Microstructure invariance in U.S. stock market trades[☆]

Albert S. Kyle[a,1], Anna A. Obizhaeva[b], Tugkan Tuzun[c,2,**]

[a] *Robert H. Smith School of Business, University of Maryland, College Park, MD 20742*
[b] *New Economic School, Moscow, Russia*
[c] *Board of Governors of the Federal Reserve System, Washington, DC 20551*

Abstract

We examine invariance relationships in tick-by-tick transaction data in the U.S. stock market. Over the period 1993–2001, monthly regression coefficients of the log of the trade arrival rate on the log of trading activity have an almost constant value of 0.666, close to the value of two-thirds predicted by market microstructure invariance. Over the 2001–2014 period, after decimalization and the increasing use of electronic order matching systems and algorithmic trading, the coefficients increase to about 0.79. The evidence suggests that changes in coefficients are due to increasing importance of minimum lots size in a world where algorithmic traders split orders into tiny pieces.

Keywords: market microstructure, invariance, transaction data, market frictions, trade size, tick size, order shredding, clustering, TAQ data, liquidity
JEL: G10, G23

We examine time-series and cross-sectional implications of market microstructure invariance hypotheses, described by Kyle and Obizhaeva (2016), for transaction frequencies and transaction sizes in the U.S. equity market. According to the invariance principles, trading in all securities is similar after accounting for differences in the flow of risk transfer among traders. The empirical results show that predictions based on this invariance principle fit the data much better than alternatives, especially for earlier years before algorithmic trading led to the extensive use of order splitting. Deviations from invariance benchmarks may be attributed to market microstructure frictions related to institutional details such as minimum tick size and minimum lot size.

Regardless of the nature of the security and absent institutional frictions, trading in securities markets can be modeled as the same trading game, but played at different speeds for dif-

[☆] We are grateful to Elena Asparouhova, Peter Bossaerts, and Stewart Mayhew for very helpful comments on an earlier draft of this paper. Joseph Saia and Zach Fernandes provided excellent research assistance.

[**] Corresponding author telephone number: +1-202-736-5646

Email addresses: akyle@rhsmith.umd.edu (Albert S. Kyle), aobizhaeva@nes.ru (Anna A. Obizhaeva), tugkan.tuzun@frb.gov (Tugkan Tuzun)

[1] Kyle has worked as a consultant on finance topics for companies, banks, stock exchanges, and various U.S. federal agencies including the Securities and Exchange Commission, the Commodity Futures Trading Commission, and the Department of Justice. He also serves as a non-executive director of a U.S.-based asset management company.

[2] The views expressed herein are those of the authors and do not necessarily reflect the views of the Federal Reserve Board, the Federal Reserve System or their staff.

ferent assets. Asset managers place bets or meta-orders, which approximately represent uncorrelated decisions to buy or sell specific numbers of shares; the arrival times of bets are random. Each bet, of institutional size, may be executed as a sequence of many smaller orders submitted milliseconds apart. The expected arrival rate of new bets determines the pace of business time in each market. Trading in liquid securities occurs at fast speeds, with new bets arriving over short horizons, perhaps only a few seconds or minutes. Trading in illiquid securities takes place slowly, with new bets arriving over longer horizons, perhaps a few days or even months.

Market microstructure invariance has useful practical applications. Invariance suggests a parsimonious characterization of how different securities trade. It generates operational time-series and cross-sectional predictions for many variables; it helps to understand transaction costs, trading volumes, trade counts, and trade sizes. For practitioners, models based on invariance principles allow analysts to develop transaction cost models for securities for which data are limited or unavailable. For regulators and practitioners who design markets, invariance makes it possible to predict market activity for securities that have not yet traded and to assess effects of contemplated fees and other market frictions on trading and liquidity.

Trade counts and trade sizes are predictable based on dollar volume and returns variance. Since the predictions are consistent with market microstructure invariance, our results increase confidence in the invariance hypothesis.

The invariance hypothesis implies that the dollar risks transferred by bets and the dollar costs of executing bets are the same across markets when measured per unit of *business-time*, which corresponds to the expected arrival rate of bets. Invariance implies a specific decomposition of the order flow. Let *trading activity* be a measure of aggregate risk transfer in calendar time, defined as the product of chronological (daily) dollar volume and chronological (daily) return volatility. Then, in frictionless markets, invariance implies that the number of bets (per day) is proportional to the two-thirds power of trading activity, and the distribution of bet sizes as a fraction of (daily) volume is inversely proportional to the two-thirds power of trading activity. Kyle and Obizhaeva (2016) provide empirical evidence in favor of the invariance hypothesis using a sample of portfolio transitions as proxies for bets.

We extend the invariance principle from bets (or meta-orders) to transactions (or prints) resulting from executing bets as small pieces over time in different trading venues. If one assumes that each bet is shredded into some number of trades, and this number of trades is approximately the same across time and assets, then the exponents of two-thirds will show up in the relationships between trades and trading activity as well. These predictions are then tested for the number of trades and the distribution of trade sizes in the Trades and Quotes (TAQ) dataset, which contains tick-by-tick transactions over the period 1993–2014 for the stocks listed in the U.S. market, and compared with the fit of predictions generated by alternative models. The empirical results supplement previous findings because they are based on a much broader sample of all publicly reported transactions in the U.S. stock market, rather than a small subset of transactions related exclusively to portfolio transitions. This advantage comes at the expense of having to deal with transaction data affected by order splitting practices, intermediation, market frictions, and institutional features of trading and trade reporting.

Over the 1993–2001 subperiod, a time series of month-by-month cross-sectional regression coefficients of the log of trade arrival rates on the log of trading activity shows that the estimated coefficients remained virtually constant. The estimated coefficient of 0.666 is indeed strikingly close to the benchmark invariance prediction of two-thirds. After 2001, the monthly estimates

increase from about 0.690 in 2001 to about 0.77 in 2014; this breakdown in the invariance relationships is both statistically and economically significant. The evidence suggests that changes in fit in the recent period are due to the increasing importance of minimum lots sizes in a world where algorithmic traders shred their orders. Our results also indicate further changes in trends in 2008–2009; identifying why these changes occurred is left for future research.

For the years 1993, 2001, and 2014, the empirical distributions of logs of scaled print sizes for stocks sorted into ten dollar-volume groups and four price-volatility groups tell a similar story. In 1993, consistent with invariance, all 40 empirical distributions resemble a bell-shaped normal density function with common mean and variance across the forty subgroups. In 2001 and 2014, the shape of the distributions looks much less like the shape of a normal distribution than in 1993. Furthermore, average scaled trade sizes decrease during the 1993–2014 period by a factor of about 7. The results clearly reject the hypothesis that scaled trade sizes are distributed as a common log-normal random variable. The rejection arises due to clearly visible microstructure effects related to tick size, censoring of trades at the minimum round-lot threshold, and clustering of trades at round-lot sizes such as 100, 1,000, and 5,000 shares. These results are consistent with the findings of Alexander and Peterson (2007) and O'Hara et al. (2014).

Invariance principles explain a substantial fraction of the variation in trade arrival rates and average trade sizes across stocks, especially in the first half of the sample.

Specifically, when the slope coefficient is restricted to be $\pm 2/3$ as predicted by the invariance hypothesis and only intercepts are estimated in separate month-by-month regressions, the time series of R^2s fluctuates around 0.88. Glosten and Harris (1988) find that average trade size (in shares) is negatively related to market depth. Brennan and Subrahmanyam (1998) regress average trade sizes on return volatility, standard deviation of trading volume, market capitalization, number of analysts following a stock, number of institutional investors holding a stock, and the proportion of shares institutional investors hold. Although their samples of dependent variables are not exactly the same as ours, the R^2 of 0.92 in their cross-sectional regressions with multiple explanatory variables is only modestly larger than the average R^2 of 0.88 in our regressions with coefficients fixed at levels predicted by invariance. This small difference suggests that other variables offer only limited improvement in explanatory power over the invariance hypothesis.

In practice, trade sizes and transaction counts may depend on structural issues such as tick sizes, on size-based secondary precedence rules, on fixed trade-ticket charges by clearing entities, and on minimum lots sizes that provide incentives to execute large trades. The unexplained time-series and cross-sectional variations in the variables of interest are likely to be related to complex interactions between market friction parameters and stock characteristics, such as volume, volatility, and stock price. These market frictions are studied by Harris (1994), Angel (1997), Goldstein and Kavajecz (2000), and Schultz (2000).

A new perspective on their results about market frictions comes from examining the data through the lens of the invariance hypothesis. We introduce two new measures of frictions that take into account that trading games in different financial markets run at different speeds. We define *effective price volatility* as price volatility in business time. *Effective tick size* is defined as the ratio of tick size to price volatility in business time. *Effective lot size* is defined as a ratio of lot size to a median bet, or equivalently to volume in business time. Both of these measures are closely related to effective price volatility in the U.S. equity market, where tick sizes and lot sizes do not vary much across assets. In addition to the 88% of the variation in print arrival rates and

average print sizes explained by the invariance principle, an additional 4.5% and 5.5% can be attributed to variations in effective price volatility during the 1993–2000 and 2001–2014 periods, respectively. Disentangling the effects of lot size and tick size further is difficult because they operate in opposite directions. Yet, among the two main frictions, minimum lot size seems to become more important and tick size seems to become less important in the later period.

Building theoretical microfoundations for the invariance hypothesis is an important issue. Kyle and Obizhaeva (2017a) show that invariance predictions for bets can be derived in the context of a dynamic equilibrium infinite-horizon model with informed trading, noise trading, market making, endogenous production of information, and linear market impact. The key assumptions in the model is that each informed trader must pay some cost for generating a bet, and this cost is the same across time periods and traders. Kyle and Obizhaeva (2018a) and Kyle and Obizhaeva (2018b) clarify that invariance predictions for bets come from several generic properties of theoretical models rather than specific modeling details. Building a theoretical model delivering invariance predictions for trades is difficult because it requires modeling how traders strategically shred their bets over time, for example, along the lines of Kyle et al. (2018). Yet, invariance predictions for trades can be justified using the approach of Kyle and Obizhaeva (2017b), which is based on dimensional analysis arguments, leverage neutrality, and a bet-invariance assumption.

In the remainder of the paper, we provide the implications of the invariance hypothesis for transaction sizes and counts, discuss the design of our empirical tests and institutional details related to microstructure of trading, present our results, and examine the effects of market frictions.

1. Testable implications of the invariance hypothesis

We first review the empirical hypothesis of market microstructure invariance developed by Kyle and Obizhaeva (2016), show how to formulate a similar hypothesis and develop predictions for transaction frequencies and transaction sizes in the TAQ dataset, and then briefly discuss possible alternative hypotheses.

An invariance hypothesis. According to market microstructure invariance, trading in all assets is similar after accounting for differences in the flow of risk transfer among traders, which measures business time. Asset managers buy and sell securities by placing bets that represent decisions to acquire long-term positions of a specific size, distributed approximately independently from other such decisions. Intermediaries with short-term trading strategies clear markets by taking the other side of bets. The same trading game is played at a fast pace in active, liquid markets and at a slow pace in inactive, illiquid markets. The expected arrival rate of bets determines the speed of business time in each market.

Suppose bets arrive at an expected rate of γ_{jt} bets per day and their sizes are random variables $\tilde{Q}_{jt}$ shares for asset j at time t. The random variable $\tilde{Q}_{jt}$ has a zero mean; positive values represent buying, and negative values represent selling. Let V_{jt} be the expected daily share volume per day. This variable is equal to the product of the expected number of bets per day γ_{jt} and their average size $\mathrm{E}\big[|\tilde{Q}_{jt}|\big]$, adjusted for an intermediation multiplier ζ_{jt}:

$$V_{jt} = \frac{\zeta_{jt}}{2} \cdot \gamma_{jt} \cdot \mathrm{E}\big[|\tilde{Q}_{jt}|\big]. \tag{1}$$

Total volume is divided by the factor 2 because each unit of volume has both a buy side and a sell side.

The parameter ζ_{jt} in equation (1) measures short-term intermediation trading by market makers, high-frequency traders, and other arbitragers who intermediate among long-term bets. Intuitively, it reflects the typical length of intermediation chains in the market; the longer the intermediation chains, the larger the ζ_{jt}. Without intermediation, $\zeta_{jt} = 1$. If each bet is intermediated by a single market maker, similar to a New York Stock Exchange (NYSE) specialist intermediating every bet in the market, then $\zeta_{jt} = 2$. If each bet is intermediated by two market makers, who lay off positions trading with one another, similar to NASDAQ dealers in the early 1990s, then $\zeta_{jt} = 3$. If each bet goes through the hands of multiple short-term intermediaries before finding its place in portfolios of long-term traders, then $\zeta_{jt} > 3$.

Let W_{jt} denote daily trading activity, or an aggregate risk transfer per day. Let P_{jt} be the price in dollars per share and σ_{jt} be returns volatility per day. Then, trading activity W_{jt} is defined as the product of returns volatility σ_{jt} and expected dollar volume $P_{jt} \cdot V_{jt}$ per day:

$$W_{jt} = \sigma_{jt} \cdot P_{jt} \cdot V_{jt}. \tag{2}$$

Plugging equation (1) into equation (2) yields another equation for trading activity:

$$W_{jt} = \frac{\zeta_{jt}}{2} \cdot \sigma_{jt} \cdot P_{jt} \cdot \mathrm{E}\big[|\tilde{Q}_{jt}|\big] \cdot \gamma_{jt}. \tag{3}$$

This equation shows that trading activity W_{jt}, which is easy to calculate from the data on prices and trading volume, can be decomposed into the product of much less readily obtainable characteristics such as expected bet arrival rates γ_{jt} and average bet sizes $\mathrm{E}\big[|\tilde{Q}_{jt}|\big]$. The invariance principle helps to predicts how these characteristics γ_{jt} and $\mathrm{E}\big[|\tilde{Q}_{jt}|\big]$ vary across markets with different levels of trading activity W_{jt}.

According to market microstructure invariance, the dollar risks transferred by bets per unit of business time $1/\gamma_{jt}$ are the same across assets and time. More specifically, the random variable $\tilde{I}_{jt}$, defined as

$$\tilde{I}_{jt} := \frac{P_{jt} \cdot \tilde{Q}_{jt} \cdot \sigma_{jt}}{\gamma_{jt}^{1/2}} \overset{d}{=} \tilde{I}, \tag{4}$$

has an invariant probability distribution denoted $\tilde{I}$. Here, the risk transferred by a bet $\tilde{Q}_{jt}$ per unit of business time is equal to the product of its dollar size $P_{jt} \cdot \tilde{Q}_{jt}$ and returns volatility per unit of business time $\sigma_{jt} \cdot \gamma_{jt}^{-1/2}$.

Take the expectation of equation (4) and use the result to eliminate $\sigma_{jt} \cdot P_{jt} \cdot \mathrm{E}\big[|\tilde{Q}_{jt}|\big]$ in equation (3) to get the following testable prediction concerning how expected bet arrival rate γ_{jt} varies with trading activity W_{jt}:

$$\gamma_{jt} = \mu_\gamma \cdot W_{jt}^{2/3}. \tag{5}$$

Divide the absolute value of equation (4) by equation (2) and plug in γ_{jt} from equation (5) to obtain

$$\frac{|\tilde{Q}_{jt}|}{V_{jt}} \overset{d}{=} \mu_q \cdot W_{jt}^{-2/3} \cdot |\tilde{I}|. \tag{6}$$

The parameters μ_γ and μ_q depend on the volume multiplier ζ_{jt} and moments of random variable $|\tilde{I}|$.[3] In these equations, the distribution of the random variable $\tilde{I}$ is approximately the same across assets and time. In what follows, we make an identifying assumption that the volume multiplier ζ_{jt} is also somewhat constant, and therefore μ_n does not have indices j and t. Of course, in reality the volume multiplier ζ_{jt} may change over time and assets, and—as we discus later—this variation might explain the difference between our results for samples before and after the year 2000.

Under the above assumptions, equations (5) and (6) state that the expected bet arrival rate γ_{jt} scaled by $W_{jt}^{-2/3}$ and distributions of bet sizes $|\tilde{Q}_{jt}|/V_{jt}$ scaled by $W_{jt}^{2/3}$ (that is, all of its percentiles) are the same across all markets and times. In line with these predictions, Kyle and Obizhaeva (2016) find that the distributions of logs of scaled portfolio transition orders, $(2/3)\cdot\ln(W_{jt})+\ln(|Q_{jt}|/V_{jt})$, which can be thought of as proxies for bets, are similar to each other and to a normal distribution with log-variance of 2.53.

Equations (5) and (6) characterize the flow of bets in the markets with different trading activity W_{jt}. Differences in trading activity come from both differences in average bet sizes and differences in expected bet arrival rates. For example, if ζ_{jt} and $\tilde{I}$ are indeed constant, then—fixing returns volatility σ_{jt}—it follows that a 1% increase in trading activity W_{jt} comes from an increase by two-thirds of 1% in the expected bet arrival rate γ_{jt} and an increase of one-third of 1% of the scaling of the entire distribution of unsigned bet sizes $P_{jt}|\tilde{Q}_{jt}|$. The latter statement is also equivalent to a downward shift by two-thirds of 1% in the distribution of unsigned bet sizes as a fraction of share volume, $|\tilde{Q}_{jt}|/V_{jt}$.

Kyle and Obizhaeva (2016) also discuss another invariance hypothesis related to transaction costs. In this paper, we focus only on order flow and leave testing the implications for market impact and bid-ask spreads for future research.

Invariance implications for TAQ print data. The TAQ dataset reports transaction prices and share quantities for all trades in stocks listed in the United States from 1993 to 2014. Each report of a trade execution is called a "print." We test implications of the invariance hypothesis using data on TAQ print sizes and the number of TAQ prints recorded per day.

Testing invariance in this way is not straightforward because prints are different from bets. One bet may generate multiple prints. To minimize transaction costs, traders often break bets or meta-orders into smaller pieces—as documented in Keim and Madhavan (1995), among others—and execute them at several venues, trading with multiple counterparties at multiple prices.

Let $\tilde{X}_{jt}$ denote the number of shares in a single print for asset j and at time t. Let ξ_{jt} represents into how many prints each bet generates. Then, the distribution of print sizes $\tilde{X}_{jt}$ differs from the distribution of bet sizes $\tilde{Q}_{jt}$ by a factor ξ_{jt}:

$$\tilde{X}_{jt} = \frac{|\tilde{Q}_{jt}|}{\xi_{jt}}. \tag{7}$$

Let N_{jt} denote the expected number of prints per day for asset j at time t. Each bet $\tilde{Q}_{jt}$

[3]The specific parameter values are $\mu_\gamma = \mu_{\gamma,jt} := \mathrm{E}\left[\frac{\zeta_{jt}}{2}\cdot|\tilde{I}|\right]^{-2/3}$ and $\mu_q = \mu_{q,jt} := \mathrm{E}\left[\frac{\zeta_{jt}}{2}\cdot|\tilde{I}|\right]^{-1/3}$, where the subscripts j and t can be dropped under the assumption made below that ζ_{jt} is constant across time and assets.

results on average in ξ_{jt} prints, and its execution inflates trading volume by a factor of $\zeta_{jt}/2$ due to induced intermediation volume. Thus, the expected number of prints N_{jt} differs from the expected number of bets γ_{jt} by a factor of $\xi_{jt} \cdot \zeta_{jt}/2$:

$$N_{jt} = \xi_{jt} \cdot \frac{\zeta_{jt}}{2} \cdot \gamma_{jt}. \tag{8}$$

Using predictions (5) and (6) for the number of bets and their sizes, one obtains similar testable implications for the number of prints and their sizes with the exponents $\alpha_n = -\alpha_x = 2/3$:

$$N_{jt} = \mu_n \cdot W_{jt}^{\alpha_n}, \tag{9}$$

$$\frac{|\tilde{X}_{jt}|}{V_{jt}} \stackrel{d}{=} \mu_x \cdot W_{jt}^{\alpha_x} \cdot |\tilde{I}|. \tag{10}$$

The values of the exponents α_n and α_x may be different under alternative hypotheses introduced below. The proportionality coefficients μ_n and μ_x depend on the volume multiplier ζ_{jt}, the order-splitting multiplier ξ_{jt}, and moments of $|\tilde{I}|$.[4]

As a benchmark for interpreting our results, we make two identifying assumptions. First, assume that a constant order-splitting multiplier ξ exists such that $\xi_{jt} = \xi$ for any asset j and time t. Second, assume that a constant volume multiplier ζ exists such that $\zeta_{jt} = \zeta$ for any asset j and time t. For simplicity of exposition, results may be interpreted under the identifying assumptions that $\xi = 1$ and $\zeta = 2$. This case would correspond to the hypothesis that each bet is executed as one print against a single intermediary and makes μ_n and μ_x in equations (9) and (10) constant across markets with no indices j and t. These predictions can be written in the form of regressions.

In actual markets, parameters ξ and ζ may vary in cross-section and time-series. In practice, the order-splitting multiplier ξ depends on specific details of order-splitting algorithms used by traders—as modeled by Almgren and Chriss (2000) and Obizhaeva and Wang (2013)—and it may potentially vary across stocks in a complex and systematic manner that depends on bet size, tick size, lot size, and perhaps other factors. For example, tiny orders may be executed as one print and gigantic orders may be executed as thousands of small prints. The amount of intermediation may also fluctuate with volatility, trading volume, and other factors. We discuss these microstructure issues in detail later.

Two alternative hypotheses. We also consider two alternative hypotheses about trading processes that make different predictions about the implied exponents α_n and α_x in equations (9) and (10).

In the first alternative hypothesis, *invariance of print frequency*, we assert that the expected number of prints is the same for all stocks, implying $\alpha_n = -\alpha_x = 0$ in equations (9) and (10). If returns volatility is approximately the same across markets, this hypothesis implies that the number of prints is constant across markets, and only their average dollar sizes change. Of

[4] Specific parameter values are $\mu_n = \mu_{n,jt} := (\xi_{jt} \cdot \zeta_{jt}/2) \cdot \mathrm{E}\left[\frac{\zeta_{jt}}{2} \cdot |\tilde{I}|\right]^{-2/3}$ and $\mu_x = \mu_{x,jt} := \xi_{jt}^{-1} \cdot \mathrm{E}\left[\frac{\zeta_{jt}}{2} \cdot |\tilde{I}|\right]^{-1/3}$, where the subscripts j and t can be dropped under the assumption below that ξ_{jt} and ζ_{jt} are constant across time and assets.

course, this hypothesis is empirically unrealistic; actively traded stocks are well-known to have more prints than inactively traded stocks. This benchmark is interesting because the illiquidity measure of Amihud (2002)—the empirical application of the notion that illiquidity is the price response per dollar traded on a percentage basis—is implicitly based on the assumption that the standard deviation of order imbalances is proportional to trading volume. This assumption holds only if the number of bets is constant across markets and time, as stated in the first alternative hypothesis.[5]

In the second alternative hypothesis, *invariance of print sizes*, we assert that the number of prints is proportional to trading activity, implying that $\alpha_n = -\alpha_x = 1$ in equations (9) and (10). If returns volatility is approximately the same across markets, this hypothesis implies that average dollar sizes of prints are constant across markets, and only their number per day changes. Related models are discussed in Tauchen and Pitts (1983), Harris (1987), Jones et al. (1994), Hasbrouck (1999), and Ané and Geman (2000), among others.

