Algos gone wild: Are order cancellations in financial markets excessive? ☆

Marta Khomyn[a*] and Tālis J. Putniņš [a,b]

[a] *University of Technology Sydney, PO Box 123 Broadway, NSW 2007, Australia*
[b] *Stockholm School of Economics in Riga, Strelnieku Street 4a, Riga, LV 1010, Latvia*

Abstract

We investigate whether the explosive growth in order-to-trade ratios and order cancellation rates in financial markets is something to be concerned about. We develop a simple theoretical model (which we test and calibrate with data) of a liquidity provider in a fragmented market, who monitors several sources of information and updates quotes to avoid being picked off (trading at stale prices). We find that recent growth in order-to-trade ratios is driven by fragmentation of trading across multiple venues as well as decreasing monitoring costs, with the increase in monitoring leading to improved liquidity. Our model explains why there is considerable cross-sectional heterogeneity in order-to-trade ratios, with higher ratios in more volatile stocks, higher price-to-tick stocks, lower volume stocks, and in ETFs compared to stocks. Our findings suggest that message taxes can have adverse effects on market making in securities that already have disadvantageous conditions for liquidity providers. Furthermore, message taxes create unlevel competition between trading venues due to higher order-to-trade ratios on venues with lower volume shares.

JEL classification: G14

Keywords: order-to-trade ratio, market fragmentation, regulation, liquidity, HFT

---

1. Introduction

The rapid recent growth in order-to-trade ratios and order cancellation rates in financial markets has alarmed regulators and some market participants around the world. For example, in US equities, the order-to-trade ratio (number of order enter/amend/cancel messages to the number of trades) has increased more than ten-fold since 2000 (Committee on Capital Markets Regulation, 2016) and recent news reports highlighted 96.8% of all orders being cancelled before they trade, with 90% being cancelled within one second (US SEC[1]). A response to these concerns is message taxes, which have been proposed in some countries (such as the US) and implemented in others (e.g., Australia, Italy, Germany). Despite the concerns and proposed regulation, there is a lot that we do not yet understand about the drivers of order-to-trade ratios, whether their growth warrants concern, and the impacts of regulatory proposals such as message taxes. This paper aims to increase our understanding of these issues.

High order-to-trade ratios have been in public spotlight as they are claimed to be a symptom of predatory or manipulative behaviour of high-frequency traders. It is important to recognize that while market manipulation strategies such as spoofing or quote stuffing can generate spikes in quoting activity, high order-to-trade ratios can also arise from a number of activities that do not necessarily imply illicit behaviour or are not harmful. In fact, rising quoting traffic (as we will show in this paper) could be a result of legitimate market making activity that requires posting liquidity across multiple venues and adjusting the quotes rapidly in response to new information to minimize picking off risk. The combination of advances in technology that have lowered monitoring costs and allowed much more information to be processed by market makers and fragmentation of trading across multiple venues necessitates increasing amounts of quote revisions by market makers in order to stay competitive in liquidity provision. It is thus perhaps not surprising that the majority of liquidity provision is currently undertaken by HFT firms.

As a result of the alleged link between high order-to-trade ratios (OTTR) and illicit HFT behaviour, a number of regulators have imposed messaging taxes, effectively charging high-OTTR traders a fee for excessive message traffic. To the extent that such regulation curbs harmful HFT behaviour, the tax could improve liquidity and other measures of market quality. However, if the regulation negatively affects liquidity providers (the majority of which are actually HFT firms), market liquidity could decrease.

To investigate these issues, we develop a simple theoretical model (which we test and calibrate with data) of a liquidity provider in a fragmented market, who monitors several

[1] See https://www.sec.gov/marketstructure/research/highlight-2013-01 for instance.

sources of information and updates quotes to avoid being picked off (trading at stale prices). We find that recent growth in order-to-trade ratios is driven by fragmentation of trading across multiple venues as well as decreasing monitoring costs, with the increase in monitoring leading to improved liquidity.

Our model explains why there is considerable cross-sectional variation in order-to-trade ratios, with higher ratios in more volatile stocks, higher price-to-tick stocks, lower volume stocks, and in ETFs compared to stocks. In particular, we allow for endogenous choice of monitoring intensity by the cost-constrained market maker, who adds an additional signal to his monitoring set as long as the marginal benefit of monitoring that signal exceeds the marginal cost. The incentive to monitor arises from the picking-off risk (being hit by market orders while having stale quotes), and is positively related to the signal quality, and negatively related to the cost of monitoring. The cost of being picked off gives rise to the cost of liquidity provision and constitutes an adverse selection component of the spread.

By extending the model to include multiple trading venues, we generate theoretical predictions about the impact of fragmentation on order-to-trade ratios. The model with multiple trading venues also predicts higher OTTRs for markets with lower shares of trading volume. As markets fragment, liquidity providers have to update quotes across markets, which leads to OTTRs scaling up almost linearly with the degree of fragmentation. We find empirical evidence for this prediction in the cross-section of US stocks over 2012-2016 sample period.

Our findings suggest that message taxes can have adverse effects on market making in securities that already have disadvantageous conditions for liquidity providers. Furthermore, message taxes create unlevel competition between trading venues due to higher order-to-trade ratios on venues with lower volume shares. Finally, securities with natural signals (e.g. ETFs) always have higher OTTRs compared to common stocks, so taxing market makers in those securities would have detrimental effects on liquidity provision.

The remainder of this paper proceeds as follows. Section 2 reviews existing literature related to OTTRs, fragmentation and HFT activity. Section 3 develops a simple model of the drivers of OTTRs and outlines model propositions. Section 4 proposes empirical hypotheses based on the model predictions, tests those hypotheses through regression analysis, and discusses the empirical results. Section provides policy implications of our analysis.

2. Literature review

Academics, stock exchanges and regulators often use order-to-trade ratios as a proxy for high-frequency trading. For example, U.S. Securities and Exchange Commission, U.S.

Congressional Research Services, U.K. Government Office of Science, and European Securities and Market Authorities are among the institutions relying on OTTR in their HFT policies. Moreover, Brogaard, Hendershott, and Riordan (2014) report that some stock exchanges (e.g. NASDAQ) use OTTRs to classify HFTs. Even more important, regulatory initiatives aimed at curbing HFT activity are usually tied to OTTR. Chung & Lee (2015) survey mentions message taxes implemented in Italy, France and Norway in 2012. Germany launched an HFT regulation in 2013, aiming to decrease OTTRs by HFT firms. Australian and Canadian regulators' cost recovery programs are also based on charging messaging taxes since 2012.

However, a number of recent academic studies (Rosu, Sojli & Tham, 2017; Ye & Yao, 2015; Ye, 2017) have cast doubt on the merits of using OTTR as a proxy for HFT activity. Rosu et al. (2017) show that equilibrium OTTRs reflect a number of factors beyond HFT activity, including the asset's risk bearing capacity, dealer's inventory, cost of monitoring and monitoring precision. Ye & Yao (2015) provide empirical evidence that message-to-trade ratios are negatively related to HFT liquidity provision in a cross-section of stocks. Ye (2017) offers a theory model that explains this effect: due to their speed advantage, HFTs provide relatively higher fraction of liquidity in stocks with larger tick sizes, and once their queue priority is secured, they are less likely to cancel orders. In stocks with larger tick sizes, there are fewer HFTs, but all liquidity providers compete more on price than time priority, and hence cancel more orders.

Hence, message-to-trade ratios are not necessarily a good proxy for HFT activity, as they reflect multiple factors related to market making activity, including monitoring precision, picking-off risk, and strategic price-time priority choices by liquidity providers. From regulatory perspective, understanding the determinants of OTTR has important policy implications, as HFT-targeted messaging fees might be misdirected and harmful to market making activity. Our study investigates the determinants of OTTR recognizing the relationship between OTTR, market maker's monitoring intensity, and market fragmentation.

Existing literature can help us understand the relationship between HFT, fragmentation, messaging tax and market quality, but there are no studies we are aware of that directly tackle the question of which factors drive OTTRs. A number of recent studies explore the following related issues:

(1) the effect of HFT on liquidity and market quality;
(2) the effect of messaging tax on liquidity and market quality;

2.1. The effect of HFT on liquidity and market quality

HFT-related literature offers some rich insights into order submission and cancellation strategies of different types of traders. Van Kervel's (2015) theory model differentiates between fast and slow traders, treating the former as a source of additional adverse selection costs in fragmented markets. In their model, adverse selection costs arise due to fast traders being able to observe the order flow before slow traders. Their paper does not explicitly consider fragmentation as a model parameter, but it shows empirically that trading across multiple venues leads to trades on one venue being followed by cancellations of limit orders on competing venues. We consider this finding in light of the link between fragmentation and OTTR, but propose a different reason for order cancellations: market making across trading venues rather than competition for order flow.

