An investigation of market invariance for the Tokyo Stock Exchange*

William M. Cheung† Lewen Guo‡ Thomas H. McInish §

December 21, 2024

Abstract

We test the invariance-of-bet hypothesis from Kyle and Obizhaeva (2016) for the Tokyo Stock Exchange (TSE). The pooled regression coefficients of the logarithm of the number of trades on the logarithm of trading activities range from 0.665 to 0.669, close to the theoretical value of two-thirds predicted by the invariance-of-bet hypothesis. European markets data also confirms the two-thirds relations implied by market invariance. Our results suggest that using alternative transaction data reduces the measurement errors in variables such as the number of trades and the trade sizes, explaining why the two-thirds relation might not hold using more recent U.S. data.

Keywords: Microstructure Invariance, Measurement Errors, Tokyo Stock Exchange

JEL Classification Numbers: G12, G14

*First version: October 1st, 2021. We are grateful to Pankaj K. Jain, Amy Kwan, Albert (Pete) Kyle, Oliver Randall (discussant), Arikawa Yasuhiro, Masahiro Yamada (discussant), Eduard Yelagin, and participants in Asia Meeting of the Econometric Society 2022, Tokyo, FMA Asia/Pacific 2022, 32nd Nippon Finance Association, FISC workshop at the University of Memphis, and Waseda Organizational and Financial Economics Seminar at Waseda University for their insightful comments.

†Waseda Business School, Waseda University.

‡Corresponding author. School of Commerce, Waseda University. Email: guolewen@aoni.waseda.jp

§Fogelman College of Business and Economics, Department of Finance, Insurance and Real Estate, University of Memphis. Professor Thomas H. McInish deceased on July 3, 2022. We honor his contributions to the field of empirical market microstructure.

1 Introduction

A bet is a decision to invest in an asset. Kyle and Obizhaeva (2016) show that, under certain assumptions, the number of bets is proportional to two-thirds of the product of the fiat-currency volume of the bets and the volatility of the asset. Assuming a constant order-splitting multiplier and a constant volume multiplier, testing the invariance-of-bets hypothesis could involve a simple OLS regression, examining whether the regression coefficient of the natural logarithm of the number of trades on the logarithm of trading activities equals two-thirds.[1]

Kyle et al. (2020) test the invariance-of-bets hypothesis using trades and quotes data (TAQ) for the U.S. equity market. Their results support the invariance hypothesis for the first half of their sample period (1993-2001), with coefficients of the OLS regressions closely aligning with the theoretical values of two-thirds. However, in the second half (2001 to 2014), their results deviate from expectations, as the estimated coefficients increase substantially to more than 0.80 by the end of their sample period.

Why do the estimated coefficients deviate from the theoretical values of two-thirds when using transaction data like TAQ? What might contribute to the systematic upward biases in estimated coefficients observed after 2001, beyond the changes in tick size due to decimalization? Our paper addresses these issues using alternative transaction data from the Tokyo Stock Exchange (TSE).

[1]For example, a researcher could use the trades files in trades and quotes data (TAQ) to gather daily stock-level summary statistics, such as the number of trades and the trade sizes. They could also obtain stock prices and trading volume data in CRSP to construct the measure of trading activities, which is defined as the product of dollar (fiat-currency) volume and the volatility of the asset.

We argue that the key to appropriately testing the invariance hypothesis using transaction data involves accurately measuring the "trades"—a trade should be defined as executing a marketable order against the resting orders. However, using the straightforward number of trade prints in recent TAQ data as a measure of the number of trades may result in measurement errors, as a trade executed from one marketable order against the resting orders might produce multiple trade prints due to i) order routing, ii) differences in trading reporting systems across U.S. equity exchanges (Upson et al., 2021), and iii) the execution of hidden orders. Such measurement errors, if correlated with the trading activities of the stocks, will inevitably induce a positive bias in the estimated OLS coefficients.[2]

This paper argues that measurement errors in the number of trades may not pose a significant concern when using TSE transaction data to test the invariance hypothesis, given the various favorable settings in the TSE. First, the TSE consistently employs the *Level Book* reporting system in our sample period. Under this trade reporting system, a trade of 300 shares is documented when a 300-share marketable order executes against multiple resting orders at the same price level. In contrast, under *Order Based* reporting system, the same execution would produce two separate trade prints if the resting orders consist of one order of 100 shares and another of 200 shares. Second, there are no order routing activities or the execution of hidden orders in the TSE, in contrast to the U.S. equity markets. Third, reconstructing a trade resulting from a large marketable order that walks the book is rather straightforward.[3]

[2]We will discuss the details in Section 2.2.

[3]We can merge the trade prints at the same timestamp with different trade prices to reconstruct the

Our main results support the invariance hypothesis—the estimated coefficients from regressions of the natural logarithm of the number of trades on the natural logarithm of the trading activities are close to the theoretical values of two-thirds. Pooled regression coefficients for all trades, seller-initiated trades, and buyer-initiated trades are 0.669, 0.665, and 0.665, respectively. These coefficients are not significantly different from two-thirds at the 1% level. A time-series plot of the coefficients estimated by the cross-sectional regressions echoes the same narratives—the estimated coefficients fluctuate around two-thirds throughout the sample period from 2015 to 2020.

In addition to the baseline results, our TSE data enables us to explore the invariance hypothesis for the non-marketable orders. Through a tick-by-tick comparison of changes in shares at different price levels, we can obtain the number and the sizes of non-marketable orders, categorized by their order aggressiveness (Biais et al., 1995). The coefficients estimated from the regressions of the log of the number of non-marketable orders on log trading activities yield comparable results—from 0.616 to 0.686 depending on order aggressiveness. For the most aggressive non-marketable orders—the non-marketable orders improve the best bid or the best ask quotes, the estimated coefficients are 0.665 for the bid side and 0.672 for the ask side.[4] Both coefficients are not significantly different from the theoretical values of two-thirds at the 1% level.

To investigate how measurement errors in the number of trades might affect the estimated coefficients, we impose four different levels of measurement errors on the true

large marketable order that walks the book. Our main results using the TSE data are not affected by how we treat these trade prints.

[4]We define the "improvement" as the scenario where a non-marketable order either adds liquidity at the best price or updates to a more aggressive price.

number of trades using the TSE data. Using the pseudo-generated number of trades with measurement errors, we show that the estimated coefficients are upward-biased towards the theoretical values of two-thirds. The estimated coefficients using the number of trades with measurement errors range from 0.706 to 0.757 compared with the coefficient of 0.669 using the true number of trades.

We also document stock-level frictions, such as relative tick sizes imposed on individual stocks, that may cause the estimated coefficients to deviate from the theoretical values of two-thirds, consistent with the findings in Kyle et al. (2020). In particular, when categorizing stocks into quartiles based on their relative tick sizes (O'Hara et al., 2019), we find that more tick-constrained stocks exhibit a larger deviation from the theoretical values of two-thirds. However, the trade size distributions among different relative tick size groups appear to be log-normally distributed, aligning with the invariance hypothesis.

We also consider European markets, which share similarities with the TSE, for external validity tests of the invariance hypothesis. Specifically, we include six order-driven markets: Brussels Stock Exchange, Paris Stock Exchange, Irish Stock Exchange, Amsterdam Stock Exchange, Oslo Stock Exchange, and Euronext Lisbon. The estimated coefficients range from 0.698 for the Paris Stock Exchange to 0.770 for the Brussels Stock Exchange, suggesting that the invariance hypothesis also holds in European markets. However, further investigations into why some of these markets deviate more from two-thirds than others may involve examining the institutional details at each exchange level.

We contribute to the small but growing literature in testing market invariance theories (Bae et al., 2016; Kyle and Obizhaeva, 2016; Kyle et al., 2017; Bucci et al., 2020; Kyle

and Obizhaeva, 2020; Kyle et al., 2020). First, we extend the empirical literature on the invariance hypothesis by providing strong empirical evidence that the scaling law of two-thirds holds in order-driven markets such as the TSE. We argue that using the transaction data from the TSE resolves the measurement errors in the key variables such as the number of trades and trade sizes.

Second, we are among the first to investigate the invariance hypothesis in order-driven markets. Our findings highlight potential challenges in testing invariance theories due to institutional details across markets, which, in turn, may induce measurement errors due to variations in trade reporting systems and order routing resulting from market fragmentation.

The rest of the paper is organized as follows: Section 2 discusses the background and estimation of the market invariance hypothesis. Section 3 discusses the institutional details of the Tokyo Stock Exchange. Section 4 reports the data and descriptive statistics. Section 5 presents our main results and Section 6 concludes.

2 Background and estimation of the market invariance hypothesis

2.1 Adapting the invariance-of-bet hypothesis into transaction data

The market microstructure invariance-of-bet hypothesis originates from the intuition that when market participants trade financial assets, their primary objective is to shift risks among them (Kyle and Obizhaeva, 2016). Investment decisions (bets) lead to trading. Bets take place in business time. The invariance-of-bet hypothesis suggests that trading these bets is similar when it is measured by the fiat-currency risks transferred per unit (bet) of business time, that is, by definition the product of the fiat-currency risks and the business time for one bet to complete. For example, more volatile assets are usually associated with more trading volume (Jones et al., 1994; Downing and Zhang, 2004; Giot et al., 2010; Wang and Wu, 2015), in other words, bets are taking place at a faster speed. As a result, per unit (bet) of business time is smaller for these volatile assets. In contrast, trading occurs at a slower speed for less volatile assets, and thus per unit (bet) of business time is larger for these less risky assets. Therefore, the fiat-currency risks transferred per unit (bet) of business time, which is, by definition, the fiat-currency volatility time per unit (bet) of business time, should be close to each other for both high-volume-high-volatility and low-volume-low-volatility financial assets. Kyle and Obizhaeva's invariance-

of-bet hypothesis extends this straightforward concept to a more general argument that *the distribution of the fiat-currency risk transferred by a bet is in units of business time is the same across assets and time.*

Appendix A reviews the mathematical representations of the invariance-of-bet hypothesis and its testable implications in detail. Kyle and Obizhaeva (2016) finds the following equations hold:

$$\gamma_{it} = E[\frac{\zeta_{it}}{2}|\tilde{I}_{it}|]^{-2/3} W_{it}^{2/3} \tag{1}$$

$$\frac{|\tilde{Q}_{it}|}{VLM_{it}} \stackrel{d}{=} E[\frac{\zeta_{it}}{2}|\tilde{I}_{it}|]^{-1/3} W_{it}^{2/3} |\tilde{I}_{it}| \tag{2}$$

For stock i trading on day t, γ_{it} represents the expected arrival rate of bets, $\tilde{Q}_{it}$ is the size of bets, ζ_{it} represents the volume multiplier which relates to the intermediary activities, and $\tilde{I}_{it}$ is the market microstructure invariant that equals to the distribution of a random variable $\tilde{I}$. In other words, the realized value of $\tilde{I}_{it}$ at the same percentile of the risks is a constant I across stocks and time. W_{it} is the trading activities defined as the product of the fiat-currency volume and the volatility.