Although extreme, these two hypotheses provide convenient benchmarks for thinking about the relationship between invariance and the literature on print sizes and print frequencies.

Institutional details related to the microstructure of TAQ data. Although we make the identifying assumptions that the order-splitting multiplier is constant ($\xi_{jt} = \xi$) and the volume multiplier is constant ($\zeta_{jt} = \zeta$), we do not expect these assumptions to perfectly describe the U.S. equity market. Instead, these assumptions generate benchmarks that can be used to evaluate the economic significance of various institutional arrangements related to tick size, lot size, and intermediation.

Changes in institutional arrangements in the U.S. equity market may have affected the order-splitting multiplier ξ_{jt} and volume multiplier ζ_{jt} over the 1993–2014 sample period. If these multipliers changed through time, they affect only the constant terms in equations (9) and (10), when expressed as OLS regressions, and not the slope coefficients on trading activity W_{jt}. If these multipliers are correlated with trading activity, they induce omitted-variable bias.

Progress in computing technology and regulatory changes have significantly affected trading. The degree of order splitting varies with order size, ticker symbol, prices, and time. Our results imply that less order splitting occurred in the earlier part of our sample, and large bets may result in more prints than small bets.

At the NYSE in the early 1990s, traders typically executed large bets as block trades in the "upstairs" market, in which case at least one side of a reported block trade might correspond precisely to a bet. Prior to the changes in order handling rules in 1997, NASDAQ dealers often took the other side of entire bets because customers themselves could not place their own or-

[5]Amihud (2002) defines an illiquidity measure as the ratio of the absolute value of daily returns $|r|$ to daily dollar volume $V \cdot P$. This measure can be thought of as an empirical proxy for the market impact parameter λ in Kyle (1985), scaled by P^2 to represent the percentage price change per dollar traded. The value of λ is equal to the ratio of fundamental volatility σ_v to the standard deviation of order imbalances σ_u, and also scaled by P^2 to represent the percentage price change per dollar traded. The absolute value of daily returns $|r|$ in the numerator of the Amihud illiquidity ratio is a good proxy for a theoretical variable σ_v / P. The daily volume $V \cdot P$ in the denominator of the Amihud ratio is a reasonable proxy for a theoretical variable $\sigma_u \cdot P$ only if the number of bets γ executed per day is constant across time and assets. While equations for share volume $V \cdot P = \gamma \cdot |Q| \cdot P$ and the standard deviation of order imbalances $\sigma_u \cdot P = \gamma^{1/2} \cdot |Q| \cdot P$ show that $V \cdot P$ is proportional to σ_u only when γ is constant, our tests reject the hypothesis of invariant print frequencies.

ders into a central limit order book and NASDAQ dealers were unhappy if customers "bagged" them by dumping many blocks into the market one after another. As the use of the NYSE's Direct Order Transfer (DOT) system became more commonly used by professional traders in the 1990s, the use of electronic order submission strategies and order splitting increased. For NASDAQ stocks, this practice accelerated after new order handling rules were implemented in the late 1990s. Splitting thus was more significant in the first half of our sample (2001–2014) than in the second half (1993–2000).

Another important market friction is tick size. In May and June of 1997, the major U.S. stock exchanges reduced tick sizes from 12.5 cents (1/8 of a dollar) to 6.25 cents (1/16 of a dollar). Before that, tick sizes were also reduced for some subsamples of stocks. In September 1992, the AMEX reduced tick sizes from 1/8 of a dollar to 1/16 of a dollar for stocks priced under five dollars. In 1995, NASDAQ stocks with bid quotations at or above ten dollars per share were traded with tick sizes of 1/8 of a dollar, and NASDAQ stocks with bid quotations below ten dollars per share were traded with tick sizes of 1/32 of a dollar, as mentioned in Bessembinder (2000).

In 2001, the tick size was further cut from 6.25 cents (1/16 o a dollar) to one cent. As a result, some quoted bid-ask spreads decreased, and fewer shares were shown at the best bid and best offer. Traders used electronic interfaces to place scaled limit orders of small size at adjacent price points separated by one cent, and this led to smaller print sizes for bets of the same size.

Regulation National Market System (NMS), introduced in 2005, further encouraged market fragmentation and competition among multiple trading venues based on speed and efficiency of electronic interfaces, which led to significant order shredding across both time and trading venues. Over the period 2001–2014, continued improvements in computer technology have widened the use of electronic order handling systems and have made it practical to shred bets for many thousands of shares into tiny pieces of 100 shares or fewer.

Several institutional details associated with trade reporting may have further influenced the magnitude of order splitting parameter ξ:

- Traders shred bets to hide their sizes, and this order splitting is likely to depend on the price level because prices affect the dollar value of a 100-share trade. Thus, splitting may depend more on the dollar value of a trade than the number of shares traded.

- Larger bets are likely to be matched against multiple bets of smaller sizes, possibly resulting in disproportionably more TAQ prints for larger bets than for smaller ones.

- "Tape shredding" affects trading patterns. As suggested by Caglio and Mayhew (2016), large orders may be broken up into more trades than small orders to generate additional revenue from sales of consolidated trade and quote data.

- The number of prints per order and their sizes may depend upon details of the procedures exchanges use to report transactions. Exchanges may report transactions either from the perspective of the active party placing the executable order or the passive party whose limit orders were executed. Sometimes the rules of reporting are asymmetric across buy and sell orders as well as across different trading venues and time periods.

- During the sample period, odd-lot transactions were executed through a separate odd-lot trading system, and these small trades were not reported for dissemination on the consolidated tape, as discussed by O'Hara et al. (2014).

- Although traders may have increasingly shredded orders into "odd lots" of fewer than 100 shares, some traders probably resist splitting orders into numerous odd lots.
- Under NYSE Rule 411(b), broker-dealer member firms are obligated to consolidate a customer's odd-lot orders if the share amount of such orders exceeded 100 shares; brokers-dealer member firms could combine (or "bunch") multiple odd-lot orders from different customers after prior approval. Other exchanges had similar provisions and brought enforcement cases against member firms that did not comply with those rules.
- Buy and sell orders are treated asymmetrically. According to the Consolidated Tape Association (CTA), the exchanges are required to collect and report last sale data (CTA Plan (1992) Section VIII). At the NYSE, for example, the member representing the seller has to ensure that a trade has been reported. Because the rules required reporting of "sales" and not "trades," order splitting may be intrinsically more important for intended buy orders than intended sell orders.

The volume multiplier most likely also varies with intended order size, ticker symbol, and time. In 1993, the volume multiplier was probably larger for orders traded on NASDAQ than for orders traded on the NYSE. Atkyn and Dyl (1997) claim that because NASDAQ dealers were either buyers or sellers in almost every trade on NASDAQ, the NASDAQ trading volume was inflated by at least a factor of 2 relative to the number of trades actually occurring between end investors. Over time, these patterns may have changed, as dealers' participation rates in trade facilitation has decreased and trades from other trading systems have begun to be reported on the consolidated tape through the NASDAQ system.

Technological developments have most likely increased the amount of intermediation in securities markets during 2003–2014. The number of TAQ prints has soared because of order splitting and intermediation by high-frequency traders, who now account for a significant share of trading volume, as described by Chordia et al. (2011) and Hendershott et al. (2011). For example, Kirilenko et al. (2017) find that high-frequency traders account for more than 30% of stock index futures trading volume but hold their inventories for only a few minutes. Smaller print sizes could also be the result of a larger number of liquidity suppliers providing the same aggregate size using smaller orders.

Another important market friction for the extent of intermediation is the tick size. The primary channel by which tick size influences trade sizes is by increasing the value of secondary precedence rules. When the secondary precedence rules are important, market makers may submit substantial size to obtain precedence for larger portions of their orders than they would if the secondary precedence rules were not important. Changes in tick size affect trading decisions by changing incentives to provide liquidity and shred orders, as discussed in Harris (1994). When volatility is high and stock price is high, the tick size is small relative to a typical day's trading range, and better opportunities thus exist for order splitting and intermediation. Although firms can implement stocks splits to adjust percentage tick sizes, these adjustments occur infrequently and with long time lags, as noted by Angel (1997).

2. Data

We use the NYSE TAQ database, which contains trades and quotes reported on the consolidated tape by each participant in the Consolidated Tape Association (CTA) for all stocks listed

on all exchanges during the entire 1993–2014 sample period. The data are accessed through Wharton Research Data Services (WRDS),

2.1. Data description

Because we do not attempt to sign trades as buys (positive) or sells (begative) based on whether they are executed at the bid or ask price, our analysis employs only data on trades, not quotes. For each trade, the dataset contains the time, exchange, ticker symbol, number of shares traded, execution price, trade condition, and other parameters. There are about 50 billion records, with the number of data entries exponentially increasing over time from over 5 million records per month in 1993 to over 400 million records per month in 2014.

We transform the very large raw data files into a smaller dataset. First, bad records are removed using standard filters. The TAQ database provides information about the quality of recorded trades using condition and correction codes. We eliminate prints with condition codes 8, 9, *A*, *C*, *D*, *G*, *L*, *N*, *O*, *R*, *X*, *Z* or with correction codes greater than 1. The correction code 8 indicates, for example, that the trade was canceled.

Second, the remaining prints are aggregated in a specific way to reduce the size of the data set while preserving information about the monthly distributions of trade sizes. For each ticker symbol and each day, each print is placed into one of 55 bins constructed based on the number of shares traded. Letting X denote the size of a print in shares, "even" bins correspond to prints of the following exact "even" sizes of $X = 100$, $X = 200$, $X = 300$, $X = 400$, $X = 500$, $X = 1,000$, $X = 2,000$, $X = 3,000$, $X = 4,000$, $X = 5,000$, $X = 10,000$, $X = 15,000$, $X = 20,000$, $X = 25,000$, $X = 30,000$, $X = 40,000$, $X = 50,000$, $X = 60,000$, $X = 70,000$, $X = 75,000$, $X = 80,000$, $X = 90,000$, $X = 100,000$, $X = 200,000$, $X = 300,000$, $X = 400,000$, and $X = 500,000$ shares. "Odd" bins correspond to prints with trade sizes X between adjacent even bins—that is, $X < 100$, $100 < X < 200$, ..., $400,000 < X < 500,000$, and $500,000 < X$. Note that the size of bins grows approximately at a log-rate. Prints with even sizes are considered separately because trades tend to cluster in round-lot sizes.

The result is a much smaller dataset storing the number of trades by day, ticker symbol, and bin. To simplify the subsequent analysis, we make the approximate assumption that the assumed average print size in a bin (in shares) is equal to a midpoint of that bin. If print size is larger than 500,000 shares, we assign it to the 55th bin and assume its size to be 1,000,000 shares. This simplified aggregation makes it possible to capture the most important properties of print-size distributions while implementing our analysis efficiently. The convenience comes, however, at the expense of introducing some additional noise, which may slightly affect results. As a robustness check in the Appendix, we present the results of our tests for different percentiles for three periods: April 1993, April 2001, and April 2014 for the data aggregated by bins in Table A.7 and for the raw data in Table A.8; all results are quantitatively similar to each other. Also, Table A.17 shows that the dollar volume computed from the aggregated data and the raw data are very similar to each other for all bins; Table A.18 provides further results.

For each day and each ticker symbol, the small database also stores the open price; the close price; the number of trades per day; the dollar volume per day; the share volume per day; the close-to-close return; and the volatility, defined as the daily standard deviation of returns over the past 20 trading days from the TAQ data.

Then, we use the data collected for each stock, each size bin, and each day to average the frequency distributions of scaled print sizes and to construct an empirical distribution of print

sizes (in shares) for each stock and each month in the sample. Aggregation by month makes it possible to build better empirical approximations to theoretical distributions of bet sizes, given that many inactively-traded stocks have very few daily prints.

In addition to calculating the average number of prints per day, we calculate several statistics describing the possibly complicated shape of the distribution of print sizes. We consider the average print size and various percentiles of trade-size distributions. We refer to these percentiles as *trade-weighted percentiles*. For example, the xth trade-weighted percentile corresponds to a print size such that prints with sizes below this threshold constitute x% of all prints for a given stock in a given month. Note that trade-weighted percentiles effectively put the same weight onto prints of different sizes, which tends to emphasize small trades. For example, if the sample contains 99 prints of 100-share lots and one print of 100,000 shares, then the distribution of print sizes is mostly concentrated at the 100-share level. All trade-weighted percentiles below the 99th percentile are equal to 100 shares. The total trading volume and average print size, however, are largely determined by one big print of 100,000 shares.

Because large trades are economically more important than small trades, we also investigate the right tail of print-size distributions in more detail by examining *volume-weighted percentiles* based on trades' contributions to total volume. The contribution to the total volume by trades from a given print-size bin is calculated based on its midpoint.

The volume-weighted distributions give the percentage of trading volume resulting from prints of different sizes. The xth volume-weighted percentile corresponds to a trade size such that trades with sizes below this threshold constitute x% of total trading volume. In the example in the previous paragraph, percentiles 1–9 are 100 shares and percentiles 10–99 are 100,000 shares.

We report empirical results for both trade- and volume-weighted distributions. Of course, if we know a trade-weighted distribution of print sizes, then we can easily calculate a volume-weighted distribution as well. For the purpose of comparing trade-weighted and volume-weighted distributions, the log-normal is a useful benchmark. One can show (using a change in probability measure) that if the log of trade-weighted print size is distributed as $\mathcal{N}(\mu,\sigma^2)$, then the log of volume-weighted print size is distributed as $\mathcal{N}(\mu+\sigma^2,\sigma^2)$. The only difference between the two distributions is the shift in mean; the log-variance remains the same.

To acquire share and exchange codes for stocks in our sample, the monthly data are matched with the Center for Research and Security Prices (CRSP) data, which is accessed through WRDS. Only common stocks listed on the NYSE, AMEX, and NASDAQ from 1993 through 2014 are included in our study. Stocks that had splits or reverse splits in a given month are eliminated from the sample for that month. For each stock and each month, using the TAQ data, we compute average daily volume (in dollars and in shares), average price, and the standard deviation of close-to-close daily returns.

As robustness checks, we also rerun our tests for two alternative measures of daily returns volatility. First, we use an ARCH(5) model to predict one-day ahead volatility. Second, we compute daily volatility by using the standard deviation of the 5-day log return cumulative from the previous 50 trading days. In both cases, the average daily volume is also recalculated over the same corresponding trading window (5-day or 50-days) as volatility for consistency. We convert daily volatility and daily volume measures to monthly numbers by averaging them within the month. As a robustness check, we also implemented our tests on the data aggregated over annual frequency rather than monthly frequency. Tables A.9, A.10, A.11, and A.12 report the

robustness results for volatility estimations, elimination of 10-share stocks, and annual data aggregation. These results are very similar to our main results.

In Table A.19 and Figure A.13 in the Appendix, we also present our results for alternative measures of volatility adjusted for tick size following Harris (1990). The results remain quantitatively similar.[6]

Our final sample includes 1,383,857 stock-month observations, covering on average about 5,262 stocks for each of the 263 months between February 1993 and December 2014.

2.2. Descriptive statistics

We present descriptive statistics in Table 1. Panel A reports statistics for the 1993–2000 subperiod. Panel B reports statistics for the 2001–2014 subperiod. Statistics are calculated for all securities in aggregate, as well as separately for 10 groups of stocks sorted by average dollar volume. Instead of dividing the securities into 10 deciles with the same number of securities, volume break points are set at the 30th, 50th, 60th, 70th, 75th, 80th, 85th, 90th, and 95th percentiles of dollar volume for the universe of stocks listed on the NYSE with CRSP share codes of 10 and 11. Group 1 contains stocks in the bottom 30th percentile. Group 10 contains stocks in the top 5th percentile. Group 10 approximately covers the universe of S&P 100 stocks. The top five groups approximately cover the universe of S&P 500 stocks. Smaller percentiles for the more active stocks make it possible to focus on the stocks that are the most economically important. For each month, the thresholds are recalculated and stocks are reshuffled across groups.

Summary statistics before 2001. Panel A of Table 1 reports statistical properties of securities and prints in the sample before 2001. For the entire sample of stocks, the average trading volume is \$6.186 million per day, ranging from \$0.15 million for the lowest volume decile to \$176.99 million for the highest volume decile. The average volatility for the entire sample is 4% per day. The volatility tends to be higher for smaller stocks. The volatility is 4.5% for the lowest-volume decile and 2.9% for the highest-volume group. Thus, the measure of trading activity, equal to the product of dollar volume and volatility, increases from $0.15 \cdot 0.045$ to $176.99 \cdot 0.029$—that is, by a factor of 760 from the lowest- to the highest-volume group.

Before 2001, the average print size is equal to \$23,629, ranging from \$11,441 for low-volume stocks to \$89,338 for high-volume stocks. This corresponds to a decrease from 7.6% to 0.05% of daily volume from the lowest- to the highest-volume group. The median is much lower than the mean, as large prints make the distribution of print sizes positively skewed. The trade-weighted median ranges from \$5,682 for low-volume stocks to \$28,567 for high-volume stocks, corresponding to a decrease from 3.8% to 0.016% of daily volume. Note that the invariance hypothesis predicts that the shape of the distributions of trade sizes as a fraction of daily volume must be similar across stocks, with the only difference that their log-means are shifted downward by two-thirds of the increase in a log-trading activity (equation (10)). Because trading activity increases by a factor of 760 from the lowest to the highest deciles, a back-of-the-envelope calculation suggests that the distributions of trade sizes as a fraction of volume should be shifted downward by a factor of $760^{2/3} \approx 80$. This estimate is less than the observed differences in means of $7.6\%/0.05\% \approx 150$ and medians of $3.8\%/0.016\% \approx 240$ between the highest-

[6]The variance adjustment is $\sigma^2 - \frac{1}{6} \cdot \left(\frac{\text{Tick Size}}{P}\right)^2$.

and lowest-volume groups.

In the 1993–2001 subperiod, the average number of prints recorded per day is 142 for the entire sample, increasing monotonically from 17 to 2,830 from the first to the tenth volume group. The number of prints increases by a factor of $2,830/17 \approx 166$. Using the invariance hypothesis, we predict that the expected number of prints should increase by two-thirds of the increase in trading activity (equation (9))—that is, $760^{2/3} \approx 80$. This back-of-the envelope calculation suggests that the number of prints increases more than predicted, potentially reflecting a more intensive order splitting in high-volume groups where the stocks also tend to have higher price levels, but further investigation is certainly warranted.

Some print sizes are unusually common in the TAQ data. Before 2001, even-sized trades account for over 61% of volume traded and 80% of trades executed. The fraction of even-prints is stable across volume groups. The prevalence of these prints validates our choice of bins with even-share bins considered separately. About 16% of all transactions and 2% of volume traded are executed in 100-share prints. These trades represent 15% of transactions for low-volume stocks and 25% of transactions for high-volume stocks. The number of 1,000-share prints is also significant. The large fraction of 1,000-share prints for low-volume stocks relative to high-volume stocks—18% versus 14%—probably reflects the regulatory rule according to which the NASDAQ market makers had to post quotes for at least 1,000 shares prior to 1997.

It may also be partially related to the difference in typical price levels for low-volume and high-volume stocks, which makes 1,000-lot trades to be of a smaller dollar value for stocks with lower volume.

Summary statistics after 2001. Panel B of Table 1 provides statistical properties of the sample after 2001. Daily volume more than tripled from $6.18 million before 2001 to over $24.4 million after 2001. Volatility decreased from 4.0% to 3.1% per day. These numbers imply that trading activity doubled from the 1993–2001 to 2001–2014 subperiods. The average number of prints increased by a factor of 21 from 143 to 3,013, and the average print size decreased by more than a factor of 3 from $23,629 to $6,424. Back-of-the-envelope calculations implied by the invariance hypothesis suggest that the changes in print arrival rates and print sizes cannot be explained only by differences in levels of trading activity between the two subperiods but must be attributed to other factors. One of these factors is the order splitting that became increasingly prevalent over time, especially after the reduction in tick size to one cent on January 29, 2001, for NYSE stocks and on April 9, 2001, for NASDAQ stocks. The growth of algorithms and tape shredding have likely influenced patterns of trading as well. During the 2001–2014 subperiod, for example, 100-share trades account for 57% of all transactions and 25% of volume traded, with these numbers reaching their peaks at 73% and 41%, respectively, in 2013. Trades of 1,000 shares became less important in this half of the sample. Evident order splitting is likely to significantly affect tests of the invariance hypothesis using TAQ data after 2001.

Frequency and sizes of TAQ prints during the 1993–2014 period.

In Figure 1, we plot the 263 scaled numbers of monthly total prints, calculated as $N_{m,i} \cdot W_i^{-2/3}$, where $N_{m,i}$ is the total number of prints per month and W_i is trading activity. Figure 1 also shows the averages of the 20th, 50th, and 80th percentiles of the trade- and volume-weighted distributions of logs of scaled print sizes, calculated as $\ln(|\tilde{X}_i|/V_i \cdot W_i^{2/3})$, where $|\tilde{X}_i|$ is the print size. To facilitate comparison across stocks and across time, all variables are scaled by

$W_i^{-2/3}$ and $W_i^{2/3}$ as implied by the invariance hypothesis. The left panel shows the scaled variables averaged across low-volume stocks (group 1). The right panel shows the scaled variables averaged across high-volume stocks (groups 9 and 10).

A comparison of the left column to the right column of Figure 1 reveals that the profiles of volume-weighted percentiles of scaled trade sizes and the scaled number of prints per month are very similar across stocks. This similarity is evidence in favor of the invariance hypothesis.