Other HFT studies typically address the question of HFT impact on some aspect of market quality (liquidity, price discovery, institutional execution costs, liquidity co-movement etc.), and use either exogenous entry of HFTs (Brogaard & Gariott, 2015; Mlceniece, Malcenieks & Putnins, 2016), explicit HFT identifiers (Van Kervel & Menkveld, 2016; Goldstein, Kwan & Philip, 2017), or OTTR as a proxy for high-frequency trading (Malinova, Park, and Riordan, 2013; Hoffman, 2014; Conrad, Wahal, and Xiang, 2015; Brogaard, Hendershott and Riordan, 2016; Subrahmanyam & Zheng, 2016). HFT studies typically cite speed as a source of quote flickering that accompanies high OTTRs: Jovanovich & Menkveld (2015) show that fast and well-informed HFTs increase gains from trade if their quoting activity reduces the information asymmetry between other traders. Empirically, a number of studies document high OTTRs being related to HFTs undercutting each other as a result of market orders consuming liquidity from the order book (Hasbrouck, 2015), episodic bursts of HFT quoting activity not related to market orders (Eggington, Van Ness and Van Ness, 2016), higher variance ratios in quotes (Hasbrouck, 2015), higher noise to information ratios in order flow (Yueshen, 2015).

O'Hara (2015) mentions market fragmentation and increasing trading speeds as two core features of modern financial markets. At the same time, no studies to date have explored the link between fragmentation and one of the key manifestations of speed – order-to-trade ratios. Our paper addresses the question of how market making in fragmented markets affects order-to-trade ratios, and whether the high message traffic should be a matter of concern to regulators.

2.2. The effect of messaging tax on liquidity and market quality

Literature addressing the effects of regulatory restrictions on excessive order submissions and cancellations generally finds negative or neutral effects of messaging taxes on liquidity and market quality. For example, Van Kervel (2015) provides evidence from the sample of ten FTSE 100 stocks that imposing a cancellation fee discourages competition among trading venues and harms liquidity.

A number of studies investigate the effects of messaging taxes introduced in European countries in 2012. Caivano et al. (2012), Friedrich and Payne (2015), and Capelle-Blancard (2014) study the effect of taxing traders with excessive OTTRs (above 100:1) on Borsa Italiana (Italy's largest stock exchange). The former two studies find the tax to be detrimental to market quality (in the time span of four months), while the latter study found no effect (in the time span of three years). Similarly, Colliard and Hoffmann (2015) find no effect on market quality from the French messaging tax levied on HFTs OTTRs above 5 across all stocks. Jorgensen et al. (2014) find that Norwegian messaging tax (imposed on traders with OTTR above 70) had no harmful effects on the stocks in the treatment group, as relative spreads decreased slightly, while depth and turnover did not change. In Germany, Haferkorn (2015) finds that the price dispersion across trading venues has increased after implementation of the German HFT Act, which charges HFTs based on their OTTRs. Canadian regulator (IIROC) imposed a messaging tax as part of its cost recovery program, and charges traders proportionally to their share of submitted messages. Malinova et al. (2013) show that these measures resulted in increasing quoted and effective spreads in the Canadian market. Similarly, Lepone and Sacco (2013) find that IIROC's cost recovery program coincided deterioration in liquidity on Chi-X Canada.

While the studies mentioned above provide empirical evidence on negative to neutral effects of messaging taxes, they do not address two relevant concerns: firstly, they do not offer formal theoretical models for why taxing messaging is harmful to liquidity; secondly, they do not investigate the heterogeneity of these effects in the cross-section of stocks. Thus, we fill the gap in existing literature by investigating both of these issues.

3. A simple model of what drives the order-to-trade ratio

3.1. Baseline model structure

Consider a simple model in which a market maker posts quotes (bid and ask prices and quantities) for a given asset in a given market. The market maker could monitor one or more signals from a set of signals, $\{s_1, s_2, \ldots, s_N\}$. Each signal is a time-series (e.g., a price in a related security, price of the same security in another market, an order book state, etc.) that changes at stochastic times given by Poisson processes with intensity λ_i for the i^{th} signal. The quality of signal i, q_i, is the probability that when there is a change in that signal ("information" arrival), the market maker will want to update his posted quoted price(s) or quantities (we term such events "relevant information" arrivals), resulting in a "cancel and enter" or "amend" message from the market maker.[2]

There is a cost to monitoring a signal, with the cost per unit time being proportional to the intensity of information conveyed by the signal (changes in the signal), $\lambda_i c$. This cost can be interpreted as the additional processing capacity that is required to interpret information arrivals and determine whether to respond, without delaying reactions to other signals.

Market orders arrive and trade against the market maker's posted quotes at stochastic times given by a Poisson process with arrival rate λ_m. The market maker's benefit from monitoring comes from avoiding having stale quotes picked off. When a market order arrives after a relevant information arrival but the market maker has not updates their quotes in response to the information (this occurs when relevant information arrives for a signal that is not monitored by the market maker) then the market maker's (stale) quotes are picked off and he incurs a picking-off cost, k. The more signals the market maker monitors, the lower the probability (frequency) of his quotes being picked off, because the more of the relevant information he has through his monitoring. For a given monitoring intensity, the picking-off cost per unit time increases with the asset's fundamental volatility (frequency of useful information arrivals) because of more frequent relevant information that makes quotes stale unless monitored.

The market maker chooses which signals (if any) to monitor by weighing up the costs of monitoring, $\lambda_i c$, against the benefits of monitoring, namely reducing picking off risk. The benefits depend on the arrival intensity of market orders and the arrival intensity of relevant information. Hence, the choice of monitoring intensity is endogenous in the model.

---

[2] To be more precise, two messages, if the market maker adjusts both the bid and the ask.

We define a signal's usefulness, u_i, as the arrival intensity of relevant information from the signal (signal changes that cause the market maker to want to revise his quotes): $u_i = \lambda_i q_i$. The expected benefit (per unit time) from monitoring a given signal i is the saved losses from having avoided having quotes picked off. That benefit is the expected number of times the market maker's quotes would be hit by a market order when he would have wanted to revise them had he seen the signal, multiplied by the cost of getting hit by a market order without having updated quotes, k. In one unit of time, the expected number of market order arrivals is λ_m and the probability that a given market order is preceded by useful information from signal i is $\frac{\lambda_i q_i}{\lambda_m + \lambda_i q_i}$. Therefore, the benefit per unit time of monitoring signal i is $\lambda_m \left(\frac{\lambda_i q_i}{\lambda_m + \lambda_i q_i}\right) k$.

As a result of monitoring signals, executing trades, and updating quotes, the market maker generates messaging activity (messaging includes order entry, cancelation, and amendment messages) at an expected rate of $Q = 2\sum_{i\in\{MonitoredSignals\}} \lambda_i q_i + 2\lambda_m$ messages per unit time. The first term, $2\sum_{i\in\{MonitoredSignals\}} \lambda_i q_i$ is due to quote updates in response to relevant information arrivals on monitored signals, and the second term, $2\lambda_m$, is due to re-posting liquidity after being hit by a market order (re-entering one quote and amending the other).[3]

Recognising that the expected number of trades per unit time is just the market order arrival intensity, λ_m, the ***order-to-trade ratio***[4] for the asset is given by $OTTR = \frac{2\sum_{i\in\{MonitoredSignals\}} \lambda_i q_i + 2\lambda_m}{\lambda_m}$.

3.2. Equilibrium

To solve for the endogenous choice of monitoring, we set the marginal benefit of monitoring i^{th} signal, $\lambda_m \left(\frac{\lambda_i q_i}{\lambda_m + \lambda_i q_i}\right) k$, equal to the marginal cost of monitoring, $\lambda_i c$.

Recall the cost per unit time of monitoring signal i is $\lambda_i c$, giving a net benefit of $\lambda_m \left(\frac{\lambda_i q_i}{\lambda_m + \lambda_i q_i}\right) k - \lambda_i c$ from monitoring the signal. The market maker adds signals to his "monitored list" from greatest to least net benefit until the marginal expected net benefit of adding the next signal is less than or equal to zero. The market maker therefore monitors all

[3] Both terms ($\sum_{i\in\{MonitoredSignals\}} \lambda_i q_i$ *and* λ_m) are multiplied by two reflecting the fact that after observing useful information or being hit by a market order, the market maker updates his view of the fundamental value and thus adjusts both bid and ask prices.

[4] We define the order-to-trade ratio as the total number of messages (order entry, cancellation, and amendment) divided by the total number of trades. In some industry settings, this ratio is referred to as the message-to-trade ratio.

signals for which $\lambda_m \left(\frac{\lambda_i q_i}{\lambda_m + \lambda_i q_i}\right) k - \lambda_i c > 0$ with the set of monitored signals denoted $\{MonitoredSignals\}$. This condition determines ***monitoring intensity***. Monitoring intensity is calculated as $M = \sum_{i \in \{MonitoredSignals\}} m_i$, where $m_i = 1\ \forall i \in \{MonitoredSignals\}$, and $m_i = 0\ \forall i \notin \{MonitoredSignals\}$.

