

Algos gone wild: Are order-to-trade ratios excessive? [☆]

Marta Khomyn^{a*} and Tālis J. Putniņš^{a,b}

^a *University of Technology Sydney, PO Box 123 Broadway, NSW 2007, Australia*

^b *Stockholm School of Economics in Riga, Strelnieku Street 4a, Riga, LV 1010, Latvia*

December 5, 2017

Abstract

We investigate the drivers and implications of the rapid growth in order-to-trade ratios (OTTRs). We develop and test a simple model of liquidity provision in which the OTTR is determined by a tradeoff between information monitoring costs and picking off risk (trading at stale prices). Our model explains the cross-sectional heterogeneity in OTTRs, with higher ratios in stocks that have higher volatility, more fragmented trading, higher price-to-tick, and lower volume. We find that recent growth in OTTRs is driven largely by fragmentation of trading across multiple venues and decreasing monitoring costs. Calibration reveals that OTTRs on a typical day are within levels that are consistent with market making activity, but occasionally spike beyond such levels. Our findings imply that regulatory measures designed to curb OTTRs (e.g., messaging taxes) are likely to harm liquidity provision, in particular in certain stocks, and create unlevel competition between trading venues.

JEL classification: G14

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

[☆] Marta Khomyn acknowledges funding from the Capital Markets CRC Limited and is a visiting researcher at Chi-X Australia during the time spent on this study.

* Corresponding author. Tel.: +61 481314349.

Email addresses: marta.khomyn@student.uts.edu.au (M. Khomyn) and talis.putnins@uts.edu.au (T. Putniņš).

1. Introduction

The rapid recent growth in order-to-trade ratios (OTTR) and order cancellation rates in financial markets has alarmed regulators and market participants in many countries. For example, in US equities, the average OTTR (number of order enter/amend/cancel messages divided by the number of trades) has 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.¹ A response to these concerns is message taxes, which have been proposed in some countries (such as the US) and already implemented in others (e.g., Australia, Italy, Germany). This paper aims to increase our understanding of the drivers of OTTRs, whether their growth warrants concern, and the impacts of regulatory proposals such as message taxes.

High OTTRs have been in the public spotlight, with concerns that they are a symptom of predatory or manipulative behavior of high-frequency traders (SEC, 2010; Biais and Woolley, 2011). While market manipulation strategies such as spoofing or quote stuffing can generate spikes in quoting activity (Egginton, Van Ness, and Van Ness, 2016), high OTTRs can also arise in various circumstances from trading strategies that are neither illegal nor harmful. In fact, as we will show in this paper, market making can result in high OTTRs, in particular when it requires posting quotes across multiple trading venues and adjusting the quotes rapidly in response to new information to minimize picking off risk. The combination of advances in technology, which have lowered monitoring costs and allowed much more information to be processed by liquidity providers, and fragmentation of trading across multiple venues necessitates increasing amounts of quote revisions by liquidity providers to remain competitive. It is thus not surprising that the majority of liquidity provision is currently undertaken by HFT firms (Carrion, 2013; Jarnecic and Snape, 2014; Hagströmer, Nordén, and Zhang, 2014).

As a result of the alleged link between high OTTR and illicit HFT behavior, a number of regulators have imposed message taxes, effectively charging high-OTTR traders a fee for excessive message traffic (Friedrich and Payne, 2015). To the extent that such regulation curbs harmful HFT behavior, the tax could improve liquidity and other measures of market quality.² However, if the

¹ US SEC Market Structure Research highlights: <https://www.sec.gov/marketstructure/research/highlight-2013-01>

² We are not aware of any existing studies finding a positive effect of messaging taxes on liquidity or market quality. However, some studies have found a neutral effect: for example, Capelle-Blancard (2017) in Italian market, Colliard and Hoffmann (2017) in French market, Jørgensen, Skjeltorp, and Ødegaard (2014) in Norwegian market.

regulation negatively affects liquidity providers, market liquidity could deteriorate.³ The model that we develop in this paper helps resolve this debate by characterizing the relation between liquidity provision and OTTRs. Moreover, the model can explain why OTTRs are naturally expected to be higher in certain stocks and time periods compared to others, suggesting a useful regulatory tool to detect *abnormal* messaging traffic.

We develop and test a simple theoretical model of a liquidity provider that posts and updates quotes in a fragmented market. The liquidity provider monitors several sources of information (“signals”) and updates quotes to avoid being picked off (trading at stale prices). Monitoring intensity by the liquidity provider in our model is endogenous—the liquidity provider decides how many and which signals to monitor by weighing up the benefit (reduced “picking off risk” or likelihood of being hit by market orders while having stale quotes) and the cost (the computing, telecommunications, and data feed costs). Consequently, the OTTR emerges endogenously in our model as a function of monitoring costs, market conditions (e.g., volume, volatility), and stock characteristics (e.g., how closely correlated the stock is with other securities). Our approach is related to other models of the behavior of modern liquidity providers (Foucault, Röel, and Sandås, 2003; Liu, 2009; Foucault, Kadan, and Kandel, 2013; Lyle and Naughton, 2015), but unlike previous literature, we seek to answer the question of what drives OTTRs and whether they are excessive.

By extending the model to include multiple trading venues, we characterize the impact of fragmentation on OTTRs. As trading fragments across multiple venues, liquidity providers have to post and adjust quotes across all of them, causing OTTRs to scale up almost linearly with the number of trading venues. The model with multiple venues also predicts higher OTTRs for markets with lower shares of trading volume. Intuitively, the quotes on a market with a low share of trading volume must be updated to track the quotes of other markets to avoid arbitrage opportunities (leading to a similar number of quote messages across markets) but because there are fewer trades on markets with low volume shares, the denominator of the OTTR is smaller, resulting in larger OTTRs. Incorporating fragmentation as one of the drivers of OTTR is novel, as previous studies have mostly overlooked the effects of fragmented markets on OTTRs, and only considered factors related to stock’s risk-bearing capacity and dealer’s inventory (Rosu, Sojli and Tham, 2017), price-time priority (Ye and Yao, 2015; Ye, 2017), and limit order profitability (Dahlström, Hagströmer, and Nordén, 2017).

³ A number of empirical studies have found messaging taxes to be detrimental to liquidity and market quality: for example, Caivano et al. (2012) and Friedrich and Payne (2015) in Italian market, Haferkorn (2015) in German market, Malinova and Riordan (2016) and Lepone and Sacco (2013) in Canadian market.

Our model links OTTRs to the key features of the modern financial markets — fragmentation, technology, and regulation (as mentioned in O’Hara, 2015). We help explain why OTTRs have increased over time (from around two in 2000s to around ten in 2016), because our modelling approach recognizes the concurrent occurrence of HFT and fragmentation (Menkveld, 2016) as a result of regulatory changes embedded in Rule 611 (trade-through rule) of Regulation National Market System (Reg NMS) (Mahoney and Rauterberg, 2017). In line with the model, empirical data suggest that long-term trends in OTTRs are related to increasing market fragmentation and decreasing monitoring costs, while short-term dynamics can be explained by market volatility (OTTRs spike around the same time as VIX index).

We find empirical evidence for the model predictions in the cross-section of US stocks over 2012–2016 sample period. The model explains why there is a considerable cross-sectional variation in OTTRs, with higher ratios in more volatile markets, higher price-to-tick stocks, lower volume stocks, and in ETFs compared to stocks. Also, the empirical data corroborate the positive linear relation between OTTRs and fragmentation, as well as the inverse relation between OTTR and market share. The intuition behind these effects is as follows. In more volatile markets, monitoring intensity of liquidity providers increases, as they try to avoid being picked off by informed traders, and this in turn increases OTTR. Similarly, liquidity providers update quotes more often in high price-to-tick stocks, because for stocks with less constrained spreads even less important signals might have value implications (hence picking-off risk is higher). ETFs naturally have higher OTTR (compared to stocks) due to the greater number of highly relevant signals available for monitoring.

Applying the model to the most recent period (2016) of our sample reveals that in most cases, empirically observed OTTRs are in line with or below those that would be expected from liquidity provision in a fragmented market, even under conservative assumptions of one liquidity provider and one signal monitored. The distribution of empirical OTTRs is right-skewed, with 7% of observations above the theoretical level. This suggests a useful tool for regulators to detect abnormal quoting activity in certain securities and penalize illicit behavior in cases which actually require intervention.

Our findings suggest that the recent levels of OTTRs do not necessarily warrant concern, as legitimate market making would result in OTTRs that are similar or above those observed in the market data. Therefore, regulatory measures aimed at curbing quoting activity (e.g., message taxes) can have adverse effects on market making in securities that already have disadvantageous conditions for liquidity providers. Furthermore, message taxes create unlevel competition between trading venues due to higher OTTRs on venues with lower volume shares. Finally, securities with

natural signals (e.g., ETFs) always have higher OTTRs compared to common stocks, so taxing liquidity providers in those securities would have detrimental effects on liquidity provision.

2. A simple model of what drives the OTTR

2.1. Baseline model structure

Consider a simple model in which a liquidity provider posts quotes (bid and ask prices and quantities) for a given asset in a given market. The liquidity provider could monitor one or more signals from a set of signals, $\{s_1, s_2, \dots, s_N\}$. Each signal is a time-series (e.g., a price in a related security, price of the same security in another market, an order book state, and so on) that changes at stochastic times (termed “information arrivals”) given by Poisson processes with intensity λ_i for the i^{th} signal. The quality of signal i , q_i , is the probability that when there is a change in that signal (an “information arrival”), the liquidity provider will want to update his posted quoted price(s) or quantities (we term such events “relevant information arrivals”), resulting in a “cancel and enter” or “amend” message from the liquidity provider.⁴

There is a cost to monitoring a signal, with the cost per unit time being proportional to the intensity of information arrivals (changes in the signal), $\lambda_i c$. 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).

Market orders arrive at stochastic times given by a Poisson process with arrival rate λ_m and trade against the liquidity provider’s posted quotes. The liquidity provider’s benefit from monitoring comes from avoiding having stale quotes picked off. When a market order arrives after a relevant information arrival but the liquidity provider 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 liquidity provider) then the liquidity provider’s (stale) quotes are picked off and he incurs a picking-off cost, k . The more signals the liquidity provider monitors, the lower the probability (frequency) of his quotes being picked off, because the more of the relevant information he has through his monitoring. For a given monitoring intensity, the picking-off cost per unit time increases with the asset’s fundamental volatility (frequency of useful information arrivals) because of more frequent relevant information that makes quotes stale unless monitored.

⁴ To be more precise, two messages, if the liquidity provider adjusts both the bid and the ask.

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

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

As a result of monitoring signals, executing trades, and updating quotes, the liquidity provider generates messaging activity (messaging includes order entry, cancellation, and amendment messages) at an expected rate of Q messages per unit of time:

$$Q = 2 \sum_{i \in \{MonitoredSignals\}} \lambda_i q_i + 2\lambda_m \quad (1)$$

The first term, $2 \sum_{i \in \{MonitoredSignals\}} \lambda_i q_i$ is due to quote updates in response to relevant information arrivals on monitored signals, and the second term, $2\lambda_m$, is due to reposting liquidity after being hit by a market order (reentering one quote and amending the other).⁵

Recognizing that the expected number of trades per unit time is just the market order arrival intensity, λ_m , the *OTTR* for the asset is given by Eq. (2):⁶

$$OTTR = \frac{2 \sum_{i \in \{MonitoredSignals\}} \lambda_i q_i + 2\lambda_m}{\lambda_m} \quad (2)$$

2.2. Equilibrium

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

⁵ Both terms ($\sum_{i \in \{MonitoredSignals\}} \lambda_i q_i$ and λ_m) are multiplied by two reflecting the fact that after observing useful information or being hit by a market order, the liquidity provider updates his view of the fundamental value and thus adjusts both bid and ask prices or bid and ask quantities.

⁶ We define the *OTTR* as the total number of messages (order entry, cancellation, and amendment) divided by the total number of trades. In some industry settings, this ratio is referred to as the message-to-trade ratio.

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

$$\lambda_m \left(\frac{\lambda_i q_i}{\lambda_m + \lambda_i q_i} \right) k - \lambda_i c > 0, \quad (3)$$

with the set of monitored signals denoted $\{MonitoredSignals\}$. This condition determines *monitoring intensity* (the number of monitored signals).

