Algos gone wild: What drives the extreme order cancellation rates in modern markets?*

Marta Khomyn[1], Tālis Putniņš[1,2]

[1] University of Technology Sydney

[2] Stockholm School of Economics in Riga

March 10, 2021

Journal of Banking and Finance, forthcoming

Abstract

97% of orders in US stock markets are cancelled before they trade, straining market infrastructure and raising concerns about predatory or manipulative trading. To understand the drivers of these extreme cancellation rates, we develop a simple model of liquidity provision and find that growth in order-to-trade ratios (OTTRs) is driven by fragmentation of trading and tech-nological improvements that lower monitoring costs. High OTTRs occur legitimately in stocks with high volatility, fragmented trading, small tick sizes, and low volume. OTTRs are usually within levels consistent with market making, but occasionally spike to levels that may indicate illegitimate trading such as spoofing.

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

JEL Codes: G12

*We thank two anonymous referees, Shane Miller, Maureen O'Hara, Terry Walter, Zhuo Zhong, Katja Ignatieva, Yong Jin, Hao Ming Chen, seminar participants at 7th SIRCA Young Researcher's Workshop, Auckland Finance Meeting 2017, FMA Asia-Pacific Conference 2018, FMA Annual Meeting 2018, Australian Securities and Exchange Commission, Chi-X Australia, and University of Technology Sydney for helpful comments. Marta Khomyn gratefully acknowledges funding from the Capital Markets CRC Limited and was a visiting researcher at Chi-X Australia during the time spent on this study. Tālis J. Putniņš gratefully acknowledges funding from the Australian Research Council (ARC DP200101445).
Address: University of Technology Sydney, PO Box 123 Broadway, NSW 2007, Australia. Fax: +61 2 9514 3088. E-mail addresses: marta.khomyn@uts.edu.au (M. Khomyn) and talis.putnins@uts.edu.au (T. Putniņš).

Algos gone wild: What drives the extreme order cancellation rates in modern markets?

Abstract

97% of orders in US stock markets are cancelled before they trade, straining market infrastructure and raising concerns about predatory or manipulative trading. To understand the drivers of these extreme cancellation rates, we develop a simple model of liquidity provision and find that growth in order-to-trade ratios (OTTRs) is driven by fragmentation of trading and technological improvements that lower monitoring costs. High OTTRs occur legitimately in stocks with high volatility, fragmented trading, small tick sizes, and low volume. OTTRs are usually within levels consistent with market making, but occasionally spike to levels that may indicate illegitimate trading such as spoofing.

1 Introduction

Who is Wall Street? Think Goldman Sachs, JP Morgan and Morgan Stanley drive today's equity markets? Fuhgeddaboudit. Today's largest trading firms are Citadel Securities, GTS, HRT, IMC, Susquehanna/G1X, and Virtu.

Larry Tabb

In today's markets, liquidity is provided almost entirely by electronic market makers, operating across fragmented markets and equipped with fast data feeds and smart market monitoring technologies. One manifestation of electronic market making is high order-to-trade ratios (OTTRs: number of order enter/amend/cancel messages divided by the number of trades). OTTRs have increased more than ten-fold since the year 2000 (Committee on Capital Markets Regulation, 2016). In 2013, the US Securities and Exchange Commission (SEC) reported that 96.8% of all orders were cancelled before they traded, with 90% being cancelled within one second. The response of policymakers has been to curb OTTRs by imposing messaging taxes, which have been proposed in some countries (such as the US) and already implemented in others (e.g., Australia, Italy, and Germany). This paper investigates the drivers of OTTRs, whether their growth warrants concern, and the impacts of constraining OTTRs with regulatory measures.

High OTTRs have been in the public spotlight, with concerns that they are a symptom of predatory or manipulative behavior of high-frequency traders (Biais, Foucault, & Moinas, 2011). These concerns are accompanied by a recent surge in prosecutions of spoofing, a trading strategy that involves misleading other market participants with orders that are cancelled before they are able to execute. For example, the US Commodity and Futures Trading Commission (CFTC) in 2018 took action against a record number of spoofing cases: five times more than the average in the previous eight years.[1] It is true that market manipulation strategies such as spoofing, layering, or quote stuffing generate spikes in order activity and high cancellation rates. Therefore, high OTTRs are used by regulatory surveillance systems to identify spoofing and have been used in courts as evidence of spoofing.[2] However, high OTTRs can also arise from trading that is neither illegal nor harmful. For example, as we show in this paper, market making can result in high OTTRs,

[1]See https://www.wsj.com/articles/u-s-market-manipulation-cases-reach-record-1540983720.
[2]For example, see Egginton, Van Ness, & Van Ness (2016) and Putnins (2020).

in particular when posting quotes across multiple exchanges and adjusting the quotes rapidly in response to new information. It is therefore crucial to be able to distinguish between legitimate and illegitimate levels of OTTRs.

High OTTRs also place considerable strain on market infrastructure due to the massive amounts of data generated in the trading process. In 2013, NYSE received four to five terabytes of data a day, according to Forbes.[3] That number (4×10^{12} bytes) represents a substantial fraction of the total global daily data usage (estimated at 10^{18} bytes as of 2018).[4] The costs include the immense processing power needed by exchanges and market participants, the bandwidth consumed in data transmission, and the burden on regulators to conduct surveillance over such large datasets. In response to the concerns about illicit trading activity and the costs of the large volumes of market data, a number of regulators have imposed taxes on high-OTTR traders (see Friedrich & Payne, 2015; Jørgensen, Skjeltorp, & Ødegaard, 2018). If such regulation curbs harmful behavior of high-frequency traders (HFTs), the tax could improve liquidity and other measures of market quality. However, if the regulation negatively affects market makers by constraining legitimate liquidity provision, market liquidity could deteriorate. Therefore, the optimal design of message traffic taxes requires an understanding of the levels and variation in OTTRs that result from legitimate market making.

To understand the drivers of the high OTTRs in today's markets and to provide a benchmark against which to evaluate whether an OTTR is excessive, we develop and calibrate a simple illustrative model of liquidity provision. In the model, a market maker in a fragmented market monitors several sources of information ("signals") and updates quotes to avoid being "picked off" (trading at stale prices). His monitoring intensity is endogenous: the market maker decides which signals to monitor by weighing up the benefit (reduced "picking-off risk") against the cost (the computing and data feed costs). Consequently, the OTTR emerges endogenously as a function of monitoring cost, market conditions (e.g., volume, volatility), stock characteristics (e.g., how closely correlated the stock is with other securities), and the extent of fragmentation of trading across multiple exchanges.

Starting with the time-series drivers, our empirical tests suggest that the long-term growth in

[3] See https://www.forbes.com/sites/tomgroenfeldt/2013/02/14/at-nyse-the-data-deluge-overwhelms-traditional-databases/4493d9805aab.

[4] See https://www.forbes.com/sites/bernardmarr/2018/05/21/how-much-data-do-we-create-every-day-the-mind-blowing-stats-everyone-should-read/268546af60ba.

OTTRs is largely driven by the introduction of automatic quoting (Autoquote), increasing market fragmentation, and decreasing monitoring costs. For example, while the average US stock at the start of our sample (year 1998) was traded on around two stock exchanges, by 2018 that number had increased to ten. Microprocessor speeds increased 14-fold during our sample period and data storage costs fell by a factor of 2 000, collectively reflecting a substantial reduction in the costs of monitoring a large number of digital signals. Yet around these longer term trends, we find that the short-term dynamics of OTTRs are closely related to market volatility, consistent with those being periods of large changes in security values and therefore high picking-off risk.

Our analysis also sheds light on why OTTRs are considerably higher in some securities compared to others. Our empirical tests support the model predictions that OTTRs are higher in more volatile stocks, higher price-to-tick stocks, lower volume stocks, and in ETFs compared to stocks. The intuition for these effects is as follows. In more volatile markets, information arrives more frequently. Frequent information causes market makers to update quotes (amend or cancel and replace) more often to avoid being picked off by informed traders, which increases OTTRs. Similarly, market makers update quotes more often in stocks that have a small minimum price increment relative to their price level (high price-to-tick stocks) because the higher resolution in prices allows smaller changes in valuations to be reflected with quote updates. ETFs naturally have higher OTTRs compared to stocks, because ETFs have a number of highly relevant signals available for monitoring, such as the prices of the underlying stocks or the index.

The third dimension of variation in OTTRs explained by our model is across markets. We find that OTTRs are naturally much higher on markets with lower shares of total trading volume. The intuition is as follows. When a security trades on multiple markets, a liquidity provider often duplicates their quotes in each (or at least some) of the different markets. Each time the market maker updates their quoted prices or volumes, they send order amendment or cancellation messages to all of the markets in which they are quoting, and consequently the amount of messaging activity (the OTTR numerator) across the different markets is approximately equal. Yet if the number of trades in a market with a low share of total trading volume is considerably lower, the smaller OTTR denominator results in a higher OTTR for that market.

In the last part of the paper we apply a calibrated version of the model to assess how often OTTRs spike to levels beyond what is expected under regular market making. In this analysis we account for the large number of factors that drive natural variation in OTTRs. We find that in most stocks and on most days, the empirically observed OTTRs in 2018 are consistent with levels that would be expected from liquidity provision in a fragmented market, even under conservative assumptions. However, the empirical distribution of OTTRs is right-skewed and 17% of observations exceed the theoretical natural level. This exercise of benchmarking actual OTTRs against the natural OTTR expected given the characteristics of the stock, day, and market, is an approach regulators could use to detect abnormal order activity that warrants further investigation. By accounting for the substantial natural variation in OTTRs, this approach is likely to lead to more efficient allocation of regulatory resources and surveillance effort than using arbitrary unconditional OTTR thresholds to trigger investigation.

Our findings have implications for policy measures and regulatory proposals that aim to curb OTTRs, such as messaging taxes and OTTR limits. Because the recent levels of OTTRs, in most cases, are within the levels expected under legitimate liquidity provision, restricting order activity is likely to have adverse effects on market making. The adverse effects are likely to be more severe in securities that would naturally have higher OTTRs, for example, securities with low volume and high volatility. As a result, securities with already low liquidity are likely to be disproportionately impacted by such policy measures. Similarly, if the restrictions were applied uniformly across securities, ETFs would be disproportionately affected, because of their high natural OTTRs in ETFs due to the availability of precise, frequent signals. Furthermore, messaging taxes create a non-level playing field between exchanges, because venues with a lower share of total volume will naturally have higher OTTRs.

Our paper is related to empirical studies of OTTRs, such as those that examine the effects of policies designed to reduce OTTRs. For example, Friedrich & Payne (2015) find that a fee on high OTTRs introduced in the Italian Stock Exchange resulted in less deep and less resilient markets, although spreads were not substantially affected. The underlying mechanism seems to be that the regulation constrained the order activity of HFTs, thereby negatively affecting their ability to provide liquidity. In contrast, Jørgensen, Skjeltorp, & Ødegaard (2018) show that a fee on OTTRs introduced in the Oslo Stock Exchange did not substantially impact measures of market quality. The difference in the case of Oslo seems to be that the regulation was accompanied by incentives for liquidity provision. Therefore, these empirical studies are consistent with our model's implication that a restriction in OTTRs without accompanying incentives for liquidity provision is likely to harm liquidity.

The theoretical underpinnings of our analysis rely on models of modern market makers (e.g., Foucault, Röell, & Sandås, 2003; Liu, 2009; Foucault, Kadan, & Kandel, 2013; Lyle & Naughton, 2018), but unlike the previous work, we address the issue of what drives OTTRs and whether they are excessive. We also contribute to the small number of studies that have identified specific factors related to OTTRs, including price-time priority (Yao & Ye, 2018; Wang & Ye, 2017), cross-venue strategies by fast and slow traders (van Kervel, 2015), limit order profitability (Dahlström, Hagströmer, & Nordén, 2018) and the difficulty to value a stock (Rosu, Sojli, & Tham, 2020). We characterize a broader set of drivers of OTTRs, including fragmentation, monitoring costs, and technological advances, which we argue are the most salient drivers of the growth in OTTRs. Further, we build on studies of cancellations (Nikolsko-Rzhevska, Nikolsko-Rzhevskyy, & Black, 2020; Jain & Jordan, 2017), but we focus on measuring messaging activity overall, as modern market making is manifested through high overall OTTR levels.

There is currently no consensus in the literature on how market fragmentation affects quoting activity and liquidity. For example, recent work by Budish, Lee, & Shim (2020) implies that fragmentation is irrelevant for the equilibrium quoting behavior. At the same time, multiple studies suggest that fragmentation matters for liquidity (e.g., Chowdhry & Nanda, 1991; Biais, Martimort & Rochet, 2000; Pagnotta & Philippon, 2018; Baldauf & Mollner, 2019).

Finally, by providing a novel benchmark for identifying abnormal OTTRs, our paper contributes to the recent literature on layering and spoofing in markets (see Putnins (2020) for an overview). In particular, the benchmark for abnormal OTTRs that we propose may be useful in detecting or confirming the presence of layering or spoofing in markets.