3.3. Model with fragmented markets

If the number of markets increases from 1 to N, the single market maker posts liquidity across multiple venues. The market order arrival rate, λ_m, is assumed to be the same as in one-market case: the trade volume fragments across multiple venues, but stays unchanged from the overall market perspective. The overall quoting activity of the market maker consists of two components: (a) quote updates resulting from signal monitoring, $2N \sum_{i \in \{MonitoredSignals\}} \lambda_i q_i$ (market maker updates quotes on all N markets in response to monitored signals), and (b) reposting liquidity after getting a fill on market orders, $2\lambda_m$. Note that market fragmentation does not affect the signal monitoring problem of the market maker, who chooses the set of signals to monitor in the same manner as in a single-market case. The resulting order-to-trade ratio for the market overall is therefore $OTTR = \frac{2N \sum_{i \in \{MonitoredSignals\}} \lambda_i q_i + 2\lambda_m}{\lambda_m}$.

Consider the order-to-trade ratio of individual markets $k = 1 \dots N$. The market share of each individual market k is ρ_k. We assume that market orders are divided across markets proportionally to their respective market shares. The market maker updates his quotes on market k every time there is a signal update or after being hit by market order. Then, the order-to-trade ratio for market k is $OTTR_k = \frac{2 \sum_{i \in \{MonitoredSignals\}} \lambda_i q_i + 2\rho_k \lambda_m}{\lambda_m \cdot \rho_k}$.

3.4. Propositions

We now derive theoretical propositions about the relations between order-to-trade ratios, monitoring intensity and fragmentation. First, we establish the link between the two key variables of interest: order to trade ratios and fragmentation (propositions 1a and 1b). Second, we show how order to trade ratios are related to fragmentation (proposition 2). Third, we relate order to trade ratios to all the model parameters that affect OTTR directly, via monitoring intensity or both. In the next section, we build on these propositions to develop the testable hypotheses.

Proposition 1a. As markets fragment, market-wide order-to-trade ratio for a given security increases with the extent of fragmentation, if there is at least one non-zero quality signal in the monitoring set.

Proof. See Appendix 1A.

The intuition for this result follows from the nature of market making across multiple venues. As markets fragment, a market maker has to post quotes across several exchanges, hence for a given level of trading activity, his quoting activity will increase, driving order-to-trade ratios up. This occurs as long as the market maker has a reason to update quotes: arrival of useful information about the fundamental value of the asset (aka non-zero quality signal to act on) or new fills on market orders that require re-posting liquidity. Because we assume trading activity to be non-zero in every state of the world ($\lambda_m > 0$ by the properties of Poisson process), the only condition for this proposition to hold is non-zero quality of the signals. In practical terms, if this condition is not satisfied, and market makers' signals are too noisy to be useful (e.g. in market crash events), the market maker withdraws from the market, and the order to trade ratio becomes irrelevant.

Proposition 1b. As markets fragment, order-to-trade ratio for a given security on a given market increases as the market share of that market decreases.

Proof. See Appendix 1B.

When trade volume fragments across multiple trading venues, it is natural to expect higher order-to-trade ratios for the venues with lower volumes, if we keep overall market-wide trading activity and quoting activity constant. This is another way of saying that other things equal, venues with lower share of trading volume will naturally have higher order-to-trade ratios.

Proposition 2. Order-to-trade ratio for a given security increases with monitoring intensity.

Proof. See Appendix 2.

Monitoring intensity and order-to-trade ratios are closely related, because the market maker posts quotes as a result of his monitoring activity. If his cost-benefit analysis leads the market maker to monitor more and hence react to more signals, he will post more quote updates per unit of time. This means that order-to-trade ratio increases with more monitoring, hence parameters that affect monitoring intensity also affect order-to-trade ratios, and the effect is in the same direction. In further propositions, we will rely on this result to derive predictions about how the model parameters affect order-to-trade ratio.

Proposition 3. Order-to-trade ratio for a given security increases with the quality of signals available for monitoring.

Proof. See Appendix 3.

When a market maker gets access to better quality signals, his monitoring becomes more profitable and he has an incentive to monitor more. This effect follows from higher probability of observing a useful signal as the signal quality improves. With higher monitoring intensity, the market maker posts more quote updates and hence the order-to-trade ratio increases.

Note that it is the signal quality, not the number of signals available for monitoring, that that drives this result. Because the potential number of signals that can be monitored is infinite, signal quality rather than quantity determines how many signals the market maker chooses to monitor.

Proposition 4. Order-to-trade ratio for a given security increases with picking-off risk.

Proof. See Appendix 4.

When faced with higher frequency of being picked off, the market maker has an incentive to monitor more signals in order to minimize the costs of being hit by market orders without having updated quotes. Therefore, higher picking-off risk leads to higher monitoring intensity and higher order-to-trade ratios.

Proposition 5. Order-to-trade ratio for a given security decreases with monitoring cost.

Proof. See Appendix 5.

When the market maker faces higher cost per signal monitored, his marginal costs increase, hence leading him to decrease the monitoring intensity and order-to-trade ratios. Market maker's marginal costs are proportional to signal intensity, so the effect on monitoring intensity and OTTR is higher for more higher intensity signals.

Proposition 6. Order-to-trade ratio for a given security decreases with the trading frequency, holding the monitoring intensity constant.

Proof. See Appendix 6.

The effect of trading frequency on OTTR is two-fold. On one hand, higher intensity of market order arrivals increases monitoring intensity, as the market maker has an incentive to monitor more in order to avoid picking-off costs. Therefore, he posts more quote updates based on signals monitored, which drives up order-to-trade ratio. On the other hand, higher market orders intensity decreases OTTR every trade is associated with fewer quote updates on average.

Hence, if we keep the number of signals in the monitoring set constant (aka constant monitoring intensity), only the second effect takes place: OTTR decreases with trading frequency.

4. Empirical analysis

We structure the empirical analysis part of this paper around two objectives: firstly, we propose the testable hypotheses based on the theoretical propositions outlined in the previous section; secondly, we use regression analysis to test the hypotheses and propose policy implications.

4.1. Hypothesess

Empirically, order-to-trade ratios are directly observable in the order book data and vary both over time and in a cross-section of stocks. At the same time, the degree of fragmentation in a given stock can be proxied by the number of markets a stock trades on, as well as by Herfindahl-Hirschman index. based on volume or number of trades (as in Degryse, de Jong and van Kervel, 2015 and Malceniece, Malcenieks & Putnins, 2016). It is therefore straightforward to test the relationship between OTTR and fragmentation empirically. Going forward, we refer to observations for a given stock on a given day as stock-day observations, and observations for a given market on a given day as market-day observations. To extract the relationship between quoting activity and fragmentation that is not contaminated by other cross-sectional dependencies, we have to control for trade volume, stock and market characteristics. Hypotheses 1a and 1b follow directly from propositions 1a and 1b:

Hypothesis 1a. Order-to-trade ratios are higher for stock-days with higher degrees of fragmentation.

Hypothesis 1b. Order-to-trade ratios are higher for markets[5] with lower market shares.

The empirical counterpart of signal quality could be seen as the degree of co-movement between related securities. For example, to the extent that prices of two securities co-move, one security could be used as a signal for another security's value. One example could be the relationship between ETFs and their underlying stocks: because ETFs derive their value from the underlying components, their signal quality is always higher than signal quality for stocks (which do not have such precise signals to be monitored). Hence, based on proposition 3, we would expect higher order-to-trade ratios for ETFs than for their underlying components.

[5] In regression analysis, we use market-day units of observation, because market shares, as well as stock and market characteristics vary over time, as well as across markets.

Hypothesis 2a. ETFs have higher order-to-trade ratios compared to the common stocks.

From the market maker's perspective, a broad market index is the most obvious signal for security's value. Hence, to the extent that the market index constitutes a better signal (as is the case for more correlated securities), the market maker will update his quotes more often in response to changes in the index.

Hypothesis 2b. Securities with higher correlation with the broad market index have higher order-to-trade ratios.

The market maker has an incentive to update his quotes more often if he risks losing more per each picking-off event from market orders, and with higher frequency. More frequent value updates and wider value fluctuations are the case for stocks with higher fundamental volatility, and also during days with higher market volatility; hence, based on proposition 5, we would expect order-to-trade ratios to be higher in such cases.

Hypothesis 3a. Order-to-trade ratios are higher for stock-days with higher market volatility.

Hypothesis 3b. Order-to-trade ratios are higher for stock-days with higher stock volatility.