2.3. Model with fragmented markets

If the number of markets increases from 1 to N , the single (representative) liquidity provider posts liquidity across multiple 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. The overall quoting activity of the liquidity provider consists of two components: (a) quote updates resulting from relevant information received by monitoring signals, $2N \sum_{i \in \{MonitoredSignals\}} \lambda_i q_i$ (liquidity provider updates quotes on all N markets in response to monitored signals), and (b) reposting liquidity / revising quotes on all markets after getting a fill on market orders, $2N \lambda_m$. Note that market fragmentation does not affect the signal monitoring decision of the liquidity provider, who chooses the set of signals to monitor in the same manner as in a single-market case.⁷ The resulting OTTR for the market overall (aggregating across venues) is therefore:

$$OTTR = \frac{2N(\sum_{i \in \{MonitoredSignals\}} \lambda_i q_i + \lambda_m)}{\lambda_m} \quad (4)$$

Consider the OTTR of individual markets $k = 1 \dots N$. The market share of trading volume (market orders) for each individual market k is ρ_k . The liquidity provider updates his quotes on market k every time relevant information is received from the monitored signals and after being hit by market order. Then, the OTTR for market k is

$$OTTR_k = \frac{2 \sum_{i \in \{MonitoredSignals\}} \lambda_i q_i + 2 \rho_k \lambda_m}{\lambda_m \cdot \rho_k} \quad (5)$$

⁷ We assume market order arrivals constitute useful signals, from the liquidity provider’s viewpoint.

2.4. Propositions

We now derive theoretical propositions about the relations between OTTRs, monitoring intensity and fragmentation. First, we establish the link between OTTRs and fragmentation (Proposition 1). Second, we show how OTTRs are related to market shares (Proposition 2). Third, we relate OTTRs to all the model parameters that affect the OTTR. In the next section, we build on these propositions to develop the testable hypotheses.

Proposition 1. As trading fragments across multiple venues, the market-wide (aggregate) OTTR for a given security increases with the extent of fragmentation, if there is at least one non-zero quality signal in the monitored set.

Proof. See Appendix 1.

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

Proposition 2. As trading fragments across multiple venues, the OTTR of a given security on a given market increases as the market share of volume for that market decreases.

Proof. See Appendix 1.

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

Proposition 2a. The OTTR for a given security increases with monitoring intensity.

Proof. See Appendix 1.

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

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

Proof. See Appendix 1.

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

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

Proposition 4. The OTTR for a given security increases with picking-off cost.

Proof. See Appendix 1.

When faced with a higher cost of being picked off, the liquidity provider has an incentive 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.

Proposition 5. The OTTR for a given security decreases with monitoring cost.

Proof. See Appendix 1.

When the liquidity provider faces higher cost per signal monitored, his marginal costs increase, hence leading him to decrease the monitoring intensity and the OTTR. The liquidity

provider's marginal costs are proportional to signal intensity, so the effect on monitoring intensity and OTTR is higher for more higher intensity signals.

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

Proof. See Appendix 1.

The effect of trading frequency on OTTR is two-fold. On one hand, higher intensity of market order arrivals increases monitoring intensity, as the liquidity provider has an incentive to monitor more to avoid picking-off costs. Therefore, he posts more quote updates based on signals monitored, which drives up the OTTR. On the other hand, higher market orders intensity decreases OTTR every trade is associated with fewer quote updates on average. Hence, if we keep the number of signals in the monitoring set constant (aka constant monitoring intensity), only the second effect takes place: OTTR decreases with trading frequency.

3. Empirical analysis

We use regression analysis to test the model's theoretical predictions. This section discusses the data and regression results as they relate to model propositions and empirical hypotheses. Table 1 summarizes the mapping between model propositions, empirical hypotheses and variables used in the regression analysis.

< Table 1 here >

3.1. Data and descriptive statistics

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

Our sample period spans from January 1, 2012 (the starting date of MIDAS dataset) to December 31, 2016 (the latest date for CRSP dataset). The combined dataset contains daily frequency observations, with stock- and exchange-level granularity. We aggregate the data to

stock-day level for the first part of our analysis (exploring how OTTR varies over time in the cross-section of securities), and to exchange-day level for the second part of analysis (exploring how OTTR varies over time across markets). The main advantages of MIDAS data are that (i) they cover the whole universe of traded US securities, (ii) they offer both exchange-date and stock-date granularity, and (iii) they provide the key variables at daily frequency (unlike TRTH, which requires intraday data processing,). Figure 1 plots the time series of OTTRs computed from SEC MIDAS and TRTH databases, and shows that the two time series co-move closely, although the magnitude of OTTRs captured by the two sources of data differs.⁸ We use SEC MIDAS data in all the following regression analysis.⁹

< Fig. 1 here >

The descriptive statistics for the stock-date panel is presented in Table 2. The dataset contains just under 5,922,424 daily observations for 7,114 securities, 75% of them stocks, and the rest — exchange-traded funds (ETFs). At stock-date frequency, we have a dummy variable for ETFs that lets us control for ETF-specific characteristics beyond those suggested to drive OTTRs based on the theory model. We also account for stock-days affected by the SEC Tick Size Pilot program, and apply the wider tick sizes accordingly.¹⁰

< Table 2 here >

⁸ Note that we compute OTTRs following the SEC methodology: dividing the order volume by lit volume. By SEC definition, order volume is sum of order volume (in number of shares) for all add order messages; lit volume is sum of trade volume for trades that are not against hidden orders. We compute TRTH OTTRs by dividing the number of order updates (price or quantity) at best quotes by the number of trades. These differences in computation arise due to the data series available in the two databases.

⁹ We use MIDAS data rather than TRTH for regressions for two reasons: (i) intraday processing (required in case of TRTH data) for the universe of all US securities over multi-year data samples is not computationally feasible; (ii) MIDAS captures quote revisions at multiple depth levels, while TRTH — only at best quotes.

¹⁰ The Tick Size Pilot program affects 1,400 small capitalization stocks by widening their tick sizes from \$0.01 to \$0.05. The rollout of the program started on October 3, 2016, and occurred in several phases for three groups of securities affected. We use the official data from The Financial Industry Regulatory Authority (FINRA) web-site to identify the affected securities and effective rollout dates.

The exchange-date panel contains 27,454 observations. In exchange-date analysis, we distinguish between the markets with different fee structures by introducing a dummy variable for taker-maker markets (Edge-A, Bats-Y and Boston stock exchange).¹¹

We also use two variables with only time variation (no cross-sectional variation), which are proxies for market volatility. The descriptive statistics for those is presented in Table 2. The first proxy for market volatility is computed from daily high-low range of SPY ETF daily prices, while the second proxy is a log-level measure of daily closing VIX index.

3.2. Regression results

Our regression specifications follow from the hypotheses outlined in Table 1. To account for within-cluster correlations (i.e., correlations within exchange-date groups and stock-date groups), we use double-clustered standard errors. Regression models are estimated for stock-date (Eq. 6) and exchange date (Eq. 7) regressions accordingly.

$$\begin{aligned} \ln(1 + OTTR_{it}) = & \alpha + \beta_1 Frag_{it} + \beta_2 \ln(Volume_{it}) + \beta_3 \ln(MarketCap_{it}) + \\ & \beta_4 MarketVolatility_t + \beta_5 StockVolatility_{it} + \beta_6 CorrelationS\&P_{it} + \\ & \beta_7 \ln(TickToPrice_{it}) + \beta_8 D_{it}^{ETF} + \varepsilon_{it} \end{aligned} \quad (6)$$

$$\begin{aligned} \ln(1 + OTTR_{jt}) = & \alpha + \beta_1 Frag_{jt} + \beta_2 \ln(Volume_{jt}) + \beta_3 MarketVolatility_t + \\ & \beta_4 CorrelationS\&P_{jt} + \beta_5 \ln(TickToPrice_{jt}) + \beta_6 D_{jt}^{taker} + \beta_7 MarketShare_{jt} + \varepsilon_{jt} \end{aligned} \quad (7)$$

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

Regression results generally corroborate the predictions of our theory model. We find evidence that OTTRs increase with fragmentation (in line with Hypothesis 1a), and are higher for stocks with lower volumes (in line with Hypothesis 6), larger market cap (in line with Hypothesis 5a), higher correlations with the market index (in line with Hypothesis 3c), and higher price-to-tick ratios (in line with Hypothesis 4c). OTTRs for ETFs are higher than those for stocks, controlling for other security characteristics (in line with Hypothesis 3a). Stock and market volatility are also

¹¹ “Maker-taker” market refers to the market that compensates “liquidity makers” (i.e., those posting limit orders) and charges “liquidity takers” (i.e., those posting market orders). “Taker-maker” market refers to the trading venue that does the opposite (i.e., charges for limit orders and compensates for market orders). In our sample, nine trading venues apply maker-taker fee structure: Amex, Arca, Bats-Z, CHX, Edge-X, NSX, NYSE, NASDAQ, PHLX; three trading venues — taker-maker fee structure: Edge-A, Bats-Y, and Boston stock exchange.

positively associated with OTTRs of a given stock on a given day (in line with Hypotheses 4a and 4b). Empirical results for stock-date and exchange-date regressions are reported in Tables 3 and 4 respectively.

< Table 3 here >

The empirical result that OTTRs increase with the degree of fragmentation confirms the prediction from our theory model (see Proposition 1). This result is expected, as higher fragmentation means posting liquidity across multiple venues. This in turn leads to order revisions increasing proportionally to the number of venues, because liquidity providers revise quotes across multiple exchanges in response to monitored signals, and after getting a fill on a market order. As an illustration, consider two securities: Delta Apparel Inc. (DLA), a clothing manufacturer, and Wage Works Inc. (WAGE), a service sector firm administering consumer-directed benefit plans. DLA trades on five markets, and WAGE — on 10, with other empirical characteristics (market capitalization, tick-to-price, volume, correlation with the market) reasonably similar between the two stocks. Regression estimates imply OTTR of 23.12 for DLA, and 50.92 — for WAGE, suggesting that OTTR scales up almost linearly with fragmentation, as predicted by our model. As shown in Figure 2 (Panel B), 54% of the difference between these securities' OTTRs arises from the difference in fragmentation.

< Fig. 2 here >

The positive relation between fragmentation and OTTR indeed holds on average in the stock-day panel, as suggested by regression results in Table 3: the coefficient on fragmentation is positive and significant for all three fragmentation proxies (number of markets for a given stock on a given day, Herfindahl-Hirschman index based on share volume, and Herfindahl-Hirschman index based on number of trades). The effect is also economically significant: one standard deviation increase in the number of venues that trade a stock on a given day corresponds to 27% increase in OTTR (see Figure 5).¹² Our fragmentation proxies follow those used in Degryse, De Jong, and Van Kervel (2014) and Malceniene, Malceniaks, and Putnins (2016).

¹² To be precise, one standard deviation increase in the number of venues that trade a stock on a given day corresponds to 27% increase in $(1+OTTR)$. However, the difference is negligible in most cases, so we report the effect on OTTR for the simplicity of interpretation.

To better understand the shape of OTTR-fragmentation relation, we also regress OTTRs on dummy variables of different degrees of fragmentation, controlling for other stock characteristics. Figure 3 shows that OTTRs increase almost linearly as the number of markets increases, in line with the model predictions. This is a novel finding, as no studies to date have investigated the relation between fragmentation and OTTR.

< Fig. 3 here >

The quality of signals available for monitoring also affects OTTRs. While multiple studies have explored the link between HFT quoting and monitoring activity (e.g., Liu, 2009; Conrad, Wahal, and Xiang, 2015; Lyle and Naughton, 2015; Blocher, Cooper, Seddon, and Vliet, 2016), the reasons for monitoring in our model are related to market making and avoiding picking-off risk rather than speed competition among HFTs. Empirically, we find evidence for this effect by examining OTTRs in ETFs: the latter have high quality signals available for monitoring, unlike stocks. This leads to more intense monitoring activity by liquidity providers, keeping all other security characteristics constant. As an illustration, consider two securities — Uranium Resources Inc. (URRE), a uranium mining company, and Consumer Discretionary Select Sector SPDR Fund (XLY ETF). For XLY, the signal quality is 0.12, suggesting that market makers in XLY update their quotes 12 times for every 100 quote updates in SPY ETF. At the same time, monitoring the market is not as useful: market makers in URRA only update their quotes 0.043 times for every 100 quote updates in SPY. This wide difference in signal quality is reflected in regression-implied OTTRs: 7.18 for URRE, and 385.11 for XLY ETF. The ETF dummy variable accounts for 45% of the difference between OTTRs of these two securities (See Panel A in Figure 2).

We find evidence of the positive relation between monitoring and OTTRs, and the effect of monitoring is economically meaningful (one standard deviation increase in ETF dummy corresponds to 230% increase in OTTR, as shown in Figure 5). Another measure of monitoring — absolute value of correlation with S&P 500 index — is also positively associated with OTTRs, and highly significant. One standard deviation increase in correlation with S&P500 corresponds to 39% increase in OTTR (see Figure 5). Indeed, to the extent that liquidity providers can derive highly useful information from the available benchmarks, they will have an incentive to update the quotes more frequently to avoid the picking-off risk.

