lee-Ready. Panel A shows rate of midpoint pricing; B shows rate of price improvement of just 0.0001 exclusive of midpoint pricing; and C shows rate of no price improvement. Estimates based on TAQ data. See text for details.

Figure 4: Hypothetical Stale Quote Arbitrage As Reflected in the TAQ Data

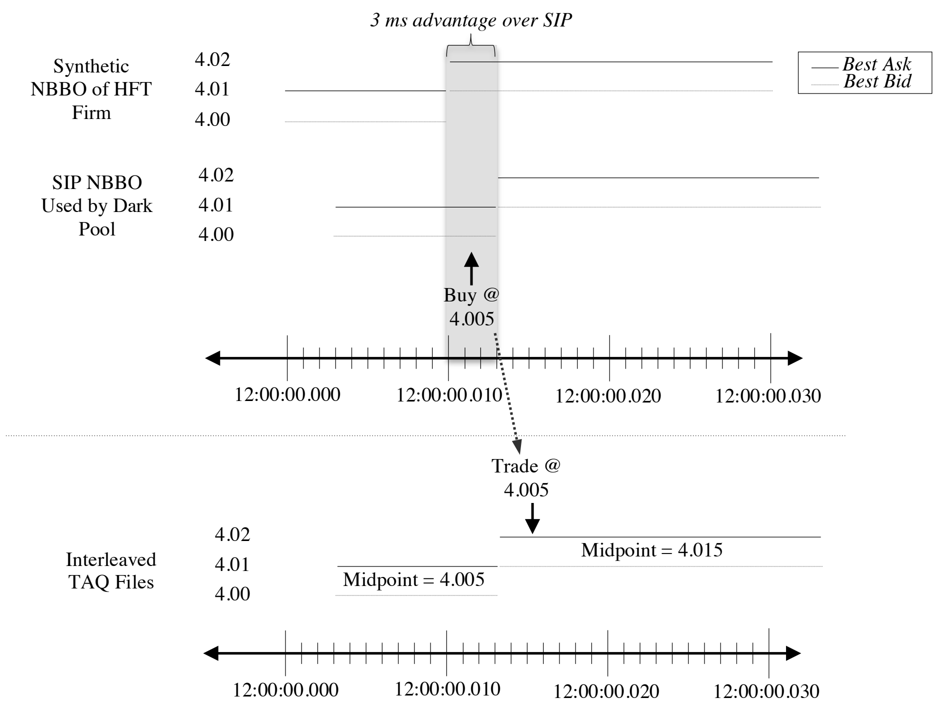
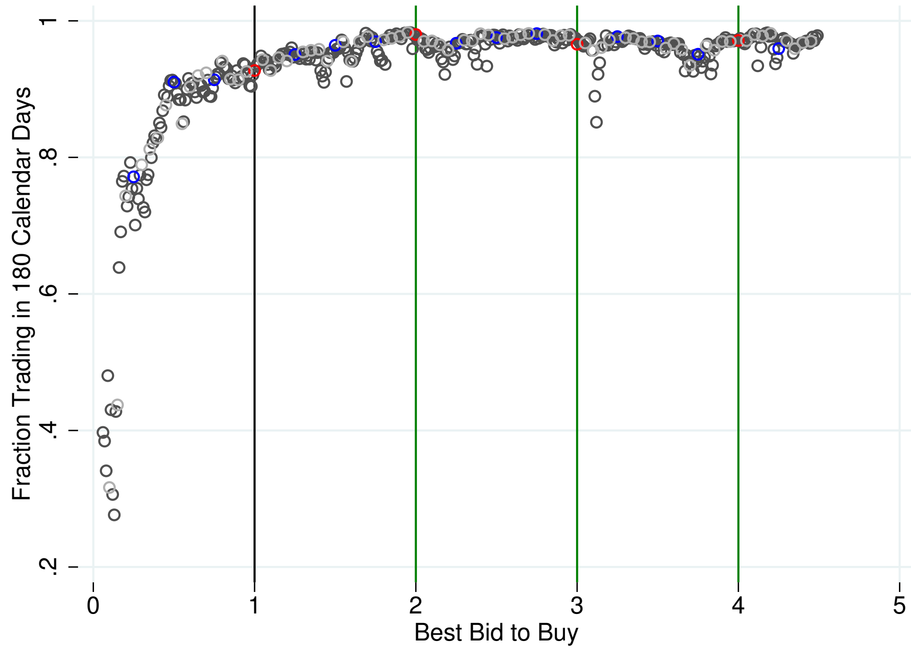
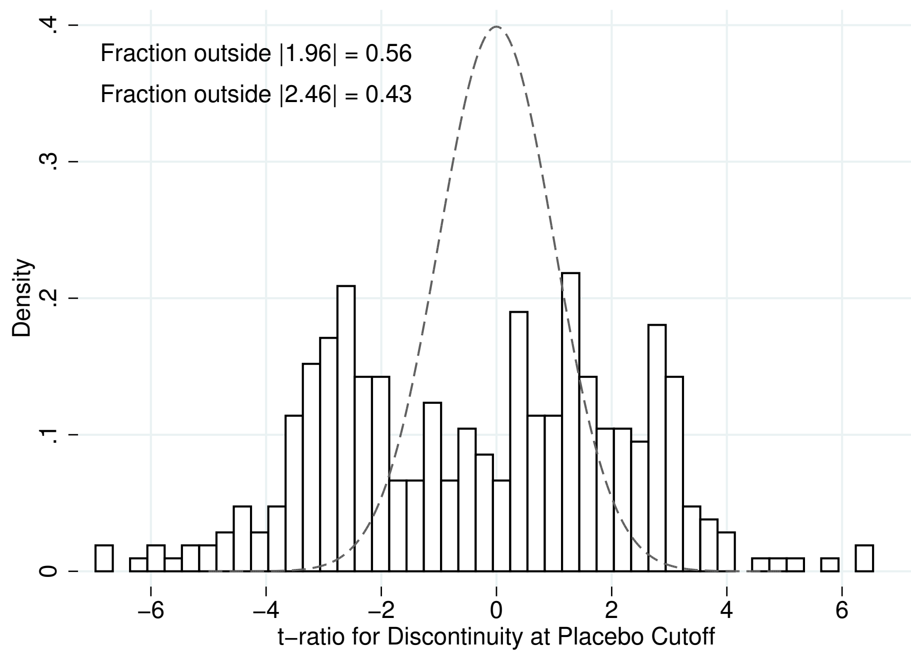