The tick size constrains the degree of granularity at which market makers can update their quotes based on the new information. In practice, the same signal might yield different usefulness for two stocks with different price ranges (and otherwise similar characteristics). For example, for low-priced stocks that have artificially constrained relative spread due to the minimum tick size, the market maker would be less likely to update quotes, as the value effect on the stock price could lie within the spread. On the opposite side of the spectrum, for high-priced stocks with narrow relative spread, the same signal could induce the market maker to update quotes, as the difference in valuation would be more likely to lie outside the spread. Because market makers face the risk of being picked off every time they do not update their quotes in response to useful information about the stock value, based on proposition 4, we would expect higher OTTR in stocks with smaller tick-to-price ratios.

Hypothesis 3c. Order-to-trade ratios are higher for stocks with higher tick-to-price ratios.

Monitoring costs faced by the market maker affect his choice of optimum monitoring. The market maker will choose to monitor more (and post more quote updates as a result) if his marginal costs of monitoring are lower. Hence, to the extent that HFTs face lower the costs of monitoring, proposition 5 suggests that order-to-trade ratios will be lower for stock-days and exchange-days that attract more HFT activity. O'Hara (2015) and Rosu et al. (2017) suggest that high market cap stocks attract more HFT activity, hence we would expect lower monitoring costs and higher OTTRs in those stocks.

Hypothesis 4a. Order-to-trade ratios are higher for stocks with higher market capitalization.

Similarly, O'Hara (2015) argues that taker-maker markets attract relatively fewer HFTs, hence we would expect higher monitoring costs and lower OTTRs for the average stock traded on those markets.

Hypothesis 4b. Order-to-trade ratios are lower on markets with taker-maker fee structure.

Trading volume is one of the key stock characteristics that affects both monitoring intensity and OTTR. Interestingly, monitoring intensity increases with the frequency of market orders to the extent that the market maker chooses to add new signals to the monitoring set. However, OTTR decreases with trading frequency as more market orders are executed. Empirically, it is important to control for the extent that trading frequency (proxied by daily trading volumes) affects OTTR to disentangle the effect of other factors. In line with proposition 7, we expect higher OTTR for stocks with lower trading volumes.

Hypothesis 5. Order-to-trade ratios are inversely related to the trading volumes, controlling for fragmentation, stock and market characteristics.

4.2. Regression analysis

Our regression specifications follow from the hypotheses specified in the previous subsection. We estimate separate regression models for stock-date level observations and for exchange-date level observations. To account for within-cluster correlations (i.e. correlations within exchange-date groups and stock-date groups), we use double-clustered standard errors. Regression models (1) and (2) are estimated for stock-date and exchange date regressions accordingly.

$$\log(1 + OTTR_{it}) = \boldsymbol{\alpha} + \boldsymbol{\beta_1} Frag_{it} + \boldsymbol{\beta_2} \log(Volume_{it}) + \boldsymbol{\beta_3} \log(MarketCap_{it}) + \boldsymbol{\beta_4}\, MarketVolatility_t + \boldsymbol{\beta_5}\, StockVolatility_{it} + \boldsymbol{\beta_6}\, CorrelationS\&P_{it} + \boldsymbol{\beta_7} TickToPrice_{it} + \boldsymbol{\beta_8}\, D_{it}^{ETF} + \varepsilon_{it}$$
(1)

$$\log(1 + OTTR_{jt}) = \boldsymbol{\alpha} + \boldsymbol{\beta_1} Frag_{jt} + \boldsymbol{\beta_2} \log(Volume_{jt}) + \boldsymbol{\beta_3}\, MarketVolatility_t + \boldsymbol{\beta_4}\, CorrelationS\&P_{jt} + \boldsymbol{\beta_5}\, TickToPrice_{jt} + \boldsymbol{\beta_6}\, D_{jt}^{taker} + \boldsymbol{\beta_7}\, MarketShare_{jt} + \varepsilon_{jt} \quad (2)$$

To prevent our results from being driven by a few extreme observations, we winsorize the $OTTR_{it}$ variable and obtain a logarithmic transformation of it to be used in regression analysis. Further, we also obtain logarithmic transformations of market cap, volume and the VIX index. See table 1 for detailed variable definitions.

< Table 1 here >

4.3. Data and descriptive statistics

We use SEC MIDAS (Market Information Data Analytics System) database and CRSP Daily Stocks as two primary data sources. The MIDAS data covers the universe of US stocks and ETFs traded across 12 major lit markets (Arca, Bats-Y, Bats-Z, Boston, CHX, Edge-A, Edge-X, NSX, PHLX, Amex, NYSE), and contains the variables necessary to compute order to trade ratios and fragmentation measures[6]. We obtain daily data on stock characteristics from CRSP to complement the MIDAS data, and use Thomson Reuters Tick History to obtain the daily values of VIX index.

Our sample period spans from January 1st 2012 (the starting date of MIDAS dataset) till December 31st 2016 (the latest date for CRSP dataset). The combined dataset contains daily frequency observations, with stock- and exchange-level granularity. We aggregate the data to stock-day level for the first part of our analysis (exploring how OTTR varies over time in the cross-section of securities), and to exchange-day level for the second part of analysis (exploring how OTTR varies over time across markets).

The descriptive statistics for the stock-date dataset is presented in Table 2. The dataset contains just under 6 million daily observations for 7114 securities, 75% of them stocks, and the rest – ETFs. At stock-date frequency, stocks and ETFs are part of the same dataset, and we have a dummy variable for ETFs that lets us control for ETF-specific characteristics beyond those suggested to drive OTTRs based on the theory model. We also create dummy variables

[6] Note that we compute order to trade ratios following the SEC methodology: dividing the order volume by lit volume. By SEC definition, order volume is sum of order volume (in number of shares) for all add order messages; lit volume is sum of trade volume for trades that are not against hidden orders.

for stock-days affected by the SEC Tick Size Pilot program, and apply the wider tick sizes accordingly.[7]

< Table 2 here >

The descriptive statistics for exchange-date dataset is presented in table 3. We split this dataset into two to run the regression analysis for stocks and ETFs separately. This is because in exchange-date analysis we weight security-level variables by dollar volume (e.g. fragmentation measures, correlations with S&P500 and tick to price measures are all dollar volume weighted to obtain their exchange-date equivalents). Hence, to the extent that ETFs and stocks have inherently different characteristics as security classes, we cannot combine them at exchange-date granularity.

In exchange-date analysis, we distinguish between the markets with different fee structures by introducing a dummy variable for taker-maker markets (Edge-A, Bats-Y and Boston stock exchange)[8]. Among stocks, 17% of exchange-date observations belong to the taker-maker group, and among ETFs – 20%.

< Table 3 here >

We also use two variables with only time variation (no cross-sectional variation), which are proxies for market volatility. The descriptive statistics for those is presented in table 4. The first proxy for market volatility is computed from daily high-low range of SPY ETF daily prices, while the second proxy is a log-level measure of daily closing VIX index.

< Table 4 here >

4.4. Regression results

Regression results at security-date level test the empirical predictions of our theory model. To preview the results, we find evidence that order-to-trade ratios increase with fragmentation (in line with Hypothesis 1a), and are also higher for stocks with lower volumes (in line with Hypothesis 5), larger market cap (in line with Hypothesis 4a), higher correlations

[7] The Tick Size Pilot program affects 1400 small capitalization stocks by widening their tick sizes from \$0.01 to \$0.05. The rollout of the program started on October 3rd, 2016, and occurred in several phases for three groups of securities affected. We use the official data from FINRA (The Financial Industry Regulatory Authority) web-site to identify the affected securities and effective rollout dates.

[8] "Maker-taker" market refers to the market that compensates "liquidity makers" (i.e. those posting limit orders) and charges "liquidity takers" (i.e. those posting market orders). "Taker-maker" market refers to the trading venue that does the opposite (i.e. charges for limit orders and compensates for market orders). In our sample, 9 trading venues apply maker-taker fee structure: Amex, Arca, Bats-Z, CHX, Edge-X, NSX, NYSE, NASDAQ, PHLX; 3 trading venues – taker maker fee structure: Edge-A, Bats-Y, and Boston stock exchange.

with the market index (in line with Hypothesis 2b), and smaller tick-to-price ratios (in line with Hypothesis 3c). Order-to-trade ratios for ETFs are higher than those for stocks, controlling for other security characteristics (in line with Hypothesis 2a). Stock and market volatility are also positively associated with order-to-trade ratios of a given stock on a given day (in line with Hypotheses 3a and 3b). Empirical results for stock-date and exchange-date regressions are reported in Tables 5 and 6 respectively.