< Fig. 5 here >

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

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

Empirically, we find evidence supporting the prediction of higher picking-off risk (higher price-to-tick ratio) being associated with higher OTTRs (one standard deviation increase in tick-to-price leads to OTTRs being on average 26% lower, as shown on Figure 4). This is in line with evidence in Ye and Yao (2015), although the theoretical argument proposed by Ye (2017) points towards the speed vs price competition by HFTs as a theoretical mechanism for this effect. Our model suggests a different mechanism — picking-off risk — although the two need not be mutually exclusive. In fact, our model might help explain why, as suggested by Ye (2017) HFTs compete more on price rather than time priority in high price-to-tick stocks: it is because their speed advantage allows them to more effectively avoid being picked off by reacting rapidly to information arrivals through adjusting their quotes. This, in turn, leads to higher OTTRs.

Monitoring cost is one of the key drivers of the endogenous monitoring intensity in our model. Hence, to the extent that lower monitoring cost increases the net marginal benefit of monitoring, the liquidity provider will monitor more and hence increase his OTTR. Since liquidity provider's costs are not directly observable, we use two proxies that previous studies have shown

to be highly correlated with the HFT activity: stock's market cap and trading venue's maker-taker fee structure. Because HFT's investment in technology enables them to achieve low marginal costs of monitoring, relative to other market participants, the prevalence of HFTs should come together with low monitoring costs.

As shown in O'Hara (2015), and Rosu et al. (2017), large-cap stocks attract more HFT activity, which in turn suggests lower cost of monitoring. Empirically, we find log market cap is strongly positively related to OTTR (see Table 3). One standard deviation increase in market cap leads to OTTRs being on average 34% higher (see Figure 4).

OTTRs are also negatively related to the trading frequency, which we proxy by the number of shares traded in a day. The corresponding theoretical parameter is market order arrival intensity (λ_m), which we compute as number of trades in a given security per second. As an illustration, consider two ETFs tracking S&P 500 Index: IVV and SPY. Otherwise similar, they are vastly different in trading frequencies: on a randomly selected day, we observe IVV traded on average 1.18 times per second, while SPY — 20.09 times. It is worth noting that SPY is the most frequently traded security in the world (Balchunas, 2016). The regression-implied OTTR for IVV is 522.87, while for SPY — 139.11, reflecting the higher trading frequency of the latter. This difference in trading frequencies accounts for 79% of the difference in OTTRs between these two securities (see Panel C in Figure 2).

Controlling for trading volume is also important to view the results from the standpoint of quoting activity and avoid them being driven by the mechanical division by trading volume.

< Table 4 here >

Exchange-day analysis (results reported in Table 6) allows us to explore the effects of market characteristics (e.g., fees structures, market shares etc.) on OTTRs. We find that empirical results support the predictions of our theoretical model, as OTTRs are positively related to fragmentation, prevalence of ETFs on the trading venue and market volatility, and negatively related to the venue's market share. The effect of market share is non-linear (the coefficient on squared market share variable is positive and significant), as predicted by the model. In fact, the shape of the empirical relation closely resembles that suggested by the model (see Figure 4 for illustration).

< Fig. 4 here >

The degree of fragmentation at exchange-date level is best measured as the number of markets the average security trades on, and this measure is positively significantly related to OTTRs. In line with Hypothesis 1a, we observe higher OTTRs for markets that trade more fragmented securities.

We also find that trading venues with lower shares of dollar volume traded (smaller market shares) experience higher OTTRs. This effect is second largest in terms of economic significance: one standard deviation increase in market share leads to 64% decrease in OTTR (see Figure 4). This is in line with Hypothesis 2a, as our theory model suggests that liquidity providers scale up their quoting activity across venues, but lower trading volumes on a smaller venue leads to higher OTTRs. Using alternative proxies for market share (log dollar volume and log share volume) confirms this result. This finding is interesting in view of introducing competition in financial markets, as it suggests the reason why new trading venues naturally have higher OTTRs than incumbent stock exchanges. It also suggests that messaging taxes disproportionately burden new entrants as compared to incumbents, thus creating an unlevel playing field from the competition point of view.

The nature of securities traded on a particular market also contributes to the venue's OTTR profile. To the extent that a market trades more securities with high-quality signals, it is expected to have higher OTTRs. For example, ETFs, unlike stocks, have a natural basket of securities that can be monitored to derive inferences about ETF value. In regression analysis, we find empirical support for Hypothesis 3b, suggesting that the higher dollar volume share of ETFs on a particular market, the higher that market's OTTR.

Fee structure also affects OTTRs, primarily by attracting specific types of liquidity providers. As suggested by O'Hara (2015), maker-taker markets have higher prevalence of HFT liquidity providers. As the HFT speed advantage allows them to monitor signals cheaply, HFTs should have higher OTTRs, translating into higher OTTRs on maker-taker markets. In line with this theoretical prediction (Hypothesis 5b), we find that OTTRs are higher on maker-taker venues, controlling for other market characteristics. Empirically, one standard deviation increase in the taker-maker dummy corresponds to 27% lower OTTR for an average exchange-date in our sample.

Similar to stock-day regressions, we control for time-varying market volatility, which is positively related to OTTRs (corroborating Hypothesis 4a). This finding also has interesting regulatory implications, as market making on high-volatility days is important for market stability. If liquidity providers are charged disproportionately more in high-OTTR times, it could exacerbate the problem of fleeting liquidity that's common in modern market making (Menkveld, 2013).

Overall, we find that the market making motivated model of OTTRs has empirical support in the data. We present the summary of empirical hypotheses mapped against regression results in Table 5.

< Table 5 here >

4. Time-series trends in OTTRs

To understand why OTTRs have increased over time, we examine the relation between OTTRs and the key variables suggested by our theory model as drivers of quoting activity by liquidity providers. We structure this discussion along three main themes: the relation between OTTRs and (1) technology, (2) fragmentation and its enabler — regulatory changes, and (3) market conditions.

< Table 6 here >

The summary statistics for historical data are presented in Table 6. The sample period extends from January 1, 2000 to December 31, 2016, and covers 100 securities (stocks and ETFs) from SEC MIDAS database. We construct the sample by randomly selecting 10 stocks from each market capitalization decile. The daily data on OTTRs is from Thomson Reuters Tick History (TRTH) database, as is the VIX index and volume (in number of shares traded) used to compute the market fragmentation measure.¹³ TRTH provides intraday counts of trades and quotes, where quote counts capture order submissions and amendments at best bid and offer. Hence, TRTH-based OTTR measure is more conservative than the MIDAS-based measure, as the latter accounts for all order submissions, amendments and cancellations. The two measures show similar dynamics during the period covered by SEC MIDAS data (2012–2016).

Liquidity providers' monitoring costs are not readily observable; hence we use two variables — CPU costs and bandwidth costs — as proxies for liquidity providers' costs of monitoring.¹⁴ Our monitoring costs proxy accounts for both of these components by forming a first

¹³ The market fragmentation measure is computed using Herfindahl-Hirschman index: $Frag_{it} = (1 - \sum_{i=1}^N (\frac{Vol_{it}}{Vol_t})^2)$, where Vol_{it} is the share volume traded on market i on day t . It is based on share volumes of 10 randomly selected high market cap stocks which are traded throughout the sample period January 2000 to December 2016.

¹⁴ We do not incorporate data feed costs into our analysis due to lack of data. However, to the extent that data feed costs are not strongly correlated with other cost components (e.g., CPU costs), they are not likely to alter the cost trend significantly.

principal component of bandwidth costs and CPU costs. CPU costs capture the extent to which computing power has become cheaper over time, allowing modern liquidity providers to reduce the cost of processing market signals. Bandwidth prices reflect the dynamics of signal transmission costs, which are another part of monitoring activity. CPU costs are in \$/MIPS (million operations per second) from the CPU Price Performance dataset by John McCallum.¹⁵ Bandwidth prices are annual leasing prices of 10 Gbps broadband circuit links between Chicago and New York. The data on bandwidth prices are obtained from Telegeography database starting from the year 2002. Both types of costs are available at quarterly frequency.

Recall that one of the model predictions is the negative relation between OTTRs and monitoring costs faced by the liquidity provider (see Proposition 5). If liquidity provider's costs per signal monitored decrease, he has an incentive to increase his monitoring intensity, which in turn increases his quoting activity and OTTRs.

< Fig. 6 here >

As shown in Figure 6, technology costs have decreased substantially over time. The first drop in technology costs also coincides with the run-up in OTTRs, corroborating our theoretical predictions. One can argue that pre-2006 (before Reg NMS) growth in OTTRs was largely driven by liquidity providers' technology costs going down, as regulatory changes enabling market fragmentation were not yet introduced.

The introduction of decimalized quoting and autoquote on NYSE in April 2001 and May 2003 respectively are arguably part of the technologically-enabled run-up in OTTRs, too. After NYSE reduced the minimum tick size to one penny, depth at best quotes decreased substantially. In response, autoquote was proposed to allow trading in large size (typically 15 000 shares) at a firm quote. This innovation provided an incentive for liquidity providers to invest in technology that would offer the most up to date view of the market. For example, Hendershott, Jones, and Menkveld (2011) argue that the introduction of autoquote on NYSE was an early incentive for algorithmic traders, as automated quote updates created the speed advantage in monitoring the terms of trade.

The theory model suggests that OTTRs increase with fragmentation (see Proposition 1). As markets fragment, liquidity providers have to update quotes across multiple venues, as well as

¹⁵ Obtained from the internet appendix of Nordhaus (2007).

re-post liquidity after being hit by a market order. That leads to OTTRs scaling up with fragmentation.

< Fig. 7 here >

In the US, fragmentation was driven by regulatory changes, specifically the order protection rule (Rule 611 of Regulation National Market System — Reg NMS). Because the order protection rule (also known as the trade-through rule) effectively levelled the playing field for competition across trading venues, the fragmentation measure spiked up in the aftermath. As shown in Figure 6, the increase in fragmentation also coincides with the run-up in OTTRs, in line with our theoretical predictions.

The model also suggests that OTTRs are higher when liquidity providers face higher picking-off risk (see Proposition 4). This is the case in more volatile market conditions, as the frequency of information updates increases, leading to more intense monitoring and higher OTTRs.

< Fig. 8 here >

As shown in Figure 8, the spikes in VIX index indeed coincide with the spikes in OTTRs, suggesting that the short-term OTTR dynamics is to a large extent driven by market volatility. However, long-term trends in OTTR are not really related to VIX (consider the opposite direction of OTTR vs VIX movements in early 2000s). Spikes in OTTRs during the global financial crisis are also in line with our model predictions.

5. Analysis of recent OTTRs

Our theory model provides a simple way to examine whether current OTTRs are justified. We estimate the model parameters to reflect most recent market data and compute OTTRs that would arise from market making alone. Then, we compare theoretical OTTRs to those observed in the market, as well as investigate how the distributions of the two sets of OTTRs differ.

5.1. Data used for estimating theoretical OTTR

We compute theoretical OTTR estimates using the most recent year of our sample — 2016, and rely on SEC MIDAS database to select 20 random stocks per market cap decile. To make the sample representative in terms of varying market conditions, we sort the dates by VIX quintiles

and select five dates, each representing the median VIX per quintile. After eliminating stock-date observations with fewer than 100 trades, we end up with the sample size of 761 stock-days.

We compute both the model parameters and the empirical OTTRs using the data from Thomson Reuters Tick History database (TRTH). We rely on TRTH rather than MIDAS for two reasons. Firstly, TRTH provides the intraday data necessary to compute the theory model parameters (e.g., number of trades per second). Secondly, TRTH-based empirical OTTRs better reflect the logic of our model than MIDAS-based OTTRs. Consider that quoting activity in the theory model occurs for two reasons: order updates due to signal arrivals, and order re-posts after market order executions. This is in line with how TRTH data capture quoting activity: based on new orders, and order updates. MIDAS data, on the other hand, record order updates as cancel and enter, generating two messages instead of one, which would inflate empirical OTTRs relative to what the model logic suggests.

5.2. Method for estimating theoretical OTTR

Our theory model offers a straightforward way to estimate OTTRs for a liquidity provider who posts liquidity across multiple venues, and responds to signals monitored. The formula derived in Section 3 captures the computation logic: $OTTR_{it} = \frac{2N_{it}(\lambda_{SPYt}q_{it} + \lambda_{mit})}{\lambda_{mit}}$, where $OTTR_{it}$ is the OTTR in stock i on day t , N_{it} is the number of markets stock i trades on during day t , λ_{SPYt} is the signal intensity, q_{it} is a measure of signal quality, λ_{mit} is market order arrival intensity. The following simplifying assumptions are made:

- We assume the liquidity provider watches only one signal — midquote updates in S&P500 ETF. This is a simplifying assumption that allows us to estimate OTTRs without solving liquidity provider’s optimization problem that involves screening all potential signals and selecting those for which the marginal benefit of monitoring exceeds the marginal costs. The monitoring set consisting of only one signal is likely to lead to theoretical OTTRs biased downwards, which is a conservative way of estimating market making justified OTTRs.
- Another conservative assumption is the presence of only one liquidity provider posting liquidity across trading venues. This assumption follows from the model itself. If we accounted for multiple liquidity providers posting liquidity across trading venues, OTTRs would be higher.