We cannot test the scale-law implications from the invariance theories directly using Equation 1 and Equation 2 with transaction data because the unobservable bets are not equivalent to the observable trades. In modern electronic trading, one bet, say a portfolio manager's decision to buy 5,000 shares of a stock, usually results in multiple small 100-share trades. As documented in various literature (Bertsimas and Lo, 1998;

Huberman and Stanzl, 2005; Obizhaeva and Wang, 2013), the optimal trading strategies, which aim at minimizing the trading costs, will break large parent bets into small child orders. Following Kyle and Obizhaeva (2016) and Kyle et al. (2020), we assume that the distribution of trade sizes $\tilde{X}_{it}$ is the same as the distribution of bet sizes $\tilde{Q}_{it}$ divided by an order-splitting multiplier δ_{it} for each stock i at day t. In this setting, each large parent order is on average splitted into δ_{it} orders. As a result, we can obtain the relation between the (expected) number of trades NT_{it} and the expected bet arrival rate γ_{it} along with volume multiplier ζ_{it} as follows:

$$NT_{it} = \delta_{it}\frac{\zeta_{it}}{2}\gamma_{it} \tag{3}$$

Substituting γ_{it} from Equation 3 into Equation 1 and Equation 2, we obtain directly testable implications for the relation between the trading activity and the number of trades, and the trade sizes using transactions data. These implications are:

$$NT_{it} = \mu_0 W_{it}^{\alpha_1} \tag{4}$$

$$\frac{|\tilde{X}_{it}|}{VLM_{it}} \overset{d}{=} \mu_x W_{it}^{\alpha_x} \tag{5}$$

where the invariance-of-bet hypothesis implies $\alpha_1 = -\alpha_x = 2/3$, $\mu_0 = \delta_{it}\frac{\zeta_{it}}{2}E[\frac{\zeta_{it}}{2}|\tilde{I}|]^{-2/3}$ and $\mu_x = \delta_{it}^{-1}E[\frac{\zeta_{it}}{2}|\tilde{I}|]^{-1/3}$.

By taking the natural logarithm of the Equation 5, we can run a simple OLS regression using the following equation:

$$Ln(NT_{it}) = Ln(\mu_0) + \alpha_1 Ln(W_{it}) + \varepsilon_{it} \tag{6}$$

2.2 Possible measurement errors in the number of trades

While estimating Equation 6 is straightforward, the dependent variable—the number of trades—may contain measurement errors, especially in the U.S. equity markets.

Table 1 illustrates the origin of the measurement errors. Panel A and Panel B report the trades and quotes prints for the Japanese markets and the U.S. equity markets, respectively. We select the stock of Kyokuyo Co Ltd (Ticker: 1301) and the stock of Apple Inc (Ticker: AAPL) on two random trading days for illustrative purposes. The trades and quotes prints are presented in chronological order.[5] In the case of AAPL, the series of trade prints starting from 09:30:470520000 to 09:30:470626304, might be a buyer-initiated marketable order with a possible limit price of 177.99 USD and a size of 1,300 shares. Similarly, in the case of 1301, the last two trade prints originate from a seller-initiated marketable order with a possible limit price of 4240 JPY. Measurement errors occur if we simply count the number of trade prints as the measure of the true number of trades (NT_{it}). For instance, the buyer-initiated trade in AAPL results in eleven trade prints, and the seller-initiated trade in 1301 generates two trade prints.

[Table 1]

To understand the possible effects of such measurement errors on the estimated co-

[5]The timestamps in this example are participant timestamps for the U.S. equity markets.

efficient, α_1, from the OLS regressions using Equation 6, we denote the true number of trades as NT. For illustrative purposes, we subsume the subscripts of i and t unless explicitly specified. The calculated number of trades by simply counting the number of trade prints might be upward-biased due to the measurement errors as illustrated in Table 1. If we assume there exists a stock-day multiplicative measurement error, e, for the calculated number of trades (NT^*), then its relation with the true number of trades (NT) can be written as:

$$NT^* = e \times NT \quad \text{where} \quad e > 1 \tag{7}$$

By taking the natural logarithm on both sides and plugging it into Equation 6, we can obtain:

$$Ln(NT^*) = Ln(\mu_0) + \alpha_1 Ln(W) + \varepsilon + Ln(e) \tag{8}$$

If we run the OLS regression using the upward-biased NT^*, then the asymptotic bias in the estimated coefficient α_1 is:

$$\text{plim} \quad \hat{\alpha_1} - \alpha_1 = \frac{Cov(Ln(W), \varepsilon + Ln(e))}{Var(Ln(W))} \tag{9}$$

Even if we assume the classical OLS assumption—the error term ε is not correlated with the independent variable $Ln(W)$, the estimated coefficient is upward-biased as long as the multiplicative measurement error, e, is positively correlated with the trading activity $Ln(W)$.

As shown in the Panel B, Table 1, the occurrence of measurement errors arises from 1) the order routing, 2) hidden liquidity, and 3) the differences in the trade reporting systems among different exchanges as first documented by Upson et al. (2021). The probability of occurrence of any of the three abovementioned events is, arguably, higher for actively traded stocks than illiquid stocks. Thus, we anticipate that in the U.S. equity markets, $Cov(Ln(W), Ln(e)) > 0$. Mathematically, the degree of such bias should depend on the ratios of the variance of trading activity and the magnitude of the covariance between the trading activities and the measurement errors:

$$\text{plim}(\hat{\alpha_1}) = \alpha_1 + \frac{\overbrace{Cov(Ln(W), Ln(e))}^{>0}}{\underbrace{Var(Ln(W))}_{>0}} \tag{10}$$

Kyle et al. (2020) test the invariance hypothesis using TAQ data from 1993 to 2014 based on Equation 8 using the number of trades as the counting of the number of trades prints. While they find the estimated coefficients α_1 are close to the theoretical values of two-thirds from 1993 to 2001, the estimated coefficients in the post-decimalization era are much higher than the expected values of two-thirds.[6]

We argue that such measurement errors may be less of a concern in a centralized order-driven market such as the Tokyo Stock Exchange. First, there is no order routing among the trading venues.[7] Thus, the trades prints that we observe in the NEEDS data

[6] Kyle et al. (2020) attribute these deviations to market frictions related to institutional details such as minimum tick size, minimum lot size, and technology improvements since these market frictions influence the magnitude of the order-splitting multiplier and the volume multiplier in various ways. For example, changes in tick size may have affected trading decisions by changing the incentives to provide liquidity and, thus, possibly affecting the order-splitting and volume multipliers. Smaller lot sizes may encourage order-splitting. Technology improvements may also help traders to split parent orders more efficiently.

[7] The official off-exchange trading venue is the Tokyo Stock Exchange Trad-

are more likely to be the true trades resulting from marketable orders. Second, we contact the TSE and confirmed that the trading report system employed by the TSE is consistent throughout our sample period from 2015 to 2020. Third, for a large marketable order walking the book, we can easily reconstruct the trade by aggregating the trade prints with the same timestamp.[8] Besides, we can also construct hypothetical measurement errors using TSE data. Section 5.2. will show that the estimated coefficients are upward-biased towards the theoretical values of two-thirds if we impose hypothetical measurement errors on the true number of trades.

3 Institutional details of the Tokyo Stock Exchange

The TSE is one of the largest equity markets in the world based on market capitalization. By the end of March 2021, there were 3,760 companies listed on the TSE with a total market capitalization of 6.736 trillion USD. The first section, which comprises 2,186 domestic companies, trades most liquid stocks. The second section comprises 473 less active companies. Other sections include Mothers, JASDAQ, JASDAQ growth, and Tokyo PRO trade venture stocks or alternative investment stocks.

The TSE is a pure order-driven market with no official registered market maker. Therefore, we believe that it is reasonable to assume that the volume multiplier, ζ_{it}, is a constant across stocks and time. The TSE has a unique tick size rule. The minimum

ing NeTwork System (ToSTNeT Market) which conducts different types of trading (https://www.jpx.co.jp/english/equities/trading/tostnet/01.html).

[8] In this paper, we use both aggregated NT and simple approach—counting the number of trade prints—NT to estimate the coefficients α_1 as in Equation 6, the results are quantitatively similar.

tick size is determined by the stock's price on the TSE. When a stock's price crosses the price threshold, for example, 3000 JPY, the minimum tick size either increases from 1 JPY to 5 JPY or decreases from 5 JPY to 1 JPY.

The TSE also has a unique lot size regime. As early as 2007, the TSE started a lot-size unification regime that required stocks listed in the first section and the second section to unify their minimum lot size from eight different minimum lot sizes to a minimum lot size of 100 shares.[9] The lot-size unification regime ended in October 2018. Each firm decided when to unify its minimum lot size. Each firm split its stock simultaneously if it wanted to maintain the JPY volume for one lot when unifying its minimum lot size.

All stocks listed on the TSE employ the *Level Book* reporting system. Upson et al. (2021) define a reporting system as *Level Book* reporting if the number of trades reported is based on the size of the marketable order at a specific price regardless of the number of limit orders the marketable order executes against. In contrast, for *Order Based* reporting, the number of trades reported for a markable order depends on how many limit orders the marketable order executes against. In the U.S., NASDAQ-listed stocks employ the *Order Based* reporting while NYSE-listed and AMEX-listed stocks employ *Level Book* reporting.

[9]The unification regime involves two steps. The first step unified eight different minimum lot sizes—1-share, 10-share, 50-share, 100-share, 200-share, 500-share, 1000-share, and 2000-share to either a minimum lot size of 100-share or 1000-share. By the end of March 2014, 2257 firms and 1,124 firms had unified their minimum lot size to 100-share and 1000-share which accounts for 66.4% and 33.0% of all stocks listed on the TSE, respectively. By the end of March 2018, 94.2% of stocks had unified their minimum lot size to 100-share.

4 Data, and descriptive statistics

4.1 Sample

We use the tick-by-tick data from the Nikkei Economic Electronic Databank System (NEEDS), which comprises trades and quotes for all stocks listed on the first and second sections of the TSE. Our sample period is January 5, 2015, through December 30, 2020. For each quote record, the dataset contains the trading day, ticker code, exchange, and sections, timestamps to the microsecond, and the ten best bid and ask prices, along with their sizes and quote condition codes. For each trade record, the dataset contains additional trading information such as the trade execution price, number of shares traded, and trade condition code.

We drop exchange-traded funds and real estate investment trusts. We also drop stocks in TOPIX 100 which employ a different tick size rule compared with other stocks listed on the TSE, see Table A.1 for details. Following Aşçıoğlu et al. (2007), we drop 7,720 stock-day observations for stocks that trade at more than one minimum tick size during a given trading day. We use daily stock price data from Compustat Global to compute daily stock volatility.[10] We include firms in both NEEDS and Compustat Global databases and drop three firms not in both. Our final sample comprises at least 2,270 stocks and 3,609,547 stock-day observations with a mean of 2,467 stocks for each trading day. For each trading day t each stock i, we compute the following variables—NT, the total number of trades;

[10]The stock prices in Compustat Global are adjusted by stock splits and dividends. We use this adjusted price to calculate daily stock volatility.