Figure A.1. Probability of Trading 180 Days Later



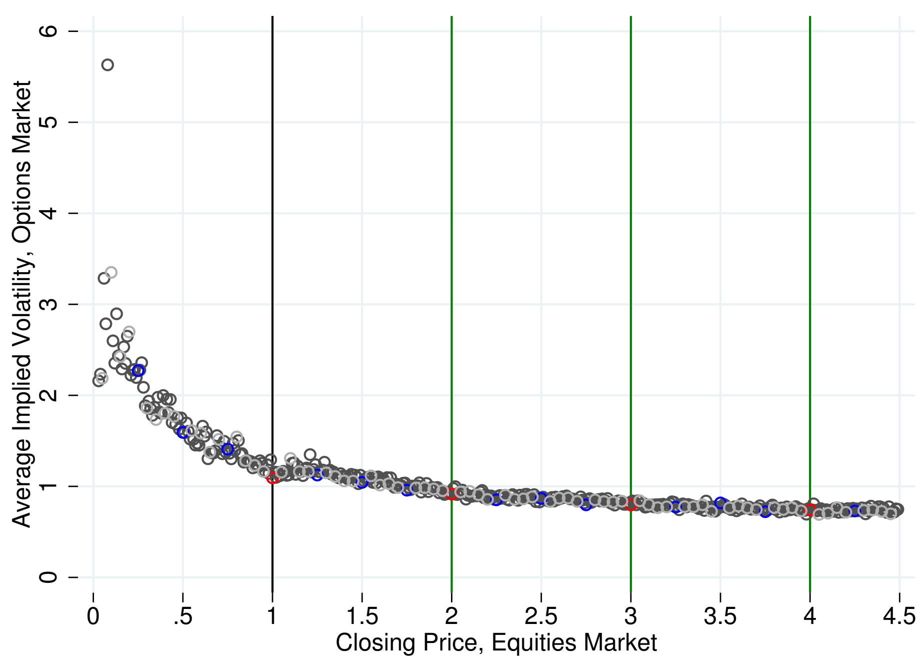
Note: Figure shows the probability, conditional on the NBB truncated to two digits in a given millisecond of date t, of the security being traded at date $t+180$. Estimates based on CRSP and TAQ data. See text for details.