We compute the relevant model parameters in the following manner:

- N_{it} is the count of how many markets stock i trades on during day t .

- λ_{SPYt} is the average number of midquote updates in SPY ETF per second during day t .
- λ_{mit} is the average number of market order arrivals per second in stock i during day t .
- q_{it} is a proportion of same-direction midquote changes in the SPY ETF and stock i during day t , out of the total number of midquote changes in SPY ETF. This measure is adjusted for coincidental same-direction midquote movements.

In estimating the signal quality parameter, we rely on the approach inspired by Dobrev and Schaumburg (2017) non-parametric measure of cross-market trading. First, we count the number of seconds per day with jointly positive or jointly negative midquote changes in the stock and SPY, and then divide this number by the total count of midquote changes in SPY. We adjust for any coincidental (i.e., not information-related) same-direction midquote returns by subtracting the same measure with the time offset of one hour. This measure captures signal quality as defined in our theoretical model, because it accounts for the frequency of quote updates in a security, given that there's a change in the signal. Signal quality is calculated as follows:

$$q_{it} = \frac{p(A_{it}^+)}{p(R_{SPYt})} + \frac{p(A_{it}^-)}{p(R_{SPYt})} - \frac{p(B_{it}^+)}{p(R_{SPYt})} - \frac{p(B_{it}^-)}{p(R_{SPYt})} , \quad (7)$$

where $p(A_{it}^+)$ and $p(A_{it}^-)$ are frequencies of positive and negative (respectively) midquote changes in security i on day t that occur within the same second as positive and negative (respectively) midquote changes in SPY ETF, $p(R_{SPYt})$ is the frequency of all non-zero midquote changes in SPY ETF on day t , $p(B_{it}^+)$ and $p(B_{it}^-)$ are frequencies of positive and negative (respectively) midquote changes in security i on day t that occur with one-hour offset compared to positive and negative (respectively) midquote changes in SPY ETF.

6.3. Estimation outcomes

Descriptive statistics on theoretical and empirical OTTRs is presented in Table 7. The theoretically justified order to trade ratios are 17 on average, compared to mean empirical OTTRs of only 6.3. This suggests that OTTRs justified by market making alone, even under the conservative assumptions of one liquidity provider and one signal watched, are over two times higher than the ratios observed in the market.

< Table 7 here >

Plotting the distributions of theoretical vs empirical OTTRs (see Figure 9) shows that for 93% of stock-days empirical OTTRs are significantly below those justified by market making

alone. However, the distribution of empirical OTTR is right-skewed, unlike the distribution of theoretical OTTRs, suggesting that the 7% of stock-days when observed OTTRs are above those justified by the model might be displaying some abnormal activity. In fact, our estimation approach can serve as a monitoring tools for regulators, as it allows to identify suspicious quoting activity beyond that justified by market making alone.

< Fig. 9 here >

6. Policy implications

Our results help explain the drivers of order-to-trade ratios (OTTRs) across markets and in a cross-section of stocks. We propose a theoretical framework that models OTTRs expected from legitimate market making activity in fragmented markets, and then verify the model predictions in empirical data from the US. We find empirical results which are well-aligned with our theoretical propositions, suggesting that the variation in OTTRs is largely driven by liquidity provision reasons. Our analysis also offers arguments for how OTTR-related regulatory measures might affect market quality. Specifically, our theory model addresses the question of messaging tax effects on market making activity across different assets. The paper has several policy implications.

Firstly, our model suggests that high OTTRs are related to market making in fragmented markets. Hence, it should not be surprising to find increasing OTTRs as technology enables ever-faster incorporation of information through quote revisions, while allowing to instantaneously revise quotes across trading venues. Importantly, as markets fragment, liquidity providers inevitably scale up their quoting activity, even in the absence of other factors (e.g., higher speed or lower cost of monitoring). Moreover, as trading volume fragments across markets, the smaller trading venues will naturally experience higher OTTRs, and hence messaging taxes would disadvantage newcomers as compared to incumbent stock exchanges. We find that theoretical predictions on fragmentation-OTTR and fragmentation-market share relations are supported in empirical evidence from the US markets.

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

more if a messaging tax were to be introduced. Also, market making activity at times of high market volatility would be harmed, as OTTRs are naturally higher in volatile markets.

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

Our model also helps explain historical dynamics of OTTRs. The run-up in OTTRs in early 2000s (before Reg NMS) coincides with the steep decrease in technology costs, suggesting that lower monitoring costs were likely at play. Further, with the introduction of Reg NMS and market fragmentation, OTTRs experience the most rapid increase. At the same time, short-term fluctuations in OTTRs are in line with spikes in market volatility, as the picking-off risk in those periods increases.

Finally, the model suggests that recent OTTRs are on average lower, not higher, than what market making in fragmented markets would require. However, the right skew in empirical OTTRs results in a small proportion of abnormally high (above theoretical level) ratios, which could be of interest to regulators monitoring the market for abnormal quoting activity. Hence our model provides a potential measure of excess quoting activity that regulators could use for market surveillance.

Appendix 1

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

Proof.

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

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

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

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

Proposition 2. *As markets fragment, OTTR for a given security on a given market increases as the market share of that market decreases.*

Proof.

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

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

Proposition 2a. *OTTR for a given security increases with monitoring intensity.*

Proof.

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

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

Proof.

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

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

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

Proposition 4. OTTR for a given security increases with picking-off risk.

Proof.

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

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

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

Proof.

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

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

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

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

Proof.

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

$$\frac{d[\lambda_m \left(\frac{\lambda_i q_i}{\lambda_m + \lambda_i q_i}\right)^{k - \lambda_i c}]}{d\lambda_m} = \frac{k \lambda_i^2 q_i^2}{(\lambda_m + \lambda_i q_i)^2} > 0$$

— this expression is strictly positive for all values of parameters except for $q_i = 0$. Thus, monitoring intensity increases with trading frequency as long as the signal quality is non-zero. Intuitively, as the expected number of market order arrivals λ_m (or trading frequency) increases, the picking-off intensity increases, thus incentivizing the liquidity provider to monitor more.

However, trading intensity also enters OTTR directly by affecting the quoting activity (quote updates after fills on market orders) and number of trades executed. Recall the expression for OTTR: $OTTR = \frac{2 \sum_{i \in \{MonitoredSignals\}} \lambda_i q_i + 2\lambda_m}{\lambda_m}$. Taking the first derivative with respect to trading

frequency (λ_m): $\frac{dOTTR}{d\lambda_m} = \frac{-2 \sum_{i \in \{MonitoredSignals\}} \lambda_i q_i}{\lambda_m^2} < 0$ for all parameter values except for $q_i = 0$.

This suggests that OTTR decreases as the trading frequency increases.

Thus, OTTR decreases with trading frequency, if we keep the monitoring intensity constant.

References

- Balchunas, E., Woolley, P., 2016, The Institutional ETF Toolbox: How Institutions Can Understand and Utilize the Fast-Growing World of ETFs. John Wiley & Sons.
- Biais, B., Woolley, P., 2011, High frequency trading, *Working paper*.
- Blocher, J., Cooper, R., Seddon, J., Vliet, B. V., 2016, Phantom liquidity and high-frequency quoting, *Journal of Trading* 11, 6–15.
- Caivano, V., Ciccarelli, S., Di Stefano, G., Fratini, M., Gasparri, G., Giliberti, M., Linciano, N., Tarola, I., 2012, High-frequency trading: Definition, effects, policy issues, Commissione Nazionale per le Società e la Borsa, *Discussion Paper №5*.
- Capelle-Blancard, G., 2017, Curbing the growth of stock trading? Order-to-trade ratios and financial transaction taxes, *Journal of International Financial Markets, Institutions and Money*.
- Carrion, A., 2013, Very fast money: High-frequency trading on the NASDAQ, *Journal of Financial Markets* 16, 680–711.
- Colliard, J. E., Hoffman, P., 2017, Financial transaction taxes, market composition, and liquidity, *Journal of Finance*.
- Committee on Capital Markets Regulation, 2016, The US equity markets. A plan for regulatory reform.
- Conrad, J., Wahal, S., Xiang, J., 2015, High-frequency quoting, trading, and the efficiency of prices, *Journal of Financial Economics* 116, 271–291.
- Dahlström, P., Hagströmer, B., Nordén, L., 2017, Determinants of limit order cancellations, *Working paper*, Stockholm Business School.
- Degryse, H., De Jong, F., Van Kervel, V., 2014, The impact of dark trading and visible fragmentation on market quality, *Review of Finance* 19, 1587–1622.
- Dobrev, D., Schaumburg, E., 2017, High-frequency cross-market trading: Model-free measurement and applications, *AFA Working paper*.
- Egginton, J. F., Van Ness, B. F., Van Ness, R. A., 2016, Quote stuffing, *Financial Management* 45, 583–608.
- Foucault, T., Kadan, O., Kandel, E., 2013, Liquidity cycles and make/take fees in electronic markets, *Journal of Finance* 68, 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, 345–384.
- Friedrich, S., Payne, R., 2015, Order-to-trade ratios and market quality, *Journal of Banking and Finance* 50, 214–223.

- Hagströmer, B., Nordén, L., Zhang, D., 2014, How aggressive are high-frequency traders? *Financial Review* 49, 395–419.
- Hendershott, T., Jones, C., Menkveld, A. J., 2011, Does algorithmic trading improve liquidity? *Journal of Finance* 66, 1–33.
- Jarnecic, E., Snape, M., 2014, The provision of liquidity by high-frequency participants, *Financial Review* 49, 371–394.
- Jørgensen, K., Skjeltorp, J., Ødegaard, B.A., 2014, Throttling hyperactive robots: Message to trade ratios at the Oslo Stock Exchange, *Journal of Financial Markets*.
- Lepone, A., Sacco, A., 2013, The impact of message traffic regulatory restrictions on market quality: Evidence from Chi-X Canada, *CMCRC working paper*, University of Sydney.
- Liu, W., 2009, Monitoring and limit order submission risks, *Journal of Financial Markets* 12, 107–141.
- Lyle, M. R., Naughton, J. P., 2015, How does algorithmic trading affect market quality? *Working paper*, University of Toronto.
- Mahoney, P. G., Rauterberg, G., 2017, The regulation of trading markets: A survey and evaluation, *Working paper*, University of Virginia School of Law.
- Malceniece, L., Malceniaks, K., Putnins, T. J., 2016, High-frequency trading and co-movement in financial markets, *Working paper*, University of Technology Sydney.
- Malinova, K., Park, A., Riordan, R., 2016, Taxing high frequency market making: Who pays the bill? *Working paper*, University of Toronto.
- Menkveld, A. J., 2013, High frequency trading and the new market makers, *Journal of Financial Markets* 16, 712–740.
- Menkveld, A. J., 2016, The economics of high-frequency trading: Taking stock, *Annual Review of Financial Economics* 8, 1–24.
- Menkveld, A. J., Zoican, M. A., 2017, Need for speed? Exchange latency and liquidity, *Review of Financial Studies* 30, 1188–1228.
- Nordhaus, W. D., 2007, Two centuries of productivity growth in computing. *Journal of Economic History* 67, 128–159.
- O’Hara, M., 2015, High-frequency market microstructure, *Journal of Financial Economics* 116, 25–70.
- Rosu, I., Sojli, E., Tham, W. W., 2017, Trades, quotes and the cost of capital, *Working paper*, University of New South Wales.
- Subrahmanyam, A., Zheng, H., 2016, Limit order placement by high-frequency traders, *Borsa Istanbul Review* 16, 185–209.

- Van Kervel, V., 2015, Competition for order flow with fast and slow traders, *Review of Financial Studies* 28, 2094–2127.
- Van Kervel, V., Menkveld, A. J., 2016, High-frequency trading around large institutional orders, *Working paper*.
- Yao, C., Ye, M., 2015, Why speed matters: A tale of queue rationing under price controls, *Working paper*.
- Ye, M., 2017, Who provides liquidity and when: An analysis of price vs speed competition on liquidity and welfare, *Working paper*.
- Yueshen, B. Z., 2017, Uncertain market making, *Working paper*, INSEAD.