< Table 5 here >

The empirical result that OTTRs increase with the degree of fragmentation confirms the prediction from our theory model (see Proposition 1). This result is expected, as higher fragmentation means posting liquidity across multiple venues. This in turn leads to order revisions increasing proportionally to the number of venues, because market makers revise quotes across multiple exchanges in response to monitored signals. The positive relation between fragmentation and OTTR indeed holds on average in the stock-day panel, as suggested by regression results in Table 5: the coefficient on fragmentation is positive and significant for all three fragmentation proxies. To better understand the shape of this relationship, we also regress OTTRs on dummy variables of different degrees of fragmentation, controlling for other stock characteristics. Figure 1 shows that OTTRs increase almost linearly as fragmentation increases, in line with the model predictions. This is a novel finding, as no studies to date have investigated the relationship between fragmentation and OTTR.

< Fig. 1 here >

The quality of signals available for monitoring also affects order-to-trade ratios. While multiple studies have explored the link between HFT quoting and monitoring activity (e.g. Liu, 2009; Conrad et al., 2015; Lyle et al., 2016; Blocher et al., 2016), the reasons for monitoring in our model are related to market making and avoiding picking-off risk rather than speed competition among HFTs. Empirically, we find evidence for this effect by examining OTTRs in ETFs: the latter have high quality signals available for monitoring, unlike stocks. This leads to more intense monitoring activity by market makers, keeping all other security characteristics constant. We find evidence of this theoretical prediction, as the coefficient on ETF dummy is positive and significant. Another measure of monitoring – correlation with S&P 500 index – is also positively associated with OTTRs, and highly significant. Indeed, to the extent that market makers can derive highly useful information from the available benchmarks, they will have an incentive to update the quotes more frequently in order to avoid the picking-off risk.

The risk of being picked off by the informed traders drives market maker's monitoring decisions and hence OTTRs. The picking-off risk is related to how often the quotes in a given stock need to be updated to keep up with the changes in fundamental value. The frequency and magnitude of such changes in fundamental value is higher for stocks with more volatile process, and also – under more volatile market conditions. Empirically, we find that to be the case, as coefficients on both market and stock volatility are positive and significant.

Another proxy for picking-off risk is tick-to-price ratio. In stocks with higher tick-to-price ratios, it would take a larger change in fundamental value to induce a market maker to update quotes, implying lower pick-off risk. To illustrate this, consider two stocks. Stock A is priced at $50, and stock B – at $5. Say, a tick size is $0.01, and stock A quotes are $49.99-$50, while stock B quotes are $4.99-$5. If a piece of news comes out, implying 2 bps improvement in the stock price, the market maker will update the quotes in A (shifting the midquote from $49.995 to $50.005, as the new bid-ask becomes $50-$50.01). However, a market maker in stock B will not update quotes, as the value change lies within the bid-ask spread (2 bps improvement translates into $0.0001 value, which is smaller than full tick size). In this simple example, market maker in security A faces higher risk of being picked off than in security B. This is the case because if he allows for stale quotes (aka does not react to the signal) in security A, the chance of losing out to informed traders is high, but in security B stale quotes are not as likely to be picked off, as it takes an event with higher value implications to move the price.

Empirically, we find evidence supporting the prediction of higher picking-off risk (lower tick-to-price ratio) being associated with higher OTTRs. This is in line with evidence in Ye & Yao (2015), although the theoretical argument proposed by Ye (2017) points towards the speed vs price competition by HFTs as a theoretical mechanism for this effect. Our model suggests a different mechanism – picking-off risk – although the two need not be mutually exclusive. In fact, our model might help explain why, as suggested by Ye (2017) HFTs compete more on price rather than time priority in low tick-to-price stocks: it is because their speed advantage allows them to more effectively avoid being picked off by reacting rapidly to information arrivals through adjusting their quotes. This, in turn, leads to higher order-to-trade ratios.

Monitoring cost is one of the key drivers of the endogenous monitoring intensity in our model. Hence, to the extent that lower monitoring cost increases the net marginal benefit of monitoring, the market maker will monitor more and hence increase his OTTR. Since market maker's costs are not directly observable, we use two proxies that previous studies have shown to be highly correlated with the HFT activity: stock's market cap and trading venue's maker-taker fee structure. Because HFT's investment in technology enables them to achieve low marginal costs of monitoring, relative to other market participants, the prevalence of HFTs should come together with low monitoring costs.

As shown in O'Hara (2015), and Rosu et al. (2017), large-cap stocks attract more HFT activity, which in turn suggests lower cost of monitoring. Empirically, we find log market cap is strongly positively related to OTTR (see Table 5).

O'Hara (2015) also argues that the prevalence of HFTs is lower on taker-maker markets, as it is more attractive for them to collect market making rebates on maker-taker markets. Therefore, the cost of monitoring is lower for stocks traded on maker-taker markets. Then, taker markets should have relatively fewer HFTs (hence higher monitoring costs and lower OTTRs). Indeed, we find that the coefficient on our taker dummy is negative and significant in exchange-date regressions (see Table 6).

Order-to-trade ratios are also negatively related to the trading frequency, which we proxy by the number of shares trades in a day. Controlling for this effect is also important in order to view the results from the standpoint of quoting activity and avoid them being driven by the mechanical division by trading volume.

Exchange-day analysis (results reported in Table 6) allows us to explore the effects of fee structures and market shares on OTTRs. As discussed earlier, taker-maker trading venues generally have lower OTTRs, in line with higher monitoring costs associated with less prevalent HFT activity. Characteristics of the average stock traded on a particular venue (weighted by dollar volume) are included as controls, and generally point in the same direction as in stock-day regression discussed above. Market volatility variables are also included and are positively related to OTTRs.

< Table 6 here >

Contrary to our theoretical prediction (Proposition 1b), market share is not significantly related to OTTRs (see Table 6) in the regression results. Further exploratory analysis (see Figure 2) reveals that when a stock trades on only a few markets (low fragmentation, e.g. bucket 19 on the graph), those tend to be low-OTTR markets (e.g. NASDAQ, Edge-X, Arca).

< Fig. 2 here >

Indeed, larger exchanges (by overall market share) tend to trade more of the least fragmented stocks, However, as a stock starts trading cross multiple venues, more high-OTTR markets are added to the list. Taker-maker maker markets seem to be the last on this pecking order: their market share is highest in the most fragmented stocks. Smallest market share markets do not necessarily follow the same dynamics, but they tend to have the highest OTTRs.

5. Conclusions and policy implications

Our results help explain the drivers of order-to-trade ratios across markets and in a cross-section of stocks. Beyond that, we also provide the theoretical and empirical arguments for how OTTR-related regulatory measures might affect market quality. Specifically, our theory model addresses the question of messaging tax effects on market making activity across different assets.

Firstly, our model suggests that high order-to-trade ratios are related to market making in fragmented markets. Hence, it should not be surprising to find increasing OTTRs as technology enables ever-faster incorporation of information through quote revisions, while allowing to instantaneously revise quotes across trading venues. Importantly, as markets fragment, market makers will inevitably scale up their quoting activity, even in the absence of other factors (e.g. higher speed or lower cost of monitoring). This prediction is supported by empirical evidence from the US markets.

Secondly, we show that heterogeneity in order-to-trade ratios is related to a number of stock and market characteristics that affect market makers' monitoring intensity: trading frequency, market volatility and correlation, market cap and tick-to-price ratios. To the extent that regulatory measures (e.g. messaging taxes) are targeted at stocks with high OTTRs, they will decrease market making activity in stocks with naturally high OTTRs. For example, stocks with low trading frequencies and high volatilities (arguably already less attractive to market makers) would be disadvantaged more if a messaging tax were to be introduced.

Thirdly, the model demonstrates that derivative securities with natural signals (e.g. ETFs) will have higher order-to-trade ratios, controlling for other factors. Because in some markets, regulators tax market makers based on messaging traffic, it might disproportionally harm liquidity provision in asset classes with naturally high order-to-trade ratios (e.g. ETFs). Similarly, to the extent that designated market makers are mandated in certain asset classes, and at the same time taxed based on messages, higher cost of liquidity provision will arise as a result of such requirements.

Appendix 1A

Proposition 1a. As markets fragment, market-wide order-to-trade ratio for a given security increases with the extent of fragmentation, if there is at least one non-zero quality signal in the monitoring set.

Proof.

Recall the expression for order-to-trade ratio from the overall market perspective: $OTTR = \frac{2N\sum_{i\in\{MonitoredSignals\}}\lambda_i q_i + 2\lambda_m}{\lambda_m}$. Taking the first derivative with respect to the number of markets:

$\frac{dOTTR}{dN} = \frac{2\sum_{i\in\{MonitoredSignals\}}\lambda_i q_i}{\lambda_m} > 0$. Below, we show that this expression is strictly positive.

One can show that $\frac{2\sum_{i\in\{MonitoredSignals\}}\lambda_i q_i}{\lambda_m} = OTTR - 2 > 0$, if $OTTR > 2$.