NT (Sell), the number of sell-initiated trades; NT (Buy), the number of buy-initiated trades;[11] $\bar{P}$, the mean of trade prices;[12] SIZ, the mean of trade sizes; VLM, the daily number of shares traded; $VOLJ$, the mean of trade prices in JPY times the number of shares traded divided by 100. Using data from Compustat Global, we calculate return volatility σ as the standard deviation of squared returns over the past 20 trading days.[13]

Following Kyle et al. (2020), we place each trade into one of the 55 bins constructed based on their trade sizes. Suppose a trade has a size of X shares, "even" bins correspond to trade sizes to the following shares: 100; 200; 300; 400; 500; 1,000; 2,000; 3,000; 4,000; 5,000; 10,000; 15,000; 20,000; 25,000; 30,000; 40,000; 50,000; 60,000; 70,000; 75,000; 80,000; 90,000; 100,000; 200,000; 300,000; 400,000; and 500,000. Odd bins correspond to trade sizes between the adjacent even bins.[14] Three statistics are computed based on these bins: $L100\%$, the percentage of trades in 100-share lots; $L1000\%$, the percentage of trades in 1,000-share lots; and $LOTE\%$, the percentage of trades in the even lots. Also, we calculate both trade- and volume-weighted percentiles of trade sizes. For example, the pth trade-weighted percentile corresponds to a trade size such that trades with size below this threshold constitute $p\%$ of all trades for stock i at day t. SIZ^{TW50} is the trade-

[11] The TSE is a pure-order driven market and except for open trades, close trades and trades outside of regular trading hours, sell-initiated trades are always executed at the bid prices and buy-initiated trades are always executed at the ask prices. We exclude open trades and close trades for both morning and afternoon sessions when we calculate the trades for sell/buy-initiated trades.

[12] For our empirical analysis, we replace P_{it} with $\bar{P}_{it}$ in computing the trading activity defined as in Equation 3. The trading activity in our empirical analysis is $W_{it} := \sigma_{it}\bar{P}_{it}VLM_{it}$.

[13] As a robustness check, we also calculate Parkinson (1980)'s volatility for stock i at trading day t as $Parkinson_{it} = \sqrt{\frac{1}{4Ln2}\frac{1}{20}\sum_{t'=t-20}^{t-1} Ln(\frac{H_{it'}}{L_{it'}})}$ where $H_{it'}$ is the highest price and $L_{it'}$ is the lowest price for stock i at trading day $t^{'}$.

[14] However, different from Kyle et al. (2020), we do not assign the midpoint size of the bin to the trades belonging to that bin.

weighted median (50th percentile of) trade sizes and SIZ^{VW50} is the volume-weighted median (50th percentile of) trade sizes. Appendix Table A.1 presents the definitions of notations and variables.

4.2 Descriptive statistics

Table 2, Panel A, presents descriptive statistics for all stock-day observations for January 5, 2015, through December 30, 2020. The mean number of trades (NT) is about 749, with 375 buyer-initiated trades and 370 seller-initiated trades. The mean of the trade prices (P) is 1,998 JPY, trade size (SIZ) is 1,366 shares, volatility (σ) is 1.9%, and daily JPY volume ($VOLJ$) is 9.609 million. All these statistics are highly right-skewed, suggesting small-size, small-price trades and small-volume stocks dominate the market. The difference between the median and the mean of the number of trades is 536, indicating that some stocks have extremely high numbers of trades. The mean of the percentage of trades in 100-share ($L100\%$), 1,000-share ($L1000\%$), and even ($LOTE\%$) lots are 51.6%, 9.5%, and 90.9%, comparable to the statistics from TAQ data after 2001 (Kyle et al., 2020).

Table 2, Panel B, reports the means for our seven tick size groups. Tick size groups are based on the minimum tick size of 1, 5, 10, 50, 100, 500, and 1,000 JPY. Each stock-day observation falls into one of these seven groups. Panel C reports the means for ten volume groups. We calculate the mean daily JPY volume, $VOLJ$, for each stock during our six-year sample periods and set the breakpoints of volume groups at 30th, 50th, 60th, 70th, 75th, 80th, 85th, 90th, and 95th percentiles. We assign each stock into one of these

ten volume groups just as in (Kyle et al., 2020).[15]

As shown in Table 2, Panel B, about 82.9% of stock-day observations fall into the tick-size-1 group, comparable to the numbers in the study of Kemme et al. (2021). These authors report the following means: the number of trades (990), the trading volume (1,187,000) and the market depth (30,500). In our paper, most of the tick-size-1 stocks are low in volume and price and are large in trade size. The larger minimum tick size groups, such as the tick-size-50 group and the tick-size-100 group, comprise high-volume, high-price, small-trade-size stocks. For example, the mean of the number of trades for the tick-size-50 group is about 664, which is a factor of $664/519 \approx 1.28$ to the tick-size-1 group. The increase in the number of trades is associated with an increase in JPY volume. The mean daily JPY volume for the tick-size-50 group is about 52.70 million. Thus, the trading activity, which is defined as the product of JPY volume and volatility, increases from 3.911×0.020 to 52.70×0.016, which is approximately by a factor of 10.8. If this factor is based on the prediction from the invariance hypothesis that the (increase in) the expected number of trades is proportional to the two-thirds of the (increase in) trading activity, then the number of trades should increase by a factor of 10.8 to the power of 2/3 which is approximately 4.88 from the tick-size-1 group to the tick-size-50 group. The fact that the number of trades only increases by a factor of 1.28 from the tick-size-1 group to the tick-size-50 group, a value less than predicted, suggests there are potentially more order-splitting activities for stock-day observations in the tick-size-1 group than those in the tick-size-50 group.

[15]We also sort stocks into ten deciles based on average daily JPY volume, $VOLJ$, for the purpose of robustness checks.

Table 2, Panel C, reports the means for ten-volume groups categorized by stocks sorted by their in-sample mean VOLJ percentiles–(1) $VOLJ < 30^{th}$, (2) $30^{th} <= VOLJ < 50^{th}$, (3) $50^{th} <= VOLJ < 60^{th}$, (4) $60^{th} <= VOLJ < 70^{th}$, (5) $70^{th} <= VOLJ < 75^{th}$, (6) $75^{th} <= VOLJ < 80^{th}$, (7) $80^{th} <= VOLJ < 85^{th}$, (8) $85^{th} <= VOLJ < 90^{th}$, (9) $90^{th} <= VOLJ < 95^{th}$, (10) $95^{th} <= VOLJ < 100^{th}$. Volume group 1 has the smallest $VOLJ$. The mean number of trades increases from about 58.4 for the lowest volume group to 5,261 for the highest volume group by a factor of $5,261/58.4 \approx 90.09$. And the mean daily JPY volume also increases from 0.161 million for volume group 1 to 109.0 million for volume group 10. Thus, the trading activity increases from 0.161×0.016 to 109.0×0.018—a factor of 761.6 from the lowest to the highest volume group. The increase (by a factor) in the number of trades predicted by the invariance-of-bet hypothesis from the lowest volume group to the highest volume group is about 83.4 is very close to the actual increase (by a factor of 90.1) in the number of trades.

In Table 2, Panel A and Panel B, show that the means of the percentage of trades in $L100\%$ lot, $L1000\%$ lot, and $LOTE\%$ are different across groups. Despite differences in market microstructures and the definition of odd-lot trade, evidence from previous studies (Bertsimas and Lo, 1998; Obizhaeva and Wang, 2013; Johnson et al., 2017), suggest that the odd-lot trades are a result of orders broken into multiple trades, and the optimal trading strategy usually breaking down the large orders into numerous non-100-share trades. Thus, we conjecture the differences in the percentage of $L100\%$, $L1000\%$ and $LOTE\%$ (odd-lot) are potentially a reflection of different intensities of order-splitting activities across volume and tick size groups but further investigation is needed.

[**Table 2**]

5 Estimating the two-thirds scaling laws

5.1 Pooled Regression Results

Table 3 presents the pooled least square estimates from the regressions of the log of the number of trades on the log of trading activity as in Equation 6. Columns 1 to 3 report the coefficients from the OLS estimates for all stock days. Column 4 reports the coefficients from the panel within estimators using day-fixed effects. Column 5 reports the coefficients from the Fama-MacBeth estimator. The sample period is from January 5, 2015, through December 30, 2020. We report robust standard errors clustered at the stock level (column 1, 2, 3, 4). Newey-West standard errors are calculated with five lags for the Fama-MacBeth regression in column 5. The point estimates of α_1, which range from 0.665 to 0.671, are close to the theoretical value of 2/3 implied by the invariance-of-bet hypothesis. Columns 2 and 3 show that when using the sell-initiated trades and buy-initiated trades as the dependent variables to run the same OLS regressions, the coefficients are closer to the theoretical value of 2/3 than those estimated using all trades. In both buy-initiated-trade and sell-initiated-trade regressions, the null hypothesis that the coefficient estimates of $\alpha_1 = 2/3$ cannot be rejected at the 1% level of significance. The Fama-MacBeth estimator in column 5 produces results similar to the OLS estimation in column 1.

In the untabulated table, the differences in estimated coefficients α_1 are primarily driven by stock-fixed effects rather than time-fixed effects. When adding stock-fixed effects to the OLS model, the estimated coefficient α_1 decreases from 0.669 to 0.580 by 13.3%. We conjecture that market frictions affect the invariance-of-bet hypothesis differentially across stocks but not that much over time because cross-sectional frictions, such as tick size and lot size, are more likely to influence the validity of the model's identification assumptions than time-series frictions like macroeconomic conditions. The regression coefficients estimated using seller-initiated trades (column 2) and buyer-initiated trades (column 3) are 0.665 and 0.665, respectively. Both coefficients are not significantly different from 2/3 at the 1% level. These results suggest that excluding trades from open and closing auctions may reduce the noise in our coefficient estimates. In addition, we employ the Wald test of Clogg et al. (1995) to test the equality of the coefficients α_1 in the regressions from column 2 and column 3. The z-statistics is 0.161, suggesting that the estimated α_1 from seller-initiated-trade regression is not significantly different at 1% level from the α_1 estimated using buyer-initiated-trade regression. Our findings are interesting because prior studies (Kraus and Stoll, 1972; Holthausen et al., 1987, 1990; Chan and Lakonishok, 1993; Gemmill, 1996; Keim and Madhavan, 1996; Saar, 2001) on block trading find that there exists information asymmetry between the trades initiated by the buyers and the traders initiated by the sellers. However, our results suggest the invariance-of-bet hypothesis is barely affected by the differences in information content between the buyer-initiated and seller-initiated trades.

[**Table 3**]

Table 4 presents the pooled least square estimates from the regressions of the log of the number of incoming non-marketable orders on the log of trading activities based on these orders, $Ln(NO_{it}^{\text{Aggressiveness}}) = \alpha_0 + \alpha_1 Ln(\frac{W_{it}}{W^*}) + \tilde{\epsilon}_{it}$. We sort the incoming non-marketable orders for each stock on each trading day into five different groups based on their price aggressiveness (Biais et al., 1995). Column 1 reports the results based on the most aggressive incoming non-marketable orders. Column 5 reports the results of the least aggressive non-marketable orders. Panel A reports the estimates from the sell non-marketable orders on the ask side, while Panel B reports the buy non-marketable orders on the bid side. The pooled least square estimates from the regressions of the number of orders on trading activities are close to 2/3 across different price aggressiveness groups ranging from 0.616 to 0.683 for the sell incoming non-marketable orders and ranging from 0.614 to 0.686 for the buy incoming non-marketable orders.