Table 1
Variables used in regression analysis

<i>Propositions and Hypotheses</i>	<i>Variable definitions</i>
<p><i>Proposition 1. As markets fragment, market-wide OTTR for a given security increases with the extent of fragmentation, if there is at least one non-zero quality signal in the monitoring set.</i></p> <p><i>Hypothesis 1a. OTTRs are higher for stock-days with higher degrees of fragmentation.</i></p> <p><i>Hypothesis 1b. OTTRs are higher for exchanges that trade more fragmented stocks.</i></p>	<p>$OTTR_{it} = OrderVol_{it}/TradeVol_{it}$ — ratio of order volume (number of shares) to trade volume (number of shares) in stock i on day t.</p> <p>$OTTR_{jt} = OrderVol_{jt}/TradeVol_{jt}$ — ratio of order volume (number of shares) to trade volume (number of shares) on market j on day t.</p> <p>$Frag1_{it} = N_{it}$ — number of trading venues that have executed trades in stock i on day t.</p> <p>$Frag2_{it} = 1 - \sum_j (\frac{TradeVol_{ijt}}{\sum_j TradeVol_{ijt}})^2$ — Herfindahl-Hirschman index of fragmentation, based on trading volume (number of shares) in stock i, on market j, on day t.</p> <p>$Frag3_{it} = 1 - \sum_j (\frac{NTrades_{ijt}}{\sum_j NTrades_{ijt}})^2$ — Herfindahl-Hirschman index of fragmentation, based on the number of trades in stock i, on market j, on day t.</p> <p>Exchange-date fragmentation variables are constructed by taking equally weighted averages of stock-date fragmentation measures for each exchange.</p>
<p><i>Proposition 2. As markets fragment, OTTR for a given security on a given market increases as the market share of that market decreases.</i></p> <p><i>Hypothesis 2a. OTTRs are higher for markets with lower market shares.</i></p>	<p>$MarketShare_{jt} = DolVol_{jt}/\sum_j DolVol_{jt}$ — dollar volume market share of market j on day t. Main proxy for market share.</p> <p>$LogVolume_{jt}$ — natural logarithm of trading volume (in number of shares) for market j on day t. Second proxy for market share.</p> <p>$LogDolVolume_{jt}$ — natural logarithm of trading volume (in dollar amount) for market j on day t. Third proxy for market share.</p>
<p><i>Proposition 3. OTTR for a given security increases with the quality of signals available for monitoring.</i></p> <p><i>Hypothesis 3a. ETFs have higher OTTRs compared to the common stocks.</i></p> <p><i>Hypothesis 3b. Exchanges that trade more ETFs have higher OTTRs.</i></p> <p><i>Hypothesis 3c. Securities with higher correlation with the broad market index have higher OTTRs.</i></p>	<p>$CorrelationS\&P_{it}$ — absolute value of the average 22-day correlation between daily returns on security i and daily returns on S&P500 index.</p> <p>$ETF\ dummy_i$ — dummy variable that takes the value of 1 if security i is an ETF, and 0 otherwise.</p> <p>$ETF\ share_{jt}$ — dollar volume weighted share of ETFs on market j on day t. Takes values between 0 (no ETFs traded on market j on day t) and 1 (only ETFs traded on market j on day t).</p>

Proposition 4. OTTR for a given security increases with picking-off risk.

Hypothesis 4a. OTTRs are higher for stock-days with higher market volatility.

Hypothesis 4b. OTTRs are higher for stock-days with higher stock volatility.

Hypothesis 4c. OTTRs are higher for stock-days with higher price-to-tick ratios.

$StockVolatility_{it} = \frac{2(High_{it}-Low_{it})}{High_{it}+Low_{it}}$ — measure of stock i 's volatility on day t is based on daily high and low prices of the respective stock.

$MarketVolatility_t = \frac{2(High_{S\&P_t}-Low_{S\&P_t})}{High_{S\&P_t}+Low_{S\&P_t}}$ — measure of S&P500 index volatility on day t based on the daily high and low prices of the S&P500 index ETF.

$LogVIX_t$ — natural logarithm of the VIX index level.

$LogTickToPrice_{it} = \log\left(\frac{TickSize_{it}}{Price_{it}}\right)$ — natural logarithm of tick to price ratio of stock i on day t (dollar tick size divided by dollar closing price).

Proposition 5. OTTR for a given security decreases with monitoring cost.

Hypothesis 5a. OTTRs are higher for stocks with higher market capitalization.

Hypothesis 5b. OTTRs are lower on markets with taker-maker fee structures.

$LogMarketCap_{it}$ — log of market capitalization for stock i on day t .

$Taker - maker\ dummy_j$ — dummy variable that takes the value of 1 if market j is a taker-maker market and 0 otherwise.

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

Hypothesis 6. OTTRs are inversely related to the trading volumes.

$LogVolume_{it}$ — natural logarithm of trading volume (in number of shares) for stock i on day t .

Table 2
Descriptive statistics

This table provides descriptive statistics for the data used in regression analysis. The sample period is January 1, 2012 to December 31, 2016. The data are at daily frequency. $LogOTTR = \ln(1 + OTTR)$. *Frag1* refers to the number of markets a stock trades on; *Frag2* refers to Herfindahl-Hirschman index based on share volume; *Frag3* refers to Herfindahl-Hirschman index based on number of trades. Detailed variable definitions are provided in Table 1.

	Mean	Standard Deviation	25th percentile	50th percentile	75th percentile
Stock-date panel					
<i>OTTR</i>	239.17	574.91	27.402	48.12	115.33
<i>LogOTTR</i>	4.76	2.11	3.4	4.02	5.36
<i>Frag1</i>	7.05	2.66	5.00	8.00	9.00
<i>Frag2</i>	0.64	0.22	0.59	0.71	0.78
<i>Frag3</i>	0.67	0.2	0.63	0.73	0.80
<i>LogVolume</i>	3.83	2.65	2.09	4.13	5.76
<i>LogMarketCap</i>	13.03	2.18	11.5	13	14.52
<i>StockVolatility</i>	0.03	0.03	0.01	0.02	0.04
<i>CorrelationS&P</i>	0.47	0.27	0.25	0.47	0.68
<i>LogTickToPrice</i>	-7.65	1.14	-8.47	-7.82	-6.99
<i>ETF dummy</i>	0.25	0.43			
Exchange-date panel					
<i>OTTR</i>	124.56	415.77	10.63	14.52	35.63
<i>LogOTTR</i>	3.25	1.32	2.45	2.74	3.60
<i>Frag1</i>	9.40	0.49	8.88	9.57	9.79
<i>Frag2</i>	0.79	0.01	0.78	0.79	0.79
<i>Frag3</i>	0.80	0.01	0.79	0.80	0.81
<i>MarketShare</i>	0.09	0.09	0.01	0.04	0.13
<i>LogVolume</i>	11.38	1.84	10.47	11.85	12.79
<i>LogDolVolume</i>	15.07	1.90	14.19	15.42	16.50
<i>ETF share</i>	0.36	0.25	0.24	0.36	0.45
<i>Taker-maker dummy</i>	0.17	0.38			
Time series variables					
<i>MarketVolatility</i>	0.0094	0.0055	0.0058	0.0081	0.0117
<i>LogVIX</i>	2.7350	0.2001	2.5885	2.6933	2.8472

Table 3
Regression results for stock-date panel

This table reports regression results for six different models, where the dependent variable is a logarithmic transformation of OTTR for stock i on day t : $\log(1 + OTTR_{it})$. *Frag1* refers to the number of markets a stock trades on; *Frag2* refers to Herfindahl-Hirschman index based on share volume; *Frag3* refers to Herfindahl-Hirschman index based on number of trades. Detailed variable definitions are provided in Table 1. Coefficient estimates are from OLS regressions with standard errors clustered by stock and date. T-statistics are reported in parentheses. ***, **, and * indicate statistical significance at 1%, 5%, and 10% levels, respectively.

	<i>OTTR (1)</i>	<i>OTTR (2)</i>	<i>OTTR (3)</i>	<i>OTTR (4)</i>	<i>OTTR (5)</i>	<i>OTTR (6)</i>
<i>Frag1</i>	0.09*** (18.28)			0.09*** (18.78)	0.09*** (18.02)	0.09*** (18.34)
<i>Frag2</i>		0.61*** (17.00)				
<i>Frag3</i>			0.53*** (12.76)			
<i>LogVolume</i>	-0.48*** (-79.44)	-0.45*** (-86.48)	-0.44*** (-82.34)	-0.49*** (-75.43)	-0.49*** (-74.26)	-0.48*** (-79.60)
<i>LogMarketCap</i>	0.14*** (17.28)	0.14*** (18.25)	0.15*** (18.81)	0.16*** (18.22)	0.16*** (18.02)	0.14*** (17.56)
<i>MarketVolatility</i>	15.80*** (8.80)	15.28*** (8.58)	15.27*** (8.54)	14.61*** (52.00)		
<i>LogVix</i>						0.55*** (24.63)
<i>StockVolatility</i>				1.79*** (12.82)	2.17*** (14.37)	
<i>CorrelationS&P</i>	1.25*** (26.82)	1.28*** (27.91)	1.29*** (28.08)	1.26*** (27.58)	1.31*** (28.40)	1.22*** (25.95)
<i>LogTickToPrice</i>	-0.14*** (-15.62)	-0.15*** (-16.71)	-0.15*** (-16.93)	-0.14*** (-15.33)	-0.14*** (-15.28)	-0.14*** (-15.76)
<i>ETF dummy</i>	2.77*** (61.57)	2.70*** (61.73)	2.69*** (61.37)	2.82*** (63.03)	2.81*** (62.25)	2.78*** (61.43)
Adjusted R ²	0.78	0.77	0.77	0.78	0.78	0.77

Table 4
Regression results for exchange-date panel

This table reports regression results for five different models, where the dependent variable is a logarithmic transformation of OTTR for exchange j on day t : $\log(1 + OTTR_{jt})$. *Frag1* refers to the number of markets a stock trades on; *Frag2* refers to Herfindahl-Hirschman index based on share volume; *Frag3* refers to Herfindahl-Hirschman index based on number of trades. Detailed variable definitions are provided in Table 1. Coefficient estimates are from OLS regressions with standard errors clustered by both exchange and date. T-statistics are reported in parentheses***, **, and * indicate statistical significance at 1%, 5%, and 10% levels, respectively.

	<i>OTTR (1)</i>	<i>OTTR (2)</i>	<i>OTTR (3)</i>	<i>OTTR (4)</i>	<i>OTTR (5)</i>	<i>OTTR (6)</i>
<i>Frag1</i>	0.32*** (3.37)			0.25** (2.26)	0.23** (2.35)	0.27*** (2.95)
<i>Frag2</i>		1.45 (0.19)				
<i>Frag3</i>			-4.81 (-0.50)			
<i>Taker-maker dummy</i>	-0.85*** (-3.78)	-0.92*** (-3.81)	-0.91*** (-3.74)	-0.56*** (-3.59)	-0.52*** (-3.86)	-0.82*** (-3.84)
<i>MarketShare</i>	-4.68*** (-3.43)	-5.04*** (-3.39)	-5.46*** (-3.69)			-12.24*** (-2.73)
<i>MarketShare Squared</i>						29.74** 1.90
<i>ETF dummy</i>	3.75*** (5.29)	3.69*** (4.87)	3.62*** (4.54)	3.83*** (5.57)	3.38*** (5.72)	3.60*** (5.26)
<i>LogVix</i>	0.38*** (3.27)	0.37*** (3.07)	0.33*** (3.03)	0.49*** (4.99)	0.58*** (5.78)	0.38*** (3.42)
<i>LogVolume</i>					-0.28*** (-4.87)	
<i>LogDolVolume</i>				-0.23*** (-3.85)		
Adjusted R ²	0.69	0.71	0.71	0.74	0.77	0.79

Table 5
Summary of empirical results

Hypotheses	Empirical support
<i>Hypothesis 1a.</i> OTTRs are higher for stock-days with higher degrees of fragmentation.	Yes
<i>Hypothesis 1b.</i> OTTRs are higher for exchanges that trade more fragmented stocks.	Yes
<i>Hypothesis 2a.</i> OTTRs are higher for markets with lower market shares.	Yes
<i>Hypothesis 3a.</i> ETFs have higher OTTRs compared to the common stocks.	Yes
<i>Hypothesis 3b.</i> Exchanges that trade more ETFs have higher OTTRs.	Yes
<i>Hypothesis 3c.</i> Securities with higher correlation with the broad market index have higher OTTRs.	Yes
<i>Hypothesis 4a.</i> OTTRs are higher for stock-days with higher market volatility.	Yes
<i>Hypothesis 4b.</i> OTTRs are higher for stock-days with higher stock volatility.	Yes
<i>Hypothesis 4c.</i> OTTRs are higher for stock-days with higher price-to-tick ratios.	Yes
<i>Hypothesis 5a.</i> OTTRs are higher for stocks with higher market capitalization.	Yes
<i>Hypothesis 5b.</i> OTTRs are lower on markets with taker-maker fee structures.	Yes
<i>Hypothesis 6.</i> OTTRs are inversely related to the trading volumes.	Yes

Table 6**Descriptive statistics for the historical data**

This table provides descriptive statistics for the time series data used in historical analysis. The 100 sample stocks are selected at random from the SEC MIDAS dataset using stratified sampling by market capitalization. The sample dates are January 1, 2000 to December 31, 2016. OTTR refers to the OTTRs computed using quotes and trades data from Thomson Reuters Tick History (daily frequency measure). Fragmentation refers to the measure of market fragmentation computed using Herfindahl-Hirschman index:

$Frag_{it} = (1 - \sum_{i=1}^N \left(\frac{Vol_{it}}{Vol_t}\right)^2)$, where Vol_{it} is the share volume traded on market i on day t . CPU costs are in \$/MIPS (million operations per second, quarterly frequency measure). Bandwidth costs refer to the costs of 1-year lease for 10 Gbps circuit on the Chicago — New York link (quarterly frequency measure, available only from the year 2002 onwards). VIX refers to the VIX index of market volatility (daily frequency measure).