2 A framework of what drives OTTRs

2.1 Primitives

Agents. Consider a single, representative **market maker** that posts bid and ask quotes for a given asset in a given market, and a continuum of **traders** that arrive at the market and trade against the market maker.

Agents' action space. The market maker can choose to monitor one or more signals from a set of signals, $\Omega_s = \{s_1, s_2, \ldots, s_N\}$. The signals could be prices of related securities, prices of the same security in another exchange, the state of the order book, and so on. Each signal is a flow of time-series data that changes at stochastic times (termed "information arrivals") given by Poisson processes with intensity λ_n for the n^{th} signal. The quality of signal n, q_n, is the probability that when there is a change in that signal (an "information arrival"), the market maker will want to update his posted quotes, resulting in a "cancel and enter" or "amend" message from the market maker. Events that trigger such quote revisions are termed "*relevant* information arrivals".

There is a cost c to the market maker to process each information arrival. Hence, the expected monitoring cost for signal n per unit time is proportional to the intensity of information arrivals: $\lambda_n c$. This cost can be interpreted as the processing capacity that is required to interpret information arrivals and determine whether/how to respond. It can be thought of as including the required technology (telecommunications bandwidth, computational capacity, and so on) and the cost of subscribing to the data feed (e.g., buying real-time streaming market data from an exchange).

Individual traders arrive at the market at stochastic times given by a Poisson process with arrival rate λ_m, and trade against the market maker's posted quotes by submitting a market order. The market maker's benefit from monitoring comes from avoiding having stale quotes picked off. When a market order arrives after a relevant information update, but the market maker has not updated 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. It follows that the more signals the market maker monitors, the less often his quotes will be picked off. It also follows that for a given monitoring intensity, the picking-off cost per unit time increases with the frequency of relevant information arrivals, i.e., with the asset's fundamental volatility.

2.2 Equilibrium OTTRs

The market maker chooses which signals (if any) to monitor by weighing up the costs of monitoring, $\lambda_n c$, 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.[5]

To help understand the optimal choice of which signals to monitor, we define a signal's usefulness, u_n, as the arrival intensity of relevant information from the signal (signal changes that cause the market maker to want to revise his quotes): $u_n = \lambda_n q_n$. The expected benefit (per unit time) from monitoring a given signal n is the saved losses that would have occurred from having quotes picked off. This 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 n is $\frac{\lambda_n q_n}{\lambda_m + \lambda_n q_n}$. Therefore, the expected benefit per unit time of monitoring signal n is $\lambda_m \left(\frac{\lambda_n q_n}{\lambda_m + \lambda_n q_n} \right) k = \lambda_m \left(\frac{u_n}{\lambda_m + \lambda_n q_n} \right) k$. As this expression reveals, the benefits of monitoring signal n are increasing with the signal's usefulness (u_n), increasing in the picking-off cost (k), and increasing in the market order arrival rate (λ_m).

Recall there is also a cost of monitoring a signal and therefore a market maker will optimally monitor all signals for which the benefit exceeds the cost. The expected cost per unit time of monitoring signal n is $\lambda_n c$, giving a net benefit of $\lambda_m \left(\frac{\lambda_n q_n}{\lambda_m + \lambda_n q_n} \right) k - \lambda_n c$ from monitoring the signal. The market maker adds signals to his "monitored list" from greatest to least expected 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 signals for which:

$$\lambda_m \left(\frac{\lambda_n q_n}{\lambda_m + \lambda_n q_n} \right) k - \lambda_n c > 0 \tag{1}$$

We denote the set of *monitored* signals Ω_{s^*}. Condition 1 determines monitoring intensity (the number of monitored signals). As a result of monitoring signals, executing trades, and updating quotes, the market maker generates messaging activity (messaging includes order entry, cancella-

[5] In the interests of maintaining a parsimonious model, we do not explicitly model the width of the bid-ask spread. One could endogenize and model the bid-ask spread by, for example, considering the spread as being set by a zero expected profit condition as is standard in models of market making under perfect competition. The width of the bid-ask spread would be a function of the picking off risk, the monitoring intensity, and trade arrival rates. However, adding this to the model would not help much in understanding the drivers of the OTTR, as the tradeoff between picking off costs and monitoring costs, which determines the OTTR in our model, would exist irrespective of the width of the spread.

tion, and amendment messages) at an expected rate of Q messages per unit time:[6]

$$Q = \sum_{n \in \Omega_{s^*}} 2\lambda_n q_n + 2\lambda_m \tag{2}$$

The first term, $\sum_{n \in \Omega_{s^*}} 2\lambda_n q_n$, is due to quote updates in response to relevant information arrivals on monitored signals, and the second term, $2\lambda_m$, is due to reposting liquidity after being hit by a market order. Both terms 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 quotes (adjusting the price and / or volume of those quotes).

Recognizing that the expected number of trades per unit time is just the market order arrival intensity, λ_m, the expected OTTR is given by:

$$OTTR = \frac{\sum_{n \in \Omega_{s^*}} 2\lambda_n q_n + 2\lambda_m}{\lambda_m} \tag{3}$$

The model makes some simplifying assumptions. We assume that in response to new information, the market maker updates both his bid and ask quotes symmetrically; that the market maker is rational; and that other traders use market orders. These assumptions allow us to focus on the key mechanism, which is the trade-off between the cost of monitoring and the benefit of not being picked off.

2.3 Model with fragmented markets

If the number of markets increases from 1 to M, the single (representative) market maker can choose to post liquidity across multiple venues. The incentive for a market maker to do so is that they want to trade against incoming market orders, however they cannot tell in which market the next market order will arrive. Having invested in monitoring signals to determine what prices to quote, in the absence of any message taxes or exchange fees, there is zero marginal cost to the market maker duplicating their quotes in multiple trading venues. Yet there is the benefit of being

[6] We treat different types of messaging traffic equally, because in our framework messaging activity is simply a way for market makers to react to new information arrivals. For characteristics of different order placement techniques and their effects on market quality, see Nikolsko-Rzhevska, Nikolsko-Rzhevskyy, & Black (2020).

able to trade with more of the incoming market orders.[7] Therefore, the representative market maker will optimally duplicate their quotes in each of the trading venues.

The aggregate market order arrival rate, λ_m, is assumed to remain the same as in the single-market case, just split across multiple venues. Therefore, the overall quoting activity of the market maker consists of two components: (a) quote updates resulting from relevant information received by monitoring signals, $2M \sum_{n \in \Omega_{s^*}} \lambda_n q_n$ (market maker updates quotes on all M markets in response to monitored signals), and (b) reposting liquidity / revising quotes on all markets after getting a fill on market orders, $2M\lambda_m$ (market order arrivals constitute useful signals, from the market maker's viewpoint). Note that market fragmentation does not affect the signal monitoring decision of the market maker, who chooses the set of signals to monitor in the same manner as in a single-market case. For a given security, the resulting expected OTTR across venues is therefore:[8]

$$OTTR = \frac{2M(\sum_{n \in \Omega_{s^*}} \lambda_n q_n + \lambda_m)}{\lambda_m} \tag{4}$$

Consider the OTTR of individual markets $j = 1 \dots M$. The market share of trading volume (market orders) for each individual market j is ρ_j. The market maker updates his quotes on market j in response to the following:

(a) every time a piece of relevant information is received from the monitored signals, which results in the number of quote updates $2 \sum_{n \in \Omega_{s^*}} \lambda_n q_n$

(b) after being hit by a market order on market j, which results in the number of quote updates $2\rho_j\lambda_m$

(c) after being hit by a market order on any other market besides market j, which results in the number of quote updates $2(1 - \rho_j)\lambda_m$, since market order arrivals constitute useful signals.

Then, the OTTR for market j is:

[7] In practice, there might be a marginal cost of posting quotes in multiple trading venues such as the infrastructure cost of connecting to a trading venue and the possibility of higher picking off risk if market orders are able to execute against duplicated resting orders before the market maker is able to revise the quotes. Such marginal costs are likely to result in market makers posting liquidity on several but not necessarily all trading venues. We investigate this possibility in the empirical tests.

[8] We define OTTR in line with regulatory reports of e.g., US SEC and FINRA. In our empirical analysis, security-day OTTRs are equally weighted averages. In principle, another way of defining average OTTRs is by computing share-weighted averages. However, to the best of knowledge, none of the existing academic and regulatory studies apply the share-weighting approach to compute OTTRs.

$$OTTR_j = \frac{2\sum_{n\in\Omega_s^*}\lambda_n q_n + 2\rho_j\lambda_m + 2(1-\rho_j)\lambda_m}{\lambda_m\rho_j} \tag{5}$$

Simplifying Eq. 5:

$$OTTR_j = \frac{2\sum_{n\in\Omega_s^*}\lambda_n q_n + 2\lambda_m}{\lambda_m\rho_j} \tag{6}$$

2.4 Comparative statics

We now derive propositions about the relation between OTTRs, monitoring intensity and fragmentation. First, we establish the link between OTTRs and fragmentation (Proposition 2.1). Second, we show how OTTRs are related to market shares (Proposition 2.2). Third, we relate OTTR to all the model parameters (Propositions 2.3 – 2.6). We provide the economic intuition for these propositions below, and the proofs in Appendix A. Corresponding empirical hypotheses are presented in Table 1.

[Table 1 about here.]

Proposition 1. *The OTTR for a given security increases with the extent of fragmentation of the security's trading across multiple exchanges.*

The intuition is simple. As markets fragment, the market maker posts quotes on several exchanges rather than one, hence every time the market maker updates his quotes (due to information arrival or a trade) it requires order messages be sent to each of the exchanges, driving OTTRs up. Although the model makes a number of simplifying assumptions, such as considering OTTRs of a single market maker that posts bids and asks across all available exchanges, a similar scaling effect (higher OTTR with higher fragmentation) will occur, if the market maker quotes across only a few exchanges or only on one side (bid or ask). In such a scenario, OTTRs would still increase with fragmentation, but at a lower rate than implied by the model.

Empirically, we should observe higher OTTRs for securities with more fragmented trading (Hypothesis 1a). Similarly, from the exchange perspective, we expect higher OTTRs on exchanges that trade more fragmented securities (Hypothesis 1b).

Proposition 2. *The OTTR for a given exchange is inversely related to its market share.*

This effect occurs, because if a market maker posts quotes on all exchanges and updates them all at the same time, the messaging activity (the OTTR numerator) will be approximately the same across venues, but the number of trades (the OTTR denominator) will be different. Therefore, empirically, we expect higher OTTRs for exchanges that have lower market shares (Hypothesis 2).

Proposition 3. *The OTTR for a given security increases with the quality of signals available for monitoring.*

Monitoring intensity and OTTR are closely related, because the market maker posts quotes as a result of his monitoring. Recall that monitoring intensity in the model is endogenous, and arises from the market-maker's cost-benefit analysis. For example, if the market maker faces higher quality signals (keeping other things constant), he enjoys higher marginal benefit of monitoring and therefore monitors more signals and posts more frequent order updates. This in turn means higher OTTRs.

Empirically, we should observe higher OTTRs in securities with high-quality signals. For example, exchange-traded funds (ETFs) should have higher OTTRs than stocks (Hypothesis 3a), as ETFs follow a predefined index and typically have easily observable high-quality signals (e.g., related trading instruments on the same index, including futures contracts, options, underlying stocks etc.). Similarly, securities that are highly correlated with the market index (including both highly positive or highly negative correlations) can be seen as having high-quality signals available (i.e., a simple data feed of market index updates would provide a fairly precise signal for the value of security that is highly correlated with the market). Therefore, we should observe higher OTTRs in securities that are highly correlated with the broad market index such as S&P500 (Hypothesis 3b).

The tick size is also likely to affect how often a market maker updates their quotes. This effect can be understood through the signal quality parameter. Recall the signal quality is the probability that when there is a change in the signal, the market maker will want to update his quotes. With a very large tick size, there has to be a very large change in the security value before the market maker would want to repost quotes at the next tick. On the contrary, with a very fine pricing grid, even very small changes in security value will warrant quote updates. Therefore, we expect higher

OTTRs for stocks with a smaller percentage tick size, i.e., a higher price-to-tick ratio (Hypothesis 3c).

For similar reasons, we also expect higher OTTRs in markets with a finer pricing grid, such as taker-maker markets (Hypothesis 3d). Because taker-maker markets charge limit orders and compensate market orders, this effectively allows trades at sub-penny increments (net of fees) thereby offering more granular pricing. The effect is therefore similar to a smaller tick size: market makers update quotes more often, because the finer pricing grid allows them to respond to small changes in value.

Proposition 4. *The OTTR for a given security increases with lower monitoring costs.*

When the monitoring cost per signal is lower, the market maker has an incentive to monitor more signals, which results in higher OTTRs. Therefore, in the cross-section of stocks, we expect higher OTTRs for stocks that have more low-cost (or free) signals, such as large stocks that have rich public information widely available to investors (Hypothesis 4a). Also, we should observe an upward trend in OTTRs through time, as data processing costs have decreased significantly in the last two decades (Hypothesis 4b).