Figure A.2. Placebo Distribution of t-Ratios for Delisting Risk



Note: Figure shows the histogram of t-ratios obtained by estimating placebo discontinuities in delisting risk from \$0.50 to \$4.00. The superimposed grey curve is the standard normal density function.

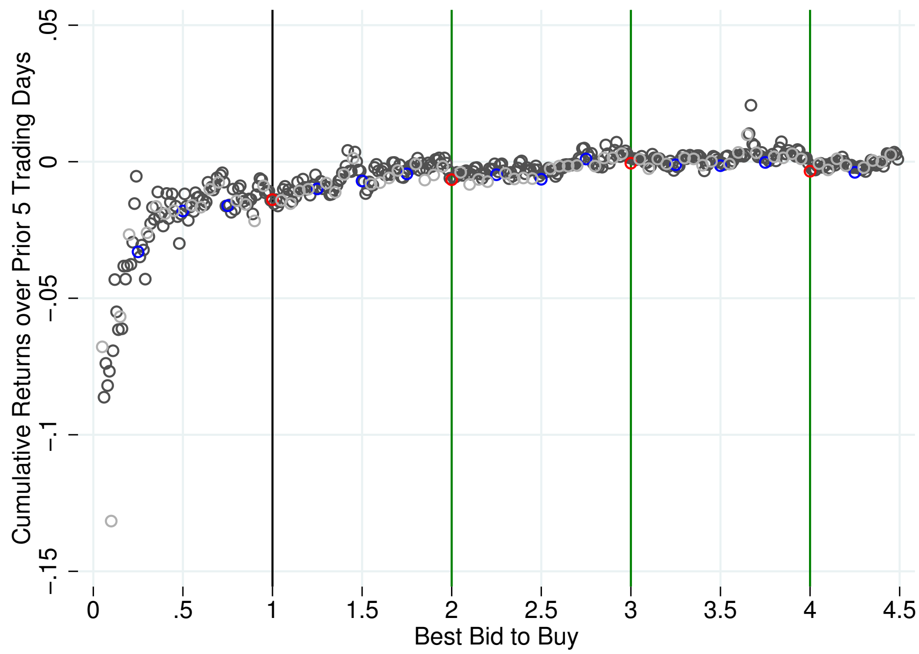
Figure A.3. Implied Volatility of Options Market Call Contracts