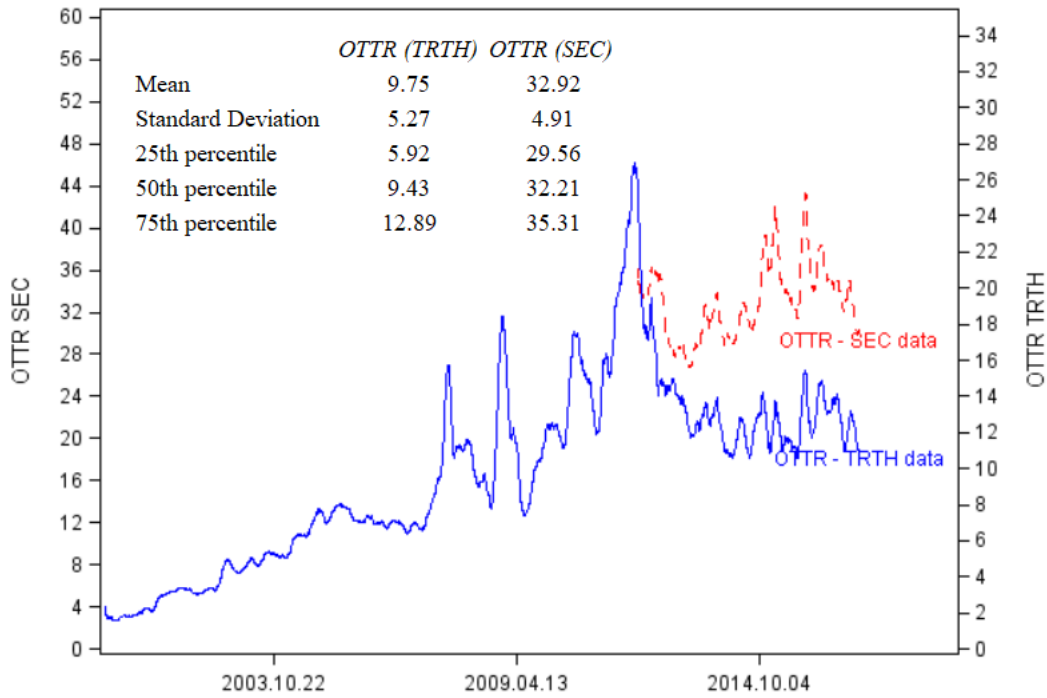
	<i>OTTR</i>	<i>Fragmentation (HHI measure)</i>	<i>CPU Costs</i>	<i>Bandwidth Costs</i>	<i>VIX</i>
<i>Mean</i>	9.75	0.54	0.13	27 178	20.36
<i>StDev</i>	5.27	0.23	0.25	28 484	8.67
<i>25th percentile</i>	5.92	0.29	0.01	6 055	14.12
<i>50th percentile</i>	9.43	0.59	0.02	12 216	18.26
<i>75th percentile</i>	12.89	0.76	0.16	43 931	23.96

Table 7**Estimation results for empirical vs theoretical OTTRs**

This table provides results from estimating theoretical OTTRs and comparing them to empirical OTTRs. The data cover 761 stock-days. The sample stocks are selected from the 2016 SEC dataset using stratified sampling by market capitalization. The sample dates are chosen to reflect median market volatility levels across five VIX quintiles. The data are at daily frequency. *OTTR empirical* refers to the OTTRs computed using quotes and trades data from Thomson Reuters Tick History. *OTTR theoretical* refers to the OTTRs computed for the same stocks and dates using the theory model under conservative assumptions of one liquidity provider monitoring one signal. Theoretical OTTRs are computed using the following formula: $OTTR_{it} = \frac{2N_{it}(\lambda_{SPY_t}q_{it} + \lambda_{mit})}{\lambda_{mit}}$, where $OTTR_{it}$ is the OTTR in stock i on day t , N_{it} is the number of markets stock i trades on during day t , λ_{SPY_t} is the signal intensity (number of midquote updates in SPY ETF per second), q_{it} is a measure of signal quality (proxied by the proportion of same-direction midquote changes in the signal and the stock, out of the total number of midquote changes in the signal), λ_{mit} is a measure of market order arrival intensity (proxied by the average number of trades in the stock i per second).

	<i>OTTR empirical</i>	<i>OTTR theoretical</i>	<i>Difference (empirical minus theoretical)</i>
<i>Mean</i>	6.31	17.02	-10.88
<i>StDev</i>	7.51	3.07	7.50
<i>25th percentile</i>	3.00	16.10	-14.26
<i>50th percentile</i>	4.18	16.60	-12.61
<i>75th percentile</i>	6.55	18.25	-10.22

Panel A: Time series of OTTRs for the period 2000–2016 using TRTH and SEC MIDAS data



Panel B: Time series of OTTRs for the period 2012–2016 using TRTH and SEC MIDAS data

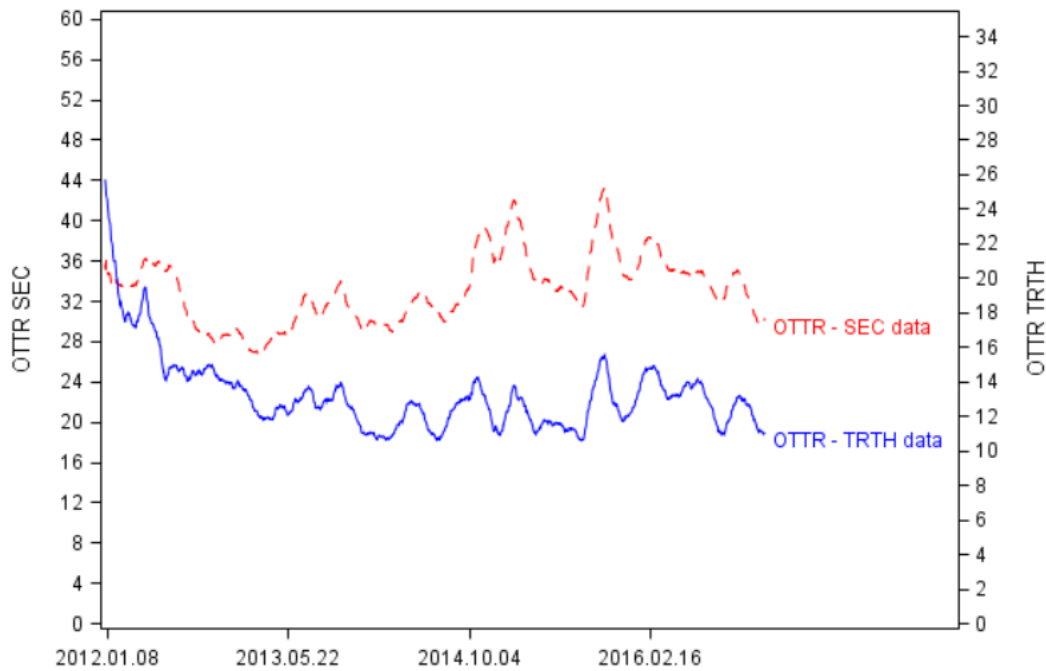
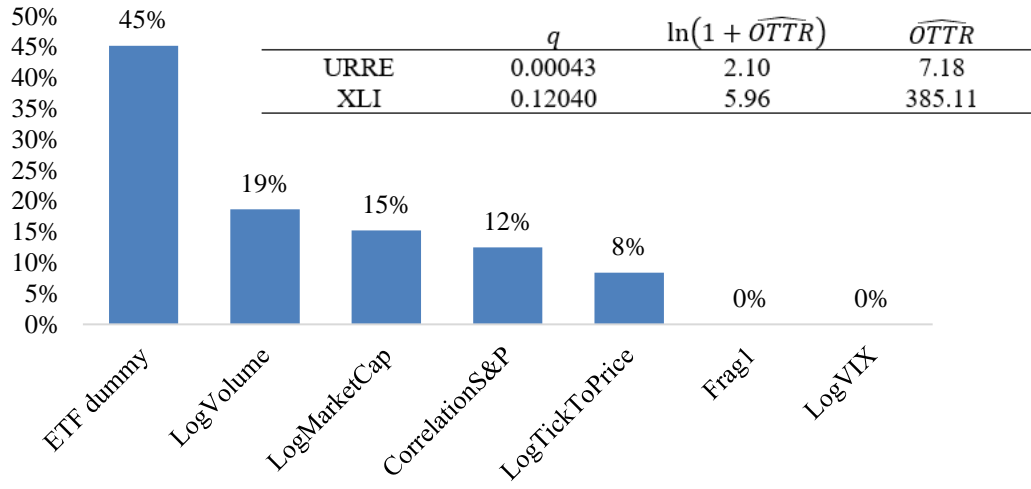


Fig. 1. Time series plot of order-to-trade ratios. The figure shows 2-month moving averages of OTTRs computed from daily data using SEC MIDAS and TRTH datasets. The SEC dataset covers the universe of all US traded stocks and ETFs. The TRTH dataset covers a random sample of 100 securities selected from MIDAS dataset. We compute SEC OTTRs following the SEC methodology: dividing the sum of order volume for all add order messages by lit volume. We compute TRTH OTTRs by dividing the number of order updates (price or quantity) at best quotes by the number of trades. Panel A uses the data for the period 2000–2016, and Panel B — for the period 2012–2016 (when the SEC MIDAS data is available).

Panel A: Empirical factors contribution to difference in $\ln(1 + \widehat{OTTR})$ between two securities with different signal quality (q)



Panel B: Empirical factors contribution to difference in $\ln(1 + \widehat{OTTR})$ between two securities with different degrees of fragmentation ($Frag1$)

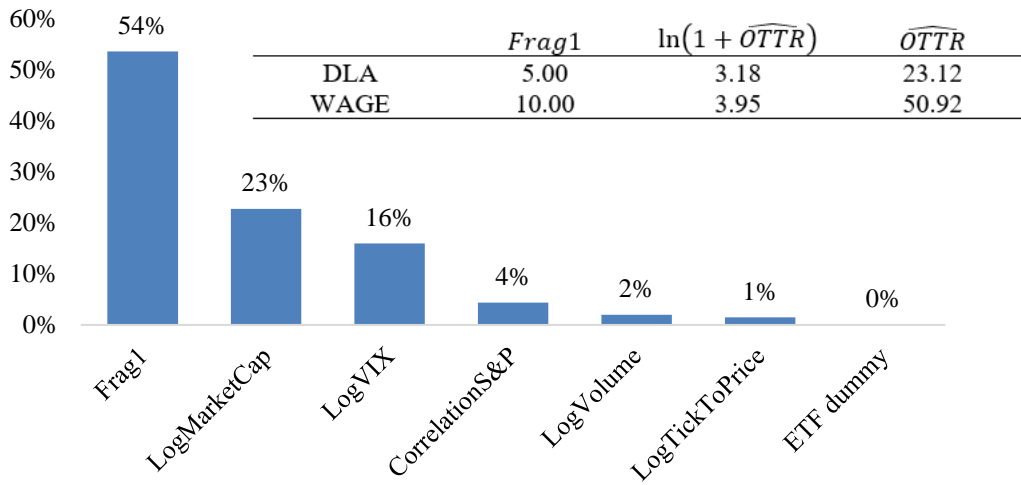


Fig. 2. Contribution to OTTR difference. Panel A presents the factors contributing to the difference between OTTRs of two securities with different signal quality. Signal quality, q , is proxied by the proportion of same-direction midquote changes in the signal (SPY ETF) and the security, out of the total number of midquote changes in the signal. The two securities are URRE (Uranium Resources) and XLY (Consumer Discretionary ETF). Panel B presents the factors contributing to the difference between OTTRs of two securities with different degrees of fragmentation (number of markets they trade on). The two securities are DLA (Delta Apparel Inc.) and WAGE (Wage Works). Panel C presents the factors contributing to the difference between OTTRs of two securities with different trading frequencies (number of trades per second). The two securities are IVV (iShares S&P 500 ETF) and SPY (SPDR S&P 500 Trust ETF). The contributions are calculated using coefficients from the regression model (6) presented in Table 3. $Frag1$ refers to the number of markets a security trades on. Detailed variable definitions are provided in Table 1.