Proposition 5. *The OTTR for a given security increases with picking-off cost.*

When faced with a higher cost of trading at stale prices, the market maker has stronger incentives to monitor more signals to minimize the costs of being hit by market orders without having updated quotes. Therefore, higher picking-off costs lead to higher monitoring intensity and higher OTTRs.

The cost of being picked off is larger when there are greater changes in the value of the underlying security. For example, if in a given period of time, the security value changes by 10 basis points (bps) and the market maker fails to adjust quotes before getting hit by a market order, they incur a loss of about 10 bps. Yet if the value changes by 100 bps and the market maker is hit before they manage to update quotes, their loss is around 100 bps. Therefore, picking-off costs (and thus OTTRs) are expected to be higher in more volatile securities and more volatile times (Hypotheses 5a and 5b).

Proposition 6. *The OTTR for a given security decreases with the trading frequency, holding the monitoring intensity constant.*

Holding market maker's monitoring intensity fixed, a higher rate of market order arrivals decreases OTTRs, as every trade is associated with fewer quote updates on average. Therefore, we expect lower OTTRs in stocks with higher trading volumes (Hypothesis 6).

3 Empirical analysis

3.1 Data and descriptive statistics

The sample starts on January 1, 1998 and ends on December 31, 2018, covering a period in which OTTRs experienced substantial growth. We analyze a representative cross-section of stocks and ETFs, chosen using random sampling with stratification by size. To construct this sample, we obtain the full cross-section of US stocks as of December 2018 from Center for Research in Security Prices (CRSP) database, sort them by size (market capitalization) and pick every 20th stock. The resulting sample contains 241 stocks in 2018. Next, we take the market capitalizations of those 241 stocks, and calculate their inflation-adjusted equivalents (using the GDP deflator) in each prior year (2017, 2016, and so on). Then, in each year (1998 – 2017), we select the 241 stocks (from the universe of stocks in CRSP) that most closely correspond to the 241 inflation-adjusted sizes (via sampling without replacement).

This approach constructs a sample that in every year is representative of today's market with respect to the distribution of market capitalizations. Hence the sample allows us to identify trends in OTTRs that happen for reasons other than changes in the composition of the market. We know from prior studies that the composition of stocks in equity markets has changed substantially through time towards fewer, but considerably larger listed companies (e.g., Doige et al., 2017). The sampling approach allows us to control for this tendency in the time series.

We also construct a random, stratified sample of ETFs. Given that ETF markets are relatively young (the first US ETF was launched in 1993) and have been growing rapidly, it is not possible to take today's distribution of assets under management (AUM) and find matching ETFs in 1998. We therefore sample ETFs by sorting them by AUM in each year, and selecting every 20th ETF. This makes sampled ETFs representative of the AUM distribution in each year, and ensures that we sample the same proportion of ETFs as stocks.[9]

[9] The number of ETFs in our sample increases from 2 in 1998 to 68 in 2018, in proportion to the total number

We obtain order book data (orders and trades) from the Thomson Reuters Tick History (TRTH) database and stock characteristics from the CRSP database. The data cover 14 US exchanges: NYSE American, NASDAQ Boston, NSX, Direct Edge X, Direct Edge A, Investors Exchange, CME Chicago, NYSE, NASDAQ, NYSE Arca, NASDAQ Philadelphia, OTC Markets, BATS-Z, and BATS-Y.

We aggregate the intraday trade and quote data to daily observations per stock and per exchange. We compute the OTTR as the number of order messages that result in changes to bid or ask prices or volumes at the best bid and ask prices, divided by the number of trades.[10] The numerator of this measure sums the number of times in a given day the best bid changes either in price or quantity and the number of times the best ask changes either in price or quantity. Each order entry, amendment, and cancellation at the best bid or ask will trigger a change to the price or quantity at the best bid and ask and therefore contribute to the numerator.[11] Table 2 provides detailed definitions of all the variables used in further regression analysis.

[Table 2 about here.]

Table 3 reports descriptive statistics. The dataset contains 1,210,225 stock-day observations, 215,748 ETF-day observations, and 48,772 exchange-day observations.[12] The sample period includes 5,284 trading days. We winsorize the OTTR variable in the security-day dataset at the 99th percentile to reduce the effects of outliers. The winsorization is done separately for the stock-day and ETF-day datasets. To compute exchange-day OTTR, we average the (winsorized) security-exchange-day observations by exchange.

[Table 3 about here.]

of ETFs available in the market. We identify ETFs as CRSP securities with code=73, and require at least 100 days with non-zero trades per year.

[10]When computing OTTRs, to capture the period of continuous trading and avoid contamination from market opening and closing mechanisms, we use the number of messages and trades between 10 am and 3 pm.

[11]The measure does not consider order messages beyond the best quotes. In unreported analysis we use SEC MIDAS data that includes order cancellations and amendments at all levels and find that the two versions of OTTR (that computed from SEC MIDAS data and that from TRTH data) are very closely correlated (correlation coefficient of 0.64), although different in their magnitude. We use the TRTH data rather than MIDAS data, because the former allows us to obtain a long time series that captures the period of growth in OTTRs, while the latter is only available starting from year 2012.

[12]We use "stock-day" and "ETF-day" to refer to the observations for stocks and for ETFs respectively. We use "security-day" to refer to the units of observation for both types of securities (stocks and ETFs) on particular days.

An average stock in the sample has daily OTTR of 16.87, daily volume of 0.99 million shares, and trades on 5.44 markets on an average day between 1998 and 2018. The average market capitalization is $5.7 billion, average absolute correlation with the S&P500 index is 0.39, and average range of daily high-low prices is 4.06%. ETFs have an order of magnitude higher OTTRs (349.41 on average), trade on slightly fewer markets (4.76 on average), but in comparable volumes (1 million shares daily), experience lower daily volatility (1.04% average high-low range, similar to daily S&P500 volatility of 1.35%), and have a higher absolute correlation with the S&P500 index (0.67) due to diversification of idiosyncratic risk. The average ETF in the sample has assets under management of $1.85 billion. ETFs also have lower average tick-to-price ratios, compared to stocks (4 bps for ETFs vs 23 bps for stocks). Note that tick-to-price ratios reflect the tick size changes in the US markets following the introduction of decimalization (effective as of April 9, 2001), as well as the tick size changes related to the SEC Tick Size Pilot 2016 – 2018.[13]

An average exchange in the sample has a daily OTTR of 70.16 across all securities it trades (both stocks and ETFs). OTTRs vary across exchanges: from the average of 41.15 on NYSE (which has an average market share of 41.03% in our sample) to 157.77 on NASDAQ Philadelphia (market share of 0.91%). Two other exchanges with high OTTRs are BATS-Y and NASDAQ Boston, which are taker-maker markets. The average exchange-day OTTR is higher than the average stock-day OTTR, because the high OTTRs of ETFs drive the mean exchange OTTR up. The average market shares of venues are 9.79% in stocks and 10.77% in ETFs.

3.2 Time series trends in OTTRs and concurrent market structure changes

The time series of OTTRs (see Figure 1) reflect key market structure changes in US markets, including the introduction of Autoquote and the order protection rule (Rule 611 of Regulation National Market System). The first increase in stock OTTRs coincides with the introduction of Autoquote in January 2003, as NYSE started disseminating order book updates automatically. As discussed in Hendershott, Jones, & Menkveld (2011), the phase-in of Autoquote resulted in a 50% increase in order message traffic almost immediately and laid the foundation for electronic market

[13]The tick size changes relevant to the sample period are as follows. In 1997, the New York Stock Exchange and NASDAQ reduced the tick size from 1/8th of a dollar to 1/16th. In April 2001, all US exchanges switched to a tick size of one cent for stocks priced above one dollar and 0.01 cents for stocks priced below one dollar. Between October 3, 2016 and September 28, 2018, as part of the SEC Tick Size Pilot, 1,400 stocks had an increase in tick size from $0.01 to $0.05.

making. Autoquote increased the arrival intensity of signals as order book information became more accessible. It also created demand for fast data feeds, in particular by market makers that drastically reduced their reaction time to new information. In other words, the arrival rates of relevant signals increased, while the marginal cost of monitoring signals decreased, resulting in more frequent quote updates by market makers and a substantially higher OTTR. The year 2003 is also the effective start of HFT in US markets according to Aitken et al. (2015) using the order cancellations and trade size approaches. HFTs use technology to bring down the marginal cost of monitoring signals and therefore, consistent with our model, they are likely to contribute to the increase in OTTRs from around 2003 onward.

[Figure 1 about here.]

OTTRs further increased in 2005, when the SEC enacted the order protection rule (OPR) of Regulation National Market System (Reg NMS). The rule forces orders to be routed to the exchange that offers the best quotes. As a result, new exchanges emerged and trading fragmented across multiple exchanges. In 2001, an average stock in the sample traded on two venues, but by 2008 it traded on four. ETF trading also fragmented around the OPR.

Stock OTTRs remain high during the period 2007–2013. These high OTTR levels coincide with several potential drivers. First, the global financial crisis of 2007–2009 was associated with very high volatility and high correlations of stocks and ETFs with the market index. Both of these factors are positively related to OTTRs, as highlighted by the model. Second, trading continued to become more fragmented, contributing to high OTTRs. Finally, this period also coincides with growth in high-frequency trading (HFT), underpinned by technological advances that decreased the costs of monitoring a large set of high-frequency data feeds.

3.3 Cross-sectional and time-series determinants of OTTRs

We test the relation between OTTRs and the drivers implied by the theory model using regressions. Broadly, we ask how the hypothesized factors such as market fragmentation, monitoring intensity, picking-off risk and so on are related to OTTRs through time and in the cross-section. The empirical design does not explicitly address endogeneity issues other than including a large number of control variables. Instead, the regressions examine non-causal relations between observed OTTR levels and hypothesized drivers of OTTR, using the theory model for guidance.

[Table 4 about here.]

[Table 5 about here.]

Tables 4 - 8 report the regression results. In Table 4, we use variables aggregated at the security-day level and thus focus mainly on security-level characteristics, Table 5 provides a robustness check by excluding ETFs from the sample. In Table 7, we use aggregation at the exchange-day level and therefore include market-level variables. Table 8 replicates the same analysis excluding ETFs.[14]

There is a positive relation between OTTRs and fragmentation, consistent with the model. For each additional market on which the security is traded ($Frag1$), the OTTR increases by approximately 8%, all else equal (see regression (1) in Table 4).[15] Using alternative proxies for fragmentation (e.g., Herfindahl-Hirschman index) confirms this result. So does analysis aggregated at the exchange level (the positive coefficient on $NumMkts$ in regression (1) of Table 7). This evidence supports Hypotheses 1a and 1b that OTTRs are higher for securities with more fragmented trading and markets that trade more fragmented securities. The model suggests that this effect occurs because liquidity providers duplicate their quotes across multiple venues, leading to more order activity for the same amount of trading.

The regression results also confirm that securities with high-quality signals have higher OTTRs. OTTRs are 135% higher for ETFs compared to stocks, after controlling for other factors (see regression (4) in Table 4).[16] OTTRs are also higher for securities that have stronger correlations with the market index. This evidence supports Hypotheses 3a–3b and is consistent with liquidity providers revising their quotes more often when there is a richer stream of relevant information about the

[14] In these robustness checks, we confirm that our results hold after excluding ETFs from the sample.

[15] Note that in discussing these results, we approximate the change in OTTR with the change in (1+OTTR). For example, the precise interpretation would be to say that for each unit increase in $Frag1$, (1+OTTR) increases by approximately 8%, all else equal. However, the difference is negligible for observed OTTR levels, so for convenience of discussion, we phrase the results with respect to effect on OTTRs, rather than on (1+OTTR). Also note that the effects described as percentage impacts on OTTRs are log growth rates because of the log dependent variable rather than simple percentage changes. For small effect sizes the log growth rates are approximately equal to simple percentage changes. The rest of the regression results are interpreted in a similar manner.

[16] Univariate analysis suggests that average ETF OTTRs are an order of maginitude higher than stock OTTRs (349.42 vs 16.87, t-statistic of 207.37).

security value. ETFs, as index-tracking securities, have a particularly frequent and relevant set of signals (e.g., underlying stock prices, index futures, and so on), leading to a considerably higher rate of order revisions and OTTRs. Similar logic applies to securities that are highly correlated with the S&P500 index because for such securities the S&P500 futures and ETFs can be used as signals for updating quoted prices.