Intuitively, $OTTR \geq 2$ as it takes at least two messages to generate a trade: posting both a bid and an ask quote. If no additional information is obtained from the signals (i.e. signal quality is 0), $OTTR = 2$, which is the case only if $q_i = 0\ \forall i \in \{MonitoredSignals\}$. As $q_i \geq 0$ by construction (signal quality cannot be negative), $OTTR > 2$ for all cases except for $q_i = 0$.

Appendix 1B.

Proposition 1b. As markets fragment, order-to-trade ratio for a given security on a given market increases as the market share of that market decreases.

Proof.

Recall the expression for order-to-trade ratio from the individual market perspective: $OTTR_k = \frac{2\sum_{i\in\{MonitoredSignals\}}\lambda_i q_i + 2\rho_k\lambda_m}{\lambda_m \cdot \rho_k}$. Taking the first derivative with respect to the market share:

$$\frac{dOTTR_k}{d\rho_k} = \frac{2}{\rho_k} - \frac{2\sum_{i\in\{MonitoredSignals\}}\lambda_i q_i + 2\rho_k\lambda_m}{\lambda_m \cdot \rho_k^2} < 0\ \forall \rho_k \in (0,1),\ \lambda_i, q_i, \lambda_m$$

Appendix 2.

Proposition 2. Order-to-trade ratio for a given security increases with monitoring intensity.

Proof.

Recall that order-to-trade ratio is calculated as $OTTR = \frac{2\sum_{i\in\{MonitoredSignals\}} \lambda_i q_i + 2\lambda_m}{\lambda_m}$, while monitoring intensity is the number of monitored signals in $\{MonitoredSignals\}$ set. Therefore, as more signals are monitored, the market maker posts proportionally more quote updates in response to those signals, which in turn increases the order-to-trade ratio.

Appendix 3.

Proposition 3. Order-to-trade ratio for a given security increases with the quality of signals available for monitoring.

Proof.

As shown in proposition 2, OTTR increases with monitoring intensity. Let us show that monitoring intensity increases with the quality of monitored signals.

Recall that monitoring intensity is a count of all monitored signals: $M = \sum_{i\in\{MonitoredSignals\}} m_i$, where $m_i = 1\ \forall i \in \{MonitoredSignals\}$, and $m_i = 0\ \forall i \notin \{MonitoredSignals\}$. A market maker monitors all signals for which the marginal benefit of monitoring exceeds the marginal cost ($\lambda_m \left(\frac{\lambda_i q_i}{\lambda_m + \lambda_i q_i}\right) k > \lambda_i c$). Because improved signal quality increases the marginal benefit of monitoring without affecting the marginal costs, the market maker will monitor more when he receives better quality signals.

Because monitoring intensity increases with signal quality, and higher monitoring intensity leads to higher OTTR, we've shown that OTTR increases with the quality of monitored signals.

Appendix 4.

Proposition 4. Order-to-trade ratio for a given security increases with picking-off risk.

Proof.

Recall that the cost of being picked off is k per each event of getting hit by market order without having updated quotes. Taking the derivative of net marginal benefit function with respect to picking off risk k:

$\frac{d[\left(\frac{\lambda_i q_i}{\lambda_m+\lambda_i q_i}\right)k_m-\lambda_i c]}{dk} = \frac{\lambda_i q_i}{\lambda_m+\lambda_i q_i} > 0$ – this expression is strictly positive for all non-zero quality signals, hence monitoring intensity (and OTTR – based on proposition 2) increases with picking off risk.

Appendix 5.

Proposition 5. Monitoring intensity for a given security decreases with monitoring cost.

Proof.

Higher monitoring cost per signal increases the marginal cost to market maker, as for every signal monitored he has to pay a higher cost, which increases with that signal's intensity. Other things equal, higher marginal cost of monitoring will induce the market maker to monitor fewer signals.

Taking the first derivative of marginal net benefit with respect to c:

$\frac{d[\lambda_m\left(\frac{\lambda_i q_i}{\lambda_m+\lambda_i q_i}\right)k-\lambda_i c]}{dc} = -\lambda_i < 0$ - this expression is strictly negative, as $\lambda_i > 0$ by the properties of Poisson process (signal intensity – the number of signal updates per unit of time - can only be a positive number). Hence, the market maker will be less likely to monitor signal i when the cost of monitoring cost is lower.

Appendix 6.

Proposition 6. Order-to-trade ratio for a given security decreases with the trading frequency, holding the monitoring intensity constant.

Proof.

To establish the effect of trading frequency on OTTR, we first show the effect on monitoring intensity. Taking the first derivative of marginal net benefit with respect to trading frequency:

$\frac{d[\lambda_m\left(\frac{\lambda_i q_i}{\lambda_m+\lambda_i q_i}\right)k-\lambda_i c]}{d\lambda_m} = \frac{k\lambda_i^2 q_i^2}{(\lambda_m+\lambda_i q_i)^2} > 0$ – this expression is strictly positive for all values of parameters except for $q_i = 0$. Thus, monitoring intensity increases with trading frequency as long as the signal quality is non-zero. Intuitively, as the expected number of market order arrivals λ_m (or trading frequency) increases, the picking-off intensity increases, thus incentivizing the market maker to monitor more.

However, trading intensity also enters OTTR directly by affecting the quoting activity (quote updates after fills on market orders) and number of trades executed. Recall the expression for order-to-trade ratio: $OTTR = \frac{2\sum_{i\in\{MonitoredSignals\}}\lambda_i q_i + 2\lambda_m}{\lambda_m}$. Taking the first derivative with respect to trading frequency (λ_m): $\frac{dOTTR}{d\lambda_m} = \frac{-2\sum_{i\in\{MonitoredSignals\}}\lambda_i q_i}{\lambda_m^2} < 0$ for all parameter values except for $q_i = 0$. This suggests that OTTR decreases as the trading frequency increases.

Thus, order-to-trade ratio decreases with trading frequency, if we keep the monitoring intensity constant.

References

Biais, B., Bisiere, C., Spatt, C., 2010. Imperfect competition in financial markets: An empirical study of Island and Nasdaq. Management Science 56, 2237–2250.

Biais, B., Martimort, D., Rochet, J., 2000. Competing mechanisms in a common value environment. Econometrica 68, 799–837.

Bloomfield, R., O'Hara, M., Saar, G., 2015. Hidden liquidity: Some new light on dark trading. Journal of Finance 70, 2227–2273.

Boni, L., Brown, D.C., Leach, J.C., 2013. Dark pool exclusivity matters. Unpublished working paper. University of New Mexico.

Boulatov, A., George, T.J., 2013. Hidden and displayed liquidity in securities markets with informed liquidity providers. Review of Financial Studies 26, 2095–2137.

Buti, S., Rindi, B., 2013. Undisclosed orders and optimal submission strategies in a limit order market. Journal of Financial Economics 109, 797–812.

Buti, S., Rindi, B., Werner, I.M., 2011. Diving into dark pools. Unpublished working paper. University of Toronto.

Buti, S., Consonni, F., Rindi, B., Wen, Y., Werner, I.M., 2014. Sub-penny and queue jumping. Unpublished working paper. Bocconi University.

Buti, S., Consonni, F., Rindi, B., Wen, Y., Werner, I.M., 2015. Tick size: Theory and evidence. Unpublished working paper. Bocconi University.

Chordia, T., Roll, R., Subrahmanyam, A., 2008. Liquidity and market efficiency. Journal of Financial Economics 87, 249–268.

Comerton-Forde, C., Malinova, K., Park, A., 2015. The impact of the dark trading rules. Report prepared for IIROC.

Comerton-Forde, C., Putniņš, T.J., 2015. Dark trading and price discovery. Journal of Financial Economics 118, 70–92.

Degryse, H., de Jong, F., van Kervel, V., 2015. The impact of dark trading and visible fragmentation on market quality. Review of Finance 19, 1587–1622.

Easley, D., Kiefer, N.M., O'Hara, M., 1996. Cream-skimming or profit-sharing? The curious role of purchased order flow. Journal of Finance 51, 811–833.

Foley, S., Putniņš, T.J., 2014. Regulatory efforts to reduce dark trading in Canada and Australia: How have they worked? Report prepared for the CFA Institute.

Foucault, T., Menkveld, A.J., 2008. Competition for order flow and smart order routing systems. Journal of Finance 63, 19–58.

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

Hendershott, T., Mendelson, H., 2000. Crossing networks and dealer markets: Competition and performance. Journal of Finance 55, 2071–2115.

Kwan, A., Masulis, R., McInish, T.H., 2015. Trading rules, competition for order flow and market fragmentation. Journal of Financial Economics 115, 330–348.