[**Table 4**]

Figure 1 shows the time-series coefficients of the number of trades on trading activity regressions. For each day t, we estimate the following cross-sectional regression $Ln(NT_{it}) = \alpha_0 + \alpha_{1t} Ln(\frac{W_i}{W^*}) + \varepsilon$, and obtain estimated α_{1t} from January 5, 2015, to December 30, 2020. The estimated coefficients are oscillating around 2/3, with the multiple peaks approaching around 0.750 and troughs going at 0.625. Among these peaks, one occurs in March 2020, where the coefficient jumps from 0.668 to 0.775, a 16% increase. We inspect the stationary of the estimated time series by conducting the Augmented

Dickey-Fuller test with no drift, no time trends, and five lags (Dickey and Fuller, 1979). The test statistic value is -0.207, which is larger than the 5% and the 1% critical value, -1.950 and -2.580, suggesting that the time series is stationary and exhibits reversion to the mean 0.671. While beyond the scope of this paper, macroeconomic conditions may also serve as frictions that may have induced estimation bias on the estimated coefficients.

[**Figure 1**]

Figure 2 plots the histograms of the time-series and cross-sectional coefficients of regressions of the number of trades on trading activity. The time-series coefficients are regression coefficients obtained from 1,461 daily regressions of the number of trades on trading activity across stocks. The cross-sectional coefficients are regression coefficients obtained from 2,834 individual stock time-series regressions of the number of trades on trading activity during our sample period. The distribution for cross-sectional coefficients is almost symmetric. It is centered around 0.6, but the time-series coefficients are more concentrated and are centered at around 0.666, closer to the theoretical value implied by the invariance-of-bet hypothesis. These results also suggest that individual stocks are more prone to market frictions that cause the estimated coefficients to deviate from 2/3 than the whole market over time.

[**Figure 2**]

Our results are, in general, consistent with the previous studies using portfolio transitions data, TAQ data, and trading data from Korea Exchange (Bae et al., 2016; Kyle and Obizhaeva, 2016; Kyle et al., 2020). Kyle et al. (2020) show the estimated coefficient

α_1 from a time-trend, Fama-Macbeth regression of the log of the number of trades on the log of trading activity is 0.666 for the U.S. stocks in the 1994-2001 sample period. Their monthly-based regressions also show that there is an increase in the estimated coefficient α_1 for the second half of their sample period. Bae et al. (2016) find the coefficient estimated from a regression of the log of the number of switching points on the log of trading activity is 0.675 for Korean stocks. In addition, Kyle and Obizhaeva (2016) find that the estimated coefficients α_{1x} from regressions of the log of the portfolio transition order sizes, expressed as a fraction of daily volume on the log of trading activity are -0.63 for NYSE buy orders, -0.71 for NASDAQ buy orders, and -0.59 for both NYSE and NASDAQ sell orders.

5.2 Hypothetical measurement errors

Section 2.2 discusses the issues of measurement errors in the number of trades when using the U.S. data. In this section, we impose hypothetical measurement errors on the key dependent variable, the number of trades using our TSE data.

Specifically, we impose four different levels of measurement errors on the true number of trades. We denote them as e_5, e_{10}, e_{20}, e_{40}. For each level of measurement error, we impose a pseudo-data-generating process that will report a large trade, e.g. a trade with more than 1000 shares, as one or more trades with equal probability. For instance, a real trade with 1000 shares under the pseudo-data-generating process of the level of e_5 measurement errors, could be reported as the following 5 scenarios with equal probability: 1 trade with 1000 shares, 2 trades with 500 shares, 3 trades with 333.33 shares, 4 trades

with 250 shares, and 5 trades with 200 shares. The subscripts in e represent the largest possible trades that will be reported for that level of measurement errors.

After constructing the hypothetical number of trades with measurement errors, we run the same OLS regression as in Column 1 of Table 3. Table 5 reports the pooled least estimates of Equation6 in Table 3 with different measures of the number of trades. Column 1 reports the estimated coefficient using the true number of trades as in Column 1 of Table 3. Column 2 to column 5 report the estimated coefficients using the number of trades with measurement errors generated from the pseudo-data-generating processes.

The results in column 2 to column 5 confirm our conjecture that the estimated coefficients will be upward biased if the measures of the number of trades contain measurement errors. The estimated coefficients for the number of trades with measurement errors range from 0.706 to 0.757. The magnitude of the biases is larger when the measurement errors in the number of trades (Mean of NT) are larger. These results also indicate that using the number of trades from the TSE might resolve the issues of measurement errors in the number of trades compared with using the transaction data in the U.S. (TAQ).

[Table 5]

5.3 Tick size constraints

This section discusses the estimated coefficients and the trade size distributions across different tick size constraint groups. Following O'Hara et al. (2019), we sort stocks into quartiles based on their relative tick sizes, defined as the average of bid-ask spread divided

by tick size. The cut-off values for the relative tick sizes are 1.61 ticks, 2.52 ticks, and 4.35 ticks, respectively.

Figure 3 reports the boxplot of the cross-sectional estimated coefficients of regressions of the number of trades on trading activity. We sort the stocks into quartiles based on their relative tick sizes. Q1 represents the most tick-constrained stocks where the relative tick sizes of the stocks in this quartile is smaller than 1.61. Q4 represents the least tick-constrained stocks where the relative tick sizes of the stocks in this quartile is larger than 4.35 ticks. Figure 3 suggests that stocks of different relative tick sizes might have different mean estimated coefficients. More stocks in the q4 relative tick size quartile have estimated coefficients close to two-thirds.

[**Figure 3**]

We further investigate the trade size distribution for these stocks across different relative tick size quartiles. Following the approach in Figure 2 of Kyle and Obizhaeva (2016), Figure 4 shows the distributions of $Ln(\tilde{X}/V) + 2/3Ln(W/W^*)$ for each relative tick size quartile. We find that the scaled trade sizes across different relative tick size quartiles approximately follow normal distributions, with means close to -7.5 to -8.5 and variances close to 2 to 3. These results suggest the trade sizes of stocks in the Tokyo Stock Exchange are approximately close to log-normal distributions as implied by the invariance-of-bet hypothesis.

[**Figure 4**]

5.4 External validity: The Euronext markets

Table 6 reports the pooled least square estimates from the regressions of the log of the number of trades on the log of trading activity for the Euronext equity markets which consist of the following six stock exchanges: BE (Brussels Stock Exchange), FR (Paris Stock Exchange), IR (Irish Stock Exchange), NL (Amsterdam Stock Exchange), NO (Oslo Stock Exchange), and PT (Euronext Lisbon). The sample period for the Euronext markets is from January 4, 2021, to Dec 31, 2021. We exclude stock-day observations with the number of trades less than 5. We report the descriptive statistics in Appendix B.

The estimated coefficients on α_1 range from 0.698 (FR) to 0.770 (BE), suggesting that the invariance hypothesis also holds in European markets. Though the estimated coefficients are slightly above two-thirds, the coefficients are still much closer to two-thirds compared with the second half of the sample period (2001 to 2014) in Kyle et al. (2020). These results, in contrast with the results in TSE, also suggest that measurement errors in the number of trades might also exist in the European markets if we simply count the number of trade prints as the estimates for the number of trades.

[Table 6]

5.5 Trade sizes and trading activity

Table 7 presents the pooled least square estimates from the regressions of log of the trade-weighted or volume-weighted quantiles of trade sizes on the log of trading activity based on Equation 5, f(:) denotes the functional form of the percentile distribution transformations. The left-hand side variables are the natural logarithm of the mean trade sizes, or the 20th, 50th, 80th percentiles of the distributions of logarithms of the volume-weighted and trade-weighted trade sizes, in turn, expressed as a fraction of daily volume for stock i on day t. Column 1 reports the coefficients using the mean trade sizes as the dependent variable. Columns 2 to 4 report the coefficients using the 20th, 50th, and 80th percentiles of volume-weighted trade sizes as the dependent variables. Columns 5 to 7 report the coefficients using 20th, 50th, and 80th percentiles of trade-weighted trade sizes as the dependent variables, in turn.

Table 7, Column 1 reports the results of estimates using the log of the mean trade sizes as the dependent variable, the estimated coefficient α_{1x} is -0.586, which is not far from the theoretical value of -2/3 implied by the invariance-of-bet hypothesis. For the volume-weighted distributions, the estimated coefficients α_{1x} are -0.597, -0.526, and -0.469 from the regressions using the 20th, 50th, and 80th percentiles of trade sizes as the dependent variables, respectively. For the trade-weighted distributions, the estimated coefficients are -0.673, -0.639, and -0.561, respectively. The absolute values of estimated coefficients α_{1x} decrease as the percentiles increase for both distributions implying that trade sizes, as a fraction of daily volume, tend to decrease slower with trading activity.

The estimated coefficients α_{1x} using trade sizes are highly correlated with the percentile of the distributions. In both trade-weighted and volume-weighted specifications, a smaller percentile of the distribution leads to a larger absolute value of the coefficient α_{1x}.

The estimated intercept α_{0x} from the regressions of the mean trade sizes on trading activity, -5.017, implies that the average of mean trade sizes of the benchmark stock is equal to $e^{-5.017}$ or about 0.66% of the daily volume. The estimated intercepts of -5.517, -4.504, and -3.042 for the volume-weighted percentiles imply that the means of the 20th, 50th, and 80th percentiles of trade sizes are about 0.38%, 1.11%, and 4.77% of the daily volume for the benchmark stock. Similarly, for trade-weighted distributions, the means of the 20th, 50th, and 80th percentiles of trade sizes are about 0.24%, 0.30%, and 0.73% of the daily volume for the benchmark stock. These results are not surprising because as Table 2, Panel A, shows that the mean of the 50th percentile of volume-weighted trade sizes (25,620 shares), SIZ^{VW50}, is 74.7 times the mean of the 50th percentile of trade-weighted trade sizes (343 shares), SIZ^{TW50}. Therefore, if both volume-weighted and trade-weighted distributions are assumed to be log-normal distributions with similar variances, then for the same percentile the mean of volume-weighted trade sizes will account for a larger proportion of the daily volume than the mean of trade-weighted sizes.

For the same quantile, the adjusted R-squared is higher in regressions using trade-weighted quantiles of trade sizes as dependent variables than in regressions using volume-weighted trade sizes. For the regressions using volume-weighted quantiles as the dependent variables, the values of the adjusted R-squared decrease monotonically from 0.614

for the 20th percentiles to 0.217 for the 80th percentiles. The values of the adjusted R-squared in regressions using trade-weighted quantiles tell a similar story. These figures suggest that the invariance-of-bet hypothesis explains more variation to small trade sizes than in large trade sizes. For the regression using the mean trade sizes as the dependent variable, trading activity explains about 71.5% of the variation in the trade sizes expressed as a fraction of the daily volume.

[**Table 7**]

6 Conclusions

Using the transaction data from one of the largest pure-order driven markets - the Tokyo Stock Exchange (TSE), we test the invariance-of-bet hypothesis proposed by Kyle and Obizhaeva (2016). Our estimated coefficients from the regressions of the number of trades on trading activity defined as the product of the fiat-currency volume and volatility are close to the theoretical value of two-thirds implied by the invariance hypothesis.

Compared with the TAQ data, using the TSE transaction data will reduce the measurement errors in variables such as the number of trades and trade sizes. Therefore, our results suggest what might cause the coefficients in the second half of Kyle et al. (2020)'s sample period to deviate from two-thirds.

Our paper shows that empirically testing the invariance hypothesis is challenging. It requires researchers to carefully assess the appropriateness of the market setting and choose suitable measures for the variables corresponding to the invariance theories.

References

A. Aşçıoğlu, C. Comerton-Forde, and T. H. McInish. Price clustering on the tokyo stock exchange. *Financial Review*, 42(2):289–301, 2007.