The relative tick size (the minimum price increment scaled by the price) also impacts the OTTR because it determines how often arriving information will be substantial enough to warrant revising the quotes by an entire tick. To illustrate, consider two stocks. Stock A is priced at $50, and stock B at $5. Suppose the tick size is $0.01 and stock A's quotes are $49.99 – $50.00, while stock B's quotes are $4.99 – $5.00. If news arrives implying a 3 bps increase in the stock fundamental value, the change in midquote that reflects that change in fundamental value is ($\$49.995 \times 0.0003 = \$0.015$) for stock A, but only ($\$4.995 \times 0.0003 = \$0.0015$) for stock B. Because the change in fundamental value is greater than the minimum tick size for stock A ($\$0.015 > \$0.01$), the market maker in A will update his quotes (shifting the mid-quote from $49.995 to $50.005, as the new bid-ask becomes $50 – $50.01). However, the market maker in stock B will not update quotes, as the value change lies within the bid-ask spread (3 bps improvement translates into $0.0015 value, which is smaller than the minimum price increment). If the two securities have the same volatility in the fundamental value, the market maker in security A (low tick-to-price security) will revise his quotes more often due to the finer pricing grid than the market maker in security B (high tick-to-price security).

Empirically, regression (4) in Table 4 suggests that a 1% higher tick-to-price ratio is associated with 13.14% lower OTTRs (in line with Hypothesis 3c). This finding complements evidence in Yao & Ye (2018) that a large relative tick size constrains competition in prices and results in lower OTTRs.

The regression results show that OTTRs are higher for large stocks, in line with Hypothesis 4a, and consistent with evidence in O'Hara (2015) and Rosu et al. (2020). As shown in regression model (4) in Table 4, a 1% higher market capitalization is associated with a 4% higher OTTR, all else equal. Because information about large stocks is more frequently updated and more readily available (e.g., more analyst coverage, more news stories, better data coverage) large stocks have higher monitoring intensity by liquidity providers and therefore higher OTTRs.

We also find that OTTRs increase with microprocessor speeds (positive coefficient on the CPU speed variable) as data processing costs decrease, supporting Hypothesis 4b. The regression results also confirm that OTTRs increase with the introduction of Autoquote in the US even after controlling for the other factors.

As an alternative to Autoquote, we include a dummy variable that takes the value of one after the effective date for HFT introduction in the US following Aitken et al. (2015). The significant positive coefficient indicates that OTTRs are higher in the period of HFT activity, controlling for other factors. These results are consistent with Aitken et al. (2015), confirming that HFT market makers are among the leaders in exploiting faster data processing speeds and the availability of rich, real-time market data. Since HFTs tend to be more active towards the end of a trading day, as they seek to offload their holdings, we also test whether OTTRs are higher in the last hour before the stock market close. We find that consistent with a positive relation between HFTs and OTTRs, average OTTRs in the last hour of trading are 52.89, compared to 16.87 between 10 am and 3 pm. The difference is statistically significant with a t-statistic of 307.19.

[Figure 2 about here.]

The results also indicate a positive relation between volatility and OTTRs, consistent with increased monitoring when picking-off risks are high. Regression (3) in Table 4 estimates that a one percentage point increase in market volatility (high-low range) increases OTTRs by 3.05%. This results supports the hypothesis (5a) that as the risk of trading at stale prices increases with a rise in volatility, markets makers increase their monitoring intensity, leading to higher OTTRs. It is also consistent with a higher arrival intensity of new information during periods of high volatility. Regression (4) in Table 4 confirms a similar (though lower in magnitude) relation between OTTR and individual stock volatility (in line with Hypothesis 5b).

To assess the economic significance and relative importance of the various factors driving OTTRs, we standardize the coefficients and plot their relative magnitude in Figure 3. The plot quantifies the effect on OTTRs of a one standard deviation increase in each of the explanatory variables. Panel A confirms that the ETF dummy, traded dollar volume, autoquote, and the security's correlation with the S&P500 are economically important determinants of OTTRs. For example, a one standard deviation increase in volume is associated with a 42.18% decrease in OTTR, while a one standard deviation increase in a security's correlation with the S&P500 index is associated with a

22.25% increase in OTTRs.

[Figure 3 about here.]

As a further test of how a stock's information environment (the availability of relevant signals) affects its OTTR, we analyze a natural experiment in which the NYSE introduced order book transparency. On January 24, 2002, NYSE introduced a new system that allowed market participants to see the orders in the limit order book (see Boehmer et al. (2005)). Importantly, Nasdaq did not change its level of transparency at that time, making stocks that traded predominantly on Nasdaq a natural control group. To examine how this exogenous increase in the availability of high-frequency, stock-specific information (corresponding to relevant signals in our model) affected OTTRs we estimate difference-in-differences models. The dependent variable is the log OTTR, while the key independent variable is the interaction of a post transparency dummy and the fraction of the stock's trading volume that is executed on NYSE as a measure of which stocks were most affected by the change.

The results in Table 6 support the model's prediction that OTTRs are higher when there is a greater availability of relevant information for monitoring. The interaction term coefficient is positive and statistically significant in all specifications. In this analysis, we also control for the same OTTR determinants as in our baseline regressions in Table 5. Our findings are consistent with results in Boehmer et al. (2005), who also find that post-transparency cancellation rates on NYSE increase, and time-to-cancellation of limit orders shortens.

[Table 6 about here.]

3.4 How OTTRs vary across markets

A third dimension in which OTTRs vary (in addition to time-series and cross-sectional variation) is across markets. The theory predicts that markets with a lower share of trading volume have higher OTTRs (see Equation (6)). To test this conjecture, we relate the average OTTR observed on a given exchange to the exchange's market share while controlling for other factors, such as fragmentation, volatility, and a CPU speeds.

We find that the average OTTR for an exchange is inversely related to its share of stock trading volume, consistent with Hypothesis 2. Regression model (2) in Table 7 suggests that OTTRs decrease by 2% for each 1% increase in an exchange's market share. This result is consistent with market makers executing trades on small venues less frequently, but updating quotes across venues at similar frequency. The estimated coefficient for ETF market share is also negative in Table 7, but is not statistically significant.

[Table 7 about here.]

[Table 8 about here.]

Differences in exchange fee structures also lead to variation in OTTRs across markets. We distinguish between the "taker-maker" markets (Edge-A, Bats-Y, and NASDAQ Boston) and the "maker-taker" markets (the rest). "Maker-taker" is a trading fee structure that charges a higher fee to "liquidity takers" (those submitting market orders) than to "liquidity makers" (those posting limit orders), with the latter sometimes being rewarded with a rebate for executed limit orders. "Taker-maker" markets do the opposite: they charge limit orders and compensate market orders. The taker-maker fee model effectively allows trades at sub-penny increments (net of fees) thereby offering more granular pricing. The taker-maker fee structure is therefore expected to have a similar effect to a smaller tick size: market makers update quotes more often, because the finer pricing grid allows them to respond to small changes in value. The univariate analysis indicates that OTTRs are higher on taker-maker markets (103.28 vs 64.87, with t-statistics of 91.02). Multivariate regression results (e.g., regression (3) in Table 7,) suggest that OTTRs are 32% higher on markets with taker-maker fee structures, controlling for other covariates, consistent with Hypothesis 3d.

The economic significance of the exchange-level results is presented in Panel B of Figure 3. The effects of CPU speeds and market share are very strong at the exchange level: OTTRs for exchanges increase at an average rate of 42.21% per standard deviation of CPU speed. Market share is the second most important factor: a one standard deviation increase in market share is associated with a 23.21% decrease in OTTRs.

3.5 The non-linear effects of fragmentation and market share

Recall that Equation (4) of the theory suggests that OTTRs scale up linearly with the number of markets on which a security is traded. The key assumption that generates this prediction is that market makers post quotes on all of the markets that trade a security. The regressions in Table 4 confirm that OTTRs are positively related to fragmentation. However, whether the relation is linear or not depends on whether liquidity providers post quotes across all markets or only some.

To examine the potential non-linearity in how OTTRs relate to the number of markets, we introduce a squared term in the regression model (7) in Table 4. We then plot the fitted values of OTTRs at different levels of fragmentation and compare them to the theoretical, linear OTTR-fragmentation relation to see whether the non-linearity is meaningful. The fitted OTTR is computed using coefficients from regression model (7) in Table 4, with explanatory variables (apart from $Frag1$) fixed at their average levels from Table 3. Therefore, the approach controls for the effects of other confounding variables when examining the empirical relation between fragmentation and OTTRs.

The plotted theoretical relation between OTTR and fragmentation uses Equation (4) and varies the number of markets from 1 to 14, keeping other parameters fixed. We conservatively assume that liquidity providers monitor only one signal (the S&P500 ETF, SPY), and compute the signal quality as the daily proportion of same-direction mid-quote changes in SPY and a given stock, out of the total number of mid-quote changes in SPY (similar to Dobrev & Schaumburg (2017)). Trading (quoting) intensity in a given stock is just the average daily number of trades (quote updates).

The result, in Figure 4, shows that for moderate levels of fragmentation (2–4 markets), the empirically observed OTTRs from the regression are broadly in line with the theoretically predicted OTTRs. However, for higher levels of fragmentation (5–14 markets), the empirically observed OTTRs are lower than those implied by the linear theoretical relation. This result is consistent with the notion that market makers post their orders on several exchanges but not all 14 and therefore OTTRs increase with fragmentation, but at a decreasing marginal rate.

[Figure 4 about here.]

Next, we examine the relation between an exchange's market share and its OTTR. The theory predicts that an exchange's OTTR is inversely related to its share of trading. The empirical results in Table 7 confirm the inverse relation between market share and OTTR. Similar to the approach taken in examining the non-linear effects of fragmentation, we augment the regression models by adding squared market share (see model (7) in Table 7). We then plot the fitted values of the non-linear regression, varying market share from 0.02 to 0.32 and holding other variables at their means. Finally, we compare this empirical relation between OTTR and market share with the theoretic relation obtained from Equation (6).

The result in Figure 4 shows that the empirical relation between OTTRs and market share is not as non-linear as predicted by the theory. In particular, at very low levels of market share such as 0-0.1, the empirically observed OTTRs do not spike as much as predicted by the theory. The theoretical OTTRs spike for very low market shares because of the assumption that liquidity providers replicate their quotes across *all* exchanges that trade a security. Under this assumption, the number of order messages (quote updates) across the various exchanges is approximately constant and so an exchange with a very low amount of trading will have a much higher OTTR. The empirically observed relation between OTTRs and market shares suggests that the number of order messages is not constant across exchanges, but instead it tends to decrease with market share.

4 A benchmark for regulatory surveillance of OTTRs

Abnormally high OTTRs can be a sign of illegal trading strategies such as spoofing. But what is an "abnormal" OTTR? How can regulators tell whether a security with twice the OTTR of another security should raise concerns about potential manipulation versus simply being a security with a naturally higher OTTR? How can regulators tell whether a spike in OTTRs on a particular day reflects suspicious trading or just different market conditions? The challenge is that OTTRs vary substantially through time, across stocks, and across markets for perfectly legitimate reasons.

The theory model provides a simple way to examine whether the OTTR in a given setting is abnormal by considering the drivers of natural variation in OTTRs. For example, by using Equation (4), and the empirical proxies for the theoretical parameters, we can estimate the expected OTTR of a given security on a given day. This provides a benchmark for a "normal" OTTR, against which to gauge whether an observed OTTR is abnormal and to what extent. A regulator can use such an OTTR benchmark to monitor markets and identify suspicious trading activity.

To demonstrate this approach, we use a conservatively calibrated version of the theory model to estimate theoretical OTTRs for 715 stock-days in November 2018 (every 5th stock in the sample by size). We then compare the distribution of theoretical OTTRs to the distribution of observed OTTRs for those same stock-days.

We obtain the conservatively calibrated theoretical OTTRs in two steps. First, we calculate the following OTTR base rate for each stock-day:

$$OTTR_{it} = \frac{2M_{it}(\lambda_{SPY_t} q_{it} + \lambda_{mit})}{\lambda_{mit}} \tag{7}$$

where M_{it} is the number of markets trading stock i on day t, λ_{SPY_t} is the number of mid-quote updates in the SPY ETF, which is assumed, conservatively, to be the only signal monitored by liquidity providers. The SPY, being one of the most liquid market-tracking securities, is likely to provide a market maker with one of the most relevant indications of changes in market-wide valuations. We estimate the quality of this signal, q_{it}, as the proportion of same-direction mid-quote changes in SPY and the given stock i, out of the total number of mid-quote changes in SPY, similar to Dobrev & Schaumburg (2017). Recall that theoretically, the quality of a signal is the probability that a change in the signal value results in a revised quote for the stock. Finally, λ_{mit} is the number of trades in stock i on day t.

In the second step, we account for the natural variation in OTTRs across stocks and in different market conditions. To do this, we start with the "conservative" OTTR base rate calculated above and add the variation implied by the estimated drivers of OTTRs. That is, we take the fitted (predicted) OTTRs for each stock-day from regression (1) in Table 4, normalize them to have zero mean (subtract their average), and then add them to the theoretical OTTR base rate. This approach effectively adds the natural variation in OTTRs to the conservatively estimated mean levels. Using these estimates of the natural or expected OTTRs as a benchmark, we obtain the excess OTTR for each stock-day by subtracting the benchmark from the actual OTTR.