Trading patterns differ significantly across the 1993–2000 and 2001–2014 subperiods. For high-volume stocks, the percentiles of print sizes and print rates do not change much prior to the beginning of decimalization in 2001. Afterward, percentiles of print size decrease steadily, and the average number of prints correspondingly increases. For low-volume stocks, similar changes started to occur even before 2001. Because most low-volume stocks are NASDAQ stocks, the pre-2001 decrease in print sizes and increase in print arrival rates may be explained by the reduction in tick size from 12.5 cents (1/8 of a dollar) to 6.25 cents (1/16 of a dollar) announced at NASDAQ in 1997. Also, for some subsets of low-priced NASDAQ and AMEX stocks, tick sizes were reduced from 1/8 to either 1/16 or 1/32 as earlier as in 1992 and 1995. With the exception of the largest print sizes in high-volume stocks, the downward trend in scaled print sizes and the upward trend in scaled number of prints seems to end at around 2007, and all variables stabilize at some constant levels. A similar pattern can be found in figures in Hendershott et al. (2011). In the following sections, we examine these patterns in more detail.

3. Results

The invariance hypothesis and the two alternative models make distinctively different predictions—all three nested in equations (9) and (10)—concerning the differences in the distributions of print sizes and their frequencies across stocks and time. We run our tests based on both the number of prints and distributions of their sizes to determine which models provides the most reasonable description of the data.

3.1. Tests based on print frequency

Comparison of three models. According to each model, the scaled number of prints $W_i^{-\alpha_n} \cdot N_i$ per day is constant across stocks. The three models differ only in the exponent a used to normalize the average number of prints. The invariance hypothesis implies $\alpha_n = 2/3$, the model of invariant print size implies $a_n = 0$, and the model of invariant print frequency implies $\alpha_n = 1$.

Figure 2 shows how the log-number of prints $\ln(N \cdot W^{-\alpha_n})$ changes with trading activity W, when the number of prints are scaled in three different ways, respectively, according to predictions of the three models. The figure has three columns and four rows. Each of the three columns contains plots of the log of the average number of prints per day N against the log of trading activity W, where the average number of prints is scaled according to each of the three models, respectively. The four rows present results for the NYSE in April 1993, for NASDAQ in April 1993, for all exchanges in April 2001, and for all exchanges in April 2014. Results are presented for different periods because trading has changed dramatically over time. Also, trading patterns of the NYSE- and NASDAQ-listed stocks differed historically because of differences in regulatory rules across exchanges during the earlier part of the sample; therefore, our results are presented separately for the NYSE- and NASDAQ-listed stocks in April 1993. We choose the month of April to avoid seasonality, as trades tend to cluster much less before the

end of the calendar quarter, as shown by Moulton (2005). Each observation corresponds to the average number of prints per day for a given stock in a given month. Each subplot has about 6,000 observations. If the model is correctly specified, the points are expected to line up along a horizontal line.

In subplots for the invariance hypothesis, observations are scattered around horizontal lines for each of the three years. The invariance hypothesis explains the data very well, especially for the NYSE stocks traded in April 1993. The levels of the horizontal lines move up from the top to bottom figures, showing that the average number of prints has increased over time. For April 1993, the average number of prints is slightly higher for the NYSE stocks than for the NASDAQ stocks. Some NASDAQ stocks with low trading activity have outliers with a very small number of prints. A few of the most illiquid NASDAQ stocks also have outliers with a very high number of prints in April 1993. The outliers for the small stocks may be due to the short sample period, as one month may be not enough to get a good estimate of mean volume for small illiquid stocks.

In subplots for the model of invariant print frequency, observations are lined up across a line with a positive slope. The model attributes all differences in trading activity entirely to differences in print sizes. Because changes in trading activity are also partially explained by changes in print arrival rates, the model tends to underestimate the number of prints for high-volume stocks and overestimate it for low-volume stocks.

In subplots for the model of invariant print size, observations are lined up along a line with a negative slope. The model attributes all differences in trading activity entirely to differences in trading rates. Because some part of these differences is actually explained by differences in print sizes, the model tends to overestimate the number of prints for high-volume stocks and underestimate it for low-volume stocks.

OLS estimates of the number of TAQ prints.

The distinctly different predictions of the models can be nested into a simple linear regression,

$$\ln(N_i) = \mu_n + a_n \cdot \ln\left(\frac{W_i}{W_*}\right) + \tilde{\epsilon}_i, \tag{11}$$

where W_* is an arbitrary scaling constant specified below. The equation relates the log of the average daily number of prints N_i for each month for stock i to the level of trading activity W_i, where W_i is average daily notional volume times standard deviation of daily returns for stock i for each month. The invariance hypothesis predicts $a_n = 2/3$, the model of invariant print frequency predicts $a_n = 0$, and the model of invariant print size predicts $a_n = 1$. For each month, we estimate the parameters μ_n and a_n using an OLS regression with one observation per stock per month. The constant term μ_n is scaled to represent the log of the expected number of prints for a benchmark stock with trading activity W_*. The scaling constant $W_* = (40)(10^6)(0.02)$ measures trading activity for an arbitrary benchmark stock with a price of \$40 per share, trading volume of 1 million shares per day, and daily volatility of 2% per day.

Figure 3 shows the time series of coefficients from the monthly regressions (11) between February 1993 and December 2014. The time series is constructed with 263 month-by-month regression coefficients a_n from monthly regression equations (11). We denote the estimate of coefficients as $\hat{\mu}_{n,T}$ and $\hat{a}_{n,T}$ for month T. The superimposed horizontal line represents $a_n = 2/3$, the value predicted by the invariance hypothesis. The figure shows two distinctive sub-periods, 1993–2000 and 2001–2014. Both the constant term μ_n and the coefficient a_n

changed over time and may have time trends that are different over the 1993–2000 and 2001–2014 subperiods. Over the 1993–2000 subperiod, all estimated coefficients a_n remained virtually constant. The average value of 0.666 is strikingly close to the predicted values of 2/3. Over the 2001–2014 subperiod, the estimates began to drift up from the values implied by the invariance hypothesis, increasing from about 0.65 to about 0.77 by the end of 2014. Around 2009, the upward trend in the time-series estimates slowed down and even reversed its direction; the explanation for this change remains unclear.

Following the approach of Fama and MacBeth (1973), Table 2 presents estimates of the coefficients in the pooled-over-time regression (11), adjusted for time trends visible in Figure 3. The six columns show the results for all stocks, the subsets of NYSE/AMEX-listed stocks, and the subset of NASDAQ-listed stocks, each shown during the 1993–2000 and 2001–2014 subperiods. To account for time trends in monthly coefficients, we calculate the Newey–West standard errors with three lags relative to a linear time trend, estimated by OLS regressions from the estimated coefficients $\hat{\mu}_{n,T}$ and $\hat{a}_{n,T}$ for each month T. The linear time trend is estimated from the equations $\hat{\mu}_{n,T} = \mu_{n,0} + \mu_{n,t} \cdot (T - \bar{T})/12 + \tilde{\epsilon}_T$ and $\hat{a}_{n,T} = a_{n,0} + a_{n,t} \cdot (T - \bar{T})/12 + \tilde{\epsilon}_T$, where T is the number of months from the beginning of the subsample and $\bar{T}$ is the mean month in the subsample. For the subperiod February 1993 to December 2000, $T = 1$ for February 1993 and $T = 95$ for December 2000—that is, $\bar{T} = 48$. For the subperiod January 2001 to December 2014, $T = 1$ for January 2001 and $T = 168$ for December 2014—that is, $\bar{T} = 84.5$. Table 2 contains estimates of $\mu_{n,0}$, $\mu_{n,t}$, $a_{n,0}$, and $a_{n,t}$.

For the 1993–2000 subperiod, the point estimate of $a_{n,0}$ is equal to 0.666, statistically indistinguishable from the predicted value of 2/3. For the subperiod 2001–2014, the point estimate of $a_{n,0}$ is equal to 0.79. The standard errors of these estimates are 0.002 and 0.005, respectively. Note also that the alternative models predicting $a_{n,0} = 0$ and $a_{n,0} = 1$ are clearly rejected. For the period 1993–2000, the estimated time trend coefficient $a_{n,t}$ of −0.001 per year is statistically insignificant. The estimates for $a_{n,0}$ of 0.626 and 0.76 for NYSE stocks during the two sub-periods 1993–2000 and 2001–2014, respectively, are smaller than the corresponding estimates for $a_{n,0}$ of 0.679 and 0.816 for NASDAQ stocks.

For the 2001–2014 period, the estimated time trend coefficient $a_{n,t}$ of 0.007 per year is statistically significant; it approximately corresponds to the increase in a_n by 0.11, from about 0.66 to about 0.77, over the 13-year period.

The point estimate of intercept $\mu_{n,0}$ is equal to 6.147 for the 1993–2000 subperiod and 8.513 for the 2001–2014 subperiod, indicating an increase in the average number of prints over time. The constant terms $\mu_{n,t}$ of 0.093 and 0.148 for those subperiods also show statistically significant upward time trends, corresponding to growth rates in the number of prints per year of about 9.3% and 14.8% for the two subperiods, respectively.

The combined results for both subperiods are consistent with the following interpretation: Over the 1993–2000 subperiod, the invariance hypothesis held due to a reasonably close correspondence between TAQ prints and bets. Over the 2001–2014 subperiod, the invariance relationship broke down because order splitting and intermediation have increased over time and affected high-volume stocks more than low-volume stocks after decimalization in 2001. These results may be also related to differences in degrees of order splitting and intermediation across low-price and high-price stocks. We discuss these patterns further in Subsection 3.3.

3.2. Tests based on TAQ print sizes

Comparison of three models. We now examine the trade- and volume-weighted distributions of print sizes scaled for differences in trading activity as suggested by the three models. The three models predict that the distributions of $W^{\alpha_x} \cdot |\tilde{X}|/V$ are constant across stocks and time, but the models make different assumptions about the exponent α_x. The invariance model predicts $\alpha_x = 2/3$, the model of invariant print frequency predicts $\alpha_x = 0$, and the model of invariant print size predicts $\alpha_x = 1$.

To approximate the distribution, we calculate print size $|X|$ based on the mid-point of a print-size bin where it was placed. For each month and for each volume group, the stock-level distributions of scaled print sizes are combined by averaging across stocks in each volume group the frequency distributions of the number of prints in each bin. The results are plotted in Figures 5 and 6. For illustrative purposes, only results for April 1993 are presented, but the data for the entire 1993–2014 period are examined more closely below.

Figure 5 shows the distributions of logs of scaled print sizes for the NYSE stocks. The figure has three rows and six columns. The three rows contain plots for low-volume stocks in volume group 1, medium-volume stocks in volume groups 2 through 8, and high-volume stocks in volume group 9–10, respectively. The first three columns contain plots of the trade-weighted distributions, with the density of logs of scaled print sizes on the vertical axis. The second three columns contain plots of the volume-weighted distributions, with the volume contribution of these trades on the vertical axis. In each of the three columns, print sizes are scaled according to the three models. If one of the three models is correct, then the three distributions in the column corresponding to that model should be the same across rows.

To make it easier to interpret results, we superimpose the bell-shaped densities of a normal distribution with the common means and variances equal to the means and variances of trade- and volume-weighted distributions of scaled print sizes based on the entire sample. As previously discussed, if trade-weighted distributions are log-normally distributed, then volume-weighted distributions are log-normally distributed as well. If scaled sizes are distributed as a log-normal, then the three plots in each column of plots are expected to coincide with the superimposed common normal density.

The first column presents the three trade-weighted distributions implied by the invariance hypothesis; they have similar means, variance, and supports. The shapes of the distributions bear some resemblance to the superimposed normal density, but the fit is by no means exact. The low-volume group matches the superimposed normal better than the medium- and high-volume groups. The fourth column presents the volume-weighted distributions implied by the invariance hypothesis; they are even more similar to the superimposed common normal density. Thus, the invariance hypothesis explains a substantial part of variation in the distribution of print sizes, especially in the distribution of economically important large trades. The print sizes seem to be distributed similarly to a log-normal; Kyle and Obizhaeva (2016) find that the distribution of portfolio-transition orders is close to a log-normal as well.

For the model of invariant print frequency, the trade-weighted densities are in the second column and the volume-based densities are in the fifth column; they are much less stable across volume groups. In both columns, the distributions shift to the left as trading volume increases, which suggests that the first alternative model understates print sizes for high-volume stocks and overestimates them for low-volume stocks. The model fails to account for the fact that some variation in trading activity is explained by variation in the number of prints.

For the model of invariant print size, the trade-weighted densities are in the third column and the volume-based densities are in the sixth column; they are clearly unstable across volume groups as well. In both columns, the distributions shift to the right as trading volume increases. The second alternative model overstates print sizes for high-volume stocks and underestimates them for low-volume stocks. The alternative models clearly provide worse explanations for the observed variations in print sizes than the invariance hypothesis.

Figure 6 shows our results for the sample of NASDAQ stocks. Similar to the NYSE stocks, the distributions of print sizes are more stable across volume groups when print sizes are scaled according to the invariance hypothesis. Compared with the NYSE distributions, the NASDAQ distributions are less smooth and have more spikes, especially the trade-based densities. We attribute these patterns to a regulatory rule that required NASDAQ dealers to quote prices for at least 1,000 shares, leading to a disproportionably large number of 1,000-share NASDAQ trades recorded on the consolidated tape before 1997.

Implications for log-normal distributions. The invariance hypothesis makes predictions about the mean of the distribution of bet size, not the shape of the distribution. Calibrating the shape of distributions is nevertheless important. Kyle and Obizhaeva (2016) find that the distribution of portfolio transition orders closely resembles a log-normal with a large log-variance and therefore heavy tails. Kyle and Obizhaeva (2018c) predict the size and frequency of stock market crashes based on extrapolating the frequency of tail events in the distribution of portfolio transition orders to bets on the market as a whole.

Comparing distributions of bet size and trade size may provide insights about the order splitting algorithms used by traders. Therefore, we examine whether the distribution of trade size resembles a log-normal by superimposing onto the graphs of the distributions of log of trade size a normal distribution with mean consistent with invariance and common log variance. Figures 5 presents the results for NYSE stocks, and Figure 6 presents the results for NASDAQ stocks. As previously discussed, if the log of trade-weighted scaled print sizes is distributed as $\mathcal{N}(\mu,\sigma^2)$, then the log of volume-weighted scaled print sizes is distributed as $\mathcal{N}(\mu+\sigma^2,\sigma^2)$.

For NYSE stocks, the constraint implies that the volume-weighted mean of 0.97 should be the same as the sum of the trade-weighted mean of -1.15 and its variance of 1.90. As $-1.15+1.90=0.75\neq 0.97$, we see that the constraint fails to hold by a margin of only about 25%. The log-variance of 3.04 for the volume-based distribution is much larger than the log-variance of 1.90 for the trade-weighted distribution. This discrepancy is inconsistent with log-normality, which implies that these log-variances should be the same.

For NASDAQ stocks, the volume-weighted mean of 1.2 should be the same as the sum of the trade-weighted mean of -0.19 and its variance of 1.93. As $-0.19+1.93=1.72\neq 1.2$, we see that this constraint fails to hold by a margin of about 30%. The log-variance of 1.98 for the volume-based distribution is similar to the log-variance of 1.93 for the trade-weighted distribution, consistent with the predictions of log-normality. Because these moment restrictions are not perfectly satisfied in the data, the hypothesis of log-normality can be valid only as a very rough approximation at best. Deviations from log-normality include clustering of trades in even-lot sizes (especially prints of 1,000 shares on NASDAQ), censoring and rounding of odd lots, clustering of 100-share trades, and the possibility that very large trades follow a fatter-tailed power-law distribution rather than a log-normal distribution.

OLS regression estimates of TAQ print sizes, February 1993 to December 2014. We test implications of the invariance hypothesis for print sizes using OLS regressions in which the left-side variable is either a mean or a percentile of either the trade- or volume-weighted distributions of logs of print sizes. For each stock in a given month, the trade- and volume-weighted distributions of logs of print sizes are constructed. Letting $f(.)$ denote a functional that corresponds to either the mean or the pth percentiles (20th, 50th, 80th) of these distributions, these variables are regressed on logs of trading activity:

$$f\left(\ln\left(\frac{\tilde{X}_i}{V_i}\right)\right) = \mu_x + a_x \cdot \ln\left(\frac{W_i}{W_*}\right) + \tilde{\epsilon}_i. \tag{12}$$

Equations (11) and (12) are closely related. Expected trading volume V is equal to the product of the expected number of prints N and expected print size $\mathrm{E}\left[\tilde{X}\right]$, $V = N \cdot \mathrm{E}\left[\tilde{X}\right]$, implying that the left side of equation (11) is $\ln N = -\ln\left(\mathrm{E}\left[\tilde{X}\right]/V\right)$. Thus, the left-hand side variable in equation (12) is similar to reversing the sign on the left-hand-side variable in the regression equation (11). For example, if $\tilde{X}/V$ were distributed log-normally with the same variance across stocks, then the coefficient estimates for a_n and a_x would be the same in absolute value and opposite in sign in all of the regressions in equations (12) and (11), but the constant terms μ_n and μ_x would be different. In our data (as we show below), $\tilde{X}/V$ deviates from a log-normal distribution sufficiently to make the coefficients a_n and a_x vary across regression specifications.

Figure 4 shows the time series of coefficients of the monthly regressions from February 1993 to December 2014. Plotted are the time series of 263 month-by-month regression coefficients a_x from regression equation (12) for the 20th, 50th, and 80th percentiles of print sizes over the 1993–2014 period. Panel A presents results for trade-weighted distributions. Panel B presents results for volume-weighted distributions. Superimposed horizontal lines represent the level of $-2/3$, the benchmark predicted by the invariance hypothesis.

As before, the figure shows two distinct subperiods. Over the 1993–2000, all estimated coefficients remain close to $-2/3$. The estimates of a_x are lower than the predicted value of $-2/3$ for the 20th, 50th, and 80th percentiles of trade-weighted distributions; they fluctuate between -0.70 and -0.82, implying that small print sizes as a fraction of volume decrease faster with trading activity than predicted by the invariance hypothesis. For the 50th and 80th percentiles of volume-weighted distributions, the estimates of a_x fluctuate between -0.45 and -0.70, somewhat higher than predicted by the invariance hypothesis, and all estimates for the 20th volume-weighted percentiles are close to $-2/3$.

Over the 2001–2014 subperiod, the estimates begin to drift away from their initial levels. For the volume-weighted percentiles, the estimates of a_x decrease from about -0.45 to -0.82 for the 80th percentile, from about -0.54 to -0.89 for the 50th percentile, and from -0.645 to -0.75 for the 20th percentile. For the trade-weighted percentiles, the estimates for the 20th and 50th percentiles do not exhibit any definite patterns, but the estimates for the 80th percentile decrease from about -0.71 to -0.77. As before, the behavior of estimated coefficients changes in 2009. Overall, changes in the large print sizes (right tails) are more significant than changes in the small print sizes (left tails).

Table 3 reports the estimates from regressions in equation (12) pooled over the 1993–2000 period. The first four columns show estimates for the means and percentiles of the trade-weighted distributions. The last four columns show estimates for the means and percentiles of the volume-weighted distributions. Because the monthly estimates of $\hat{\mu}_{x,T}$ and $\hat{a}_{x,T}$ for

each month T are changing over time, we add a linear time trend. The table reports Fama–MacBeth estimates of the coefficients, with Newey–West standard errors calculated with three lags relative to a linear time trend estimated by OLS regressions from the estimated coefficients $\hat{\mu}_{x,T}$ and $\hat{a}_{x,T}$ for each month. As before, the equations estimated for the time trend are $\hat{\mu}_{x,T} = \mu_{x,0} + \mu_{x,t} \cdot (T - \bar{T})/12 + \tilde{\epsilon}_T$ and $\hat{a}_{x,T} = a_{x,0} + a_{x,t} \cdot (T - \bar{T})/12 + \tilde{\epsilon}_T$, where T is the number of months from the beginning of the subsample and $\bar{T}$ is the median month in the subsample.

For the trade-weighted distributions, the estimate of $a_{x,0}$ is equal to -0.741 for the means. This estimate is larger in absolute value by 0.075 than the estimate of 0.666 for the number of prints in Table 2. For the trade-weighted percentiles, the estimated coefficients range from -0.781 for the 20th percentile to -0.725 for the 80th percentile. All of these estimates are larger in absolute value than the value of $-2/3$ predicted by the invariance hypothesis, which implies that print sizes as a faction of volume tend to decrease with trading activity faster than implied by the theory. For the volume-weighted distributions, the estimated coefficient $a_{x,0}$ is equal to -0.56 for the means; the estimates range from -0.661 for the 20th percentile to -0.481 for the 80th percentile. Across means and percentiles, the standard errors of estimates $a_{x,0}$ range from 0.002 to 0.003; these values are similar in magnitude to the averages of standard errors of a_x from the cross-sectional monthly regressions (12), in which those values range from 0.002 to 0.065. The data suggests that the invariance hypothesis $a_{x,0} = -2/3$ explains the data much better than the alternatives $a_{x,0} = 0$ and $a_{x,0} = -1$, but all three models are statistically rejected.

For the trade-weighted distributions, the estimated time trend coefficient $a_{x,t}$ ranges from 0.008 to 0.012 per year and is statistically significant. For the volume-weighted distributions, the time-trend coefficient is either statistically insignificant or negative.

The estimated intercept $\mu_{x,0}$ of -7.238 in the regression for the trade-weighted means implies that the median print size for the benchmark stock is equal to $\exp(-7.238)$, or 0.07% of daily volume. The estimated intercepts of -8.490 and -6.260 in the regressions for the 20th and 80th percentiles suggest that the average 20th and 80th print-size percentiles are equal to 0.02% and 0.19% of daily volume for the benchmark stock, respectively. Under the assumption of log-normality, Kyle and Obizhaeva (2016) note that the fraction of volume generated by trades larger than z standard deviations above the log-mean (which equals the median) is given by $1 - \mathcal{N}(z - \sigma)$, where σ is the standard deviation for the distribution of the log of trade sizes; based on the trade-weighted variance of 1.90 in Figure 5, log-normality would imply that about 91% of volume occurs in print sizes larger than 0.07% of daily volume (median trade). The standard errors of $\mu_{x,0}$ in cross-sectional regressions are similar in magnitude and range between 0.014 and 0.031, across means and percentiles. The negative and statistically significant estimates of time trend $\mu_{x,t}$ indicate that the print sizes have been gradually decreasing during the 1993–2000 subperiod, with a downward drift that is especially pronounced in the right tails of the distributions.

The R^2 is lower in regressions based on volume-weighted distributions than in regressions based on trade-weighted distributions. For the means, the values of R^2 are 0.90 and 0.69, respectively. The difference in R^2 increases monotonically from a difference between 0.87 and 0.83 for the 20th percentiles to a difference between 0.88 and 0.52 for the 80th percentiles. These numbers show more unexplained variation in large print sizes than in small print sizes. Note that some of this variation may result from the rounding of large odd-size trades to the mid-point of bins or from the small number of observations in the largest bins.