K.-h. Bae, A. S. Kyle, E. J. Lee, and A. A. Obizhaeva. Invariance of buy-sell switching points. *Working Paper (October 26, 2016)*, 2016.

D. Bertsimas and A. W. Lo. Optimal control of execution costs. *Journal of Financial Markets*, 1(1):1–50, 1998.

B. Biais, P. Hillion, and C. Spatt. An empirical analysis of the limit order book and the order flow in the paris bourse. *the Journal of Finance*, 50(5):1655–1689, 1995.

F. Bucci, F. Lillo, J.-P. Bouchaud, and M. Benzaquen. Are trading invariants really invariant? trading costs matter. *Quantitative Finance*, 20(7):1059–1068, 2020.

L. K. Chan and J. Lakonishok. Institutional trades and intraday stock price behavior. *Journal of Financial Economics*, 33(2):173–199, 1993.

C. C. Clogg, E. Petkova, and A. Haritou. Statistical methods for comparing regression coefficients between models. *American Journal of Sociology*, 100(5):1261–1293, 1995.

D. A. Dickey and W. A. Fuller. Distribution of the estimators for autoregressive time series with a unit root. *Journal of the American statistical association*, 74(366a):427–431, 1979.

C. Downing and F. Zhang. Trading activity and price volatility in the municipal bond market. *The Journal of Finance*, 59(2):899–931, 2004.

G. Gemmill. Transparency and liquidity: A study of block trades on the london stock exchange under different publication rules. *The Journal of Finance*, 51(5):1765–1790, 1996.

P. Giot, S. Laurent, and M. Petitjean. Trading activity, realized volatility and jumps. *Journal of Empirical Finance*, 17(1):168–175, 2010.

R. W. Holthausen, R. W. Leftwich, and D. Mayers. The effect of large block transactions on security prices: A cross-sectional analysis. *Journal of Financial Economics*, 19(2): 237–267, 1987.

R. W. Holthausen, R. W. Leftwich, and D. Mayers. Large-block transactions, the speed of response, and temporary and permanent stock-price effects. *Journal of Financial Economics*, 26(1):71–95, 1990.

G. Huberman and W. Stanzl. Optimal liquidity trading. *Review of Finance*, 9(2):165–200, 2005.

H. Johnson, B. F. Van Ness, and R. A. Van Ness. Are all odd-lots the same? odd-lot transactions by order submission and trader type. *Journal of Banking & Finance*, 79: 1–11, 2017.

C. M. Jones, G. Kaul, and M. L. Lipson. Information, trading, and volatility. *Journal of Financial Economics*, 36(1):127–154, 1994.

D. B. Keim and A. Madhavan. The upstairs market for large-block transactions: Analysis and measurement of price effects. *The Review of Financial Studies*, 9(1):1–36, 1996.

D. M. Kemme, T. H. McInish, and J. Zhang. Market fairness and efficiency: Evidence from the tokyo stock exchange. *Journal of Banking & Finance, forthcoming*, 2021.

A. Kraus and H. R. Stoll. Price impacts of block trading on the new york stock exchange. *The Journal of Finance*, 27(3):569–588, 1972.

A. S. Kyle and A. A. Obizhaeva. Market microstructure invariance: Empirical hypotheses. *Econometrica*, 84(4):1345–1404, 2016.

A. S. Kyle and A. A. Obizhaeva. Market microstructure invariance: A dynamic equilibrium model. *Working Paper (March 23, 2020)*, 2020.

A. S. Kyle, O. A. Obizhaeva, and A. A. Obizhaeva. Order shredding, invariance, and stock returns. *Working Paper (November 17, 2017)*, 2017.

A. S. Kyle, A. A. Obizhaeva, and T. Tuzun. Microstructure invariance in us stock market trades. *Journal of Financial Markets, forthcoming*, 49, 2020.

A. A. Obizhaeva and J. Wang. Optimal trading strategy and supply/demand dynamics. *Journal of Financial Markets*, 16(1):1–32, 2013.

M. O'Hara, G. Saar, and Z. Zhong. Relative tick size and the trading environment. *The Review of Asset Pricing Studies*, 9(1):47–90, 2019.

M. Parkinson. The extreme value method for estimating the variance of the rate of return. *Journal of Business*, pages 61–65, 1980.

G. Saar. Price impact asymmetry of block trades: An institutional trading explanation. *The Review of Financial Studies*, 14(4):1153–1181, 2001.

J. Upson, T. McInish, and B. H. Johnson IV. Order based versus level book trade reporting: An empirical analysis. *Journal of Banking & Finance, forthcoming*, 125, 2021.

J. Wang and C. Wu. Liquidity, credit quality, and the relation between volatility and trading activity: Evidence from the corporate bond market. *Journal of Banking & Finance*, 50:183–203, 2015.

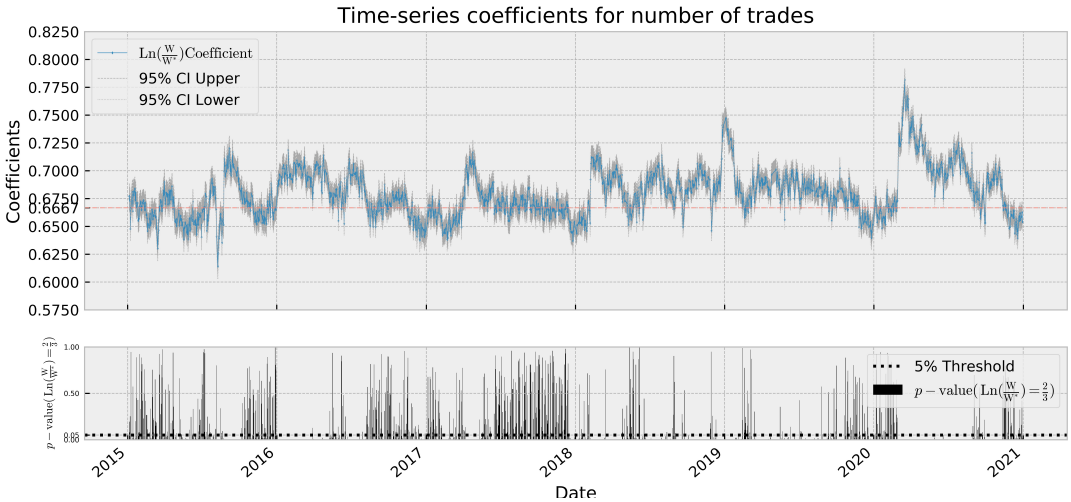


Figure 1: Time series coefficients for the number of trades
This figure shows the time-series coefficients of regressions of the number of trades on trading activity. For each day t, we estimate the following cross-sectional regression: $Ln(NT_i) = \alpha_0 + \alpha_{1t} Ln(\frac{W_{it}}{W*}) + \varepsilon_i$ and obtain an estimated α_{1t} from January 5, 2015, through December 30, 2020. The dashed line and dot line represent the upper and lower 95% confidence intervals of the estimated coefficients. The subplot at the bottom shows the dynamics of p-values of null hypotheses that the coefficient estimates α_{1t} equal 2/3. The black dot line denotes the 5% threshold of the corresponding t-test to reject the null hypothesis of 2/3.

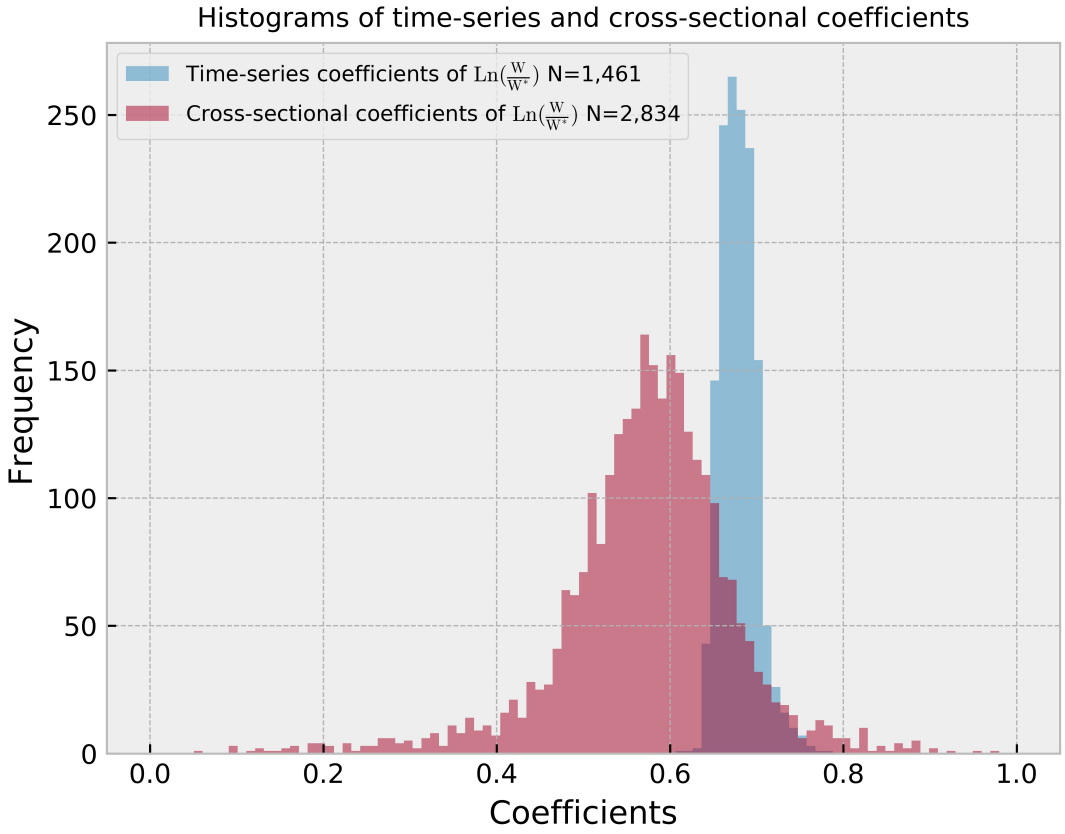


Figure 2: Histograms of time series and cross-sectional coefficients
This figure plots the histograms of time-series (in blue) and cross-sectional (in red) coefficients of regressions of the number of trades on trading activity. The time-series coefficients are regression coefficients obtained from 1,461 daily regressions of the number of trades on trading activity across stocks. The cross-sectional coefficients are regression coefficients obtained from 2,834 individual stock time-series regressions of the number of trades on trading activity during our sample period.