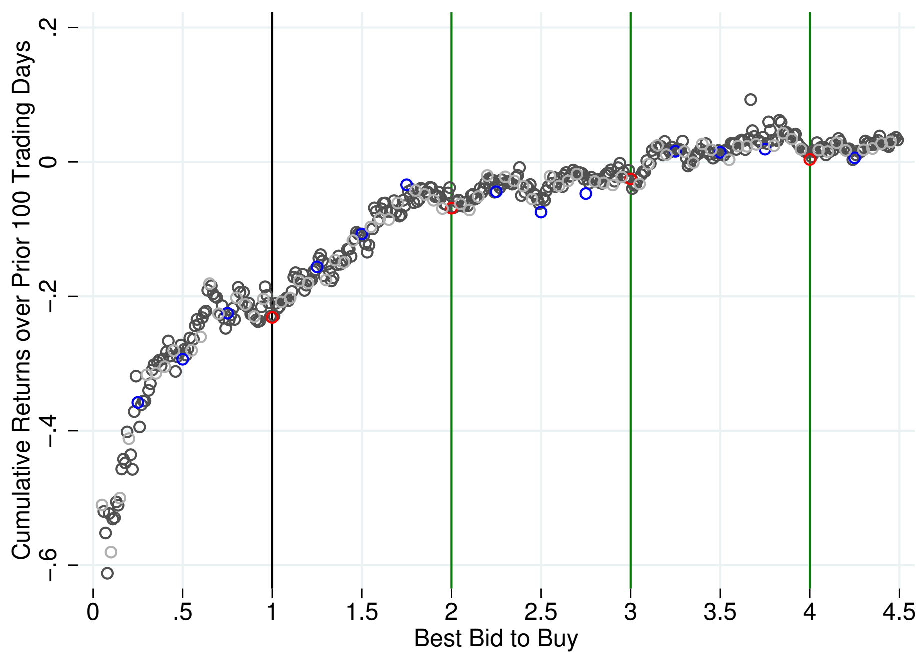
Note: Figure shows the conditional expectation of the implied volatility of options market call contracts given the corresponding closing price in the equities market. Implied volatility based on Black-Scholes using a risk-free rate of 0.5%. Estimates based on OptionMetrics data. See text for details.

Figure A.4. Relationship Between Current Prices and Past Cumulative Returns

A. Cumulative Returns over Past Week

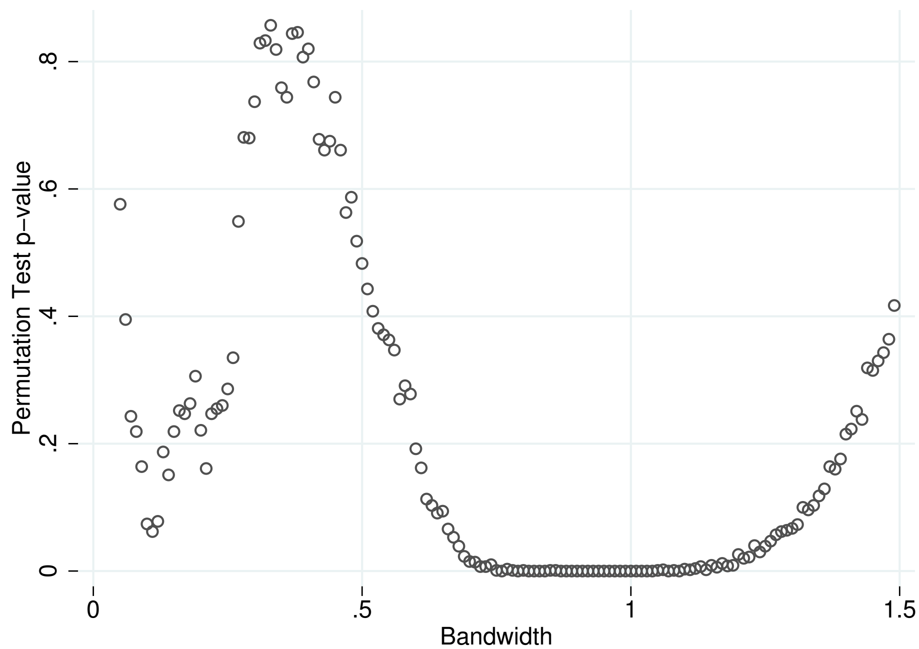


B. Cumulative Returns over Past 100 Trading Days



Note: Figure shows conditional expectations of prior returns given the NBB. Panel A (B) shows the conditional expectation of the cumulative returns over the past 5 (100) trading days, calculated as $-1+\exp(\sum_{t=1}^{T}\ln(1+R_{\tau-t}))$, where T is either 5 or 100 and R_τ is the daily return. Estimates based on CRSP and TAQ data. See text for details.

Figure A.5. Randomization Inference p-value as a Function of Bandwidth



Note: For bandwidths over a grid of $\{0.05, 0.06, ..., 1.5\}$, figure shows randomization inference p-value associated with Hotelling's T^2 and the four covariates from Figures 1, 2, 3A, and 3B.