Lee, C.M.C., Ready, M.J., 1991. Inferring trade direction from intraday data. Journal of Finance 46, 733–746.

Lo, A., MacKinlay, C., 1988. Stock market prices do not follow random walks: Evidence from a simple specification test. Review of Financial Studies 1, 41–66.

O'Hara, M., Saar, G., Zhong, Z., 2014. Relative tick size and the trading environment. Unpublished working paper. Cornell University.

O'Hara, M., Ye, M., 2011. Is market fragmentation harming market quality? Journal of Financial Economics 100, 459–474.

Ready, M.J., 2014. Determinants of volume in dark pool crossing networks. Unpublished working paper. University of Wisconsin.

Rindi, B., 2008. Informed traders as liquidity providers: Anonymity, liquidity and price formation. Review of Finance 12, 497–532.

Rosu, I., Sojli, E., & Tham, W. W. (2017). Trades, Quotes and the Cost of Capital.

Yao, C., & Ye, M. (2015). Why trading speed matters: A tale of queue rationing under price controls.

Ye, M. (2017). Who Provides Liquidity, and When: An Analysis of Price vs Speed

Table 1
Variables used in regression analysis

Propositions and Hypotheses	Variable definitions
Proposition 1a. As markets fragment, market-wide order-to-trade ratio for a given security increases with the extent of fragmentation, if there is at least one non-zero quality signal in the monitoring set. *Hypothesis 1a. Order-to-trade ratios are higher for stock-days with higher degrees of fragmentation.*	$OTTR_{it} = OrderVol_{it}/TradeVol_{it}$ – ratio of order volume (number of shares) to trade volume (number of shares) in stock i on day t. $Frag_{it}$ – measure of fragmentation, for which the following three proxies are defined: $Frag1_{it} = N_{it}$ – number of trading venues that have executed trades in stock i on day t; $Frag2_{it} = 1 - \sum_j (\frac{TradeVol_{ijt}}{\sum_j TradeVol_{ijt}})^2$-Herfindahl-Hirschman index of fragmentation, based on trading volume (number of shares) in stock i, on market j, on day t; $Frag3_{it} = 1 - \sum_j (\frac{NTrades_{ijt}}{\sum_j NTrades_{ijt}})^2$-Herfindahl-Hirschman index of fragmentation, based on the number of trades in stock i, on market j, on day t. $Frag_{jt}$ – dollar volume weighted average fragmentation measure $Frag1_{it}$ for the average stock trading on market j on day t.
Proposition 1b. As markets fragment, order-to-trade ratio for a given security on a given market increases as the market share of that market decreases. *Hypothesis 1b. Order-to-trade ratios are higher for markets with lower market shares.*	$MktShare_{jt} = DolVol_{jt}/\sum_j DolVol_{jt}$ – dollar volume-based market share of market j on day t.
Proposition 3. Order-to-trade ratio for a given security increases with the quality of signals available for monitoring. *Hypothesis 2a. ETFs have higher order-to-trade ratios compared to the common stocks.* *Hypothesis 2b. Securities with higher correlation with the broad market index have higher order-to-trade ratios.*	$CorrelationS\&P_{it}$ – average 22-day correlation between daily returns on security i and daily returns on S&P500 index. $CorrelationS\&P_{jt}$ – dollar volume weighted average 22-day correlation between daily returns on securities traded on market j and daily returns on S&P500 index.

Proposition 4. Order-to-trade ratio for a given security increases with picking-off risk. *Hypothesis 3a. Order to trade ratios are higher for stock-days with higher market volatility.* *Hypothesis 3b. Order to trade ratios are higher for stock-days with higher stock volatility.* *Hypothesis 3c. Order to trade ratios are higher for stock-days with higher tick-to-price ratios.*	$StockVolatility_{it} = \frac{2(High_{it} - Low_{it})}{High_{it} + Low_{it}}$- - measure of stock i's volatility on day t is based on daily high and low prices of the respective stock. $MarketVolatility_t$ – measure of market volatility on day t is proxied by the following measures: $MarketVolatility1_t = \frac{2(High_{S\&Pt} - Low_{S\&Pt})}{High_{S\&Pt} + Low_{S\&Pt}}$- measure of S&P500 index volatility on day t based on the daily high and low prices of the S&P500 index ETF. $MarketVolatility2_t = \log(VIX)$- natural logarithm of the VIX index level. $TickToPrice_{it} = \frac{TickSize_{it}}{Price_{it}}$ – tick to price ratio of stock i on day t (dollar tick size divided by dollar closing price). $TickToPrice_{jt}$ – dollar volume weighted average tick to price ratio for market j on day t.
Proposition 5. Order-to-trade ratio for a given security decreases with monitoring cost. *Hypothesis 4a. Order to trade ratios are higher for stocks with higher market capitalization.* *Hypothesis 4b. Order to trade ratios are lower on markets with taker-maker fee structures.*	$\log(MarketCap_{it})$ – log of market capitalization for stock i on day t. D_{jt}^{taker} – dummy variable that takes the value of 1 if market j is a taker-maker market and 0 otherwise.
Proposition 6. Order-to-trade ratio for a given security decreases with the trading frequency, holding the monitoring intensity constant. *Hypothesis 5. Order-to-trade ratios are inversely related to the trading volumes.*	$\log(Volume_{it})$ – natural logarithm of trading volume (in number of shares) for stock i on day t. $\log(Volume_{jt})$ – natural logarithm of trading volume (in number of shares) for market j on day t.

Table 2
Descriptive statistics for stock-date panel

This table provides descriptive statistics for the stock-date panel used in regression analysis. The sample period is January 1, 2012 – December 31, 2016. The data is at daily frequency. Variable definitions are provided in Table 1.

	OTTR_log	*Frag1*	*Frag2*	*Frag3*
N	5922424	5922424	5922424	5922424
Mean	4.7620	7.0520	0.6364	0.6685
StDev	2.1092	2.6629	0.2156	0.2008
Skewness	1.5459	-0.9053	-1.7378	-2.0589
Kurtosis	1.9782	-0.3237	2.2817	3.9745

	LogVolume	*LogMarketCap*	*StockVolatility*	*CorrelationS&P*	*TickToPrice*
N	5922424	5899597	5922424	5920996	5922424
Mean	3.8328	13.0263	0.0310	0.4055	0.0010
StDev	2.6460	2.1820	0.0344	0.3583	0.0016
Skewness	-0.4097	0.0706	5.4994	-0.8629	4.2484
Kurtosis	-0.1462	-0.2021	80.1652	1.0098	65.0061

N obs stocks	4467256	75%
N obs ETFs	1455168	25%

Table 3
Descriptive statistics for exchange-date panel

This table provides descriptive statistics for the exchange-date panel used in regression analysis. The sample period is January 1, 2012 – December 31, 2016. The data is at daily frequency. Variable definitions are provided in Table 1.

	Stocks at exchange-date level				*ETFs at exchange-date level*	
	OTTR_log	*Frag1*	*Frag2*	*Frag3*	*OTTR_log*	*Frag1*
N	14638	14638	14638	14638	12816	12816
Mean	2.1330	9.5457	0.7859	0.7981	4.6623	9.2069
StDev	1.0948	0.5379	0.0150	0.0163	1.2661	0.5119
Skewness	1.8164	0.1409	-1.5814	-0.3875	0.2495	-0.1493
Kurtosis	3.9781	1.0505	27.4876	0.7401	1.9738	-0.2986

	Stocks at exchange-date level				*ETFs at exchange-date level*			
	Log Volume	*Corr S&P*	*Tick To Price*	*Mkt Share*	*Log Volume*	*Corr S&P*	*Tick To Price*	*Mkt Share*
N	14638	14638	14638	14638	12816	12816	12816	12816
Mean	10.94123	0.51649	0.00034	0.08587	9.8497	0.6969	0.0002	0.0980
StDev	2.24970	0.11467	0.00020	0.08898	2.2037	0.2034	0.0001	0.0985
Skewness	-1.45989	-0.22743	7.08678	1.03274	-1.9149	-3.5200	11.1384	0.9723
Kurtosis	1.56480	0.41792	91.28564	0.20660	4.0875	14.7798	332.4552	-0.3569

	Stocks		*ETFs*	
N obs maker-taker	12128	83%	10306	80%
N obs taker-maker	2510	17%	2510	20%

Table 4
Descriptive statistics for market volatility measures

This table provides descriptive statistics for the time series of market volatility used in regression analysis. The sample period is January 1, 2012 – December 31, 2016. The data is at daily frequency. Variable definitions are provided in Table 1.