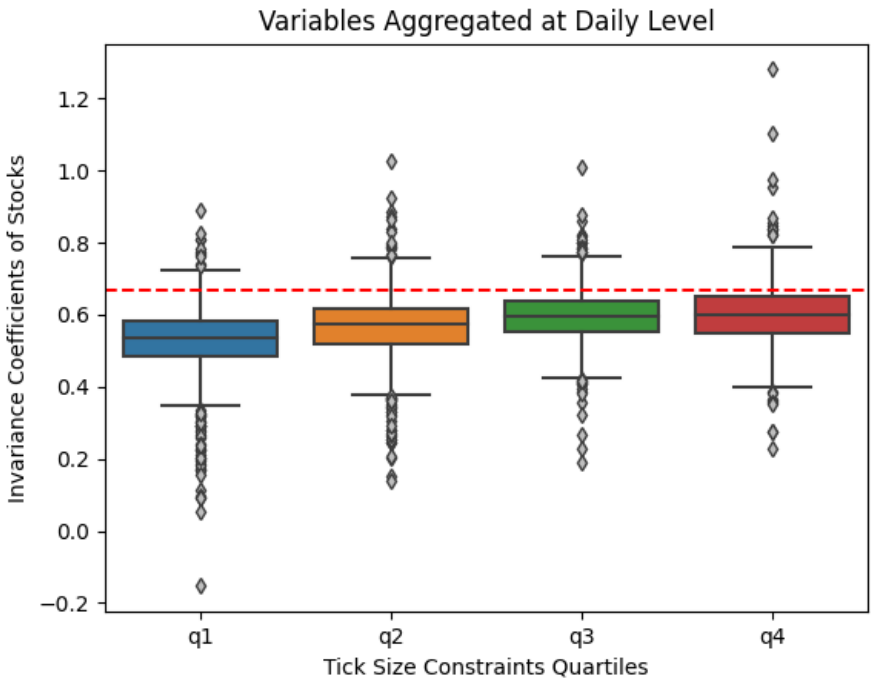


Figure 3: This figure shows the boxplot of the estimated coefficients of regressions of the number of trades on trading activity for different tick size constraint quartiles.

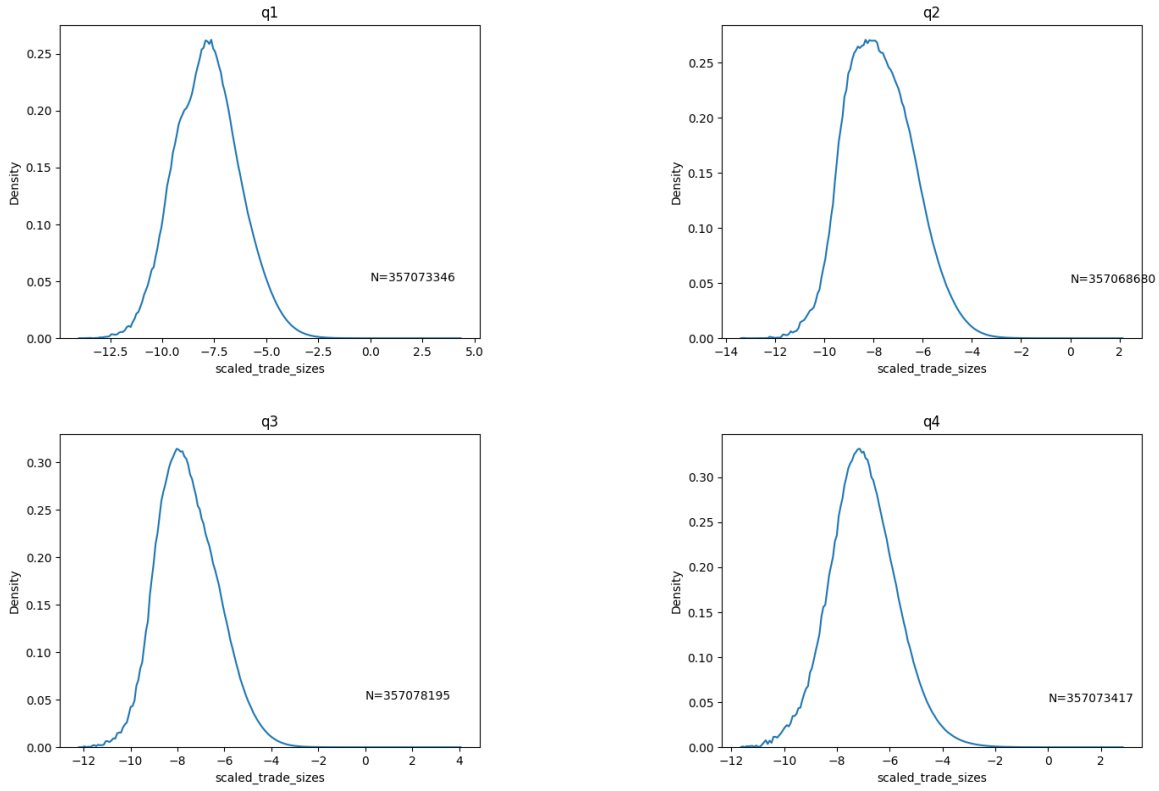


Figure 4: This figure shows the scaled trade size distributions, $Ln(\tilde{X}/V)+2/3Ln(W/W^*)$. The q1 represents the stocks with smallest relative tick sizes which are defined as the average of bid-ask spread divided by tick size. N represents the number of observations.

Table 1. An illustration of trades and quotes: the Japanese equity markets and the U.S. equity markets
This table presents the trades and quotes prints for both the Japanese equity market (the Tokyo Stock Exchange) and the U.S. equity markets (all exchanges aggregated). Panel A shows a selection of trades and quotes prints for the ticker 1301 on Jan 4, 2018 in the Tokyo Stock Exchange. Specifically, we show the detailed timestamp to microseconds (Detailed), the count of the same timestamp (Count), the ticker number (Ticker), quote prices and quote sizes (Price@Size) at the best bid and the best ask, trade prices and trading volume (Price@Volume), and the notes (Notes) on the trades. Panel B shows a selection of trades and quotes prints for the ticker symbol of AAPL on October 25, 2016. Specifically, we show the participant timestamp (Participant), the count (Count) of the same participant timestamp, the ticker symbol (Symbol), quote prices and sizes (Price@Size) at the National Best Bid (NBB), and the National Best Offer (NBO), the best exchanges (Best Exchange) quoting at the NBB and NBO, trade prices and trading volume (Price@Volume), the executed exchange of the trades (Exchange), and the notes (Notes) on the trades. Trades prints in dashed boxes may be treated as one trade.

Panel A: Trades and quotes in the Tokyo Stock Exchange

Timestamp		Ticker	Best Bid	Best Ask	Trade	
Detailed	Count		Price@Size	Price@Size	Price@Volume	Notes
09:05:02.064162	1	1301	4245@200	4250@300		
09:05:02.066696	1	1301	4245@100	4250@300	4245@100	Seller-Initiated Marketable Order
09:05:03.181948	1	1301	4245@200	4250@300		
09:05:04.061576	1	1301	4245@200	4255@300	4250@300	Buyer-Initiated Marketable Order
09:05:04.063104	1	1301	4250@100	4255@300		
⋮	⋮	⋮	⋮	⋮	⋮	⋮
09:10:01.357086	1	1301	4245@100	4255@200		
09:10:01.357839	1	1301	4240@600	4255@200	4245@100	Seller-Initiated Market Order
09:10:01.357839	2	1301	4240@500	4255@200	4240@100	Seller-Initiated Market Order

Panel B: Trades and quotes in the U.S. equity markets (all exchanges aggregated)

Timestamp		Symbol	NBB		NBO		Trade		
Participant	Count		Price@Size	Best Exchange	Price@Size	Best Exchange	Exchange	Price@Volume	Notes
09:30:02.462245426	1	AAPL	177.95@1100	P	117.99@1100	Q, K, P			
09:30:02.464000000	1	AAPL	177.95@1100	P	117.99@1100	Q, K, P	D	117.97@15	Off-Exchange
09:30:02.470520000	1	AAPL	177.95@1100	P	117.99@700	Q, P	K	177.97@100	Hidden
09:30:02.470520000	2	AAPL	177.95@1100	P	117.99@700	Q, P	K	177.99@300	Regular
09:30:02.470520000	3	AAPL	177.95@1100	P	117.99@700	Q, P	K	177.99@100	Regular
09:30:02.470602060	1	AAPL	177.95@1100	P	117.99@600	Q, P	Q	177.97@28	Routed & Hidden
09:30:02.470602060	2	AAPL	177.95@1100	P	117.99@500	P	Q	177.97@100	Hidden
09:30:02.470602060	3	AAPL	177.95@1100	P	117.99@500	P	Q	177.99@8	Hidden
09:30:02.470602060	4	AAPL	177.95@1100	P	117.99@500	P	Q	177.99@36	Hidden
09:30:02.470602060	5	AAPL	177.95@1100	P	117.99@500	P	Q	177.99@100	Regular
09:30:02.470602060	6	AAPL	177.95@1100	P	117.99@500	P	Q	177.99@28	Hidden
09:30:02.470626304	1	AAPL	177.95@1100	P	118.00@143700	Q, K, P	P	117.99@100	Routed & Regular
09:30:02.470626304	2	AAPL	177.95@1100	P	118.00@143700	Q, K, P	P	117.99@400	Regular

Table 2. Descriptive statistics
This table reports descriptive statistics for stock-day observations calculated from transaction data. Panel A reports statistics for all observations from January 5, 2015, through December 30, 2020. We present the mean, standard deviation, percentiles 1, 50, and 99 percentiles of the variables listed in column one for each stock for each trading day. Panel B reports the means for seven tick-size groups–1, 5, 10, 50, 100, 500, and 1000 JPY. Panel C reports the means for ten-volume groups categorized by stocks sorted by their in-sample mean VOLJ percentiles–(1) $VOLJ < 30^{th}$, (2) $30^{th} <= VOLJ < 50^{th}$, (3) $50^{th} <= VOLJ < 60^{th}$, (4) $60^{th} <= VOLJ < 70^{th}$, (5) $70^{th} <= VOLJ < 75^{th}$, (6) $75^{th} <= VOLJ < 80^{th}$, (7) $80^{th} <= VOLJ < 85^{th}$, (8) $85^{th} <= VOLJ < 90^{th}$, (9) $90^{th} <= VOLJ < 95^{th}$, (10) $95^{th} <= VOLJ < 100^{th}$. Volume group 1 has the smallest $VOLJ$. $DEPTH$ has n=189,302.

Panel A: Descriptive statistics for pooled sample (N = 3,609,547)

Variables	Mean	STD	P1	P50	P99
NT	749	1,783	3	213	7,981
NT (Sell)	375	896	0	107	4,015
NT (Buy)	370	895	0	101	3,989
$\bar{P}$	1,998	8,113	71.1	1,207	10,899
SIZ	1,366	24,025	111	281	9,079
SIZ^{TW50} ($\tilde{X}_{it}$)	294	708	100	100	2,000
SIZ^{VW50} ($\tilde{X}_{it}$)	25,901	780,950	100	500	83,200
VLM	727,018	4,658,984	600	72,000	10,500,000
$VOLJ$ (1,000)	9,609	40,770	5.869	769.2	142,556
$VOLJ$ (Sell 1,000)	3,947	17,523	0	323.3	57,632
$VOLJ$ (Buy 1,000)	3,904	17,378	0	294.2	57,489
$DEPTH$	52,267	1,299,772	258.5	1,262	481,262
σ	0.019	0.012	0.004	0.017	0.061
$L100\%$	51.6	25.3	0	57.0	92.3
$L1000\%$	9.5	22.3	0	0.7	88.4
$LOTE\%$	90.9	9.0	61.1	93.7	100.0