Panel C: Empirical factors contribution to difference in $\ln(1 + \widehat{OTTR})$ between two securities with different trading frequencies (λ_m)

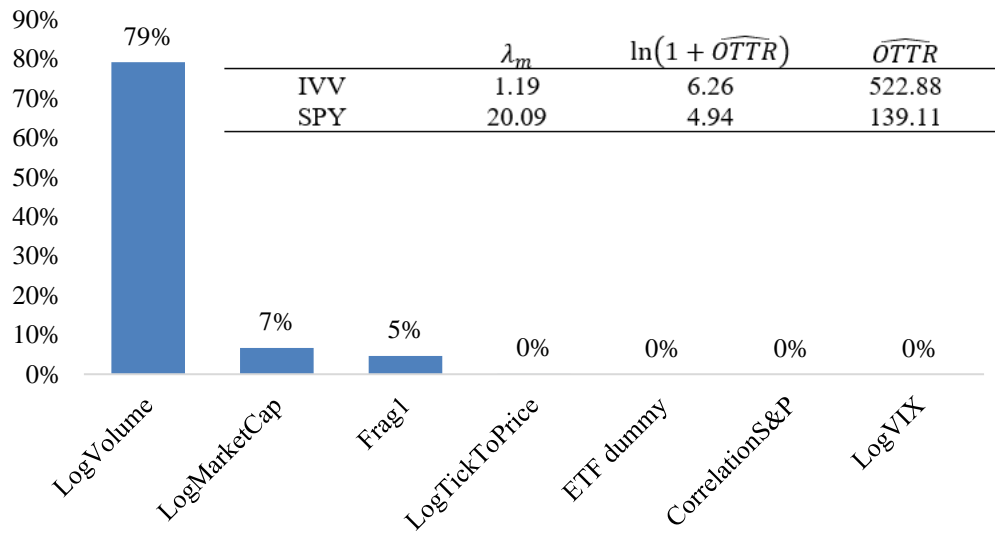


Fig. 2. Contribution to OTTR difference. Panel A presents the factors contributing to the difference between OTTRs of two securities with different signal quality. Signal quality, q , is proxied by the proportion of same-direction midquote changes in the signal (SPY ETF) and the security, out of the total number of midquote changes in the signal. The two securities are URRE (Uranium Resources) and XLY (Consumer Discretionary ETF). Panel B presents the factors contributing to the difference between OTTRs of two securities with different degrees of fragmentation (number of markets they trade on). The two securities are DLA (Delta Apparel Inc.) and WAGE (Wage Works). Panel C presents the factors contributing to the difference between OTTRs of two securities with different trading frequencies λ_m (number of trades per second). The two securities are IVV (iShares S&P 500 ETF) and SPY (SPDR S&P 500 Trust ETF). The contributions are calculated using coefficients from the regression model (6) presented in Table 3. *Frag1* refers to the number of markets a security trades on. Detailed variable definitions are provided in Table 1.

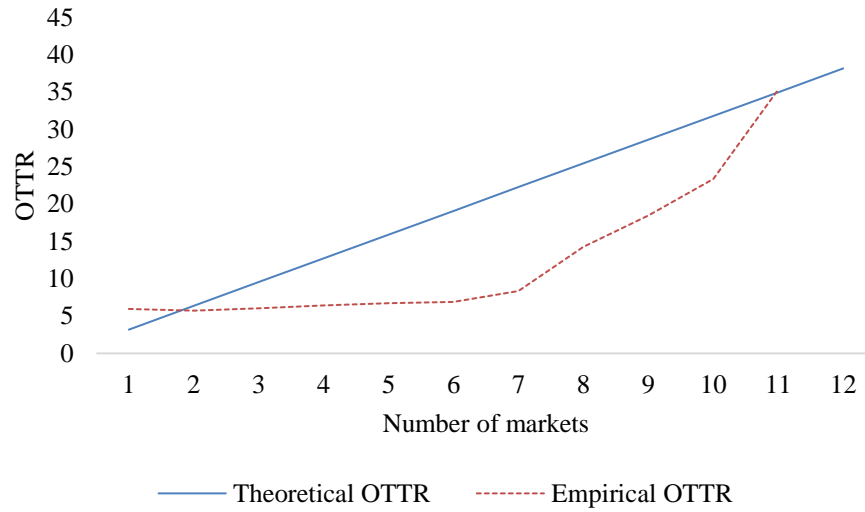


Fig. 3. Theoretical vs empirical relation between OTTR and fragmentation. The figure shows the relation between OTTR and fragmentation (i) predicted by the theory model and (ii) measured in the data. The empirical measurement is done by regressing OTTR on control variables and 11 dummy variables for fragmentation (where fragmentation measure is the number of markets). The control variables are volume, market capitalization, VIX, correlation with the market, tick-to-price ratio, and ETF dummy.

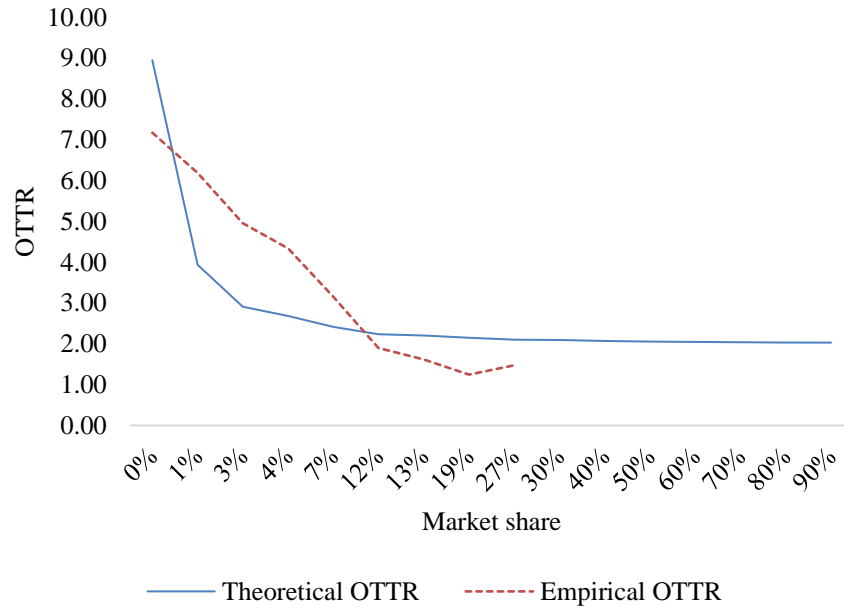
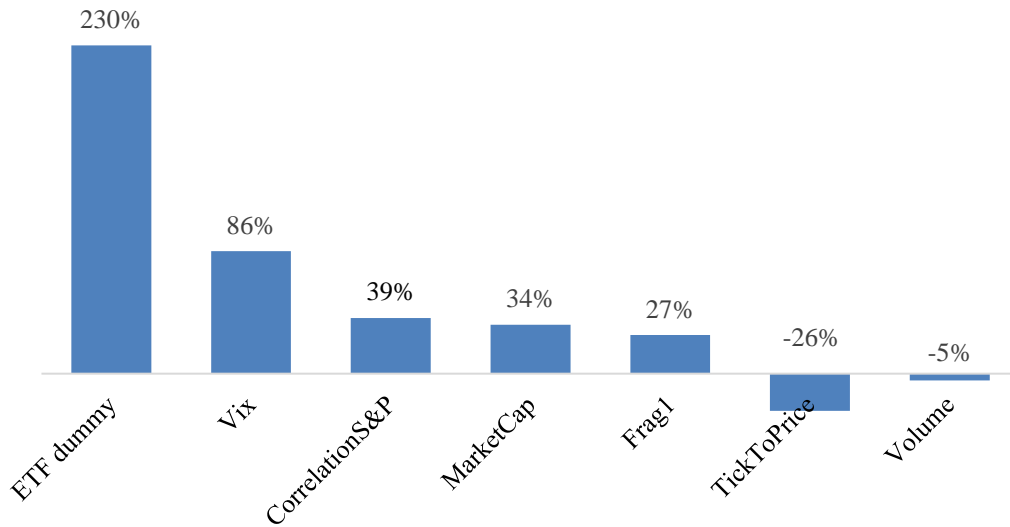


Fig. 4. Theoretical vs empirical relation between OTTR and market share. The figure shows the relation between OTTR and market share (i) predicted by the theory model and (ii) measured in the data. The empirical measurement is done by regressing OTTR on control variables and 9 dummy variables for market share deciles (where market share measure is measured using dollar volume), and the omitted dummy is for the lowest decile market share. The control variables are taker-maker dummy, ETF dummy, and VIX. The horizontal axis plots mean % market shares for each decile.

Panel A: Percentage change in (1+OTTR) resulting from one standard deviation change in explanatory variables at stock-date level



Panel B: Percentage change in (1+OTTR) resulting from one standard deviation change in explanatory variables at exchange-date level

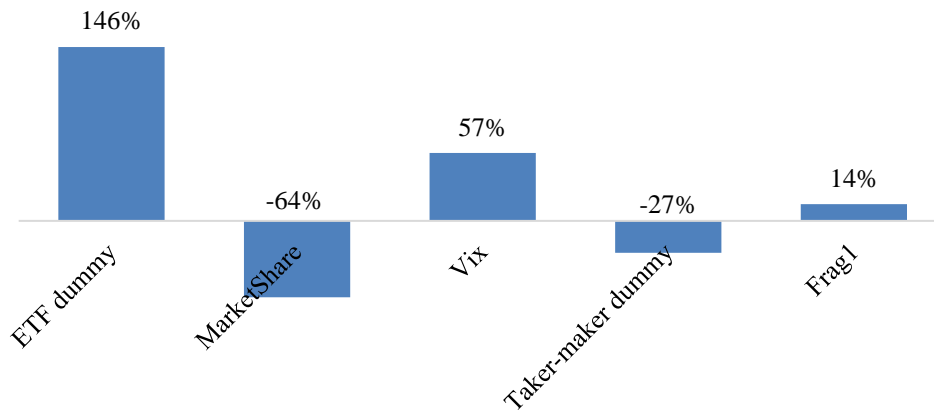


Fig. 5. Standardized coefficients from regression models. The figure presents standardized coefficients from the regression model (6) presented in Table 3 (Panel A) and Table 4 (Panel B). The vertical axis shows percentage change in $(1 + OTTR)$ that results from one standard deviation change in explanatory variables on the horizontal axis. For example, one standard deviation change in the VIX index leads to 86% increase in $(1 + OTTR)$ on an average stock-day in our sample. *Frag1* refers to the number of markets a stock trades on. Detailed variable definitions are provided in Table 1.

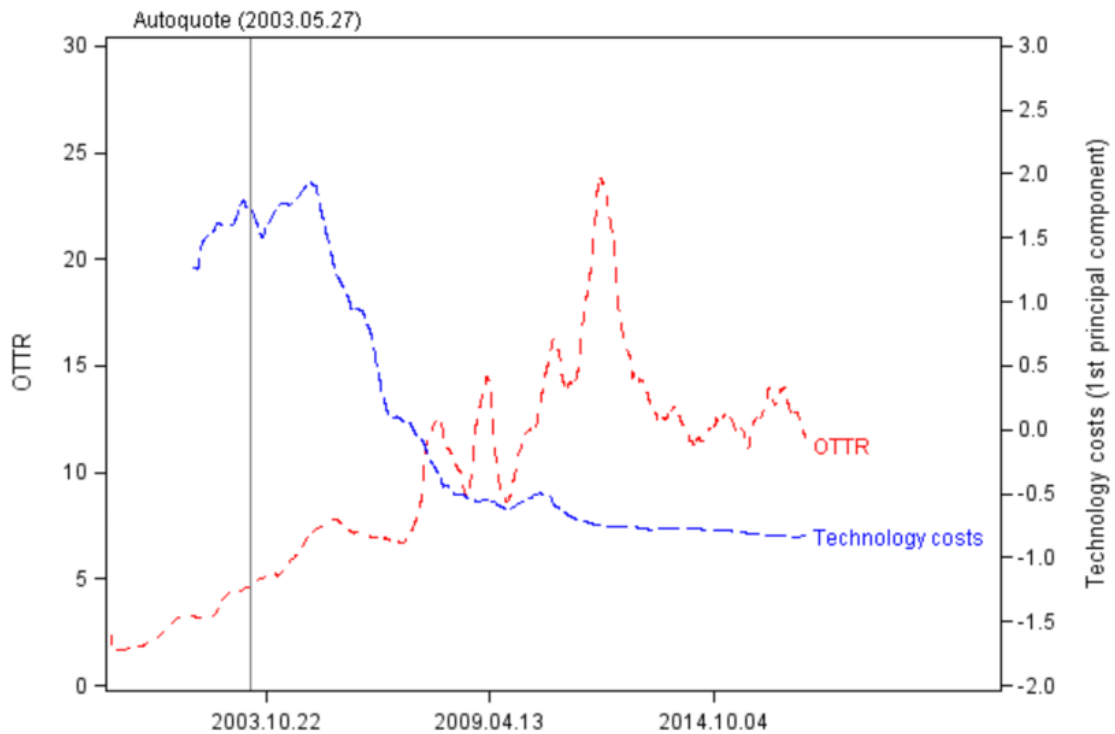


Fig. 6. OTTR vs technology costs. The figure reports 6-month moving averages for historical OTTRs and technology costs. OTTRs are computed using intraday Thomson Reuters Tick history data for the period January 1, 2000 to December 31, 2016 for 100 stocks. Sample stocks used are selected at random from SEC MIDAS database using stratified sampling by market capitalization deciles. Technology costs proxi is constructed as first principal component of two types of costs — CPU costs (in \$/MIPS - million operations per second - from the CPU Price Performance dataset by John McCallum) and broadband costs (in \$/ annum - annual leasing prices of 10 Gbps broadband circuit links between Chicago and New York — from Telegeography database). Technology costs data are available for the period 2002–2016. Autoquote refers to the final date of the phase-out of autoquote on NYSE.