	MarketVolatility	*LogVIX*
N	1257	1257
Mean	0.0094	2.7350
StDev	0.0055	0.2001
Skewness	3.1699	0.9007
Kurtosis	25.2791	0.7458

Table 5
Regression results for stock-date panel

This table reports regression results for six different models, where OTTR is the dependent variable. For definitions of independent variables (column 1), see Table 1. Coefficient estimates are from OLS regressions with double-clustered standard errors. T-statistics are reported in parentheses. Coefficients significant at 1% level are reported with ***, at 5% level – with **, and at 10% level – with *.

	OTTR (1)	*OTTR (2)*	*OTTR (3)*	*OTTR (4)*	*OTTR (5)*	*OTTR (6)*
Frag1	0.0964***					
	(17.81)					
Frag2		0.6073***		0.5991***	0.5866***	0.6051***
		(15.5564)		(15.5989)	(14.9934)	(15.5160)
Frag3			0.4898***			
			(11.2794)			
LogVolume	-0.4894***	-0.4512***	-0.4453***	-0.4615***	-0.4598***	-0.4531***
	(-75.7877)	(-80.7781)	(-76.4078)	(-76.3907)	(-74.9163)	(-81.2612)
LogMarketCap	0.1798***	0.1879***	0.1922***	0.2036***	0.2023***	0.1919***
	(23.5104)	(24.7659)	(25.3483)	(24.3647)	(24.0917)	(25.1527)
MarketVolatility	18.2883***	17.7522***	17.7424***	16.7594***		
	(9.0046)	(8.7907)	(8.7566)	(55.5077)		
LogVix						0.6237***
						(28.6612)
StockVolatility				1.5119***	1.9385***	
				(10.3987)	(12.5767)	
CorrelationS&P	0.6653***	0.6886***	0.6916***	0.6923***	0.7160***	0.6728***
	(14.6215)	(15.2296)	(15.1979)	(15.2517)	(15.7299)	(14.8598)
TickToPrice	-54.9522***	-61.2770***	-63.7547***	-63.7769***	-65.2938***	-61.4609***
	(-12.9599)	(-14.2885)	(-14.8188)	(-15.3563)	(-15.3307)	(-14.2947)
ETF dummy	3.1145***	3.0441***	3.0318***	3.0787***	3.0863***	3.0485***
	(71.5076)	(72.0443)	(71.5893)	(74.3589)	(72.9716)	(72.0122)

Table 6
Regression results for exchange-date panel

This table reports regression results for five different models, where OTTR is the dependent variable. For definitions of independent variables (column 1), see Table 1. Coefficient estimates are from OLS regressions with double-clustered standard errors. T-statistics are reported in parentheses. Coefficients significant at 1% level are reported with ***, at 5% level – with **, and at 10% level – with *.

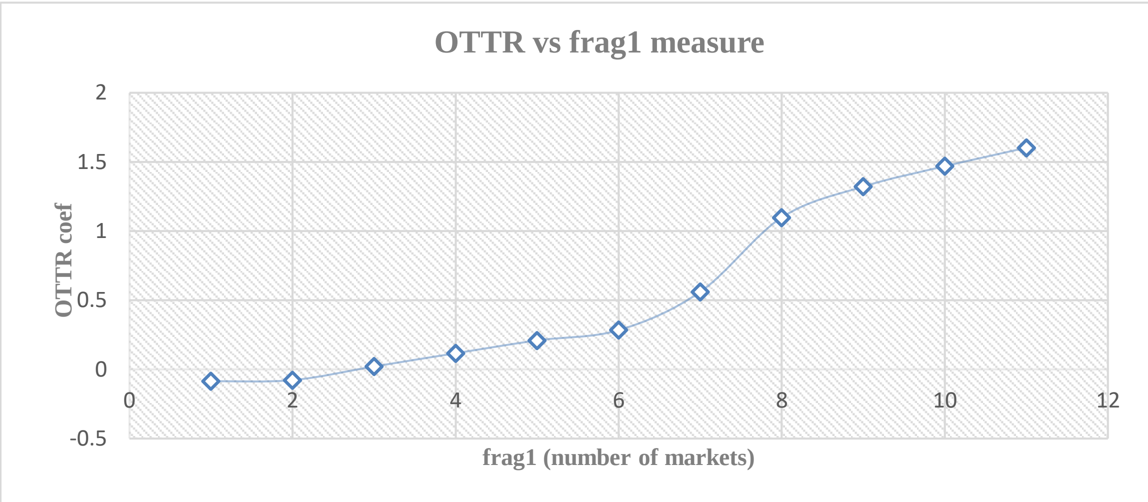
	OTTR (1)	*OTTR (2)*	*OTTR (3)*	*OTTR (4)*	*OTTR (5)*
Frag1	0.0884				
	(0.6354)				
Frag2		8.8817		8.5843	9.3489
		(1.0883)		(1.3939)	(1.5130)
Frag3			7.9219		
			(1.0883)		
LogVolume	-0.3184***	-0.3452***	-0.3424***	-0.3464***	-0.3483***
	(-4.8073)	(-4.9011)	(-4.9011)	(-9.7225)	(-9.4303)
MarketVolatility	12.7512***	14.7730***	15.7476***		
	(-5.4183)	(5.0504)	(5.0504)		
LogVix					0.7076***
					(4.8443)
CorrelationS&P	0.6529***	0.5489***	0.5470***	0.8318***	
	(2.1689)	(2.0467)	(2.0467)	(-2.6578)	
TickToPrice	-886.1485***	-909.31345***	-968.9094***	-869.4143***	-1007.3673***
	(-4.1313)	(-4.4203)	(-4.4203)	(-5.6321)	(-6.3549)
Taker-maker dummy	-0.7459***	-0.7464***	-0.7477***	-0.7410***	-0.7389***
	(-2.3532)	(-2.3399)	(-2.3399)	(-3.6559)	(-3.6362)
MktShare	-0.8495	-0.0825	0.1113		
	(-0.3593)	(-0.0424)	(-0.0424)		

Figure 1
Regression coefficients on fragmentation dummies

This figure plots regression coefficients on fragmentation dummies from the following regression:

$$\log(1 + OTTR_{it}) = \alpha + \sum_{k=1}^{K} \beta_{1k} D_{itk}^{frag1} + \sum_{n=1}^{N} \beta_{2n}\, Controls_{itn} + \varepsilon_{it}$$

Frag1 is the measure of fragmentation that counts the number of markets for a given stock on a given day. Frag2 is HHI index fragmentation measure based on the share volume. We use 11 dummies for frag1 measure (as there are 12 markets overall), and 9 dummies for frag2 measure (as there are 10 deciles overall). The omitted dummy corresponds to the lowest degree of fragmentation.



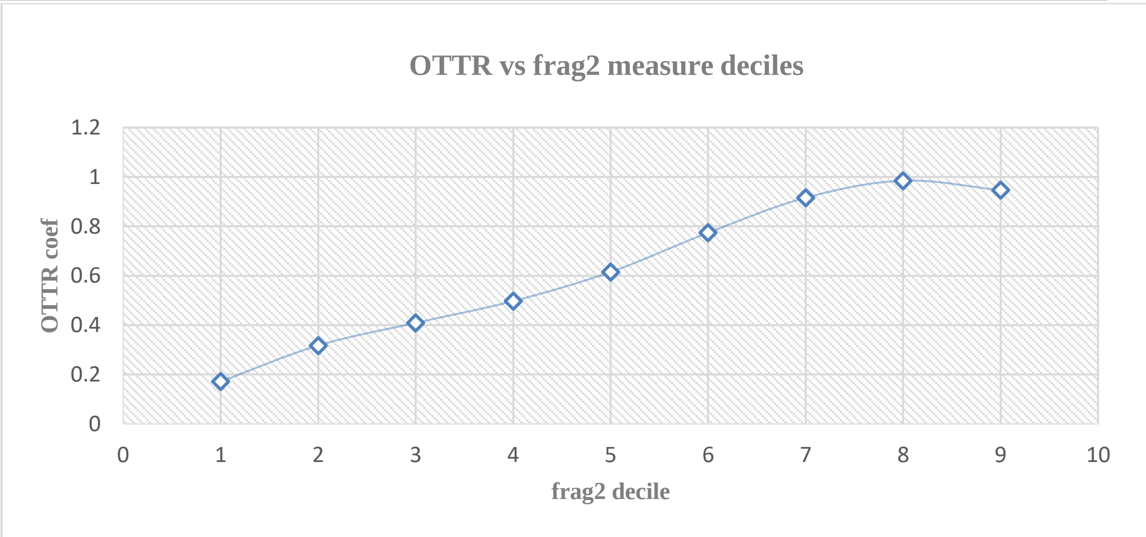


Figure 2
Market shares across different degrees of fragmentation

The figure reports market shares of 12 US trading venues across 20 “buckets” of stocks sorted by frag2 measure (share volume-based HHI fragmentation index), where 0 corresponds to the “bucket” with the lowest fragmentation, and 19 – to the bucket with the highest fragmentation.