Table 4 reports the estimates for regressions in equation (12) for the 2001–2014 subperiod.

For the means, the estimates of $a_{x,0}$ are −0.779 for trade-weighted distributions and −0.776 for volume-weighted distributions. These estimates are larger in absolute value than the corresponding estimates of −0.741 and −0.560 for the 1993–2000 subperiod in Table 3. All but one of the estimates of −0.757, −0.769, and −0.811 for the 20th, 50th, and 80th trade-weighted percentiles, respectively, and −0.796, −0.843, and −0.779 for the 20th, 50th, and 80th volume-weighted percentiles, respectively, are also higher in absolute values than the estimates of −0.781, −0.750, −0.725, −0.661, −0.579, and −0.481 for the earlier subperiod. The biggest changes occur in the estimates for the 80th percentile of trade-weighted distributions and the 50th and 80th percentiles of volume-weighted distributions, which suggests that recent technological and regulatory changes had the largest effect on the right tail of print-size distributions. The standard errors of $a_{x,0}$ are between 0.003 and 0.004, similar to the averages of standard errors of a_x in monthly regressions (12). This fact validates the adjustment for time trend in the Fama–MacBeth procedure; without inclusion of a time trend, the standard errors would have been higher.

The estimates of the intercept $\mu_{x,0}$ for the 1993–2000 subperiod are lower than for the 2001–2014 subperiod for the means as well as for all percentiles. For the pooled sample, for example, the estimate of −9.029 in Table 4 is lower than the corresponding estimate of −7.238 in Table 3; these estimates imply that a typical print size for the benchmark stock fell from 0.07% to 0.01% of daily volume over the 1993–2000 subperiod. The estimated time trend $\mu_{x,t}$ is negative and statistically significant in all columns, except for the 20th percentile of trade-weighted distributions, also implying that the distributions of print sizes have been shifting downward.

3.3. Detailed analysis of market frictions over time

We next study how market frictions such as tick size and lot size affect the trading process. More specifically, we examine whether these market frictions can help explain variations across markets in the number and size distribution of prints that cannot be explained by the invariance hypothesis.

To facilitate our analysis, we introduce several new concepts that measure restrictiveness of market frictions in the spirit of the invariance hypothesis. We also discuss why these measures are related to price volatility and share volume in business time as well as to each other.

Effective price volatility and effective tick size. Tick size imposes a restriction on the minimum price change. The tick size changed from 1/8 of a dollar to 1/16 of a dollar in the late 1990s and then to one cent in 2001. To assess the restrictiveness of this friction, one can compare it with price volatility—that is, with a typical price change. Based on the same intuition, practitioners often measure price volatility in units of tick size. For example, if a $40 stock has a volatility of 2% per day, then daily dollar price volatility is equal to $0.80 = $40 · 0.02 and a tick size of $0.01 is equal to 1/80 of price volatility.

The invariance hypothesis suggests defining the relative tick size as a fraction of price volatility in business time, not in calendar time. Business time is proportional to the expected arrival rate of bets γ_{jt}, or in terms of trading activity it is also proportional to $W_{jt}^{2/3}$, as shown in equation (5). Hence, define *effective price volatility* in asset j and time t as

$$\text{Effective Price Volatility}_{jt} := P_{jt} \cdot \sigma_{jt} \cdot \left(\frac{W_{jt}}{W_*}\right)^{-1/3}. \tag{13}$$

This measure is scaled by a constant $W_*^{-1/3}$, previously defined as trading activity in the benchmark stock, so that effective price volatility is exactly equal to daily price volatility for the benchmark stock. In comparison with calendar-time volatility, effective price volatility is lower for more liquid stocks and higher for less liquid stocks to take into account that business time runs faster in more liquid securities.

Now define *effective tick size* as the ratio of the dollar tick size to effective price volatility:

$$\text{Effective Tick Size}_{jt} := \frac{\text{Tick Size}_{jt}}{\frac{P_{jt}\cdot\sigma_{jt}}{P_*\cdot\sigma_*}\cdot\left(\frac{W_{jt}}{W_*}\right)^{-1/3}}. \tag{14}$$

The presence of $P_*\cdot\sigma_*$ in the denominator scales the definition of relative tick size to make it exactly equal to the dollar tick size for the benchmark stock. A higher effective price volatility makes the effective tick size lower. We conjecture that lower effective tick size may encourage traders to shred meta-orders into a larger number of trades and may affect the amount of intermediation.

Effective share volume and effective lot size. Round lot size imposes a restriction on the minimum number of shares in prints on the tape. For most stocks in the sample, the lot size is equal to 100 shares. An odd lot comprises orders smaller than a round lot (e.g., 30 shares) or the non-round-lot portion of larger orders (e.g., the 30-share portion of a 530-share order); these odd-lot transactions are executed according to special, often less flexible rules, and information about them is not disseminated to the tape, as described in detail by Hasbrouck et al. (1993).

To assess how restrictive this friction is, one can compare it with the size of a median bet in the corresponding market. Equation (10) implies that median bet size is also proportional to share volume in business time, which we refer to as *effective share volume*:

$$\text{Effective Share Volume}_{jt} := V_{jt}\cdot\left(\frac{W_{jt}}{W_*}\right)^{-2/3}. \tag{15}$$

This measure takes into account that business time runs faster by a factor of $W_{jt}^{2/3}$ in more liquid securities. Scaling it by the constant $W_*^{-2/3}$ makes effective share volume equal to daily share volume for the benchmark stock.

Now define *effective lot size* as the ratio of the lot size to effective share volume:

$$\text{Effective Lot Size}_{jt} := \frac{\text{Lot Size}_{jt}}{\frac{V_{jt}}{V_*}\cdot\left(\frac{W_{jt}}{W_*}\right)^{-2/3}}. \tag{16}$$

Practitioners often measure order size as a fraction of daily volume and restrict their trading rate to a fixed fraction—say, 5%—of volume in order to control transaction costs. Scaling by V_* makes the effective lot size equal to the lot size for the benchmark stock. A lower effective share volume makes the effective lot size larger and therefore more binding, in the sense that a larger fraction of bets falls below that threshold. Some small bets may be executed as odd lots and not recorded on the consolidated tape, some may not be executed at all, and some may be rounded up to round-lot size. The extent of such censoring and rounding is expected to be related to effective lot size.

Since $W_{jt} = V_{jt} \cdot P_{jt} \cdot \sigma_{jt}$, the product of effective tick size (14) and effective lot size (16) equals to the product of tick size and lot size:

$$\text{Effective Tick Size}_{jt} \cdot \text{Effective Lot Size}_{jt} = \text{Tick Size}_{jt} \cdot \text{Lot Size}_{jt} \tag{17}$$

For most stocks in our sample (with one-cent tick size and 100-share lot size), the right-hand side of this equation is a constant equal to one dollar. It relates to the minimum possible dollar cost (and profit) per transaction in the market.

For a given month in the sample, dollar tick size and lot size are usually constant across most of the U.S. stocks, implying that effective tick size and effective lot size are closely related to each other and to effective price volatility. In the following analysis, we therefore examine how trading patterns differ across stocks with different levels of effective price volatility.

The effects of the two market frictions on the number of prints and print-size distributions are difficult to separate. For example, higher effective price volatility makes effective tick size lower, but it also makes effective lot size higher; lower effective share volume has similar implications. The first effect, operating through lower effective tick size, probably encourages more intermediation and more order splitting of bets into smaller trades placed at finer adjacent price points as a strategy to avoid front-running, leading to more prints of smaller sizes. The second effect, operating through higher effective lot size, probably encourages more censoring and rounding up of odd-lot trades, thus leading to fewer prints of larger sizes in the data sample.

Distributions of print sizes over time. We next examine how market frictions affect trading in the U.S. equity market cross-sectionally and through time. After sorting stocks into 10 volume groups and 4 effective price-volatility groups, we analyze distributions of the logs of scaled print sizes $\ln\left(\frac{X_i}{V_i} \cdot W_i^{2/3}\right)$, with the scaling factor $W_i^{2/3}$ implied by the invariance hypothesis. According to the invariance hypothesis, these 40 distributions are the same in frictionless markets. To examine how these distributions change over the entire sample period, we examine the shape of the distributions for the months April 1993, April 2001, and April 2014.

Because dollar tick size and lot size are constant during selected months, sorting stocks into effective price-volatility groups allows us to study the effects of variations in both effective tick size and lot size.

Trade-weighted distributions for NYSE-listed stocks, April 1993. In April 1993, NYSE-listed stocks were priced in increments of 12.5 cents (1/8 of a dollar). Most of the stocks were traded in multiples of 100-share lots, even though some of them were traded in multiples of 10-share lots, as described in Hasbrouck et al. (1993).

Figure 7 shows the trade-weighted distributions of logs of scaled print sizes for 5 of the 10 volume groups and all 4 effective price-volatility groups for the NYSE-listed stocks in April 1993. The 100-share trades are highlighted in light gray and 1,000-share trades in dark gray. The number of stocks in each subgroup and the average number of trades per day for these stocks are also reported. On each subplot, the density of a normal distribution with the mean of -1.15 and standard deviation of 1.38 is superimposed, calculated for the pooled sample in April 1993. If the invariance hypothesis holds, identification assumptions are valid, and bet sizes are distributed as a log-normal, then all distributions are expected to be invariant across 40 subplots and coincide with the superimposed normal density. For most subgroups, distributions are indeed close to the superimposed normal.

The data are clearly truncated below the 100-share odd-lot boundary and cluster in 100-share trades, shown in light gray, in the left tails of the distributions. Because variation in levels of dollar volume within groups is small, the 100-share trades usually fall into the same bin or into two adjacent bins. The only exception is the first volume group, where large variation in trading activity makes the 100-share trades spread over more than four bins.

The distributions have spikes because of clustering of trades at round-lot levels. Spikes are visible at the 100-share level and the 1,000-share level, marked by the clustering of light gray and dark gray columns. Spikes at 200-share and 500-share levels are also visible but not explicitly marked. For example, the sample contains more trades of 5,000 shares than 4,000 or 6,000 shares, and far more than 4,900 or 5,100 shares. For the subsample of stocks with low volume and low price volatility, large variation in trading activity smooths out spikes in the distribution of print sizes.

A visual inspection suggests that, holding effective price volatility fixed, the supports of distributions stay relatively constant across volume groups, but their shapes become more skewed to the right as volume increases, especially when price volatility is low, consistent with large orders being shredded into smaller trades.

Holding dollar volume fixed, the distributions vary across effective price-volatility groups in a systematic way as well. When effective price volatility increases, the 100-share boundary becomes more binding, the truncation threshold shifts to the right, and the effects of censoring and rounding to the 100-share boundary become more pronounced. At the same time, the relative tick size decreases, thus encouraging more order splitting. The first effect seems to dominate, because the average number of prints decreases with price volatility. For high-volume stocks, for example, the average number of prints decreases monotonically from 1,139 prints recorded per day for low-volatility stocks to only 285 prints for high-volatility stocks. In the absence of any market frictions, the number of trades is expected to be relatively constant within a given volume group because volatility does not vary much.

Volume-weighted distributions for NYSE-listed stocks, April 1993. Figure 8 shows the volume-weighted distributions of logs of scaled print sizes for the NYSE stocks in April 1993. In comparison with the trade-weighted distributions in Figure 7, the volume-weighted distributions put more weight onto larger trades and allow us to see more clearly the distribution of large print sizes.

According to Hasbrouck et al. (1993), a block of 10,000 shares was not a particularly large trade in 1993. Some block trades (of 10,000 shares or more), especially block trades in illiquid stocks, were executed in the upstairs market. The average size of upstairs-facilitated blocks was about 43,000 shares.

In comparison with the trade-weighted distributions in Figure 7, the volume-weighted distributions are more stable across subgroups and more closely resemble the superimposed normal distribution. On most plots, the space below the bell-shaped density function is filled up. The truncation at the odd-lot boundary is almost invisible, because the numerous 100-share trades almost "disappear" from the left tail of the distribution, as they contribute little to overall volume.

Small gaps in the distributions relative to a log-normal can be seen in mid-range print sizes between 1,000 shares and 10,000 shares. Perhaps these gaps represent intended orders shredded into smaller trades. The strong visual resemblance of the graphs to a log-normal, as in Kyle

and Obizhaeva (2016), are consistent with the interpretation that most of the largest orders in 1993 appear to have been executed as single blocks, generating large print sizes. An exception is the low-volume group for which the largest orders appear to be shredded because the distributions are skewed to the right. Although the volume-weighted distributions are much smoother than the trade-weighted distributions, small spikes are still detectable. These spikes, which likely correspond to clusters of trades at the levels of 1,000 shares, 5,000 shares, and 10,000 shares, are clearly visible, for example, in distributions for stocks with high volume and high effective price volatility. A few spikes in the far-right tails of several distributions suggest that occasional very large prints occur in the data more often than explained by log-normality.

Trade-weighted distributions for NASDAQ-listed stocks, April 1993. In April 1993, the NASDAQ-listed stocks usually had the minimum lot size of 100 shares. Quotes were restricted to increments of 1/8 of a dollar if the bid price exceeded $10.00, but trades were permitted in finer increments of 1/64 of a dollar for all stocks, even though these prices were then rounded to the nearest eighths for reporting, as described in Christie et al. (1994) and Smith et al. (1998).

Also, in 1988, the Securities and Exchange Commission required NASDAQ market makers to have a quotation size of at least 1,000 shares for most stocks. The rule mostly affected large stocks, and, indeed, we observe larger spikes in subplots for high-volume stocks. For small stocks, the rule was slightly different. For example, orders smaller than 1,000 shares could be executed through the Small Order Execution System (SOES) in stocks that were trading at prices lower than $250 per share. After 1996, the minimum quote size restriction was gradually removed. Under the Actual Size Rule, the minimum quote size was reduced from 1,000 to 100 shares, first for 50 pilot stocks in January 1997, then for an additional 104 stocks in November 1997, and finally for all others.

Figure 9 shows the trade-weighted distributions of logs of scaled print sizes for NASDAQ stocks in April 1993. The biggest difference between the trade-weighted distributions of the NASDAQ stocks and the NYSE stocks is the very large fraction of 1,000-share trades, shown as dark gray spikes, typically in the middle of the NASDAQ distributions. These spikes can be attributed to the requirement to quote at least 1,000 shares. In line with this explanation, we do not observe the clustering at the 1,000-share level after 2001 (unreported). Apart from the clustering in the 1,000-share level and truncation at the 100-share level, the distributions bear some resemblance to the superimposed normal distribution.

Trade-weighted distributions for all stocks, April 2001 and 2014. After decimalization in 2001, U.S. stocks traded in increments of one cent. The round-lot size is 100 shares; odd-lot trades account for a large fraction of trades, but they are not reported in the TAQ dataset. The U.S. stock market is fragmented into multiple trading venues. Angel et al. (2011; 2015) discuss other recent innovations in trading.

Figures 10 and 11 show the trade-weighted distributions of logs of scaled print sizes for all stocks traded in April 2001 and April 2014, respectively. In 2001, decimalization and use of electronic interfaces led to a significant increase in order splitting, the effect of which is clearly seen in both figures.

The frequency of trades increased significantly over time. For high-volume and low-volatility stocks, for example, the number of trades has increased from about 325 trades per day in April 1993 to 16,230 trades per day in April 2001 and 41,295 trades per day in April 2014. The distributions

of scaled print sizes shifted substantially to the left during the 1993–2014 period. Based on the means of superimposed normals, for example, the median print size dropped from 0.11% of daily volume for the NYSE stocks in April 1993 and 0.08% for the NASDAQ stocks to 0.003% in 2001 and only 0.001% in 2014. The market for block trades seems almost to have disappeared, and trading is now dominated by transactions of 100 shares. Indeed, trades of 100 shares constitute about 57% of trades executed and 38% of volume traded in 2014.

Note that the estimated log-variances of 2.05 in 1993, 1.78 in 2001, and 1.21 in 2014 for trade-weighted distributions of scaled print sizes are lower than the variance of 2.50 for the distributions of portfolio-transition orders, reported by Kyle and Obizhaeva (2016). A gradual decrease in log-variance is consistent with the hypothesis that large bets in liquid stocks in 2014 result in disproportionately more prints than small bets in illiquid stocks.

Regressions with effective price volatility. Table 5 presents Fama–MacBeth estimates of μ_n and a_σ from monthly regressions:

$$\ln(N_i) = \mu_n + \frac{2}{3}\cdot\ln\left(\frac{W_i}{W_*}\right) + a_\sigma\cdot\ln\left(\frac{P_i\cdot\sigma_i}{P_*\cdot\sigma_*}\cdot\left(\frac{W_i}{W_*}\right)^{-1/3}\right) + \tilde{\epsilon}_i. \tag{18}$$

The regression effectively imposes the invariance restriction of $a_n = 2/3$ in regression (11) and adds effective price volatility as an additional explanatory variable to capture the effect of both market frictions. The table reports Fama–MacBeth estimates of the coefficients, with Newey–West standard errors calculated with three lags relative to a linear time trend estimated by OLS regressions from the estimated monthly coefficients $\hat{\mu}_{n,T}$ and $\hat{a}_{\sigma,T}$ for each month. Specifically, the specification is $\hat{\mu}_{n,T} = \mu_{n,0} + \mu_{n,t}\cdot(T-\bar{T})/12 + \tilde{\epsilon}_T$ and $\hat{a}_{\sigma,T} = a_{\sigma,0} + a_{\sigma,t}\cdot(T-\bar{T})/12 + \tilde{\epsilon}_T$, where T is the number of months from the beginning of the subsample, and $\bar{T}$ is the mean month in the subsample. The six columns show the results for the entire sample as well as subsets of NYSE/AMEX-listed stocks and NASDAQ-listed stocks during the 1993–2000 and 2001–2014 subperiods.

The point estimates for $a_{\sigma,0}$ are negative and statistically significant for all subsamples. The estimates of −0.471, −0.338, and −0.497 for the 1993–2000 subperiod are smaller in absolute terms than the corresponding estimates of −0.608, −0.475, and −0.676 for the 2001–2014 subperiod. The standard errors range from 0.003 to 0.001. The point estimates of $a_{\sigma,t}$ are equal to −0.007, −0.02, −0.007, −0.027, −0.023, and −0.013, with standard errors between 0.001 and 0.003. The number of prints is inversely related to effective price volatility. Higher effective volatility implies fewer prints in the context of the invariance hypothesis. The estimates of $\mu_{n,0}$ and $\mu_{n,t}$ are not too different from the corresponding estimates in Table 2.

The significant increase in values of R^2 in regression equations (18), relative to values of R^2 in regression (11) constrained with $a_n = 2/3$, shows that differences in effective price volatility partially explain the cross-sectional variation in the number of prints unexplained by the invariance hypothesis. For the entire sample period, adding effective price volatility as an explanatory variable increases the R^2 from 0.873 to 0.918 for the 1993–2000 subperiod and from 0.899 to 0.955 for the 2001–2014 subperiod. For NYSE stocks, the R^2 increases from 0.908 and 0.917 to 0.924 and 0.955; for NASDAQ stocks, the R^2 increases from 0.857 and 0.881 to 0.917 and 0.959 during the two subperiods.

Finally, we analyze the R^2 in the regression that imposes the invariance restriction $a_n = 2/3$ in regression (11), but allows the coefficients on the three components of trading activity W_i

(volume V_i, price P_i, and volatility σ_i) to vary freely:

$$\ln(N_i) = \mu_n + \frac{2}{3}\ln\left(\frac{W_i}{W_*}\right) + b_1 \cdot \ln\left(\frac{V_i}{(10^6)}\right) + b_2 \cdot \ln\left(\frac{P_i}{(40)}\right) + b_3 \cdot \ln\left(\frac{\sigma_i}{(0.02)}\right) + \tilde{\epsilon}_i. \tag{19}$$

Table 6 presents the results. For the 1993–2000 subperiod, the estimates are $\hat{b}_{1,0} = 0.122$ for the coefficient on volume V_i, $\hat{b}_{2,0} = -0.370$ for the coefficient on price P_i, and $\hat{b}_{3,0} = -0.514$ for the coefficient on volatility σ_i. For the entire 2001–2014 period, the estimates are $\hat{b}_{1,0} = 0.259$ for the coefficient on volume V_i, $\hat{b}_{2,0} = -0.281$ for the coefficient on price P_i, and $\hat{b}_{3,0} = -0.492$ for the coefficient on volatility σ_i. All coefficients are statistically different from zero. Similar patterns are observed for other subperiods and subsamples of the NYSE stocks and the NASDAQ stocks. The exponents for volatility $\hat{b}_{3,0}$ behave similarly to the exponents for price $\hat{b}_{2,0}$ and differently from the exponents for volume $\hat{b}_{1,0}$, but the p-values of statistical tests of the linear restriction $b_{2,0} = b_{3,0} = -2 \times b_{1,0}$ are less than 0.001. This implies that the rejection of the invariance hypothesis might depend in a subtle manner on how effective price-volatility influences incentives to shred orders and make intermediation trades. Note that for the pooled sample, the R^2 increases from 0.873 to 0.928 for the 1993–2000 subperiod and from 0.899 to 0.970 for the 2001–2014 subperiod. The R^2 of 0.928 and 0.970 are only slightly larger than the R^2 of 0.918 and 0.955 in regression equations (18), respectively. Although statistically significant, the addition of two extra degrees of freedom beyond the effective price volatility improves the R^2 by only a small amount.

4. Conclusion

The distributions of TAQ print sizes (adjusted for trading activity as suggested by invariance hypotheses) resemble a log-normal, with truncation below the 100-share odd-lot boundary. The resemblance was stronger during the earlier 1993–2001 subperiod than during the later 2001–2014 subperiod, and it shows up more clearly in volume-weighted distributions than in trade-weighted distributions.

The invariance hypothesis explains about 88% of variation across stocks in the number of prints. The unexplained 10% can be most likely attributed to other microstructure effects such as order splitting, intermediation activity, and various market frictions like lot size and tick size. For example, subtle effects of effective price volatility explain an additional 4.5% and 5.5% of variations during the 1993–2000 and 2001–2014 periods, respectively. An interesting topic for future research would be to analyze these effects at a deeper level by designing more refined econometric tests.

Alexander, G., Peterson, M., 2007. An analysis of trade-size clustering and its relation to stealth trading. Journal of Financial Economics 84, 435–471.

Almgren, R., Chriss, N., 2000. Optimal execution of portfolio transactions. Journal of Risk 3, 5–39.