[Figure 5 about here.]

[Table 9 about here.]

In Figure 5, we plot the empirical distribution of actual OTTRs in Panel A and the distribution of excess OTTRs in panel B. Table 9 provides the descriptive statistics for the same two distributions. On average, the observed OTTRs are lower than the conservative theoretical OTTRs expected to arise from liquidity provision. However, the empirical distribution of observed OTTRs is right-skewed and for 17% of stock-day observations the empirical OTTRs are higher than the theoretical benchmark values. Given we calculated the theoretical OTTRs conservatively by assuming only one signal is monitored, the results suggest that most of the time empirically observed OTTRs are well within levels that are consistent with legitimate market making, but that they occasionally spike to abnormal levels. By scrutinizing "excess OTTRs" as opposed to raw OTTRs, regulators can better identify areas of concern for further investigation.

5 Conclusions and policy implications

Our results shed new light on why OTTRs have grown so rapidly in equity markets and whether they warrant concern. We find that the growth in OTTRs through time is largely due to three major changes in US stock markets: (i) the proliferation of exchanges and fragmentation of trading following regulatory changes that encouraged competition between exchanges (e.g., Reg NMS); (ii) automation of quote dissemination, increasing the availability and timeliness of market data; and (iii) technological improvements such as increased computational power that increase the ability of market participants to monitor a large set of data feeds. Fragmentation of trading increases OTTRs because limit orders are often duplicated across markets, leading to a larger number of quote messages for the same number of trades. Automation of quote dissemination encouraged growth in algorithmic trading, which typically involves more frequent quote updates than manual market making. Finally, by increasing the ability to process data feeds, technological improvements lead to more frequent quote updates as prices or volumes change in other securities.

In light of these drivers, we show that OTTRs in the most recent part of our sample are, in most cases, well within the levels that would be expected to prevail under normal market making. Our benchmark for what is expected under normal market making is calculated under conservative assumptions (providing a lower bound on what is a normal OTTR) and accounts for the degree of fragmentation in today's markets and the high frequency of market data. We therefore conclude that the overall levels of OTTRs in the market are consistent with legitimate market making, but occasionally they spike to levels that warrant further investigation.

When investigating whether a given OTTR is abnormal, it is important to account for the substantial cross-sectional and time-series variation in what is a "normal" OTTR. The model of the determinants of OTTRs provides one way to achieve this. The results show that OTTRs tend to be higher in more volatile stocks and market conditions (due to more frequent information arrivals), higher price-to-tick stocks (due to the higher resolution of price increments), lower volume stocks (due to smaller denominator), and in ETFs (because of the availability of frequent and precise signals about an ETF's value). We find that OTTRs are naturally much higher on markets with lower market shares. All of these effects are consistent with a model of market making and therefore should be considered when evaluating whether an OTTR is abnormal, such as in a surveillance system that monitors a market for suspicious activity.

Our findings have implications for how regulatory measures that aim to constrain the level of order activity are likely to affect market quality. The model shows that the OTTR emerges endogenously as liquidity providers balance the tradeoff between the costs of monitoring signals and updating their quotes and the costs of having their quotes adversely selected when they have not updated them in response to new information. Adding a fee, a tax, or a limit on order messages raises the costs of keeping quotes up to date. Our model implies the outcome is lower monitoring and updating intensity by liquidity providers, which comes at the expense of higher adverse selection risk. Intuitively, if it is more costly to keep quotes up to date with the latest informational arrivals and changes in market conditions, those quotes will more often be "picked off" and execute when it is unfavorable for the liquidity provider. The ultimate consequence is less liquidity provision (wider spreads and/or less depth) and higher trading costs as liquidity providers recoup the higher adverse selection costs.

The reduction in liquidity due to an order message tax or limit is likely to be the largest when OTTRs would naturally be at high levels, such as periods of high volatility or in low-volume stocks. Such a tax might therefore disproportionately harm liquidity where liquidity is already scarce. Conversely, if OTTR limits or taxes are set to take effect only beyond a certain level, then as long as that level is above the natural OTTR that arises from market making, then the limits/taxes may not impede liquidity provision, and may instead help reduce excessive OTTRs. A limit/tax on OTTRs beyond a particular threshold should take into account that OTTRs will naturally tend to be much higher on exchanges that have lower market shares and for ETFs compared to stocks.

Appendix A. Proofs

Proposition 2.1. The OTTR for a given security increases with the extent of fragmentation of the security's trading across multiple trading venues.

Proof of Proposition 2.1.

Recall the expression for the OTTR for a market maker who provides liquidity in a given stock in M markets: $OTTR = \frac{2M(\sum_{n\in\Omega_{s^*}} \lambda_n q_n + \lambda_m)}{\lambda_m}$. Taking the first derivative with respect to the number of markets: $\frac{dOTTR}{dM} = \frac{2(\sum_{n\in\Omega_{s^*}} \lambda_n q_n + \lambda_m)}{\lambda_m}$. This expression is strictly positive, because $\frac{2\sum_{n\in\Omega_{s^*}} \lambda_n q_n}{\lambda_m} + 2 > 0$, since $\lambda_n q_n \geq 0$ and $\lambda_m > 0$, which means that OTTR is increasing in the number of markets (fragmentation).

Intuitively, for a single market case ($M = 1$), $OTTR \geq 2$ 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 $q_n = 0 \; \forall \; n \in \Omega_{s^*}$. As $q_n \geq 0$ by construction (signal quality cannot be negative), $OTTR > 2$ for all cases except for $q_n = 0$. As the number of markets increases (e.g., $M = 2$), the OTTR increases, because a market maker now updates quotes twice as often: (a) on two markets rather than one, when a new useful signal n arrives, and (b) on two markets rather than one, when a new market order arrives on at least one market.

Proposition 2.2 The OTTR for a given trading venue is inversely related to its market share.

Proof of Proposition 2.2

Recall the expression for the OTTR for a given trading venue j: $OTTR_j = \frac{2\sum_{n\in\Omega_{s^*}} \lambda_n q_n + 2}{\lambda_m \rho_j}$. Taking the first derivative with respect to the market share: $\frac{dOTTR_j}{d\rho_j} = -\frac{2\sum_{n\in\Omega_{s^*}} \lambda_n q_n + 2}{\lambda_m \rho_j^2} < 0 \; \forall \rho_j \in (0,1), \lambda_n, q_n, \lambda_m$. The negative first derivative of $OTTR_j$ with respect to market share means that a trading venue's OTTR decreases when its market share increases.

Proposition 2.3. The OTTR for a given security increases with the quality of signals available for monitoring.

Proof of Proposition 2.3.

Recall that a market maker's monitoring intensity is an important driver of the OTTR: $OTTR = \frac{2M(\sum_{n\in\Omega_{s^*}} \lambda_n q_n + \lambda_m)}{\lambda_m}$, where monitoring intensity is the number of *monitored* signals in the set Ω_s^*.

As more signals are monitored, the liquidity provider posts proportionally more quote updates in response to those signals, driving the OTTR up. The liquidity provider monitors all signals for which the marginal benefit of monitoring, $\lambda_m \left(\frac{\lambda_n q_n}{\lambda_m+\lambda_n q_n}\right) k$, exceeds the marginal cost, $\lambda_n c$. Because improved signal quality increases the marginal benefit of monitoring without affecting the marginal cost, the liquidity provider will monitor more when he receives better quality signals.

Proposition 2.4. The OTTR for a given security increases with lower monitoring costs.

Proof of Proposition 2.4.

The market maker's cost of processing an information arrival is c. Therefore, his cost of monitoring signal n per unit time is $\lambda_n c$. Marginal net benefit of monitoring (per unit time) is: $\left(\frac{\lambda_n q_n}{\lambda_m+\lambda_n q_n}\right) k - \lambda_n c$.

To see that the monitoring intensity (and therefore the OTTR) increases with a lower cost of monitoring, take the first derivative of marginal net benefit with respect to c: $\frac{d\left(\left(\frac{\lambda_n q_n}{\lambda_m+\lambda_n q_n}\right)k-\lambda_n c\right)}{dc} = -\lambda_n < 0$, as $\lambda_n > 0$ by the properties of Poisson processes (signal intensity, or the number of signal updates per unit time, can only be a positive number). Because a lower c reduces the marginal cost of monitoring without affecting the marginal benefit, the liquidity provider will monitor more signals when his cost of monitoring is lower.

Proposition 2.5. The OTTR for a given security increases with picking-off cost.

Proof of Proposition 2.5.

The market maker incurs the cost k each time he is hit by a market order without having updated his quotes. Taking the derivative of marginal net benefit of monitoring with respect to k: $\frac{d\left(\left(\frac{\lambda_n q_n}{\lambda_m+\lambda_n q_n}\right)k-\lambda_n c\right)}{dk} = \frac{\lambda_n q_n}{\lambda_m+\lambda_n q_n} > 0$. This expression is strictly positive for all non-zero quality signals, hence monitoring intensity (and therefore OTTR) increases with higher picking-off cost.

Proposition 2.6. The OTTR for a given security decreases with the trading frequency, holding the monitoring intensity constant.

Proof of Proposition 2.6.

For a given monitoring intensity, trading frequency only enters the OTTR expression to reflect quote updates in response to executed trades (in the numerator), and the number of executed trades (in the denominator). Taking the first derivative of the OTTR with respect to trading frequency:

$\frac{dOTTR}{d\lambda_m} = \frac{-2M\sum_{n\in\Omega_{s^*}}\lambda_n q_n}{\lambda_m^2}$. Because $\frac{dOTTR}{d\lambda_m} < 0$ for all parameter values that correspond to non-zero quality signals, OTTR decreases as the trading frequency increases.

References

Aitken, M., Cumming, D. & Zhan, F. (2015), 'High frequency trading and end-of-day price dislocation', *Journal of Banking & Finance* **59**, 330–349.

Baldauf, M. & Mollner, J. (2019), 'Trading in fragmented markets', *Journal of Financial and Quantitative Analysis* **54**(1), 1–58.

Biais, B., Foucault, T. & Moinas, S. (2011), 'Equilibrium high-frequency trading', Working paper.

Biais, B., Martimort, D. & Rochet, J.-C. (2000), 'Competing mechanisms in a common value environment', *Econometrica* **68**(4), 799–837.

Boehmer, E., Saar, G. & Yu, L. (2005), 'Lifting the veil: An analysis of pre-trade transparency at the NYSE', *Journal of Finance* **60**(2), 783–815.

Budish, E., Lee, R. S. & Shim, J. J. (2020), 'Will the market fix the market? A theory of stock exchange competition and innovation', Working paper.

Chowdhry, B. & Nanda, V. (1991), 'Multimarket trading and market liquidity', *Review of Financial Studies* **4**(3), 483–511.

Committee on Capital Markets Regulation (2016), 'The US equity markets. A plan for regulatory reform.', `https://www.capmktsreg.org/wp-content/uploads/2018/10/The-US-Equity-Markets.pdf`.

Dahlström, P., Hagströmer, B. & Nordén, L. L. (2018), 'Determinants of limit order cancellations', Working paper.

Dobrev, D. & Schaumburg, E. (2017), 'High-frequency cross-market trading: Model free measurement and applications', Working paper.

Doidge, C., Karolyi, G. A. & Stulz, R. M. (2017), 'The US listing gap', *Journal of Financial Economics* **123**(3), 464–487.

Egginton, J. F., Van Ness, B. F. & Van Ness, R. A. (2016), 'Quote stuffing', *Financial Management* **45**(3), 583–608.

Foucault, T., Kadan, O. & Kandel, E. (2013), 'Liquidity cycles and make/take fees in electronic markets', *Journal of Finance* **68**(1), 299–341.

Foucault, T., Röell, A. & Sandås, P. (2003), 'Market making with costly monitoring: An analysis of the SOES controversy', *Review of Financial Studies* **16**(2), 345–384.

Friederich, S. & Payne, R. (2015), 'Order-to-trade ratios and market liquidity', *Journal of Banking & Finance* **50**, 214–223.

Hendershott, T., Jones, C. M. & Menkveld, A. J. (2011), 'Does algorithmic trading improve liquidity?', *Journal of Finance* **66**(1), 1–33.

Jain, P. & Jordan, S. J. (2017), 'Cancellation latency: The good, the bad, and the ugly', *Financial Management* **46**(2), 377–407.

Jørgensen, K., Skjeltorp, J. & Ødegaard, B. A. (2018), 'Throttling hyperactive robots – Order-to-trade ratios at the Oslo Stock Exchange', *Journal of Financial Markets* **37**, 1–16.

Kurzweil, R. (2005), *The singularity is near: When humans transcend biology*, Penguin.

Liu, W.-M. (2009), 'Monitoring and limit order submission risks', *Journal of Financial Markets* **12**(1), 107–141.

Lyle, M. R. & Naughton, J. P. (2018), 'How does algorithmic trading improve market quality?', Working paper.

Nikolsko-Rzhevska, O., Nikolsko-Rzhevskyy, A. & Black, J. R. (2020), 'The life of U's: Order revisions on NASDAQ', *Journal of Banking & Finance* **111**, 105724.