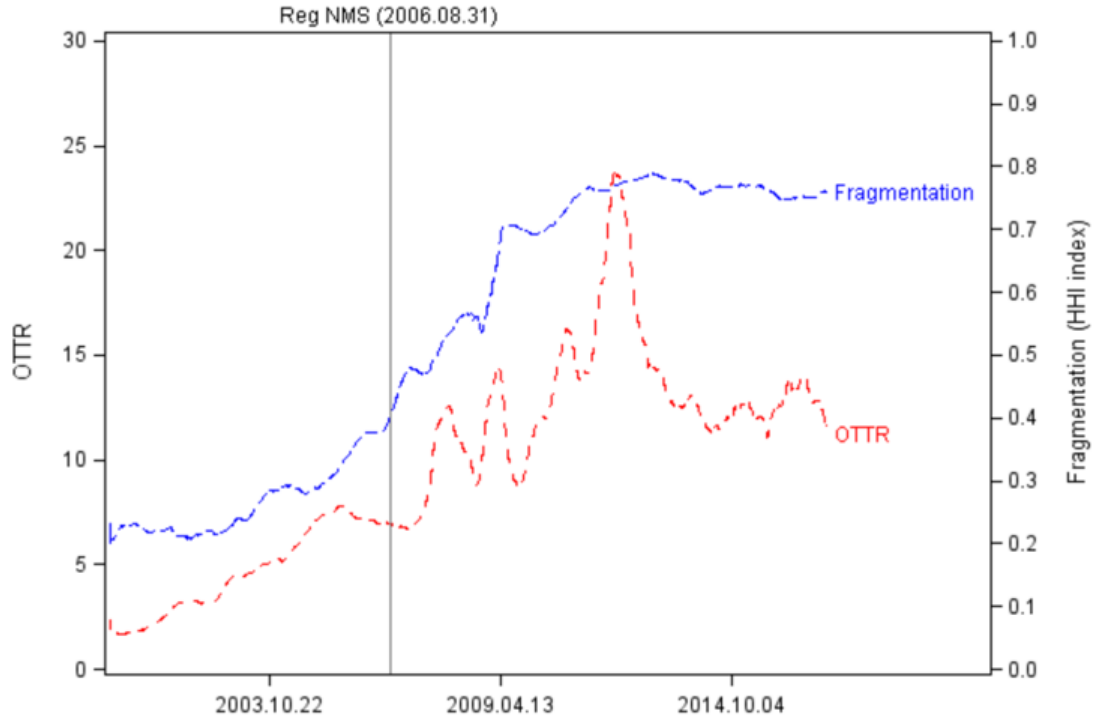


Fig. 7. OTTR vs fragmentation. The figure reports 6-month moving averages for historical OTTRs and Herfindahl-Hirschmann market fragmentation measure (scaled between 0 and 1, where higher values reflect more fragmented markets). OTTRs are computed using intraday Thomson Reuters Tick history data for the period January 1, 2000 to December 31, 2016 for 100 stocks. Sample stocks are selected at random from SEC MIDAS database using stratified sampling by market capitalization deciles. The market fragmentation measure is computed using Herfindahl-Hirschman index: $Frag_{it} = (1 - \sum_{i=i}^N (\frac{Vol_{it}}{Vol_t})^2)$, where Vol_{it} is the share volume traded on market i on day t . It is based on share volumes of 10 randomly selected high market cap stocks which are traded throughout the sample period January 2000 to December 2016. Reg NMS refers to the final compliance date with Regulation National Market System (including Rule 611 — trade-through rule).

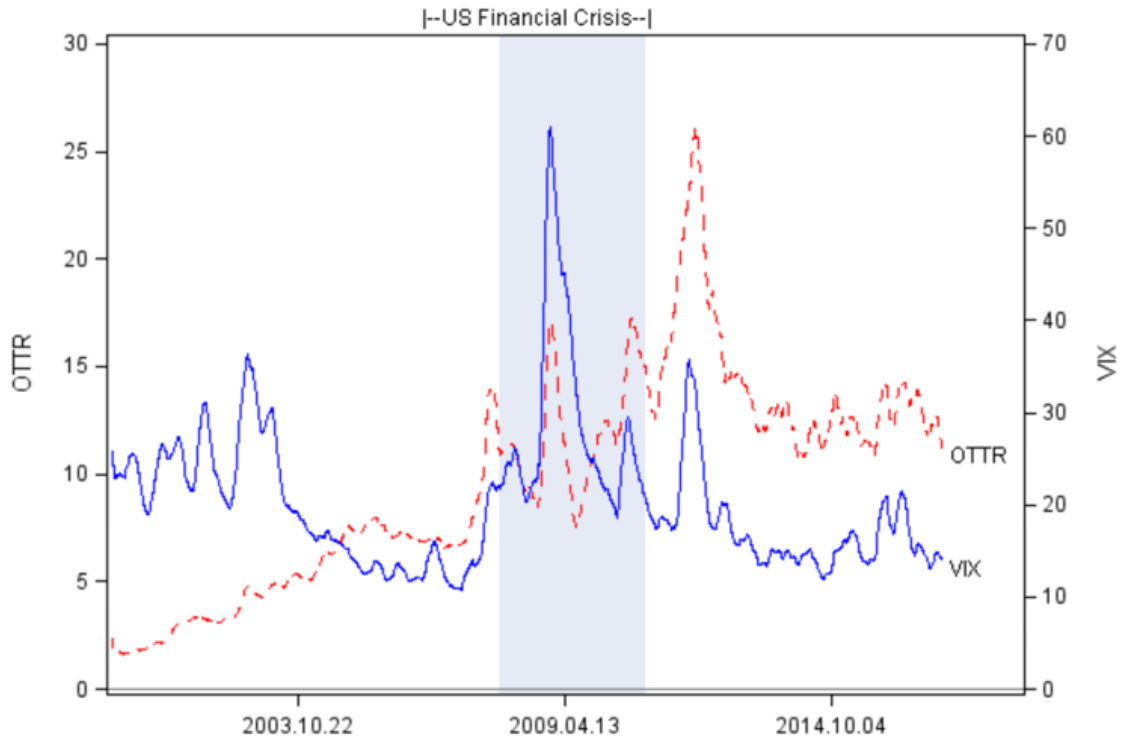
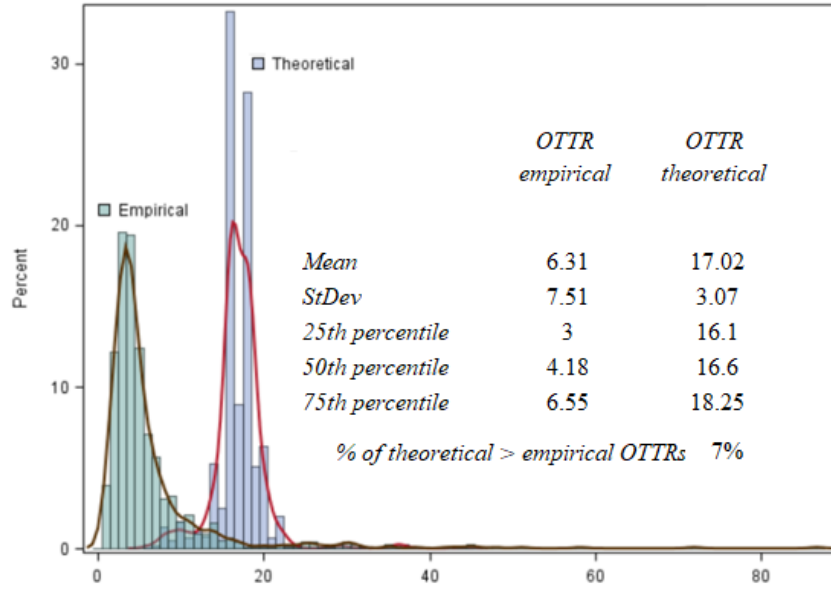


Fig. 8. OTTR vs VIX. The figure reports 3-month moving averages for historical OTTRs and VIX index of market volatility. OTTRs are computed using intraday Thomson Reuters Tick history data for the period January 1, 2000 to December 31, 2016 for 100 stocks. Sample stocks are selected at random from SEC MIDAS database using stratified sampling by market capitalization deciles.

Panel A: Distribution of theoretical and empirical OTTRs across stock-days



Panel B: Distribution of the difference (empirical minus theoretical) OTTRs across stock-days

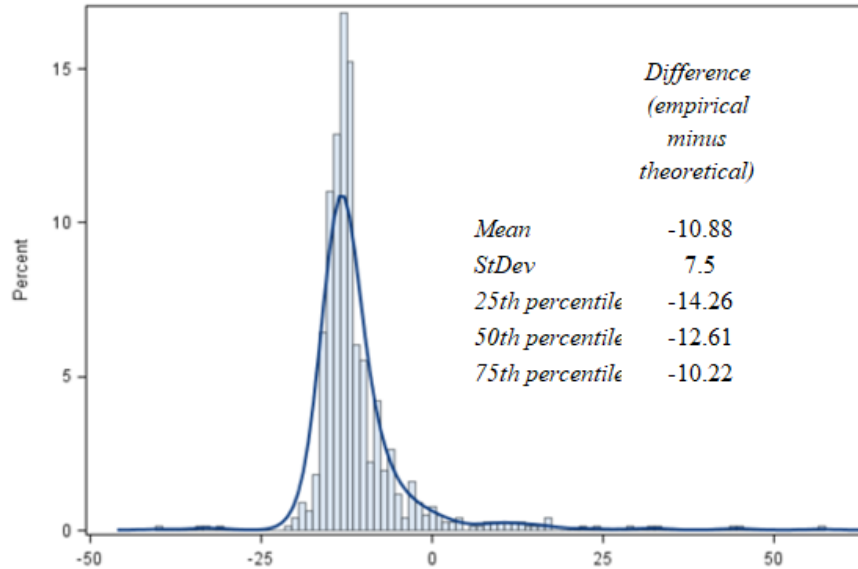


Fig. 9. Theoretical vs empirical OTTRs. The figure reports the distribution of theoretical vs empirical OTTRs (and distribution of their differences) that emerges from estimating theoretical OTTRs for 761 stock-days. The sample stocks are selected at random from the 2016 SEC dataset using stratified sampling by market capitalization. The sample dates are chosen to reflect median market volatility levels across five VIX quintiles. The data are at daily frequency. OTTR empirical refers to the OTTRs computed using quotes and trades data from Thomson Reuters Tick History. OTTR theoretical refers to the OTTRs computed for the same stocks and dates using the theory model under conservative assumptions of one liquidity provider monitoring one signal. Theoretical OTTRs are computed using the following formula: $OTTR_{it} = \frac{2N_{it}(\lambda_{SPYt}q_{it} + \lambda_{mit})}{\lambda_{mit}}$, where $OTTR_{it}$ is the OTTR in stock i on day t , N_{it} is the number of markets stock i trades on during day t , λ_{SPYt} is the signal intensity (number of midquote updates in SPY ETF per second), q_{it} is a measure of signal quality (proxied by the proportion of same-direction midquote changes in the signal and the stock, out of the total number of midquote changes in the signal), λ_{mit} is a measure of market order arrival intensity (proxied by the average number of trades in the stock i per second).