Amihud, Y., 2002. Illiquidity and stock returns: Cross-section and time-series effects. Journal of Financial Markets 5, 31–56.

Ané, T., Geman, H., 2000. Order flow, transaction clock, and normality of asset returns. Journal of Finance 55, 2259–2284.

Angel, J.J., 1997. Tick size, share prices, and stock splits. Journal of Finance 52, 655–681.

Angel, J.J., Harris, L.E., Spatt, C.S., 2011. Equity trading in the 21st century. Quarterly Journal of Finance 1, 1–53.

Angel, J.J., Harris, L.E., Spatt, C.S., 2015. Equity trading in the 21st century: An update. Quarterly Journal of Finance 5, 1550002.

Atkyn, A.B., Dyl, E.A., 1997. Transactions costs and holding periods for common stocks. Journal of Finance 52, 309–325.

Bessembinder, H., 2000. Tick size, spreads, and liquidity: An analysis of Nasdaq securities trading near ten dollars. Journal of Financial Intermediation 9, 213–239.

Brennan, M., Subrahmanyam, A., 1998. The determinants of average trade size. Journal of Business 71, 1–25.

Caglio, C., Mayhew, S., 2016. Equity trading and the allocation of market data revenue. Journal of Banking & Finance 62, 97–111.

Chordia, T., Roll, R., Subrahmanyam, A., 2011. Recent trends in trading activity and market quality. Journal of Financial Economics 101, 243–263.

Christie, W.G., Harris, J.H., Schultz, P.H., 1994. Why did Nasdaq market makers stop avoiding odd-eighth quotes? Journal of Finance 49, 1841–1860.

Fama, E.F., MacBeth, J.D., 1973. Risk, return, and equilibrium: Empirical tests. Journal of Political Economy 81, 607–636.

Glosten, L., Harris, L., 1988. Estimating the components of the bid-ask spread. Journal of Financial Economics 21, 123–142.

Goldstein, M.A., Kavajecz, K.A., 2000. Eighths, sixteenths, and market depth: Changes in tick size and liquidity provision on the NYSE. Journal of Financial Economics 56, 125–149.

Harris, L., 1987. Transaction data tests of the mixture of distributions hypothesis. Journal of Financial and Quantitative Analysis 22, 127–141.

Harris, L., 1990. Estimation of stock price variances and serial covariances from discrete observations. Journal of Financial and Quantitative Analysis 25, 291–306.

Harris, L., 1994. Minimum price variations, discrete bid-ask spreads, and quotation sizes. Review of Financial Studies 7, 149–178.

Hasbrouck, J., 1999. Trading fast and slow: Security market events in real time. Working Paper, New York University.

Hasbrouck, J., Sofianos, G., Sosebee, D., 1993. New York Stock Exchange systems and trading procedures. NYSE Working Paper 93-01.

Hendershott, Terrence, Jones, C., Subrahmanyam, A., 2011. Does algorithmic trading improve liquidity? Journal of Finance 66, 1–33.

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

Keim, D.B., Madhavan, A., 1995. Anatomy of the trading process: Empirical evidence on the behavior of institutional trades. Journal of Financial Economics 37, 371–398.

Kirilenko, A., Kyle, A.S., Samadi, M., Tuzun, T., 2017. The flash crash: High-frequency trading in an electronic market. Journal of Finance 72, 967–998.

Kyle, A.S., 1985. Continuous auctions and insider trading. Econometrica 53, 1315–1335.

Kyle, A.S., Obizhaeva, A.A., 2016. Market microstructure invariance: Empirical hypotheses. Econometrica 84, 1345–1404.

Kyle, A.S., Obizhaeva, A.A., 2017a. Market microstructure invariance: A dynamic equilibrium model. Working Paper. Available at SSRN: `https://ssrn.com/abstract=2749531`.

Kyle, A.S., Obizhaeva, A.A., 2017b. Dimensional analysis, leverage neutrality, and market microstructure invariance. Working Paper. Available at SSRN: `http://dx.doi.org/10.2139/ssrn.2785559`.

Kyle, A.S., Obizhaeva, A.A., 2018a. Adverse selection and liquidity: From theory to practice. University of Maryland.

Kyle, A.S., Obizhaeva, A.A., 2018b. The market impact puzzle. Working Paper. Available at SSRN: `https://ssrn.com/abstract=3124502`.

Kyle, A.S., Obizhaeva, A.A., 2018c. Large bets and stock market crashes. Working Paper. Available at SSRN: `http://dx.doi.org/10.2139/ssrn.2023776`.

Kyle, A.S., Obizhaeva, A.A., Wang, Y., 2018. Smooth trading with overconfidence and market power. Review of Economic Studies 85, 611–662.

Moulton, P., 2005. You can't always get what you want: Trade-size clustering and quantity choice in liquidity. Journal of Financial Economics 78, 89–119.

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

O'Hara, M., Yao, C., Ye, M., 2014. What's not there: Odd lots and market data. Journal of Finance 69, 2199–2236.

Schultz, P., 2000. Stock splits, tick size, and sponsorship. Journal of Finance 55, 429–450.

Smith, J.W., Selway, J.P., McCormick, T.D., 1998. The Nasdaq stock market: Historical background and current operation. NASD Working Paper 98-01.

Tauchen, G.E., Pitts, M., 1983. The price variability-volume relationship on speculative markets. Journal of Business and Economic Statistics 51, 485–505.

Table 1
Descriptive statistics.
This table reports descriptive statistics for securities and prints. Each observation represents averages for one security over one month. Panel A reports statistics for data from February 1993 to December 2000. Panel B reports statistics for data from January 2001 to December 2014. Both panels show the average print size, the trade-weighted median print size, the volume-weighted median print size (in dollars), the average number of prints per day, the daily dollar volume (in thousands of dollars), the average volatility of daily returns, the average price, and the percentage of trades and the percentage of volume in the 100-share lot, in the 1,000-share lot, and in the even lots for all samples as well as for 10 volume groups. Volume groups are based on average dollar trading volume with thresholds corresponding to the 30th, 50th, 60th, 70th, 75th, 80th, 85th, 90th, and 95th percentiles of the dollar volume for NYSE-listed common stocks. Volume group 1 has stocks with the lowest volume, and volume group 10 has stocks with the highest volume.

Volume groups:	All	1	2	3	4	5	6	7	8	9	10
						Panel A: 1993–2000					
Avg. Print Size ($)	23,629	11,441	27,378	36,432	43,558	49,274	53,421	60,364	67,904	78,139	89,338
Med. (VW) Print Size ($)	111,195	48,999	158,303	190,228	218,529	218,038	250,510	255,284	291,867	317,847	373,248
Med. (TW) Print Size ($)	9,363	5,682	10,842	13,407	15,402	16,910	18,024	20,049	21,909	24,599	28,567
Avg # of Prints, γ	142	17	73	126	185	257	333	409	549	852	2,830
Avg. Daily Volume ($ 1,000)	6,186	151	1,197	2,808	5,051	7,999	11,263	15,847	23,901	42,408	176,985
Avg. Daily Volatility	0.040	0.045	0.031	0.032	0.030	0.030	0.030	0.029	0.029	0.029	0.029
Avg. Price ($)	17.58	10.25	20.38	24.50	27.84	31.57	34.73	38.34	43.03	49.98	64.43
100-Shares: % Prints/ %Vol	16/2	15/2	17/2	18/2	19/2	20/2	21/2	21/2	21/2	22/2	25/3
1,000-Shares: % Prints/ %Vol	18/14	18/15	18/13	17/12	16/12	15/11	15/11	14/10	14/10	13/10	13/10
Even Lots: %Prints/ % Vol	80/61	80/63	80/60	80/59	80/58	80/57	80/57	79/56	79/55	79/56	81/58
# Obs	634,322	391,611	93,732	37,272	33,145	14,896	14,155	13,039	12,276	11,831	12,365
						Panel B: 2001–2014					
Avg. Print Size ($)	6,424	3,645	6,715	8,776	10,379	11,913	13,308	14,863	17,003	19,852	26,335
Med. (VW) Print Size ($)	29,900	27,916	23,464	26,460	29,278	27,822	41,432	36,663	42,728	53,538	83,227
Med. (TW) Print Size ($)	2,895	1,642	3,244	4,163	4,793	5,419	6,005	6,647	7,469	8,378	10,533
Avg # of Prints, γ	3,013	399	2,178	3,584	5,069	6,733	8,323	10,062	12,846	17,651	37,495
Avg. Daily Volume ($ 1,000)	24,408	937	8,295	16,961	27,429	40,200	54,090	72,872	102,484	162,619	503,229
Avg. Daily Volatility	0.031	0.035	0.026	0.025	0.024	0.024	0.023	0.022	0.022	0.022	0.022
Avg. Price ($)	21.16	12.56	26.77	32.00	34.92	37.81	40.92	44.46	48.24	51.77	63.33
100-Shares: % Prints/ %Vol	57/25	56/23	64/33	61/31	60/29	58/28	57/28	56/28	55/27	53/25	49/20
1,000-Shares: % Prints/ %Vol	3/6	4/6	2/4	2/4	2/4	2/4	2/4	2/4	2/4	3/4	3/5
Even Lots: %Prints/ % Vol	86/63	85/60	90/68	89/68	89/67	88/66	88/65	87/65	87/64	86/62	84/59
# Obs	749,535	477,127	97,347	39,485	36,846	17,455	16,780	16,222	15,858	15,815	16,600

Table 2
OLS estimates of number of TAQ prints.
This table presents Fama–MacBeth estimates μ_n and a_n from monthly regressions

$$\ln(N_i) = \mu_n + a_n \cdot \ln\left(\frac{W_i}{W_*}\right) + \tilde{\epsilon}_i.$$

Each month has one observation for each stock i. The value of N_i is the average number of prints per day. Trading activity W_i is the product of average daily dollar volume $V_i \cdot P_i$ and the percentage standard deviation σ_i of daily returns in a given month. The scaling constant $W_* = (40)(10^6)(0.02)$ corresponds to the measure of trading activity for a benchmark stock with a price of \$40 per share, trading volume of 1 million shares per day, and daily volatility of 0.02. Newey–West standard errors are calculated with three lags relative to a linear time trend estimated by OLS regressions from the estimated coefficients $\hat{\mu}_{n,T}$ and $\hat{a}_{n,T}$ for each month: $\hat{\mu}_{n,T} = \mu_{n,0} + \mu_{n,t} \cdot (T - \bar{T})/12 + \tilde{\epsilon}_T$ and $\hat{a}_{n,T} = a_{n,0} + a_{n,t} \cdot (T - \bar{T})/12 + \tilde{\epsilon}_T$, where T is the number of months from the beginning of the sample and $\bar{T}$ is the mean month. "Avg Adj-R^2" denotes the adjusted R^2 averaged over monthly regressions, and "Avg # of Obs" denotes the number of stocks averaged over monthly regressions. The estimates are reported for 1993–2000 and 2001–2014 subperiods.

	All stocks		NYSE/AMEX		NASDAQ	
	1993–2000	2001–2014	1993–2000	2001–2014	1993–2000	2001–2014
$\mu_{n,0}$	6.147	8.513	6.109	8.396	6.143	8.646
	0.017	0.041	0.012	0.040	0.020	0.048
$\mu_{n,t}$	0.093	0.148	0.042	0.178	0.141	0.115
	0.007	0.013	0.005	0.012	0.008	0.015
$a_{n,0}$	0.666	0.790	0.626	0.760	0.679	0.816
	0.002	0.005	0.001	0.005	0.002	0.006
$a_{n,t}$	-0.001	0.007	0.002	0.006	0.003	0.003
	0.001	0.001	0.000	0.001	0.001	0.002
Avg Adj-R^2	0.87	0.92	0.91	0.93	0.86	0.91
Avg # of Obs	6,621	4,452	2,189	1,759	4,432	2,694

Table 3
Regression estimates of TAQ print sizes, February 1993 to December 2000.
This table presents Fama–MacBeth estimates μ_x and a_x from the monthly regressions of the mean and percentiles of print size on trading activity W for the sample from February 1993 to December 2000. The coefficients μ_Q and a_Q are based on monthly regressions

$$\ln\left(\frac{|X_i|}{V_i}\right) = \mu_x + a_x \cdot \ln\left(\frac{W_i}{W_*}\right) + \tilde{\epsilon}_i,$$

where the left-hand side is either the mean or the pth (20th, 50th and 80th) percentile of the distribution of logarithms of (unsigned) print sizes $|X_i|$, expressed as a fraction of daily volume V_i in a given month. The means and percentiles are calculated based on both trade-weighted and volume-weighted distributions. One observation corresponds to each month and each stock i with trading activity W_i, defined as the product of the average daily dollar volume $V_i \cdot P_i$ and the percentage standard deviation σ_i of daily returns. The scaling constant $W_* = (40)(10^6)(0.02)$ corresponds to the trading activity of the benchmark stock with a price of \$40 per share, trading volume of 1 million shares per day, and volatility of 2% per day. Newey–West standard errors are calculated with three lags relative to a linear time trend estimated by OLS regressions from the estimated coefficients $\hat{\mu}_{x,T}$ and $\hat{a}_{x,T}$ for each month: $\hat{\mu}_{x,T} = \mu_{x,0} + \mu_{x,t} \cdot (T - \bar{T})/12 + \tilde{\epsilon}_T$ and $\hat{a}_{x,T} = a_{x,0} + a_{x,t} \cdot (T - \bar{T})/12 + \tilde{\epsilon}_T$, where T is the number of months from the beginning of the sample and $\bar{T}$ is the mean month. "Avg Adj-R^2" denotes the adjusted R^2 averaged over monthly regressions, and "Avg # of Obs" denotes the number of stocks averaged over monthly regressions.

	Trade-weighted distribution				Volume-weighted distribution			
	Mean	20th	50th	80th	Mean	20th	50th	80th
$\mu_{x,0}$	-7.238	-8.495	-7.289	-6.260	-4.684	-6.364	-4.887	-3.326
	0.020	0.027	0.031	0.014	0.020	0.016	0.019	0.026
$\mu_{x,t}$	-0.047	-0.039	-0.046	-0.064	-0.137	-0.101	-0.157	-0.153
	0.008	0.011	0.012	0.006	0.008	0.006	0.008	0.011
$a_{x,0}$	-0.741	-0.781	-0.750	-0.725	-0.560	-0.661	-0.579	-0.481
	0.002	0.003	0.002	0.003	0.003	0.002	0.003	0.003
$a_{x,t}$	0.008	0.012	0.007	0.005	-0.007	-0.001	-0.009	-0.009
	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001
Avg Adj-R^2	0.90	0.87	0.88	0.88	0.69	0.83	0.68	0.52
Avg # of Obs	6,621	6,621	6,621	6,621	6,621	6,621	6,621	6,621

Table 4
Regression estimates of TAQ print sizes, January 2001 to December 2014.
This table presents Fama–MacBeth estimates μ_x and a_x from the monthly regressions of the mean and percentiles of print size on trading activity W for the sample from January 2001 to December 2014. The coefficients μ_x and a_x are based on monthly regressions

$$\ln\left(\frac{|X_i|}{V_i}\right) = \mu_x + a_x \cdot \ln\left(\frac{W_i}{W_*}\right) + \tilde{\epsilon}_i,$$

where the left-hand side is either the mean or the pth (20th, 50th and 80th) percentile of the distribution of logarithms of (unsigned) print sizes $|X_i|$, expressed as a fraction of daily volume V_i in a given month. The means and percentiles are calculated based on both trade- and volume-weighted distributions. One observation corresponds to each month and each stock i with trading activity W_i, defined as the product of the average daily dollar volume $V_i \cdot P_i$ and the percentage standard deviation σ_i of daily returns. The scaling constant $W_* = (40)(10^6)(0.02)$ corresponds to the trading activity of the benchmark stock with a price of \$40 per share, trading volume of 1 million shares per day, and volatility of 2% per day. Newey–West standard errors are calculated with three lags relative to a linear time trend estimated by OLS regressions from the estimated coefficients $\hat{\mu}_{x,T}$ and $\hat{a}_{x,T}$ for each month $\hat{\mu}_{x,T} = \mu_{x,0} + \mu_{x,t} \cdot (T - \bar{T})/12 + \tilde{\epsilon}_T$ and $\hat{a}_{x,T} = a_{x,0} + a_{x,t} \cdot (T - \bar{T})/12 + \tilde{\epsilon}_T$, where T is the number of months from the beginning of the sample and $\bar{T}$ is the mean month. "Avg Adj-R^2" denotes the adjusted R^2 averaged over monthly regressions, and "Avg # of Obs" denotes the number of stocks averaged over monthly regressions.

	Trade-weighted distribution				Volume-weighted distribution			
	Mean	20th	50th	80th	Mean	20th	50th	80th
$\mu_{x,0}$	-9.029	-9.519	-9.268	-8.633	-7.420	-9.004	-8.123	-6.304
	0.030	0.024	0.038	0.039	0.058	0.047	0.059	0.059
$\mu_{x,t}$	-0.097	-0.013	-0.084	-0.173	-0.192	-0.091	-0.251	-0.251
	0.009	0.007	0.012	0.012	0.018	0.015	0.018	0.018
$a_{x,0}$	-0.779	-0.757	-0.769	-0.811	-0.776	-0.796	-0.843	-0.779
	0.003	0.003	0.003	0.004	0.007	0.006	0.009	0.010
$a_{x,t}$	0.000	0.005	0.003	-0.006	-0.018	-0.005	-0.021	-0.026
	0.001	0.001	0.001	0.001	0.002	0.002	0.002	0.003
Avg Adj-R^2	0.907	0.866	0.884	0.913	0.892	0.892	0.862	0.760
Avg # of Obs	4,452	4,452	4,452	4,452	4,452	4,452	4,452	4,452

Table 5
OLS estimates of number of TAQ prints with effective volatility.
This table presents Fama–MacBeth estimates μ_n and a_σ from monthly regressions

$$\ln(N_i) = \mu_n + \frac{2}{3}\cdot\ln\left(\frac{W_i}{W_*}\right) + a_\sigma\cdot\ln\left(\frac{P_i\cdot\sigma_i}{P_*\cdot\sigma_*}\cdot\left(\frac{W_i}{W_*}\right)^{-1/3}\right) + \tilde{\epsilon}_i.$$

One observation corresponds to each month and each stock i with trading activity W_i, defined as the product of the average daily dollar volume $V_i\cdot P_i$ and the percentage standard deviation σ_i of daily returns. Effective price volatility is defined as $P_i\cdot\sigma_i\cdot\left(\frac{W_i}{W_*}\right)^{-1/3}$, with the effective price volatility of the benchmark stocks $P_*\cdot\sigma_*$ equal to $40\cdot 0.02$. The value of N_i is the average number of prints per day. The scaling constant $W_* = (40)(10^6)(0.02)$ corresponds to the measure of trading activity for the benchmark stock with a price of \$40 per share, trading volume of 1 million shares per day, and daily volatility of 2%. Newey–West standard errors are calculated with three lags relative to a linear time trend estimated by OLS regressions from the estimated coefficients $\hat{\mu}_{n,T}$ and $\hat{a}_{\sigma,T}$ for each month: $\hat{\mu}_{n,T} = \mu_{n,0} + \mu_{n,t}\cdot(T-\bar{T})/12 + \tilde{\epsilon}_T$ and $\hat{a}_{\sigma,T} = a_{\sigma,0} + a_{\sigma,t}\cdot(T-\bar{T})/12 + \tilde{\epsilon}_T$, where T is the number of months from the beginning of the sample and $\bar{T}$ is the mean month. "Avg Adj-R^2" denotes the adjusted R^2 averaged over monthly regressions. The table also reports the average R^2 from the restricted regressions with $a_\sigma = 0$ as well as the average R^2 from unconstrained regressions

$$\ln(N_i) = \mu_n + \frac{2}{3}\ln\left(\frac{W_i}{W_*}\right) + b_1\cdot\ln\left(\frac{V_i}{(10^6)}\right) + b_2\cdot\ln\left(\frac{P_i}{(40)}\right) + b_3\cdot\ln\left(\frac{\sigma_i}{(0.02)}\right) + \tilde{\epsilon}_i.$$

"Avg # of Obs" is the number of stocks averaged over monthly regressions. The estimates are reported for 1993–2000 and 2001–2014 subperiods.

	All stocks		NYSE/AMEX		NASDAQ	
	1993–2000	2001–2014	1993–2000	2001–2014	1993–2000	2001–2014
$\mu_{n,0}$	6.270	7.952	6.316	7.998	6.247	7.942
	0.016	0.023	0.009	0.025	0.020	0.027
$\mu_{n,t}$	0.087	0.116	0.024	0.151	0.125	0.093
	0.007	0.007	0.004	0.007	0.009	0.008
$a_{\sigma,0}$	-0.471	-0.608	-0.338	-0.475	-0.497	-0.676
	0.003	0.009	0.005	0.008	0.003	0.010
$a_{\sigma,t}$	-0.007	-0.020	-0.007	-0.027	-0.023	-0.013
	0.001	0.003	0.002	0.002	0.001	0.003
Avg Adj-R^2	0.918	0.955	0.924	0.955	0.917	0.959
Avg# of Obs	6,621	4,452	2,189	1,759	4,432	2,694
	Regression with coefficient on effective price volatility $a_\sigma = 0$.					
Avg Adj-R^2	0.873	0.899	0.908	0.917	0.857	0.881
	Regression with separate coefficients for price, volume, and volatility.					
Avg Adj-R^2	0.928	0.970	0.940	0.974	0.923	0.975

Table 6
OLS estimates of number of TAQ prints with price, volatility, and volume.
This table presents Fama–MacBeth estimates μ_n and a_σ from monthly regressions

$$\ln(N_i) = \mu_n + \frac{2}{3}\ln\left(\frac{W_i}{W_*}\right) + b_1 \cdot \ln\left(\frac{V_i}{(10^6)}\right) + b_2 \cdot \ln\left(\frac{P_i}{(40)}\right) + b_3 \cdot \ln\left(\frac{\sigma_i}{(0.02)}\right) + \tilde{\epsilon}_i.$$

One observation corresponds to each month and each stock i with trading activity W_i, defined as the product of the average daily dollar volume $V_i \cdot P_i$ and the percentage standard deviation σ_i of daily returns. Effective price volatility is defined as $P_i \cdot \sigma_i \cdot \left(\frac{W_i}{W_*}\right)^{-1/3}$, with the effective price volatility of the benchmark stocks $P_* \cdot \sigma_*$ equal to $40 \cdot 0.02$. The value of N_i is the average number of prints per day. The scaling constant $W_* = (40)(10^6)(0.02)$ corresponds to the measure of trading activity for the benchmark stock with a price of \$40 per share, trading volume of 1 million shares per day, and daily volatility of 2%. Unrestricted regressions are run without any restriction on coefficients while restricted regressions introduce the restriction of $b_2 = b_3 = -0.5 \times b_1$. Newey–West standard errors are calculated with three lags relative to a linear time trend estimated by OLS regressions from the estimated coefficients $\hat{\mu}_{n,T}$ and $\hat{a}_{\sigma,T}$ for each month: $\hat{\mu}_{n,T} = \mu_{n,0} + \mu_{n,t} \cdot (T - \bar{T})/12 + \tilde{\epsilon}_T$ and $\hat{a}_{\sigma,T} = a_{\sigma,0} + a_{\sigma,t} \cdot (T - \bar{T})/12 + \tilde{\epsilon}_T$, where T is the number of months from the beginning of the sample and $\bar{T}$ is the mean month. "Avg Adj-R^2" denotes the adjusted R^2 averaged over monthly regressions.