O'Hara, M. (2015), 'High frequency market microstructure', *Journal of Financial Economics* **116**(2), 257–270.

Pagnotta, E. S. & Philippon, T. (2018), 'Competing on speed', *Econometrica* **86**(3), 1067–1115.

Putniņš, T. J. (2020), An overview of market manipulation, *in* 'Handbook of Corruption and Fraud in Financial Markets: Malpractice, Misconduct and Manipulation', Wiley, pp. 13–40.

Rosu, I., Sojli, E. & Tham, W. W. (2020), 'Quoting activity and the cost of capital', Working paper.

Van Kervel, V. (2015), 'Competition for order flow with fast and slow traders', *Review of Financial Studies* **28**(7), 2094–2127.

Wang, X. & Ye, M. (2017), 'Who provides liquidity and when: An analysis of price vs. speed competition on liquidity and welfare', Working paper.

Yao, C. & Ye, M. (2018), 'Why trading speed matters: A tale of queue rationing under price controls', *Review of Financial Studies* **31**(6), 2157–2183.

Table 1: Summary of propositions and empirical hypotheses

Propositions and empirical hypotheses	Empirical support
Proposition 2.1. The OTTR for a given security increases with the extent of fragmentation of the security's trading across multiple exchanges.	
Hypothesis 1a. *OTTRs are higher for securities with more fragmented trading.*	yes
Hypothesis 1b. *OTTRs are higher for exchanges that trade more fragmented securities.*	yes
Proposition 2.2. The OTTR for a given exchange is inversely related to its market share.	
Hypothesis 2. *OTTRs are higher for markets with lower market shares.*	yes
Proposition 2.3. The OTTR for a given security increases with the quality of signals available for monitoring.	
Hypothesis 3a. *ETFs have higher OTTRs compared to stocks.*	yes
Hypothesis 3b. *Securities with higher absolute correlation with the market index have higher OTTRs.*	yes
Hypothesis 3c. *OTTRs are higher for securities with higher price-to-tick ratios.*	yes
Hypothesis 3d. *OTTRs are higher on markets with taker-maker fee structures.*	yes
Proposition 2.4. The OTTR for a given security increases with lower monitoring costs.	
Hypothesis 4a. *OTTRs are higher for stocks with higher market capitalization.*	yes
Hypothesis 4b. *OTTRs increase as data processing speeds increase.*	yes
Proposition 2.5. The OTTR for a given security increases with picking-off cost.	
Hypothesis 5a. *OTTRs are higher on days with higher market volatility.*	yes
Hypothesis 5b. *OTTRs are higher for more volatile stocks.*	yes
Proposition 2.6. The OTTR for a given security decreases with the trading frequency, holding the monitoring intensity constant.	
Hypothesis 6. *OTTRs are inversely related to trading volumes.*	yes

Table 2: Variable definitions

Variable name	Variable definition
Security-day observations	
OTTR	Number of order messages between 10 am and 3 pm (enter, amend, and cancel) at the best quotes, divided by the number of trades. Winsorized at 99th percentile.
Frag1	Number of exchanges on which a security has at least one trade on the given day.
Frag2	One minus the Herfindahl-Hirschman index (HHI) of dollar volumes. The HHI is the sum of squared market shares of all exchanges trading a given security.
Frag3	One minus the HHI of the number of trades. The HHI is the sum of squared market shares of all exchanges trading a given security.
ETFDummy	Takes the value of one for exchange-traded funds, and zero otherwise.
AbsCorrelS&P	Absolute value of the security's 22-day correlation with S&P500 index. Calculated using daily returns.
TickToPrice	Tick size (minimum price increment) divided by the closing price.
HighLowVolat	Daily stock volatility measure, computed as the daily high minus low price, divided by the average of the high and low prices.
MktCap	Market capitalization in thousand USD. For ETFs, MktCap is AUM (as reported in CRSP).
Volume	Number of shares traded.
Exchange-day observations	
OTTR	Average of the OTTR traded on the exchange, computed from the winsorized security-day OTTRs.
NumMkts	Average of the number of exchanges ($Frag1$) of stocks and ETFs traded on the exchange.
MktShareStocks	Exchange's market share in stock trading, computed from dollar volumes.
MktShareETFs	Exchange's market share in ETF trading, computed from dollar volumes.
TakerDummy	Takes the value of one if the exchange has a taker-maker fee structure and zero otherwise.
Time series observations	
HighLowVolatMkt	Daily market volatility measure, computed as the daily high minus low S&P500 level, divided by the average of daily high and low.
VIX	Level of the VIX volatility index.
Autoquote	Takes the value of one after January 29, 2003 (the phase-in of Autoquote) and zero otherwise.
HFTdate	Takes the value of one after July 31, 2003 (the HFT effective date according to Aitken et al. (2015) using the order cancellations approach) and zero otherwise.
CPUSpeed	Microprocessor speed in Hertz, pulses per second $\times 10^{-10}$, average per year.

Table 3: Descriptive statistics

This table reports descriptive statistics for the variables used in regression analysis. The sample contains 1,210,225 stock-day observations, 215,748 ETF-day observations, 48,772 exchange-day observations, and 5,284 time-series observations. OTTR is the order-to-trade ratio. $Frag1$ is a measure of fragmentation based on number of markets. Frag2 ($Frag3$) is a measure of fragmentation based on the Herfindahl-Hirschman index of market shares by dollar volume (trades). $AbsCorrelS\&P$ is the absolute correlation of the security returns with the S&P500 index. $HighLowVolat$ is a security-level volatility measure, computed from the high/low price range. $HighLowVolatMkt$ is a market-level volatility measure, computed from the high/low range of the S&P500 index. The sample period covers years 1998 to 2018, and includes 241 stocks and 68 ETFs.

	Mean	StdDev	25th pctl	50th pctl	75th pctl
Stock-day observations					
OTTR	16.88	35.28	4.00	7.33	13.66
Frag1	5.45	3.22	3.00	5.00	8.00
Frag2	0.47	0.26	0.26	0.53	0.70
Frag3	0.53	0.26	0.33	0.61	0.75
AbsCorrelS&P	0.39	0.24	0.18	0.37	0.58
TickToPrice, %	0.23	0.89	0.03	0.07	0.21
HighLowVolat, %	4.06	4.06	1.79	2.95	4.95
MktCap, $ bn	5.70	24.15	0.11	0.54	2.36
Volume, mln	0.99	3.84	0.02	0.15	0.62
ETF-day observations					
OTTR	349.42	744.69	23.96	82.62	300.19
Frag1	4.76	2.81	2.00	4.00	7.00
Frag2	0.50	0.27	0.34	0.57	0.72
Frag3	0.54	0.26	0.43	0.61	0.75
AbsCorrelS&P	0.67	0.28	0.49	0.77	0.89
TickToPrice, %	0.04	0.07	0.01	0.02	0.04
HighLowVolat, %	1.04	2.66	0.37	0.68	1.21
MktCap, $ bn	1.85	8.26	0.04	0.19	0.89
Volume, mln	1.00	7.42	0.01	0.03	0.18
Exchange-day observations					
OTTR	70.16	48.57	34.69	60.44	98.01
NumMkts	6.53	2.74	3.57	7.50	8.84
MktShareStocks	0.09	0.14	0.01	0.05	0.13
MktShareETFs	0.11	0.17	0.00	0.03	0.14
Time series observations					
HighLowVolatMkt, %	1.35	1.00	0.71	1.09	1.68
VIX	20.21	8.50	13.89	18.55	24.05
CPUSpeed, $Hertz \times 10^{-10}$	1.22	1.18	0.27	0.79	2.05

Table 4: Determinants of OTTRs using security-day observations

This table reports OLS regression results for regression models with security-day observations. The dependent variable is $LogOTTR = Ln(1 + OTTR)$. Independent variables are in the first column. Variable definitions are in Table 2, noting that $LogTickToPrice = Ln(1+TickToPrice)$, $LogMktCap = Ln(MktCap)$, $LogVIX = Ln(VIX)$. Standard errors are clustered by security and day. T-statistics are reported in parentheses. ***, **, and * indicate statistical significance at 1%, 5%, and 10% levels, respectively. The sample period covers years 1998 to 2018, and includes 241 stocks and 68 ETFs.

	Log OTTR (1)	Log OTTR (2)	Log OTTR (3)	Log OTTR (4)	Log OTTR (5)	Log OTTR (6)	Log OTTR (7)	Log OTTR (8)	Log OTTR (9)
$Frag1$	0.08*** (25.22)						0.15*** (9.67)		
$Frag1^2$							-0.01*** (-5.29)		
$Frag2$		0.70*** (15.17)		0.65*** (13.80)	0.69*** (15.02)	0.66*** (13.96)		0.25*** (6.72)	0.24*** (6.33)
$Frag3$			0.50*** (11.67)						
$ETFDummy$	1.44*** (24.20)	1.35*** (22.58)	1.34*** (22.64)	1.35*** (22.87)	1.35*** (22.50)	1.35*** (22.83)	1.43*** (24.03)	1.38*** (23.90)	1.38*** (23.84)
$AbsCorrelS\&P$	1.13*** (20.85)	1.13*** (20.63)	1.14*** (21.47)	1.16*** (21.29)	1.15*** (20.65)	1.13*** (20.32)	1.10*** (20.00)	1.00*** (18.14)	1.00*** (18.14)
$LogTickToPrice$	-14.88* (-1.96)	-14.23** (-1.96)	-13.94** (-2.21)	-13.14* (-1.83)	-14.23** (-1.96)	-13.09* (-1.83)	-14.52* (-1.95)	-7.57 (-1.64)	-7.83* (-1.65)
$HighLowVolatMkt$	2.28*** (2.93)	3.24*** (4.51)	3.05*** (4.22)			3.25*** (4.35)	2.43*** (3.17)	6.41*** (8.49)	6.32*** (8.37)
$LogMktCap$	0.01 (1.57)	0.03*** (3.33)	0.03*** (3.08)	0.04*** (3.76)	0.03*** (3.08)	0.04*** (3.79)	0.02* (1.92)	0.07*** (7.98)	0.07*** (7.88)
$LogVolume$	-0.29*** (-38.84)	-0.29*** (-38.10)	-0.28*** (-38.55)	-0.29*** (-36.95)	-0.28*** (-37.81)	-0.29*** (-36.83)	-0.30*** (-39.67)	-0.30*** (-38.97)	-0.30*** (-38.79)
$LogVIX$					0.03 (0.98)				
$HighLowVolat$				0.63*** (3.83)		0.39** (2.33)		0.65*** (4.06)	0.64*** (3.99)
$Autoquote$								1.00*** (30.06)	
$CPUSpeed$		0.09*** (7.50)	0.11*** (9.16)	0.08*** (6.81)	0.09*** (7.16)	0.09*** (7.23)			
$HFTdate$									0.98*** (29.57)
Adj. R^2	53%	54%	54%	54%	54%	54%	53%	57%	57%
ClusteredStd Errors	Security &Day	Security &Day	Security &Day	Security &Day	Security &Day	Security &Day	Security &Day	Security &Day	Security &Day

Table 5: Determinants of OTTRs in the security-day sample excluding ETFs

This table reports OLS regression results for regression models with stock-day observations. The dependent variable is $LogOTTR = Ln(1 + OTTR)$. Independent variables are in the first column. Variable definitions are in Table 2, noting that $LogTickToPrice = Ln(1 + TickToPrice)$, $LogMktCap = Ln(MktCap)$, $LogVIX = Ln(VIX)$. Standard errors are clustered by stock and day. T-statistics are reported in parentheses. ***, **, and * indicate statistical significance at 1%, 5%, and 10% levels, respectively. The sample period covers years 1998 to 2018, and includes 241 stocks.