Table 2—continued

Panel B: Means of variables by tick size groups

Tick size group =	1	5	10	50	100	500	1000
NT	519	573	680	664	530	372	377
NT (Sell)	260	285	338	330	263	190	178
NT (Buy)	255	284	338	331	263	179	196
$\bar{P}$	1,137	3,810	7,596	34,652	131,666	455,772	556,577
SIZ	1,557	296.2	287.4	231.1	102.2	2.934	2.420
SIZ^{TW50} ($\tilde{X}_{it}$)	322.4	119.6	112.5	100.0	1.158	1.022	1.037
SIZ^{VW50} ($\tilde{X}_{it}$)	29,967	2,315	2,648	384.6	36.98	4.441	4.527
VLM	527,663	237,607	241,525	153,060	1,787	799.3	970.8
$VOLJ$ (1,000)	3,911	9,077	18,088	52,699	2,161	3,620	5,431
$VOLJ$ (Sell 1,000)	1,662	3,703	7,360	21,786	942.7	1,669	2,337
$VOLJ$ (Buy 1,000)	1,626	3,685	7,382	21,803	914	1,547	2,414
σ	0.020	0.019	0.019	0.016	0.012	0.014	0.013
$L100\%$	50.3	63.4	63.7	70.7	0.0	0.0	0.0
$L1000\%$	10.9	1.1	0.9	0.5	0.0	0.0	0.0
$LOTE\%$	91.0	92.6	92.6	94.9	0.0	0.0	0.0
N	2,992,926	313,680	156,693	332	1,441	451	298

Panel C: Means of variables by volume group

Volume group =	1	2	3	4	5
NT	58.4	199	371	552	750
NT (Sell)	29.2	101	187	278	376
NT (Buy)	26.3	95.0	181	271	370
$\bar{P}$	1,105	1,397	1,507	1,923	3,446
SIZ	602	970	1,389	1,728	935
SIZ^{TW50} ($\tilde{X}_{it}$)	299	303	281	230	258
SIZ^{VW50} ($\tilde{X}_{it}$)	1,975	6,322	22,474	32,502	17,272
VLM	32,048	131,457	285,601	339,516	462,641
$VOLJ$ (1,000)	161	675	1,467	2,581	4,091
$VOLJ$ (Sell 1,000)	66.4	292	646	1,129	1,785
$VOLJ$ (Buy 1,000)	57.2	267	608	1,084	1,736
σ	0.016	0.021	0.022	0.022	0.022
$L100\%$	54.9	54.5	54.6	55.2	51.4
$L1000\%$	12.1	10.3	7.7	6.0	6.7
$LOTE\%$	93.4	92.5	91.9	92.2	90.4
N	1,067,068	727,769	363,473	359,247	183,755

Table 2—continued

Volume group =	6	7	8	9	10
NT	915	1,121	1,445	2,180	5,261
NT (Sell)	459	561	721	1,089	2,636
NT (Buy)	452	557	720	1,086	2,621
$\bar{P}$	3,260	2,454	2,783	3,074	5,652
SIZ	2,716	4,739	2,585	1,267	1,316
SIZ^{TW50} $(\tilde{X}_{it})$	289	295	337	335	331
SIZ^{VW50} $(\tilde{X}_{it})$	60,037	130,260	123,981	13,493	22,657
VLM	661,575	985,020	1,518,871	2,214,536	6,343,498
$VOLJ$ (1,000)	6,091	9,510	15,999	29,356	109,018
$VOLJ$ (Sell 1,000)	2,636	4,009	6,523	12,028	44,266
$VOLJ$ (Buy 1,000)	2,579	3,953	6,488	11,974	43,987
σ	0.022	0.021	0.020	0.020	0.018
$L100\%$	47.0	45.6	41.8	40.6	41.5
$L1000\%$	9.7	8.3	9.7	6.9	6.9
$LOTE\%$	89.5	87.9	86.2	83.1	83.6
N	176,157	179,387	186,899	183,238	189,800

Table 3. Pooled least square estimates of the number of trades
This table reports the pooled least square estimates of Equation 6 with adding a constant W^*

$$Ln(NT_{it}) = \alpha_0 + \alpha_1 Ln(\frac{W_{it}}{W^*}) + \tilde{\epsilon}_{it}$$

where NT_{it} is the total number of trades, W_{it} is the product of stock volatility and JPY volume, W^* is the scaling constant 601,000 times 0.016, and $\tilde{\epsilon_{it}}$ is a random error term for stock i on trading day t. Columns 1 to 3 report the coefficients from the OLS estimates for all stock days. Column 2 reports the coefficients based on the sell-initiated trades. Column 3 reports the coefficients based on the buy-initiated trades. Column 4 reports the coefficients using day-fixed effects. The sample period is from Jan 5, 2015, through December 30, 2020. Numbers in parentheses are standard errors. We report robust standard errors clustered at the stock level (column 1, 2, 3, 4). Newey-West standard errors are calculated with five lags for the Fama-MacBeth regression in column 5. N (in mill.) denotes the number of observations in million. The last row reports the p-values for the null hypothesis that the coefficient estimates of $\alpha_1 = 2/3$.

	OLS (All) 1	(Sell) 2	(Buy) 3	Within 4	FMB 5
$Ln(\frac{W_{it}}{W^*})$ (α_1)	0.669	0.665	0.665	0.671	0.671
	(0.003)	(0.003)	(0.003)	(0.001)	(0.001)
Constant (α_0)	5.003	4.906	4.911	5.002	4.996
	(0.009)	(0.008)	(0.008)	(0.000)	(0.007)
Adj. R-sq.	0.888	0.893	0.894	0.893	0.892
Stock FE	NO	NO	NO	NO	NO
Day FE	NO	NO	NO	YES	NO
N (in mill.)	3,609	3,552	3,534	3,609	3,609
$\alpha_1 = 2/3$ p-value	0.505	0.679	0.510	0.000	0.000

Table 4. Pooled least square estimates of the number of non-marketable orders
This table reports the pooled least square estimates of the following equation derived from Equation 6:

$$Ln(NO_{it}^{\text{Aggressiveness}}) = \alpha_0 + \alpha_1 Ln(\frac{W_{it}}{W^*}) + \tilde{\epsilon}_{it}$$

where $Ln(NO_{it}^{\text{Aggressiveness}})$ is the number of non-marketable orders at the respective order aggressiveness, W_{it} is the product of stock volatility and JPY volume, W^* is the scaling constant 601,000 times 0.016, and $\tilde{\epsilon_{it}}$ is a random error term for stock i on trading day t. Panel A reports the estimates based on the incoming sell orders on the ask side, and Panel B reports the estimates based on the incoming buy orders on the bid side. We categorize all the non-marketable orders into five groups based on their price aggressiveness (Biais et al., 1995). Column 1 reports the estimates from the regressions of the number of incoming unmarketable orders that improve the best bid/ask prices on trading activities. Column 5 reports the estimates from the regressions of the number of incoming non-marketable orders that improve the fifth-best bid/ask prices on trading activities. We report the robust standard errors clustered at the stock level in parentheses. N denotes the number of observations. The last row reports the p-values for the null hypothesis that the coefficient estimates of $\alpha_1 = 2/3$

Panel A: Pooled least square estimates by sell order aggressiveness					
Aggressiveness	1	2	3	4	5
$Ln(\frac{W_{it}}{W^*})$ (α_1)	0.665	0.683	0.654	0.629	0.616
	(0.003)	(0.003)	(0.002)	(0.002)	(0.002)
Constant (α_0)	4.998	4.451	4.175	3.938	3.781
	(0.002)	(0.007)	(0.007)	(0.007)	(0.007)
Adj. R-sq.	0.911	0.866	0.829	0.800	0.782
Stock FE	YES	YES	YES	YES	YES
Day FE	NO	NO	NO	NO	NO
$\alpha_1 = 2/3$ p-value	0.588	0.000	0.000	0.000	0.000
N	3,507,061	2,488,805	1,940,396	1,689,791	1,566,156
Panel B: Pooled least square estimates by buy order aggressiveness					
Aggressiveness	1	2	3	4	5
$Ln(\frac{W_{it}}{W^*})$ (α_1)	0.672	0.686	0.660	0.631	0.614
	(0.003)	(0.003)	(0.002)	(0.002)	(0.002)
Constant (α_0)	5.020	4.453	4.198	3.984	3.837
	(0.002)	(0.006)	(0.007)	(0.007)	(0.007)
Adj. R-sq.	0.916	0.866	0.832	0.802	0.783
Stock FE	YES	YES	YES	YES	YES
Day FE	NO	NO	NO	NO	NO
$\alpha_1 = 2/3$ p-value	0.035	0.000	0.006	0.000	0.000
N	3,497,640	2,483,003	1,943,310	1,697,934	1,576,088

Table 5. Measurement errors and estimated coefficients
This table reports the pooled least estimates of Equation 6 as in Table 3 with different measures number of trades. We impose four different levels of measurement errors denoted as e_5, e_{10}, e_{20}, and e_{40}, respectively, on the true number of trades (NT). For instance, the level of e_5 measurement errors represents a pseudo-data-generating process that will report a large trade (a trade with more than 1000 shares) as one to five trades with equal probability. Column 2 to column 5 report the estimated coefficients using the pseudo-generated number of trades with measurement errors e_5, e_{10}, e_{20}, and e_{40}, respectively. The means of the generated number of trades (NT) are reported below the dashed line. We report standard errors clustered at the stock level. R-sq. denotes the R-squared. N=3,591,134.

	1	2	3	4	5
	TRUE	e_5	e_{10}	e_{20}	e_{40}
$Ln(\frac{W_{it}}{W^*})$ (α_1)	0.669	0.706	0.716	0.732	0.757
	(0.003)	(0.003)	(0.004)	(0.004)	(0.005)
Constant (α_0)	5.003	5.147	5.298	5.491	5.733
	(0.009)	(0.008)	(0.009)	(0.010)	(0.013)
R-sq.	0.888	0.897	0.879	0.841	0.792
Stock FE	NO	NO	NO	NO	NO
Day FE	NO	NO	NO	NO	NO
Mean of NT	749	983	1,275	1,857	3,022

Table 6. External validity: European Markets
This table reports the pooled least square estimates of Equation 6 for the European markets:

$$Ln(NT_{it}) = \alpha_0 + \alpha_1 Ln(W_{it}) + \tilde{\epsilon}_{it}$$

Where NT_{it} is the number of trades, W_{it} is the product of stock volatility and fiat-currency volume, $\tilde{\epsilon_{it}}$ is a random error term for stock i on trading day t. The sample consists of stocks in six counties across six stock exchanges: BE (Brussels Stock Exchange), FR (Paris Stock Exchange), IR (Irish Stock Exchange), NL (Amsterdam Stock Exchange), NO (Oslo Stock Exchange), and PT (Euronext Lisbon). The sample period is from Jan 4, 2021, to Dec 31, 2021. We report standard errors clustered at the stock level. R-sq. denotes the R-Squared. N denotes the number of observations

	BE	FR	IR	NL	NO	PT
$Ln(\frac{W_{it}}{W^*})$ (α_1)	0.770	0.698	0.749	0.740	0.727	0.699
	(0.002)	(0.001)	(0.005)	(0.001)	(0.001)	(0.003)
Constant (α_0)	−1.636	−0.779	−1.851	−1.236	−2.986	−0.967
	(0.015)	(0.006)	(0.050)	(0.159)	(0.013)	(0.033)
R-sq.	0.910	0.927	0.850	0.948	0.873	0.922
Stock FE	NO	NO	NO	NO	NO	NO
Day FE	NO	NO	NO	NO	NO	NO
N	25,381	132,042	6,139	22,260	59,899	6,633

Table 7. Pooled least square estimates of quantiles of trade sizes
This table reports the pooled least square estimates of the following regressions:

$$f(Ln(\frac{|\tilde{X}_{it}|}{V_{it}})) = \alpha_{0x} + \alpha_{1x} Ln(\frac{W_{it}}{W^*}) + \tilde{\epsilon}_{it}$$

where $f(.)$ denotes the functional form of the distribution transformations. The left-hand side variables are the logarithm of mean trade sizes, or the 20th, 50th, and 80th percentiles of the distributions of logarithms of the volume-weighted and trade-weighted trade sizes, in turn, expressed as a fraction of daily volume for stock i on day t. Each observation corresponds to stock i at trading day t with trading activity. W_{it} is the product of stock volatility and JPY volume. W^* is the scaling constant equal to 601000 times 0.016. Column 1 reports the coefficients using the mean trade sizes as the dependent variable. Column 2 to column 4 report the coefficients using 20th, 50th, and 80th percentiles of volume-weighted trade sizes as the dependent variables, in turn. Column 5 to column 7 report the coefficients using 20th, 50th, and 80th percentiles of trade-weighted trade sizes as the dependent variables, in turn. Figures in parentheses are standard errors. We report the robust standard errors clustered at the stock level are reported. The last row reports the p-values for the null hypothesis that the coefficient estimates of $\alpha_{1x} = -2/3$. n = 3,650,932.