	All stocks		NYSE/AMEX		NASDAQ	
	Unrestricted		Unrestricted		Unrestricted	
	1993–2000	2001–2014	1993–2000	1993–2000	1993–2000	1993–2000
$\mu_{n,0}$	6.242	8.350	6.131	8.214	6.323	8.513
	0.015	0.036	0.012	0.036	0.022	0.041
$\mu_{n,t}$	0.114	0.133	0.070	0.161	0.166	0.101
	0.007	0.011	0.006	0.010	0.009	0.013
$b_{1,0}$	0.122	0.259	0.069	0.216	0.148	0.294
	0.001	0.005	0.002	0.006	0.001	0.005
$b_{1,t}$	0.010	0.012	0.006	0.014	0.017	0.007
	0.001	0.001	0.001	0.002	0.001	0.001
$b_{2,0}$	-0.370	-0.281	-0.351	-0.230	-0.363	-0.300
	0.003	0.006	0.003	0.008	0.004	0.005
$b_{2,t}$	0.005	-0.004	0.010	-0.011	-0.002	-0.003
	0.001	0.002	0.001	0.002	0.001	0.002
$b_{3,0}$	-0.514	-0.492	-0.524	-0.493	-0.482	-0.556
	0.006	0.009	0.004	0.024	0.004	0.006
$b_{3,t}$	0.049	-0.017	0.019	-0.005	0.024	-0.013
	0.002	0.001	0.002	0.003	0.001	0.002
Avg Adj-R^2	0.928	0.970	0.940	0.974	0.923	0.975
Avg # of Obs	6,621	4,452	2,189	1,759	4,432	2,694
$b_{2,0} = b_{3,0}$	$p < 0.001$	$p < 0.001$	$p < 0.001$	$p < 0.001$	$p < 0.001$	$p < 0.001$
$b_{2,0} = -2 \times b_{1,0}$	$p < 0.001$	$p < 0.001$	$p < 0.001$	$p < 0.001$	$p < 0.001$	$p < 0.001$
$b_{3,0} = -2 \times b_{1,0}$	$p < 0.001$	$p < 0.001$	$p < 0.001$	$p < 0.001$	$p < 0.001$	$p < 0.001$
$b_{2,0} = b_{3,0} = -2 \times b_{1,0}$	$p < 0.001$	$p < 0.001$	$p < 0.001$	$p < 0.001$	$p < 0.001$	$p < 0.001$

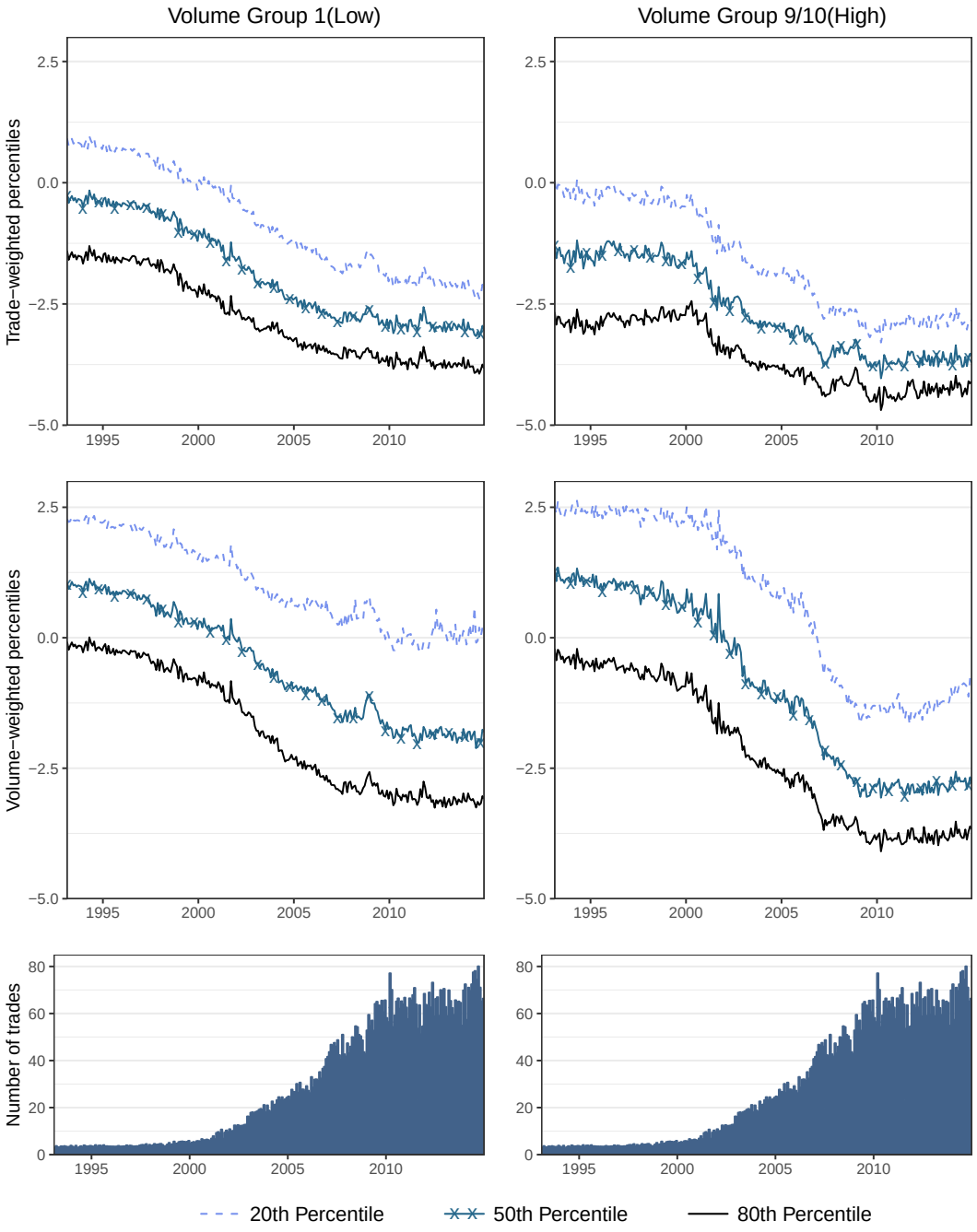

Figure 1: Time series of percentiles of scaled TAQ print size and mean number of prints, 1993–2014. The figure shows the dynamics of the 20th, 50th, and 80th percentiles for logarithms of the pooled scaled print sizes as well as the means of the scaled number of prints per month from 1993 to 2014. Volume groups are based on average dollar trading volume with thresholds corresponding to the 30th, 50th, 60th, 70th, 75th, 80th, 85th, 90th, and 95th percentiles of the dollar volume for common NYSE-listed stocks. Trade-weighted percentiles and volume-weighted percentiles are shown for stocks in volume group 1 (low volume) and volume groups 9 and 10 (high volume). For each print, the logarithm of scaled print size is calculated based on the midpoint of the print size bin, scaled according to the model of trading game invariance—that is, $\ln\left(W_i^{2/3} \cdot |X_i| / V_i\right)$, where $|X_i|$ is a midpoint of a print size bin in shares, V_i is the average daily volume in shares, and W_i is the measure of trading activity equal to the product of dollar volume and returns standard deviation. The scaled number of total monthly prints is calculated as $N_{m,i} \cdot W_i^{-2/3}$, where $N_{m,i}$ is the total number of trades per month. The stock-level distributions of scaled print sizes are averaged across stocks for volume groups 1 and 9–10 in a given month. The trade- and volume-weighted percentiles are plotted on this figure.

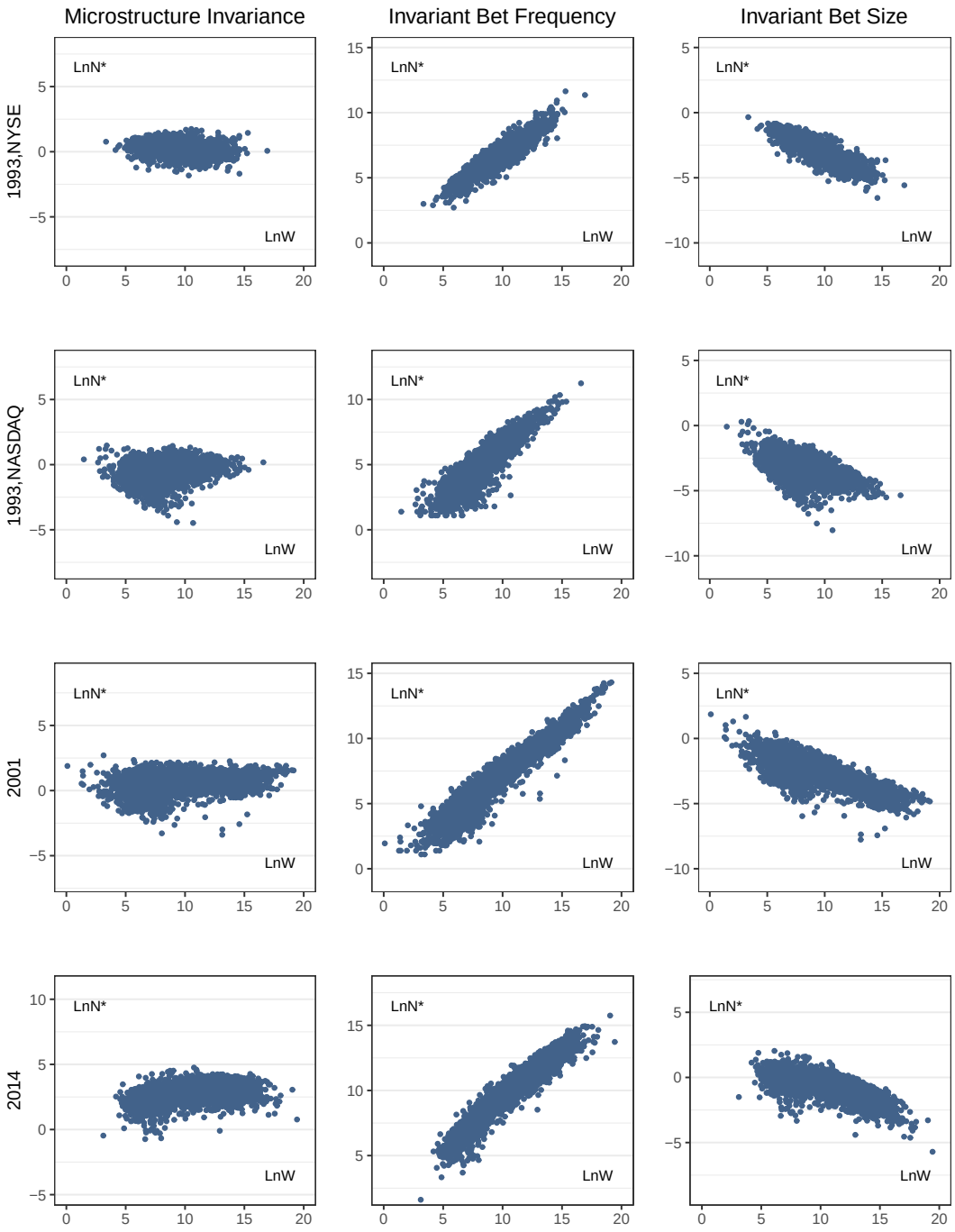

Figure 2: The scaled number of TAQ prints plotted against trading activity for three models. The figure shows the logarithm of the scaled number of prints across different levels of the logarithm of trading activity W_i. The scaled number of prints is defined by $N_i W_i^{-\alpha_n}$, with $\alpha_n = 2/3$ for the model of trading game invariance, $\alpha_n = 0$ for the model of invariant bet frequency, and $\alpha_n = 1$ for the model of invariant bet size. Four subsamples are considered: NYSE-listed stocks in April 1993, NASDAQ-listed stocks in April 1993, both NYSE and NASDAQ stocks in April 2001 and both NASDAQ and NYSE stocks in April 2014. Trading activity W_i is calculated as the product of average daily dollar $P_i \cdot V_i$ volume and the percentage standard deviation of daily returns σ_i for a given month.

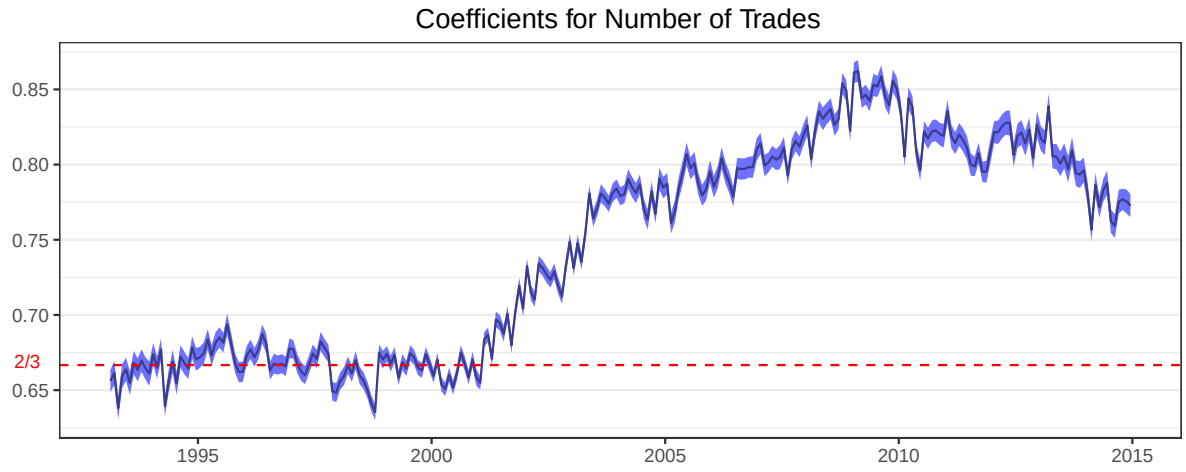

Figure 3: Time series of monthly OLS coefficient estimates for number of trades, 1993–2014. The figure shows the dynamics of coefficients from regressions of number of prints on the measure of trading activity W_i from 1993 to 2014. The coefficient a_n is calculated from monthly regressions

$$\ln(N_i) = \mu_n + a_n \cdot \ln\left(\frac{W_i}{W_*}\right) + \tilde{\epsilon}_i,$$

where N_i is the average number of prints per day in a given month. The model of trading game invariance predicts $a_n = 2/3$, and alternative models predict that $a_n = 0$ or $a_n = 1$. The model of trading game invariance predicts $a_x = -2/3$, and alternative models predict that $a_x = 0$ or $a_x = -1$. Trading activity W_i is defined as the product of dollar volume and daily percentage standard deviation of returns, and W_* measures trading activity of the benchmark stock.

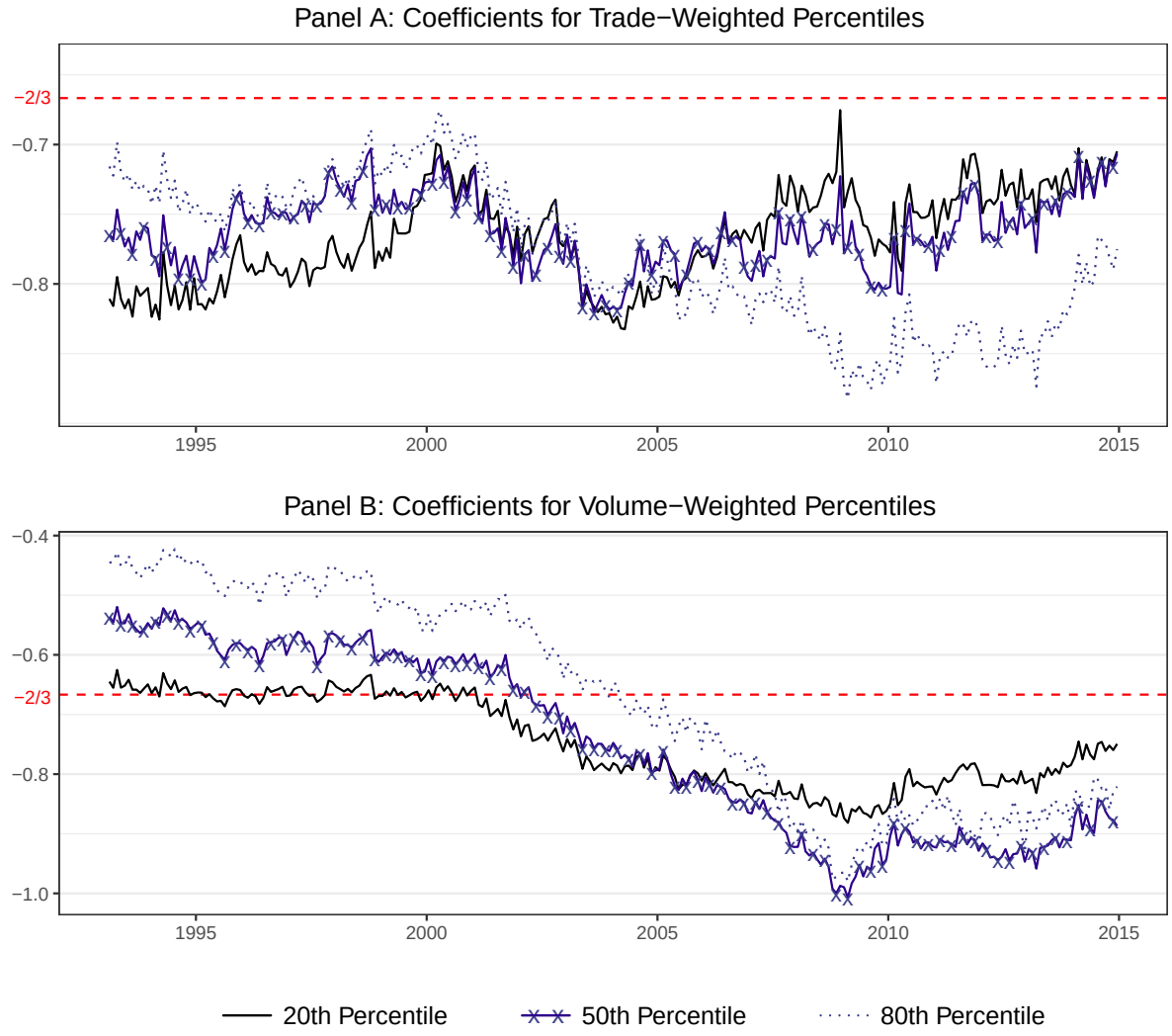

Figure 4: Time series of monthly OLS coefficient estimates for trade-weighted percentiles, and volume-weighted percentiles, 1993–2014. The figure shows the dynamics of coefficients from regressions of various percentiles on the measure of trading activity W_i from 1993 to 2014. The coefficient a_x is calculated from monthly regressions

$$\ln\left(\frac{\tilde{X}_i}{V_i}\right) = \mu_x + a_x \cdot \ln\left(\frac{W_i}{W_*}\right) + \tilde{\epsilon}_i,$$

where the left-hand side is the pth (20th, 50th and 80th) percentiles of the distribution of logarithms of print sizes $\tilde{X}_i$. The model of trading game invariance predicts $a_x = -2/3$, and alternative models predict that $a_x = 0$ or $a_x = -1$. Panel C shows the coefficient a_x from similar monthly regressions, but these regressions are based on percentiles Q_i^p, where percentiles are calculated based on the contribution to total trading volume. The model of trading game invariance predicts $a_x = -2/3$, and alternative models predict that $a_x = 0$ or $a_x = -1$. Trading activity W_i is defined as the product of dollar volume and daily percentage standard deviation of returns, and W_* measures trading activity of the benchmark stock.