	Log OTTR (1)	Log OTTR (2)	Log OTTR (3)	Log OTTR (4)	Log OTTR (5)	Log OTTR (6)	Log OTTR (7)	Log OTTR (8)	Log OTTR (9)
$Frag1$	0.08***						0.24***		
	(29.98)						(17.58)		
$Frag1^2$							-0.01		
							(-12.29)		
$Frag2$		1.02***		0.98***	1.02***	0.98***		0.41***	0.40***
		(23.14)		(21.61)	(23.16)	(21.58)		(12.31)	(11.96)
$Frag3$			0.72***						
			(17.39)						
$AbsCorrelS\&P$	0.66***	0.61***	0.68***	0.61***	0.66***	0.61***	0.59***	0.50***	0.50***
	(22.93)	(23.19)	(24.43)	(23.12)	(24.85)	(22.54)	(21.69)	(20.38)	(20.38)
$LogTickToPrice$	-12.08*	-10.93**	-11.56**	-9.84*	-10.77**	-9.83*	-11.13*	-5.37	-5.64
	(-1.95)	(-1.97)	(-2.20)	(-1.86)	(-1.97)	(-1.86)	(-1.94)	(-1.63)	(-1.64)
$HighLowVolatMkt$	0.30	0.59	0.17			0.38	0.58	4.59***	4.47***
	(0.40)	(0.91)	(0.25)			(0.58)	(0.82)	(8.00)	(7.80)
$LogMktCap$	0.05***	0.08***	0.07***	0.09***	0.07***	0.09***	0.06***	0.12***	0.12***
	(6.63)	(10.42)	(9.35)	(11.03)	(9.93)	(11.03)	(8.00)	(16.75)	(16.51)
$LogVolume$	-0.26***	-0.27***	-0.26***	-0.28***	-0.27***	-0.28***	-0.28***	-0.27***	-0.27***
	(-36.71)	(-36.58)	(-35.85)	(-35.16)	(-36.52)	(-35.19)	(-38.90)	(-38.02)	(-37.77)
$LogVIX$					-0.10***				
					(-4.07)				
$HighLowVolat$				0.62***		0.60***		0.82***	0.81***
				(4.04)		(3.79)		(5.80)	(5.66)
$Autoquote$								0.88***	
								(31.40)	
$CPUSpeed$		0.03***	0.07***	0.02**	0.02	0.02**			
		(2.58)	(5.85)	(2.15)	(1.46)	(2.20)			
$HFTdate$									0.86***
									(30.26)
Adj. R^2	27%	29%	28%	28%	29%	28%	28%	35%	35%
ClusteredStd Errors	Stock&Day	Stock&Day	Stock&Day	Stock&Day	Stock&Day	Stock&Day	Stock&Day	Stock&Day	Stock&Day

Table 6: OTTRs and information environment

This table reports results from difference-in-differences models with stock-day observations. The dependent variable is $LogOTTR = Ln(1 + OTTR)$. The key independent variable is the interaction term $D_{Post} * Share_{NYSE}$. D_{Post} is a dummy variable that takes the value of one after the NYSE introduced pre-trade transparency for its order book (January 24, 2002). $Share_{NYSE}$ is the average fraction of the stock's traded volume executed on NYSE in the period January 1, 2001 - January 24, 2002. The control variables and corresponding regression numbers are the same as in Table 5 (excluding regression (9), which requires a longer time period). The control variable for fragmentation is $Frag1$ (number of markets) in all regressions to avoid multicollinearity between $Frag2$, $Frag3$, and $Share_{NYSE}$. Standard errors are clustered by stock and day. T-statistics are reported in parentheses. ***, **, and * indicate statistical significance at 1%, 5%, and 10% levels, respectively. The sample includes 241 stocks and covers the period January 1, 2001 – December 31, 2002.

	Log OTTR (1)	Log OTTR (2)	Log OTTR (3)	Log OTTR (4)	Log OTTR (5)	Log OTTR (6)	Log OTTR (7)	Log OTTR (8)
$D_{Post} * Share_{NYSE}$	0.32**	0.32**	0.32**	0.41***	0.34**	0.41***	0.26**	0.41***
	(2.46)	(2.43)	(2.43)	(3.08)	(2.57)	(3.11)	(2.49)	(3.14)
$Share_{NYSE}$	-0.82***	-0.81***	-0.81***	-1.04***	-0.90***	-1.05***	-0.38***	-1.05***
	(-5.44)	(-5.38)	(-5.38)	(-6.49)	(-5.94)	(-6.59)	(-2.90)	(-6.64)
D_{Post}	-0.08	-0.11	-0.11	-0.15	-0.15	-0.17*	-0.08	-0.13
	(-0.71)	(-1.08)	(-1.08)	(-1.54)	(-1.47)	(-1.74)	(-0.89)	(-1.23)
Adj. R^2	21%	21%	21%	22%	23%	23%	28%	23%
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Clustered Std. Errors	Stock &Day	Stock &Day	Stock &Day	Stock &Day	Stock &Day	Stock &Day	Stock &Day	Stock &Day

Table 7: Determinants of OTTRs using exchange-day observations

This table reports OLS regression results for regression models with exchange-day observations. The dependent variable is $LogOTTR = Ln(1 + OTTR)$, where $OTTR$ is a simple average of the OTTRs of stocks and ETFs traded on a given exchange. Independent variables are in the first column. Variable definitions are in Table 2, noting that $LogVIX = Ln(VIX)$. Standard errors are clustered by exchange and day. T-statistics are reported in parentheses. ***, **, and * indicate statistical significance at 1%, 5%, and 10% levels, respectively. The sample period covers years 1998 to 2018, and includes 241 stocks and 68 ETFs traded across 14 exchanges.

	Log OTTR (1)	Log OTTR (2)	Log OTTR (3)	Log OTTR (4)	Log OTTR (5)	Log OTTR (6)	Log OTTR (7)	Log OTTR (8)
$NumMkts$	0.15***							
	(3.98)							
$TakerDummy$	0.39***	0.54***	0.32**	0.32**	0.32**	0.29**	0.34**	0.32**
	(2.78)	(3.23)	(2.31)	(2.34)	(2.32)	(2.03)	(2.40)	(2.37)
$MktShareStocks$		-2.00***	-1.75***	-1.75***	-1.75***	-1.79***	-1.33	-1.76***
		(-2.86)	(-3.43)	(-3.40)	(-3.41)	(-3.64)	(-1.34)	(-3.58)
$MktShareETFs$	-0.36	-0.92	-0.61	-0.61	-0.61	-0.55	1.58*	
	(-0.51)	(-1.00)	(-0.86)	(-0.84)	(-0.86)	(-1.02)	(1.89)	
$MktShareStocks^2$							-1.08	
							(-0.93)	
$MktShareETFs^2$							-3.47***	
							(-3.61)	
$HighLowVolatMkt$	0.80	-7.32**	2.42					
	(0.24)	(-2.06)	(0.62)					
$LogVIX$				0.18		0.28*		0.29*
				(0.87)		(1.75)		(1.91)
$Autoquote$						0.50**		
						(2.08)		
$CPUSpeed$			0.31***	0.33***	0.31***	0.27***	0.28***	0.28***
			(4.70)	(4.24)	(4.76)	(2.91)	(3.99)	(3.04)
HFTdate								0.51*
								(1.74)
Adj. R^2	28%	20%	34%	34%	34%	37%	38%	36%
Clustered Std Errors	Exch &Day	Exch &Day	Exch &Day	Exch &Day	Exch &Day	Exch &Day	Exch &Day	Exch &Day

Table 8: Determinants of OTTRs in the exchange-day sample excluding ETFs

This table reports OLS regression results for seven regression models with exchange-day observations. The dependent variable is $LogOTTR = Ln(1 + OTTR)$, where $OTTR$ is a simple average of the OTTRs of stocks traded on a given exchange. Independent variables are in the first column. Variable definitions are in Table 2, noting that $LogVIX = Ln(VIX)$. Standard errors are clustered by exchange and day. T-statistics are reported in parentheses. ***, **, and * indicate statistical significance at 1%, 5%, and 10% levels, respectively. The sample period covers years 1998 to 2018, and includes 241 stocks traded across 14 exchanges.

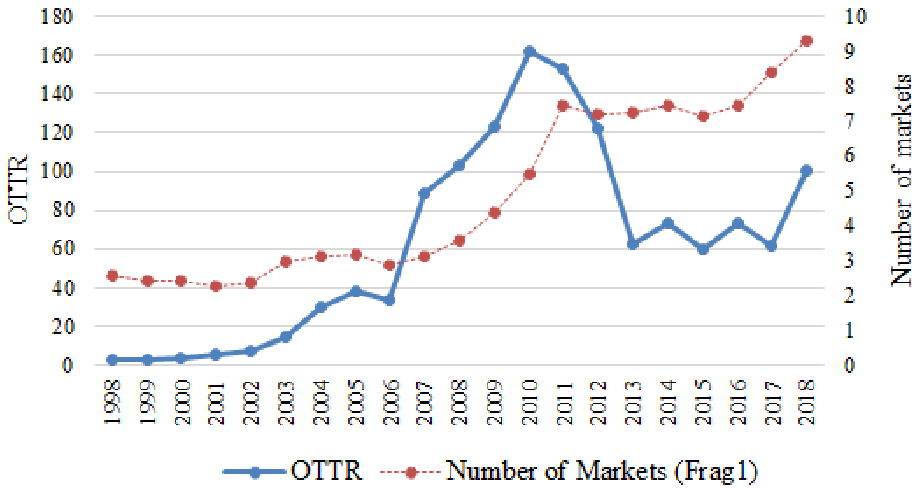
	Log OTTR (1)	Log OTTR (2)	Log OTTR (3)	Log OTTR (4)	Log OTTR (5)	Log OTTR (6)	Log OTTR (7)	Log OTTR (8)
$NumMkts$	0.16*** (4.11)							
$TakerDummy$	0.40*** (3.10)	0.60*** (3.68)	0.35*** (2.67)	0.35*** (2.69)	0.35*** (2.68)	0.32** (2.39)	0.38*** (2.71)	0.36* (1.92)
$MktShareStocks$		-1.97*** (-3.22)	-1.72*** (-3.51)	-1.72*** (-3.49)	-1.72*** (-3.51)	-1.77*** (-3.60)	-0.69 (-0.55)	-2.19*** (-3.70)
$MktShareStocks^2$							-1.90 (-1.19)	
$HighLowVolatMkt$	0.86 (0.27)	-8.00* (-1.77)	2.39 (0.61)					
$LogVIX$				0.18 (0.90)		0.29* (1.94)		0.05 (0.54)
$Autoquote$						0.52* (1.80)		
$CPUSpeed$			0.33*** (4.34)	0.35*** (4.25)	0.32*** (4.25)	0.28*** (3.15)	0.31*** (4.17)	
$HFTdate$								0.66*** (3.22)
Adj. R^2	28%	17%	33%	33%	33%	36%	33%	30%
Clustered Std Errors	Exch &Day	Exch &Day	Exch &Day	Exch &Day	Exch &Day	Exch &Day	Exch &Day	Exch &Day

Table 9: Empirical OTTRs vs a theoretical benchmark

This table reports descriptive statistics for empirically observed OTTRs of 715 stock-days (every 5th stock from the sample during November 2018) and theoretical OTTRs for these stock-days. The empirically observed OTTRs are calculated using daily data on the number of trades and quote messages. Theoretical OTTRs are calculated as described in the main text.

	Mean	StdDev	25th pctl	50th pctl	75th pctl
Empirical OTTR	18.37	18.00	9.30	14.03	21.08
Theoretical OTTR	33.77	33.16	20.85	27.54	35.84
Difference (Empirical - Theoretical)	-15.40	33.38	-20.45	-12.61	-4.64

Panel A. OTTR and fragmentation



Panel B. OTTR, correlation with S&P 500, and market volatility

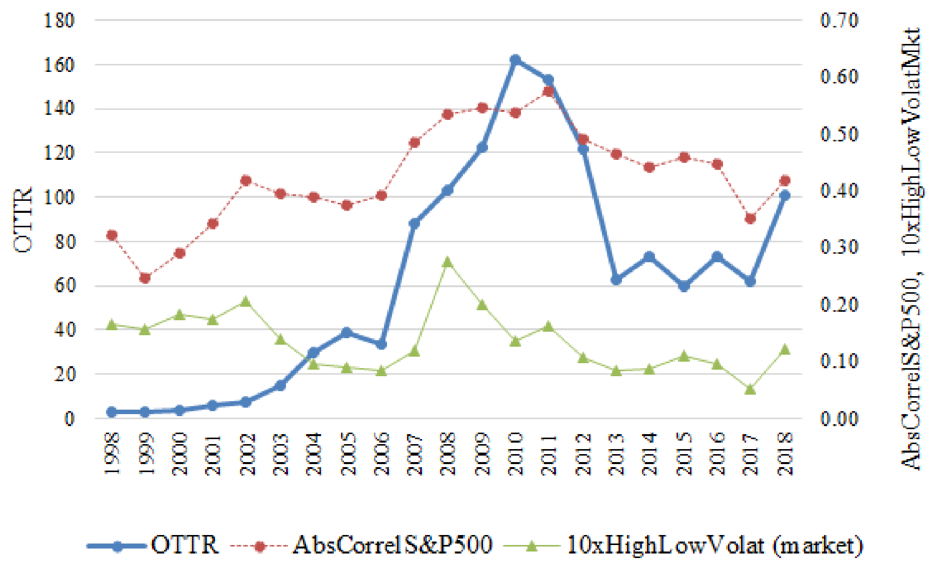


Figure 1: Time series of OTTRs and explanatory variables

This figure plots the annual averages of OTTRs and explanatory variables. $Frag1$ is the number of stock exchanges trading sample securities. $AbsCorrelS\&P500$ is the absolute correlation of the security's returns with the S&P 500 index. 10x $HighLowVolat$ (market) is daily S&P 500 volatility multiplied by 10. The sample period covers years 1998 to 2018, and includes 241 stocks and 68 ETFs.