			V.W			T.W	
	SIZ	SIZ^{VW20}	SIZ^{VW50}	SIZ^{VW80}	SIZ^{TW20}	SIZ^{TW50}	SIZ^{TW80}
	1	2	3	4	5	6	7
$Ln(\frac{W_{it}}{W^*})$ (α_{1x})	-0.586	-0.597	-0.526	-0.469	-0.673	-0.639	-0.561
	(0.002)	(0.004)	(0.004)	(0.004)	(0.003)	(0.003)	(0.003)
Constant (α_{0x})	-5.017	-5.517	-4.504	-3.042	-6.038	-5.823	-4.918
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Adj. R-sq.	0.715	0.614	0.405	0.217	0.706	0.686	0.417
Stock FE	YES	YES	YES	YES	YES	YES	YES
Day FE	NO	NO	NO	NO	NO	NO	NO
Coef = 2/3 p-value	0.000	0.000	0.000	0.000	0.018	0.000	0.000

Appendix A

A review of invariance-of-bet hypothesis

Following the notations of Kyle and Obizhaeva (2016) and Kyle et al. (2020), we review the mathematical representations of the invariance-of-bet hypothesis and its testable implications. Suppose the expected arrival rate of bets for asset i at trading day t is γ_{it} and their signed sizes are random variables $\tilde{Q}_{it}$ with $E[\tilde{Q}_{it}] = 0$. The trading volume, VLM_{it}, which measures the number of shares traded for asset i on trading day t comprises both bet volume $\tilde{V}_{it}$ and intermediary volume if bets are intermediated by market makers. Suppose that, on average, each unit of bet results in ζ_{it} units of total volume and thus $\zeta_{it} - 1$ units of intermediary volume. The bet volume $\tilde{V}_{it}$ and the trading volume VLM_{it} have the following relation:

$$\tilde{V}_{it} := \gamma_{it} \times E[\tilde{Q}_{it}] = \frac{2}{\zeta_{it}} \times VLM_{it} \tag{A.1}$$

Equation A.1 shows the relation between the bet volume and the trading volume when there is a volume multiplier, ζ_{it}. The volume multiplier is, by construction, positively correlated to the level of intermediary activities. Suppose in a market where each side of the trade corresponds to one bet and no intermediary trading occurs, then we have $\zeta_{it} = 1$. In this market, both buyers and sellers are making investment decisions (bets) simultaneously, and thus each trade is composed of two bets. As Equation A.1 indicates,

the bet volume that results from any trade in this market is twice the number of shares traded. For example, a buyer has an investment decision (bet) to buy 100 shares of a stock, and this trade can only be executed when a potential seller also has an investment decision (bet) but to sell 100 shares. The bet volume for this trade which matches both sides will be 200 because on either side of the trade is an investment decision therefore a bet.

In a market where a single monopolist market maker is intermediating the trades, then we have $\zeta_{it} = 2$. In a market where multiple market makers are intermediating the trades, then $\zeta_{it} > 2$. The choice of a given value for the volume multiplier does not alter the theoretical predictions implied by the invariance-of-bet hypothesis. However, as we will show and discuss later in this section, the identification assumption on the volume multiplier is crucial in testing the invariance hypothesis empirically.

Suppose a bet with a signed size $\tilde{Q}_{it}$ trades at an expected price P_{it} in fiat-currency for asset i on trading day t with its return volatility σ_{it}. Recall that the expected bet arrival rate is γ_{it} and thus per unit of business time should be equivalent to the inverse of the arrival rate which equals to $\frac{1}{\gamma_{it}}$. Return volatility measured per unit of business time for this bet is then $\frac{\sigma_{it}}{\gamma_{it}}$. We define the fiat-currency amount of risk for this bet transferred per unit of business time as:

$$\tilde{I}_{it} := P_{it}\tilde{Q}_{it}\frac{\sigma_{it}}{\gamma_{it}} \tag{A.2}$$

The invariance-of-bet hypothesis (Kyle and Obizhaeva, 2016) states that the distribution of risk transfers $\tilde{I}_{it}$ is a market microstructure invariant that equals to the distribution

of a random variable $\tilde{I}$. In other words, the realized value of $\tilde{I}_{it}$ at the same percentile of the risks is a constant I across stocks and time.

To obtain testable empirical implications for the invariance-of-bet hypothesis, following Kyle and Obizhaeva (2016), we define trading activity W_{it}, an aggregate measure of risks as the product of returns volatility σ_{it} and expected daily fiat-currency volume $P_{it}VLM_{it}$:

$$W_{it} := \sigma_{it} P_{it} VLM_{it} \tag{A.3}$$

Rewriting Equation A.3 using the bet volume and trading volume relation defined as in Equation A.1, we obtain $W_{it} = \frac{\zeta_{it}}{2}\sigma_{it}P_{it}E[|\tilde{Q}_{it}|]\gamma_{it}$. Taking the expectation of the absolute value of Equation A.2, we obtain $\sigma_{it}P_{it}E[|\tilde{Q}_{it}|] = E[|\tilde{I}_{it}|]\gamma_{it}^{1/2}$. Substituting $\sigma_{it}P_{it}E[|\tilde{Q}_{it}|]$ with $E[|\tilde{I}_{it}|]\gamma_{it}^{1/2}$ for the right hand side of $W_{it} = \frac{\zeta_{it}}{2}\sigma_{it}P_{it}E[|\tilde{Q}_{it}|]\gamma_{it}$, the relation between the (expected) bet arrival rate of and the trading activity can be written as:

$$\gamma_{it} = E[\frac{\zeta_{it}}{2}|\tilde{I}_{it}|]^{-2/3} W_{it}^{2/3} \tag{A.4}$$

Dividing the absolute value of Equation A.2 by Equation A.1 and substituting from Equation A.4, the bet size expressed as a fraction of trading volume can be written as:

$$\frac{|\tilde{Q}_{it}|}{VLM_{it}} \stackrel{d}{=} E[\frac{\zeta_{it}}{2}|\tilde{I}_{it}|]^{-1/3} W_{it}^{2/3} |\tilde{I}_{it}| \tag{A.5}$$

Equations A.4 and Equation A.5 summarize the testable empirical implications of market invariance for the expected bet arrival rate and the bet size $\tilde{Q}_{it}$. Kyle et al.

(2020) drop the subscripts i and t for the volume multiplier ζ_{it} and assume the values of ζ_{it} are identical across asset i and time t. By the invariance-of-bet hypothesis, we can also drop subscripts i and t for the random variable $\tilde{I}_{it}$. Dropping subscripts for both ζ_{it} and $\tilde{I}_{it}$, Equation A.4 shows that the bet arrival rate is proportional to the two-thirds power of the trading activity. Equation A.4 implies that if trading activity increases by a factor of 8, then by Equation A.4, the expected arrival rate of bets will increase by a factor of 4.

Equation A.5 shows the bet size expressed as a fraction of trading volume is proportional to the negative two-thirds power of the trading activity. If we multiply Equation 5 by $W_{it}^{2/3}$, the scaled bet size will follow the distribution of the invariant random variable $\tilde{I}$. Using portfolio transition orders data, Kyle and Obizhaeva (2016) find that the distribution of scaled portfolio orders, a proxy for the bets, is close to a normal distribution with a log-variance of 2.53. Also, Kyle et al. (2020) find that the trade-weighted distribution of scaled trade sizes are close to a normal distribution of 1.90.

Table A.1. Variable definitions
This table defines notations and variables. Each notation or variable is for stock i on the trading day t.

Panel A: Notations from Kyle and Obizhaeva (2016)	
Symbol	Definition
γ	The expected arrival rate of bets.
$\tilde{Q}$	A random variable for the signed bet size.
$\bar{V}$	The volume of bets.
ζ	The volume multiplier.
P	The expected price of bets.
$\sigma/\gamma^{\frac{1}{2}}$	Returns volatility measured in per unit of business time for a bet.
$\tilde{I}$	The distribution of fiat-currency (dollar) volume risk transferred by a bet per unit of business time.
W	The (bet) trading activity—an aggregate measure of risks—the product of risk and fiat-currency (dollar) volume.
$\tilde{X}$	The distribution of trade print sizes.
δ	The order-splitting multiplier.

Panel B: Variables definition for variables used in this paper	
Variable	Definition
NT	The (total) number of trades.
$\bar{P}$	The mean trade prices in JPY.
SIZ	The mean trade size.
SIZ^{TW50}	The 50th percentile of the trade-weighted trade sizes.
SIZ^{VW50}	The 50th percentile of the volume-weighted trade sizes.
VLM	The (total) number of shares traded.
σ	The standard deviation of squared returns over the past 20 trading days.
$VOLJ$	The JPY volume, defined as the mean trade prices in JPY ($\bar{P}$) times the total number of shares traded (VLM) divided by 100.
$L100\%$	The percentage of trades in 100-share lots.
$L1000\%$	The percentage of trades in 1000-share lots.
$LOTE\%$	The percentage of trades in even lots.
μ	The calibrated value of the order-splitting multiplier.

Tick sizes of the stocks in Tokyo Stock Exchange

Table A.2 reports the tick size rules for the stocks listed on the Tokyo Stock Exchange. Compared with other non-TOPIX 100 stocks, TOPIX 100 stocks have smaller tick sizes across all different stock prices.

Table A.2. Tick size rules for the stocks listed in the Tokyo Stock Exchange
This table reports the tick size rules of stocks traded on the Tokyo Stock Exchange. We report the tick size rules for TOPIX 100 stocks and non-TOPIX 100 stocks, respectively.

	TOPIX 100	Non-TOPIX 100
Stock Prices in JPY	Tick Size in JPY	
1 to 1,000	0.1	1
1,001 to 3,000	0.5	1
3,001 to 5,000	1	5
5,001 to 10,000	1	10
10,001 to 30,000	5	10
30,001 to 50,000	10	50
50,001 to 100,000	10	100
100,001 to 300,000	50	100
300,001 to 500,000	100	500
500,001 to 1,000,000	100	1,000
1,000,001 to 3,000,000	500	1,000
3,000,001 to 5,000,000	1,000	5,000
5,000,001 to 10,000,000	1,000	10,000
10,000,001 to 30,000,000	5,000	10,000
30,000,001 to 50,000,000	10,000	50,000
50,000,000 and more	10,000	100,000