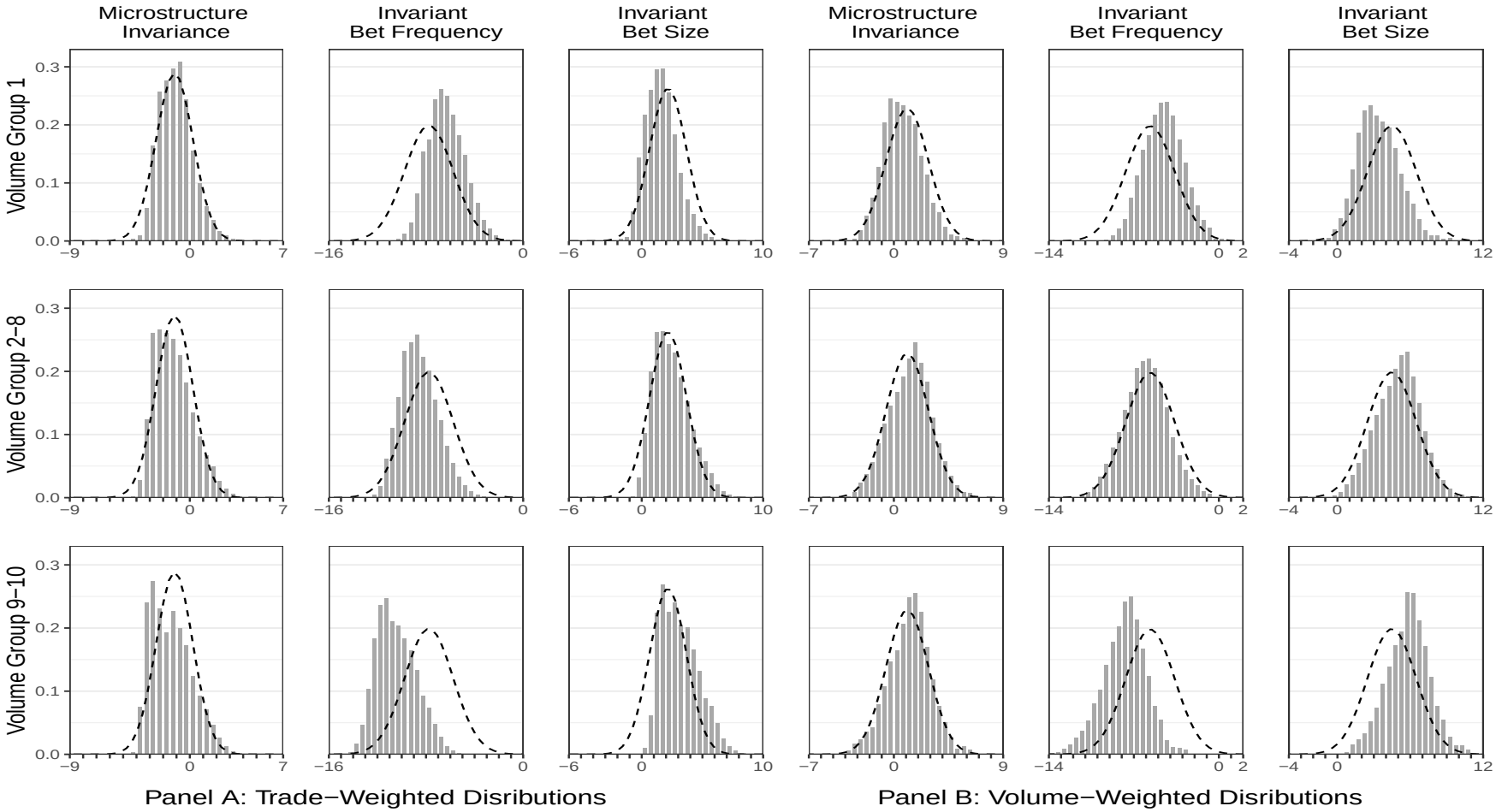

Figure 5: Trade-weighted and volume-weighted distributions of scaled TAQ print size for three models, NYSE-listed Stocks, April 1993. This figure shows the distribution of the logarithm of scaled print sizes for three different models for NYSE-listed stocks traded in April 1993. The print sizes are scaled as $W_i^{\alpha_x} \cdot |X_i|/V_i$, with $\alpha_x = 2/3$ for the model of invariant bet frequency, $\alpha = 0$ for the model of invariant bet frequency, and $\alpha_x = 1$ for the model of invariant bet size. Trading activity W_i is calculated as the product of dollar volume $P_i \cdot V_i$ and the daily percentage standard deviation of returns σ_i. Panel A shows trade-weighted distributions, and Panel B shows volume-weighted distributions. The subplots show stock-level distributions averaged across stocks in volume group 1 (low volume), volume groups 2–8, and volume groups 9–10 (high volume). Volume groups are based on average dollar trading volume with thresholds corresponding to the 30th, 50th, 60th, 70th, 75th, 80th, 85th, 90th, and 95th percentiles of the dollar volume for NYSE-listed common stocks. All subplots in each of the size columns show a fitted normal distribution (in dashed lines) with moments calculated based on a pooled data.

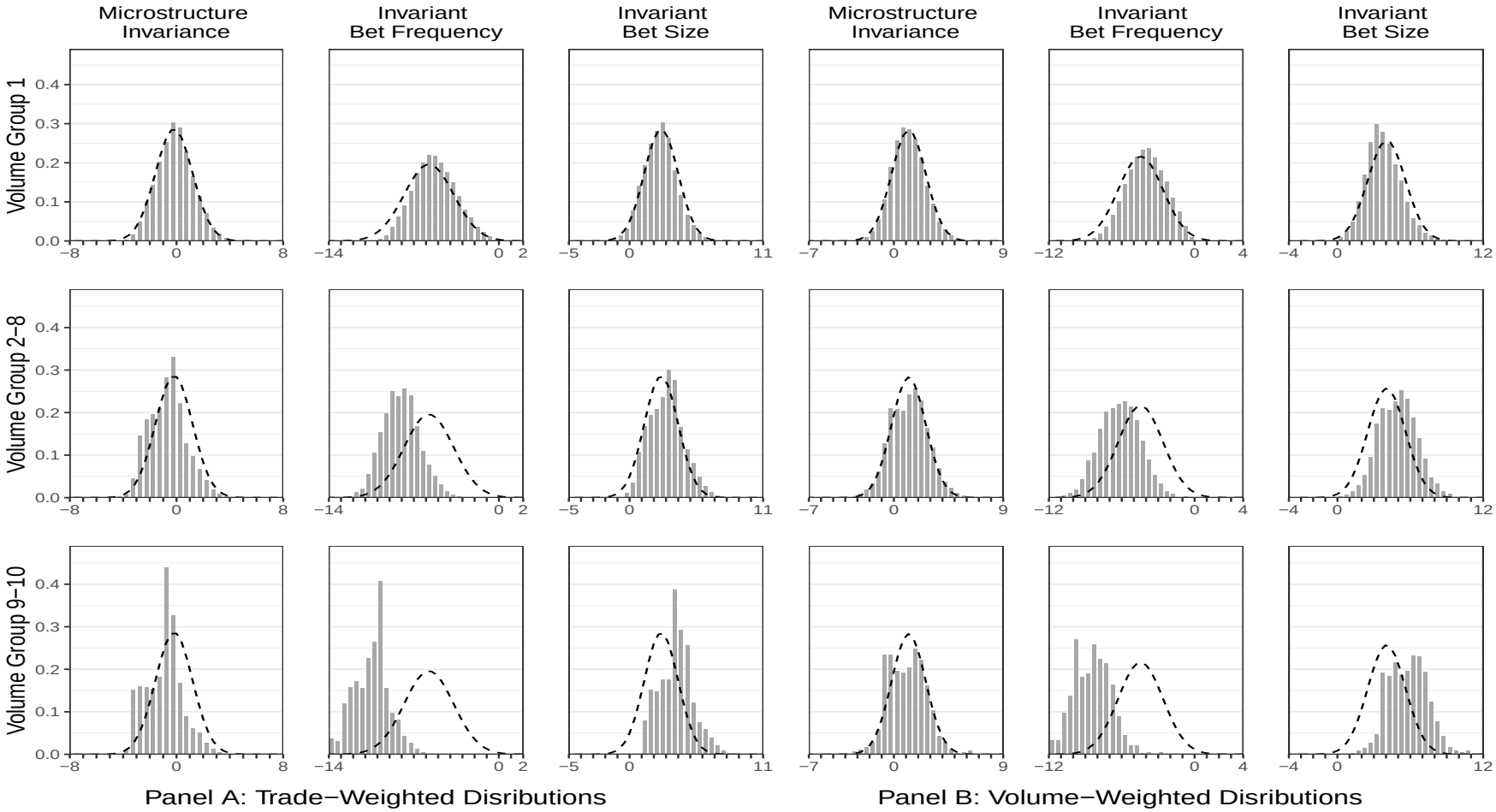

Figure 6: Trade-weighted and volume-weighted distributions of scaled TAQ print size for three models, NASDAQ-Listed Stocks, April 1993. This figure shows the distribution of the logarithm of scaled print sizes for three different models for NASDAQ-listed stocks traded in April 1993. The print sizes are scaled as $W_i^{\alpha_x} \cdot |X_i|/V_i$, with $\alpha_x = 2/3$ for the model of invariant bet frequency, $\alpha_x = 0$ for the model of invariant bet frequency, and $\alpha_x = 1$ for the model of invariant bet size. Trading activity W_i is calculated as the product of dollar volume $P_i \cdot V_i$ and the daily percentage standard deviation of returns σ_i. Panel A shows trade-weighted distributions, and panel B shows volume-weighted distributions. The subplots show stock-level distributions averaged across stocks in volume group 1 (low volume), volume groups 2–9, and volume groups 9–10 (high volume). Volume groups are based on average dollar trading volume with thresholds corresponding to the 30th, 50th, 60th, 70th, 75th, 80th, 85th, 90th, and 95th percentiles of the dollar volume for NASDAQ-listed common stocks. All subplots in each of the size columns show a fitted normal distribution (in dashed lines) with moments calculated based on a pooled data.

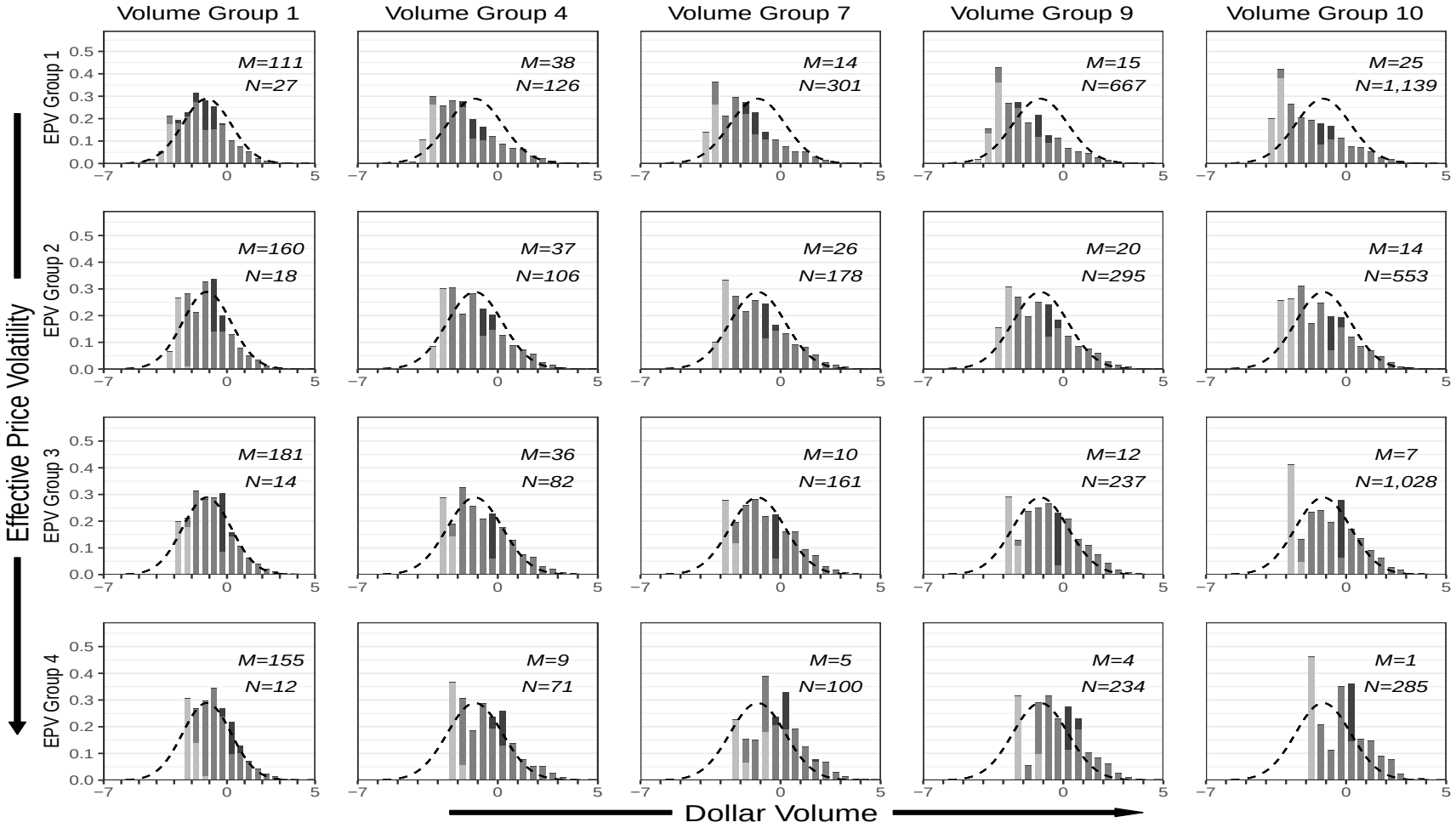

Figure 7: Trade-weighted distributions of scaled TAQ print sizes, NYSE-listed stocks, April 1993. This figure shows distributions of the logarithms of scaled print sizes for NYSE stocks in April 1993. For each trade, the scaled print size is calculated as $\ln\left(W_i^{2/3} \cdot |X_i|/V_i\right)$ based on the invariance hypothesis, where $|X_i|$ is the midpoint of the print size bin in shares, V_i is the average daily volume in shares, and W_i measures trading activity as the product of dollar volume and the daily percentage standard deviation of returns. Ten volume groups are constructed based on average dollar trading volume with thresholds corresponding to the 30th, 50th, 60th, 70th, 75th, 80th, 85th, 90th, and 95th percentiles of the dollar volume for common NYSE-listed stocks. Four equally spaced volatility groups are constructed based on effective price volatility, defined as $P_i \cdot \sigma_i \cdot (W_i/W_*)^{-1/3}$. The subplots show stock-level distributions averaged across stocks for volume groups 1 (low volume), 4, 7, 9, and 10 (high volume) and for all four price volatility groups 1 (low price volatility), 2, 3, and 4 (high price volatility). The 100-share trades are highlighted in light gray; the 1,000-share trades are highlighted in dark gray. Each subplot also shows a normal distribution with the pooled average print size mean of -1.15 and standard deviation of 1.38. M is the number of stocks, and N is the average number of prints per day for the stocks in a given subgroup.

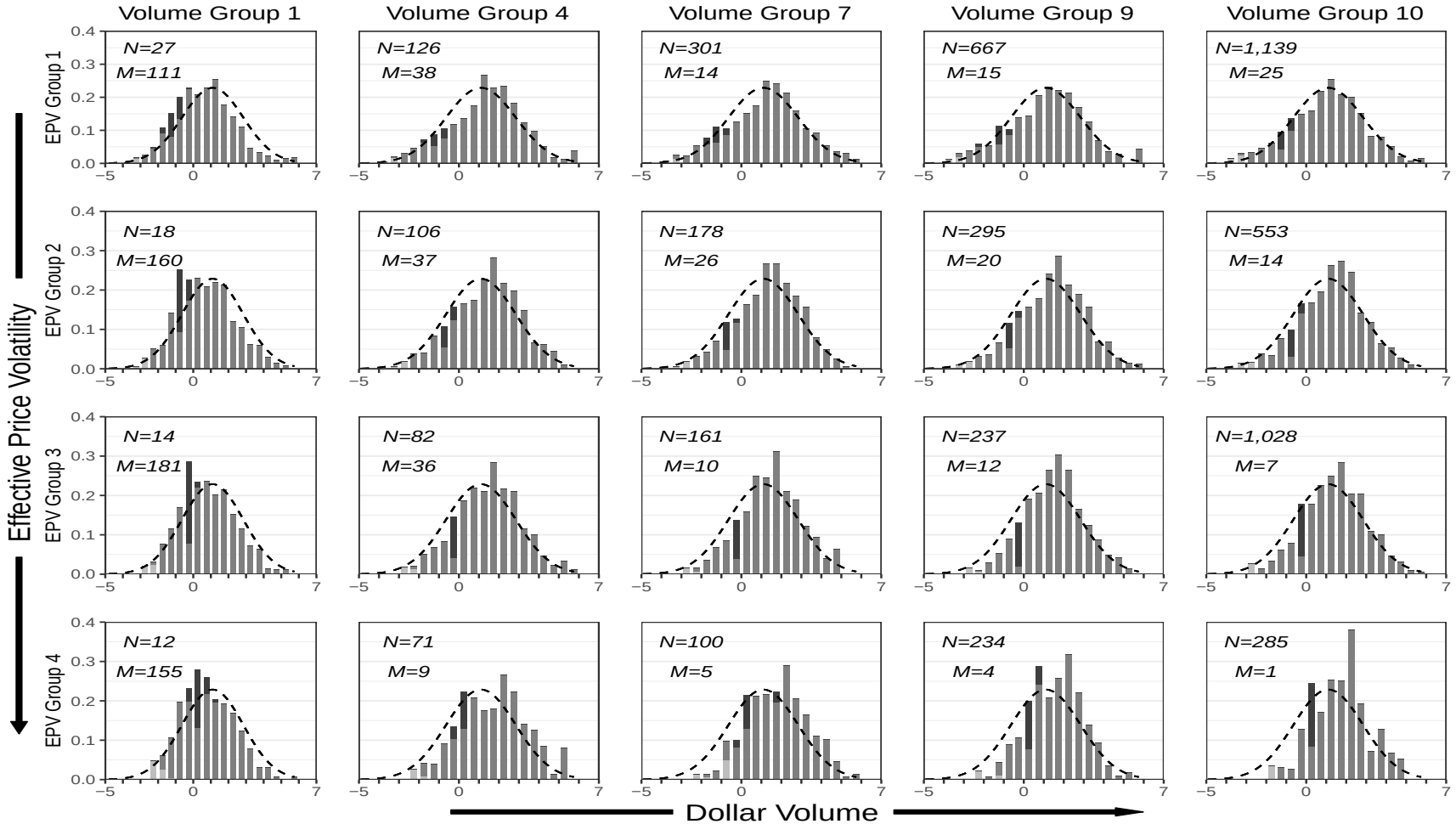

Figure 8: Volume-weighted distributions of scaled TAQ print sizes, NYSE-listed stocks, April 1993. This figure shows distributions of total volume across different scaled print size bins for the NYSE stocks in April 1993. For each stock, the volume distribution is calculated as the contribution to the total volume by trades from a given trade size bin. The x-axis is the log of scaled print sizes, defined by $\ln\left(W_i^{2/3} \cdot \frac{|X_i|}{V_i}\right)$ according to the invariance hypothesis, where $|X_i|$ is a print size in shares (midpoint of a bin), V_i is the average daily volume in shares, and W_i is the measure of trading activity equal to the product of dollar volume and returns standard deviation. Ten volume groups are constructed based on average dollar trading volume with thresholds corresponding to the 30th, 50th, 60th, 70th, 75th, 80th, 85th, 90th, and 95th percentiles of the dollar volume for NYSE-listed common stocks. Four equally spaced volatility groups are constructed based on effective price volatility, defined as $P_i \cdot \sigma_i \cdot (W_i / W_*)^{-1/3}$. The subplots show stock-level distributions averaged across stocks for volume groups 1 (low volume), 4, 7, 9, and 10 (high volume) and for all four price volatility groups 1 (low price volatility), 2, 3, and 4 (high price volatility). The 100-share trades are highlighted in light gray, and the 1,000-share trades are highlighted in dark gray. Each subplot also shows a normal distribution with the pooled average print size mean of 1.1 and standard deviation of 1.74. M is the number of stocks, and N is the average number of prints per day for the stocks in a given subgroup.

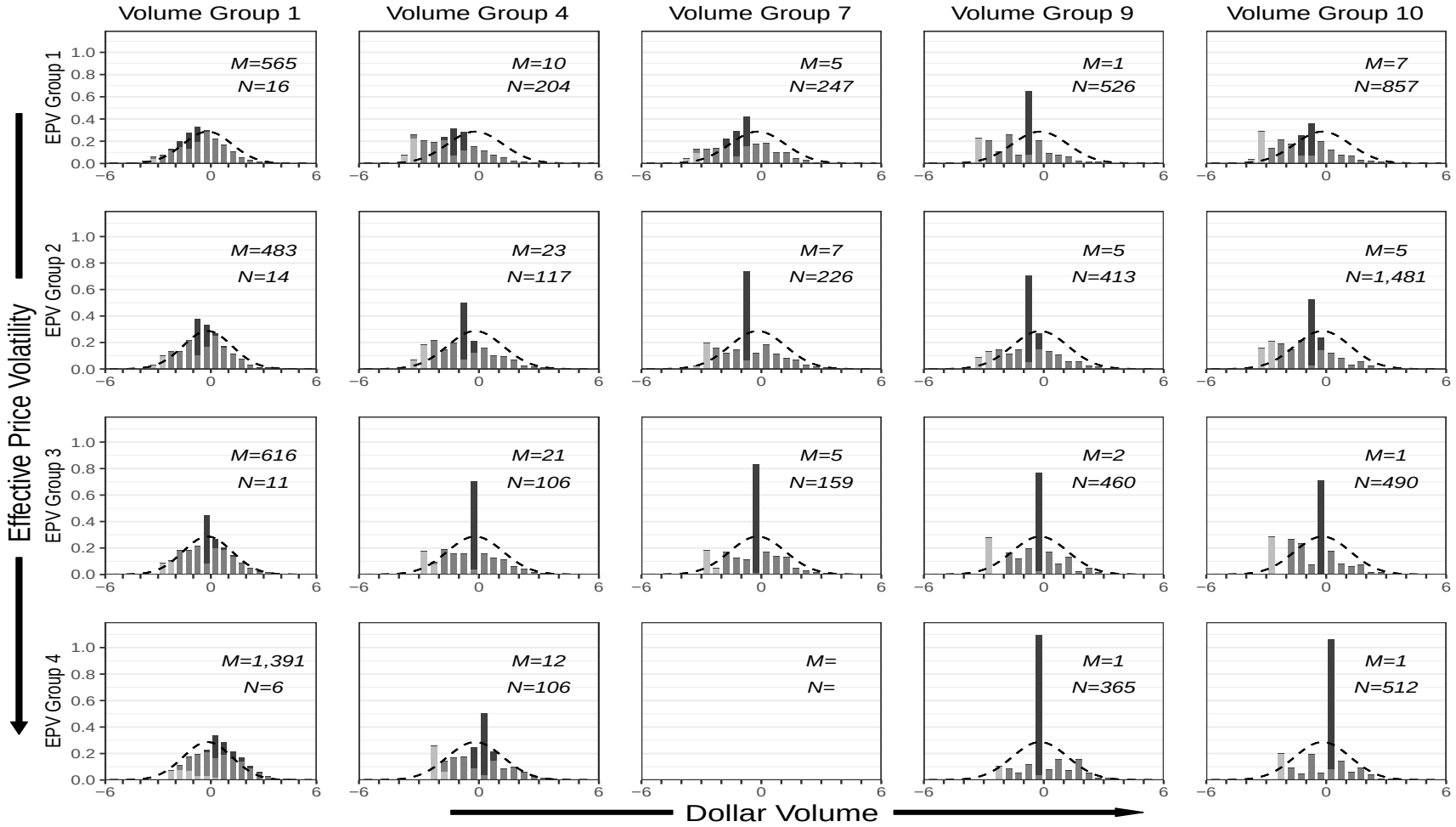

Figure 9: Trade-weighted distributions of scaled TAQ print sizes, NASDAQ-listed stocks, April 1993. This figure shows distributions of the logarithms of scaled print sizes for NASDAQ stocks in April 1993. For each trade, the scaled print size is calculated as $\ln\left(W_i^{2/3} \cdot |X_i|/V_i\right)$ based on the invariance hypothesis, where $|X_i|$ is the midpoint of the print size bin in shares, V_i is the average daily volume in shares, and W_i measures trading activity as the product of dollar volume and the daily percentage standard deviation of returns. Ten volume groups are constructed based on average dollar trading volume with thresholds corresponding to the 30th, 50th, 60th, 70th, 75th, 80th, 85th, 90th, and 95th percentiles of the dollar volume for common NYSE-listed stocks. Four equally spaced volatility groups are constructed based on effective price volatility, defined as $P_i \cdot \sigma_i \cdot (W_i/W_*)^{-1/3}$. The subplots show stock-level distributions averaged across stocks for volume groups 1 (low volume), 4, 7, 9, and 10 (high volume) and for all four price volatility groups 1 (low price volatility), 2, 3, and 4 (high price volatility). The 100-share trades are highlighted in light gray; the 1,000-share trades are highlighted in dark gray. Each subplot also shows a normal distribution with the pooled average print size mean of -0.19 and standard deviation of 1.39. M is the number of stocks, and N is the average number of prints per day for the stocks in a given subgroup.