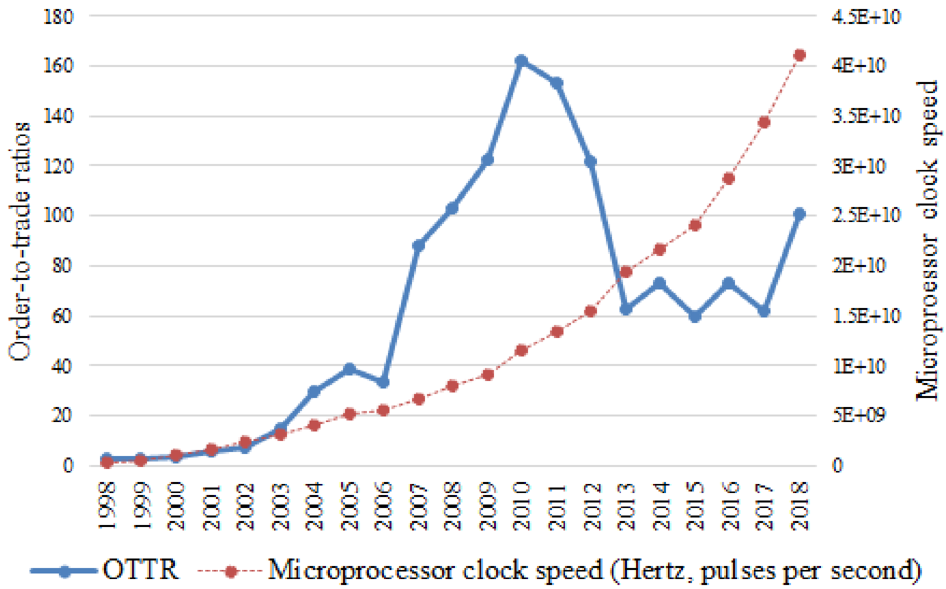
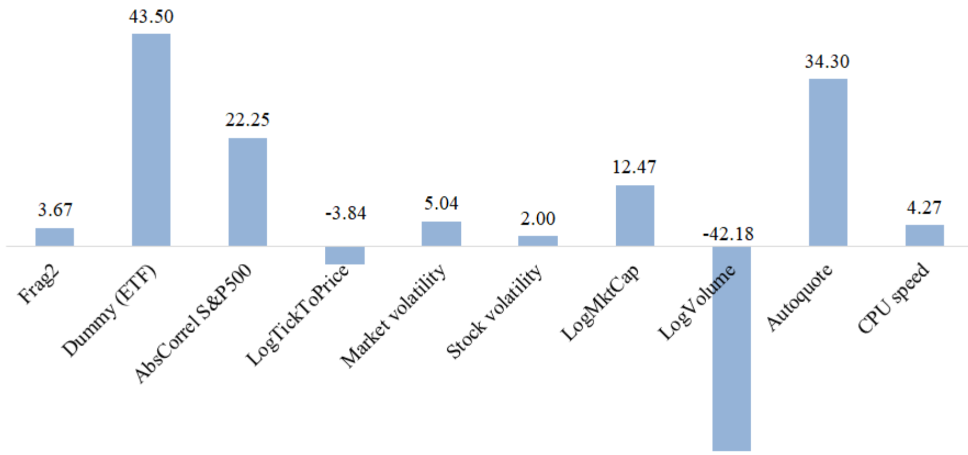


Figure 2: Order-to-trade ratios and data processing speed

This figure plots the time series of OTTRs against the microprocessor clock speed measured in hertz (a proxy for technology advancement as in Kurzweil (2005)). The variables are computed as annual averages. The sample period covers years 1998 to 2018, and includes 241 stocks and 68 ETFs.

Panel A. Security-day observations



Panel B. Exchange-day observations

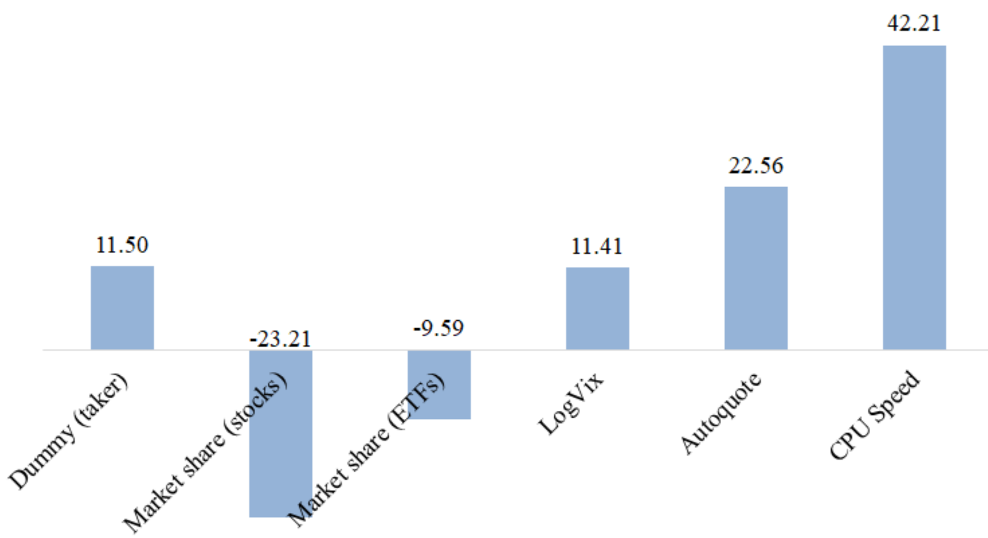
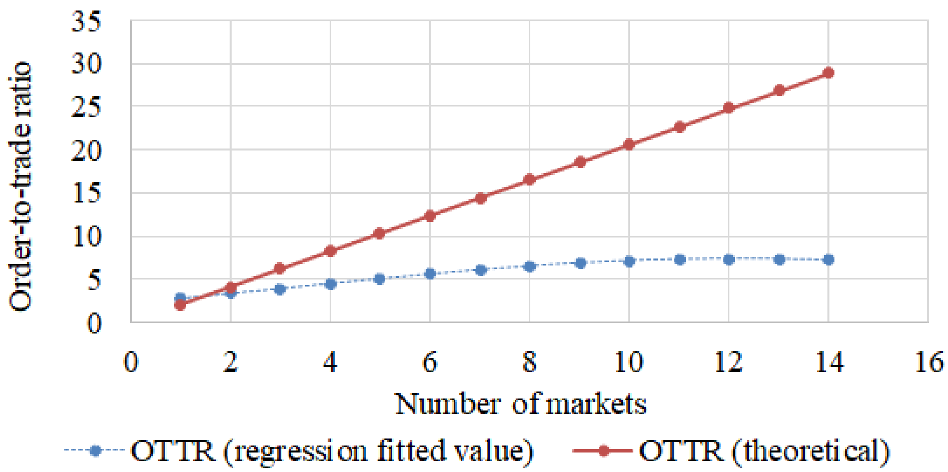


Figure 3: Standardized regression coefficients for the drivers of OTTRs

This figure plots the percentage change in $(1 + OTTR)$ associated with a one standard deviation change in the explanatory variables. The explanatory variables are on the horizontal axis. The percentage change in $(1 + OTTR)$ is on the vertical axis. For example, one standard deviation increase in the fragmentation measure ($Frag2$) is associated with 3.67% increase in $(1 + OTTR)$, holding other factors constant. The sample period covers years 1998 to 2018, and includes 241 stocks and 68 ETFs.

Panel A. Security-day observations



Panel B. Exchange-day observations

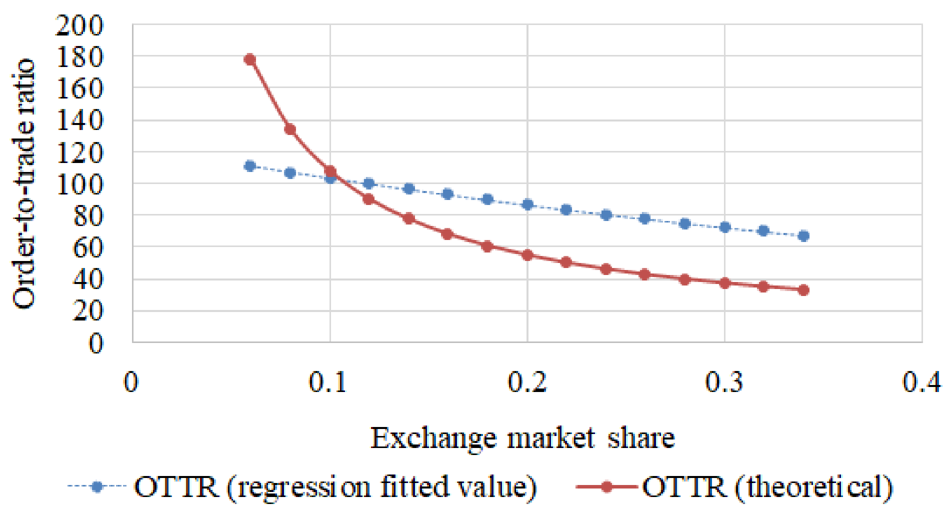
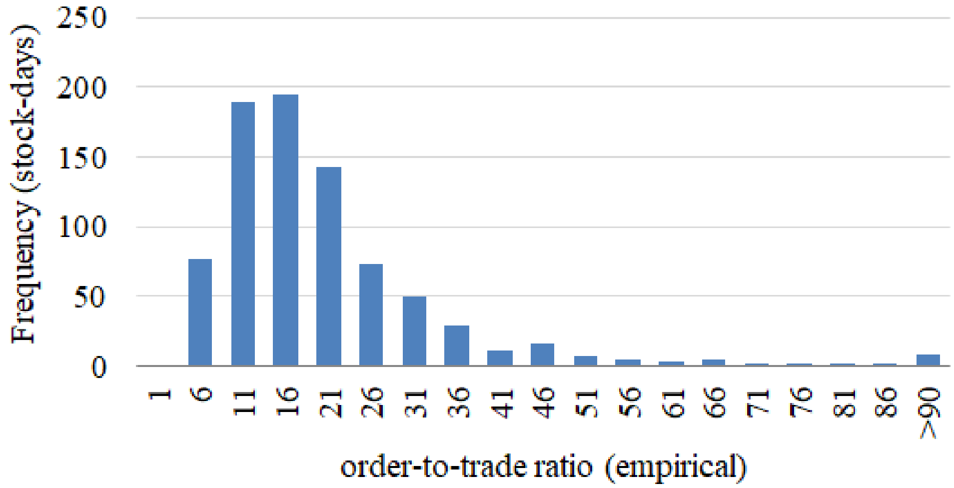


Figure 4: The relation between OTTRs, fragmentation, and market shares

Panel A plots security-day OTTRs for different levels of fragmentation. Panel B plots exchange-day OTTRs for different market shares. The dashed line is the empirically observed relation between OTTRs and fragmentation / market share. The solid line is the theoretical relation.

Panel A. Empirical OTTRs



Panel B. Empirical minus theoretical OTTRs

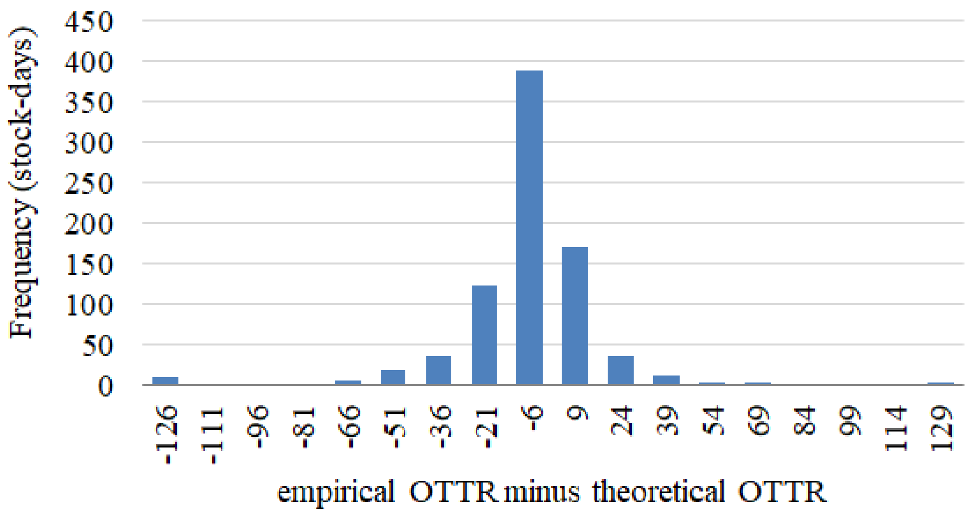


Figure 5: Distribution of theoretical vs empirical OTTRs

Panel A plots the distribution of empirically observed OTTRs on 715 stock-days (every 5th stock from the sample during November 2018) and theoretical OTTRs for these stock-days. Panel B plots the distribution of differences between the empirically observed OTTRs and the OTTRs predicted by the theory model for these same stocks. The vertical axis plots the frequency of observations in stock-days. The horizontal axis plots the values of OTTRs.