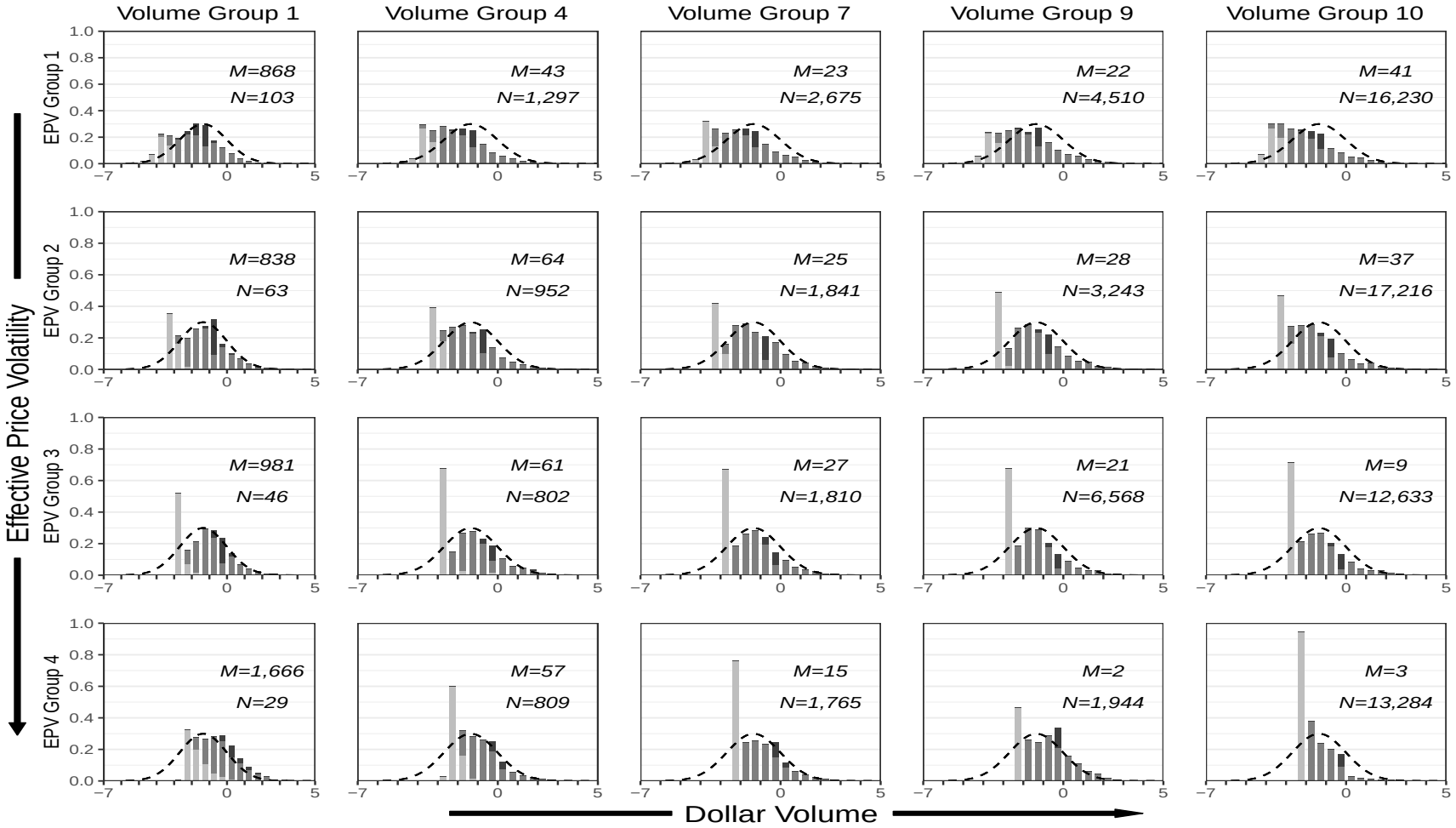

Figure 10: Trade-weighted distributions of scaled TAQ print sizes, all stocks, April 2001. This figure shows distributions of the logarithms of scaled print sizes for NYSE and NASDAQ stocks in April 2001. For each trade, the scaled print size is calculated as $\ln\left(W_i^{2/3} \cdot |X_i| / V_i\right)$ based on the invariance hypothesis, where $|X_i|$ is the midpoint of the print size bin in shares, V_i is the average daily volume in shares, and W_i measures trading activity as the product of dollar volume and the daily percentage standard deviation of returns. Ten volume groups are constructed based on average dollar trading volume with thresholds corresponding to the 30th, 50th, 60th, 70th, 75th, 80th, 85th, 90th, and 95th percentiles of the dollar volume for common NYSE-listed stocks. Four equally spaced volatility groups are constructed based on effective price volatility, defined as $P_i \cdot \sigma_i \cdot (W_i / W_*)^{-1/3}$. The subplots show stock-level distributions averaged across stocks for volume groups 1 (low volume), 4, 7, 9, and 10 (high volume) and for all four price volatility groups 1 (low price volatility), 2, 3, and 4 (high price volatility). The 100-share trades are highlighted in light gray; the 1,000-share trades are highlighted in dark gray. Each subplot also shows a normal distribution with the pooled average print size mean of −1.35 and standard deviation of 1.33. M is the number of stocks, and N is the average number of prints per day for the stocks in a given subgroup.

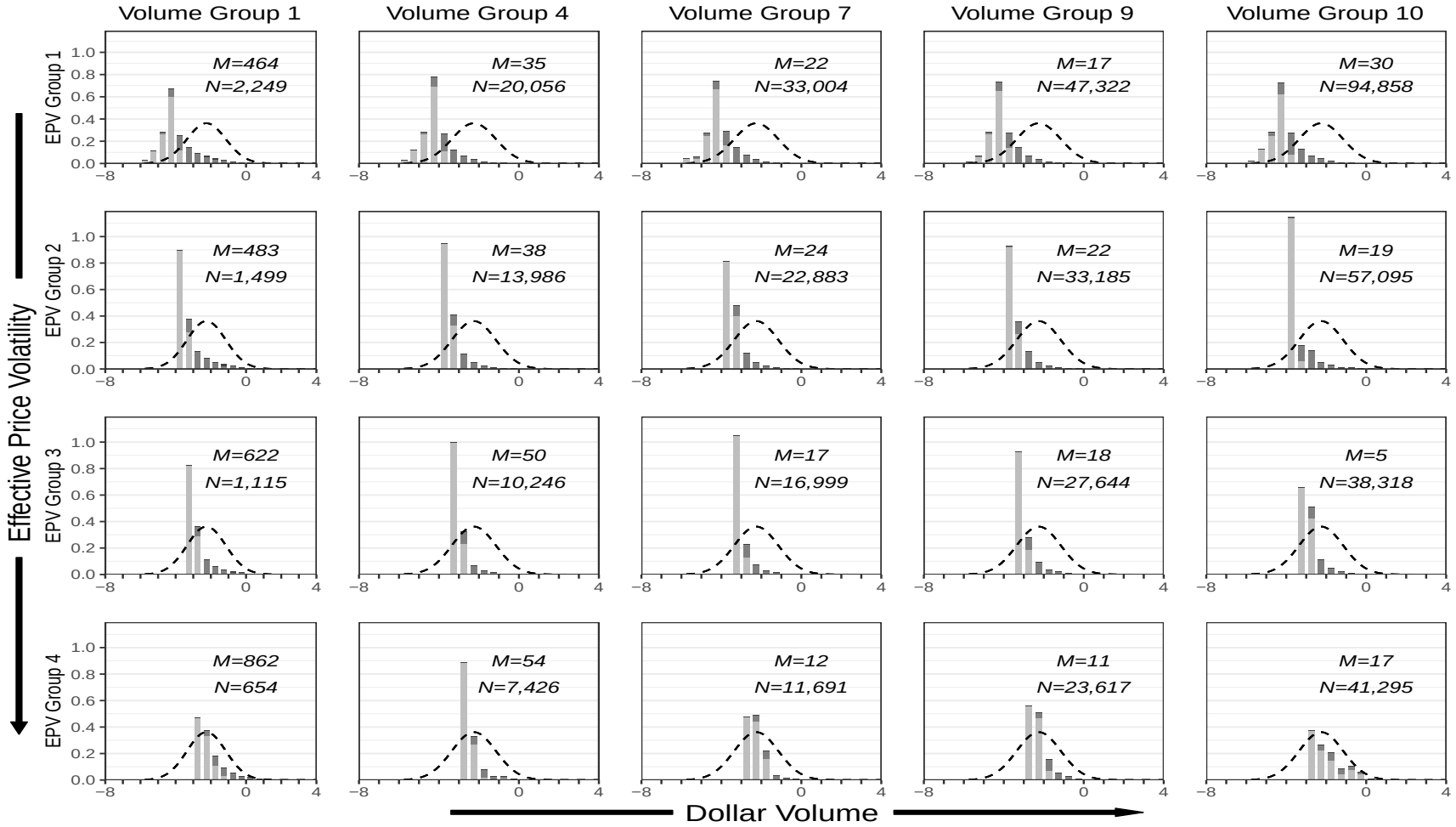

Figure 11: Trade-weighted distributions of scaled TAQ print sizes, all stocks, April 2014. This figure shows distributions of the logarithms of scaled print sizes for NYSE and NASDAQ stocks in April 2014. For each trade, the scaled print size is calculated as $\ln\left(W_i^{2/3} \cdot |X_i|/V_i\right)$ based on the invariance hypothesis, where $|X_i|$ is the midpoint of the print size bin in shares, V_i is the average daily volume in shares, and W_i measures trading activity as the product of dollar volume and the daily percentage standard deviation of returns. Ten volume groups are constructed based on average dollar trading volume with thresholds corresponding to the 30th, 50th, 60th, 70th, 75th, 80th, 85th, 90th, and 95th percentiles of the dollar volume for common NYSE-listed stocks. Four equally spaced volatility groups are constructed based on effective price volatility, defined as $P_i \cdot \sigma_i \cdot (W_i/W_*)^{-1/3}$. The subplots show stock-level distributions averaged across stocks for volume groups 1 (low volume), 4, 7, 9, and 10 (high volume) and for all four price volatility groups 1 (low price volatility), 2, 3, and 4 (high price volatility). The 100-share trades are highlighted in light gray; the 1,000-share trades are highlighted in dark gray. Each subplot also shows a normal distribution with the pooled average print size mean of −2.25 and standard deviation of 1.10. M is the number of stocks, and N is the average number of prints per day for the stocks in a given subgroup.

Appendix A. Robustness results

This appendix presents robustness results for our data aggregation, variable construction, and model estimations. [7]

Aggregating data into bins. We calculate the percentiles of trade size distributions from the raw TAQ data without aggregating data into bins. We run our print size regressions with these percentiles. Given the large volume of the raw transactions data, we produce our results for three months: April 1993, April 2001, and April 2014 and compare them with the results from our bin-aggregated dataset. Table A.7 reports for the percentiles from the aggregated data and Table A.8 reports the results for the percentiles from the raw data. In addition, we calculate dollar volume from the bins using the midpoint of the bins. Table A.17 shows the dollar volume computed from the bins and the actual dollar volume. The distribution of these two dollar volume measures are very comparable. Table A.18 shows estimates from the regression of the actual dollar volume on the dollar volume from the bins. The coefficient on the dollar volume from bins is about 0.97 in both subsample periods.

Volatility estimation. We estimate the daily stock volatility in two different ways: 1) using an ARCH (5) model to predict one-day ahead volatility; 2) computing the daily volatility by using the standard deviation of 5-day log return cumulative from the previous 50 trading days. In both cases, average daily volume is also recalculated over the same window with the volatility for consistency.

Eliminating 10-share stocks. We assess how our results are affected by 10-share stocks by aiming to remove them from our sample. For each month and stock, we calculate the ratio of trades that are below 100 shares. For each month, across stocks, we determine the 95th percentile of the ratio of trades that are below 100 shares. For each month, we remove the stocks that have higher ratio of trades that are below 100 shares than the 95th percentile of the ratio of trades that are below 100 shares.

Annual aggregation. Rather than monthly frequency, we aggregate the data to the annual frequency and rerun our regressions.

Tables A.9, A.10, A.11, and A.12 report the robustness results for volatility estimations, elimination of 10-share stocks, and annual data aggregation.

Effective volatility results without restriction on $\ln(W)$. We rerun our estimation for Table 5 without a restriction on $\ln(W)$. Table A.13 reports the regression results with and without this restriction.

Weighting coefficients with their standard errors. We weight the monthly regression coefficients with the inverse of their standard errors. Tables A.14 and A.15 report the results for weighted coefficients with their standard errors.

Stock and month fixed effects. Instead of month-by-month regressions, we run our number prints estimation with month and stock fixed effects. Table A.16 reports the results of these

[7] For simplicity, we report monthly OLS regression results without the second-stage time-trend regressions.

regressions. In Figure A.12, we plot the stock fixed effect from the first panel of Table A.16 against the stock level average value of $\ln(EPV)$. Stock fixed-effects and $\ln(EPV)$. display a negative correlation, especially when stock fixed-effects are computed after controlling for month fixed-effects. When the effective price volatility is large (the effective tick size is low and the effective lot size is large), the number of prints is low; this suggests the importance of lot size constraint in the U.S. stock market.

Adjusting volatility for tick size. Following Harris (1990), we adjust our volatility estimate for tick-size discreteness. Decimalization of the tick size happened in several stages and these stages can be different for different stocks. However, for simplicity purposes, we take the tick size as 1/8 from January 1993 to May 1997, 1/16 from May 1997 to June 2001, and 0.01 after June 2001. With the tick-size adjustment, we compute the W and rerun our number of prints regression (without time trends). Table A.19 reports the results of this regression. The coefficients on $\ln(w)$ are very similar to the ones reported in Table 2. In addition, Figure A.13 plots our estimated coefficients for $\ln(w)$ with the ones from Table 2; the difference between the two series is very small.

Table A.7
Data aggregation: Bins.

Panel A: April 1993								
	Trade-weighted				Volume-weighted			
	Mean	20th	50th	80th	Mean	20th	50th	80th
Intercept	-7.003	-8.383	-7.081	-5.883	-4.158	-5.930	-4.282	-2.803
	0.021	0.025	0.024	0.023	0.026	0.023	0.027	0.031
$\ln(W/W_*)$	-0.735	-0.795	-0.747	-0.698	-0.507	-0.625	-0.519	-0.426
	0.004	0.005	0.004	0.004	0.005	0.004	0.005	0.006
Avg Adj-R^2	0.864	0.844	0.837	0.828	0.668	0.786	0.648	0.500
Avg # of Obs	5,623	5,623	5,623	5,623	5,623	5,623	5,623	5,623

Panel B: April 2001								
	Trade-weighted				Volume-weighted			
	Mean	20th	50th	80th	Mean	20th	50th	80th
Intercept	-7.743	-8.963	-7.926	-6.763	-5.289	-6.984	-5.549	-3.894
	0.012	0.015	0.012	0.012	0.018	0.013	0.019	0.022
$\ln(W/W_*)$	-0.720	-0.733	-0.738	-0.712	-0.581	-0.674	-0.611	-0.513
	0.002	0.003	0.002	0.002	0.004	0.003	0.004	0.004
Avg Adj-R^2	0.942	0.913	0.936	0.941	0.820	0.914	0.816	0.691
Avg # of Obs	6,005	6,005	6,005	6,005	6,005	6,005	6,005	6,005

Panel C: April 2014								
	Trade-weighted				Volume-weighted			
	Mean	20th	50th	80th	Mean	20th	50th	80th
Intercept	-9.424	-9.393	-9.392	-9.285	-7.939	-9.358	-9.094	-7.468
	0.015	0.019	0.019	0.016	0.034	0.017	0.018	0.025
$\ln(W/W_*)$	-0.741	-0.712	-0.715	-0.786	-0.797	-0.751	-0.870	-0.858
	0.004	0.005	0.005	0.004	0.009	0.005	0.005	0.007
Avg Adj-R^2	0.892	0.832	0.837	0.897	0.657	0.873	0.892	0.805
Avg # of Obs	3,675	3,675	3,675	3,675	3,675	3,675	3,675	3,675

Table A.8
Data aggregation: No bins.

Panel A: April 1993								
	Trade-weighted				Volume-weighted			
	Mean	20th	50th	80th	Mean	20th	50th	80th
Intercept	-7.005	-8.389	-7.106	-5.894	-4.333	-5.936	-4.273	-2.738
	0.021	0.025	0.024	0.023	0.024	0.025	0.028	0.031
$\ln(W/W_*)$	-0.733	-0.797	-0.750	-0.697	-0.506	-0.625	-0.513	-0.404
	0.004	0.005	0.004	0.004	0.004	0.005	0.005	0.006
Avg Adj-R^2	0.860	0.840	0.834	0.823	0.698	0.769	0.626	0.471
Avg # of Obs	5,623	5,623	5,623	5,623	5,623	5,623	5,623	5,623

Panel B: April 2001								
	Trade-weighted				Volume-weighted			
	Mean	20th	50th	80th	Mean	20th	50th	80th
Intercept	-7.756	-8.965	-7.948	-6.790	-5.449	-7.004	-5.526	-3.825
	0.012	0.015	0.013	0.012	0.017	0.015	0.020	0.023
$\ln(W/W_*)$	-0.719	-0.736	-0.738	-0.711	-0.577	-0.673	-0.601	-0.487
	0.002	0.003	0.003	0.002	0.003	0.003	0.004	0.005
Avg Adj-R^2	0.939	0.911	0.931	0.936	0.835	0.898	0.794	0.657
Avg # of Obs	6,005	6,005	6,005	6,005	6,005	6,005	6,005	6,005

Panel C: April 2014								
	Trade-weighted				Volume-weighted			
	Mean	20th	50th	80th	Mean	20th	50th	80th
Intercept	-9.768	-10.095	-9.404	-9.290	-8.198	-9.360	-9.047	-7.008
	0.013	0.017	0.018	0.016	0.016	0.017	0.019	0.028
$\ln(W/W_*)$	-0.748	-0.703	-0.709	-0.786	-0.778	-0.755	-0.877	-0.824
	0.004	0.005	0.005	0.004	0.004	0.005	0.005	0.008
Avg Adj-R^2	0.920	0.864	0.849	0.898	0.896	0.870	0.885	0.758
Avg # of Obs	3,675	3,675	3,675	3,675	3,675	3,675	3,675	3,675

Table A.9
Number of prints: Robustness results.

	Panel A			
	Volatility Estimation #1		Volatility Estimation #2	
	1993-2000	2000-2014	1993-2000	2000-2014
Intercept	5.929	8.207	5.958	8.364
	0.030	0.047	0.032	0.051
$Ln(W/W_*)$	0.681	0.815	0.652	0.791
	0.001	0.003	0.002	0.003
Adj-R^2	0.879	0.932	0.827	0.904
# of Obs	6,653	4,457	6,605	4,446

	Panel B			
	Eliminating 10-share stocks		Annual Aggregation	
	1993-2000	2000-2014	1993-2000	2000-2014
Intercept	6.147	8.506	6.232	8.542
	0.024	0.051	0.101	0.144
$Ln(W/W_*)$	0.667	0.780	0.709	0.829
	0.001	0.003	0.004	0.009
Adj-R^2	0.873	0.921	0.889	0.933
# of Obs	6,618	4,331	6,281	4,253

Table A.10
1993–2000 print sizes: Robustness results.

	Trade-weighted				Volume-weighted			
	Mean	20th	50th	80th	Mean	20th	50th	80th
Panel A: Volatility Estimation #1								
Intercept	-6.997	-8.319	-7.119	-6.091	-4.567	-6.216	-4.764	-3.235
	0.023	0.020	0.022	0.019	0.035	0.027	0.039	0.039
$\ln(W/W_*)$	-0.758	-0.803	-0.771	-0.743	-0.574	-0.678	-0.594	-0.492
	0.002	0.003	0.002	0.002	0.003	0.001	0.003	0.004
Avg Adj-R^2	0.902	0.887	0.897	0.893	0.684	0.836	0.678	0.519
Avg # of Obs	6,653	6,653	6,653	6,653	6,653	6,653	6,653	6,653
Panel B: Volatility Estimation #1								
Intercept	-7.009	-8.440	-7.239	-6.225	-4.723	-6.363	-4.928	-3.397
	0.027	0.018	0.020	0.016	0.031	0.023	0.035	0.036
$\ln(W/W_*)$	-0.721	-0.763	-0.732	-0.709	-0.554	-0.650	-0.575	-0.479
	0.003	0.003	0.002	0.002	0.003	0.002	0.003	0.004
Avg Adj-R^2	0.837	0.870	0.879	0.876	0.647	0.812	0.645	0.482
Avg # of Obs	6,605	6,605	6,605	6,605	6,605	6,605	6,605	6,605
Panel C: Eliminating 10-share Stocks								
Intercept	-7.239	-8.496	-7.290	-6.261	-7.239	-6.365	-4.888	-3.327
	0.016	0.018	0.020	0.018	0.016	0.026	0.039	0.039
$\ln(W/W_*)$	-0.741	-0.781	-0.750	-0.725	-0.741	-0.661	-0.579	-0.481
	0.002	0.003	0.002	0.002	0.002	0.001	0.003	0.003
Avg Adj-R^2	0.896	0.867	0.878	0.880	0.896	0.831	0.679	0.524
Avg # of Obs	6,618	6,618	6,618	6,618	6,618	6,618	6,618	6,618
Panel D: Annual aggregation								
Intercept	-7.346	-8.606	-7.420	-6.362	-4.795	-6.485	-4.992	-3.488
	0.081	0.076	0.102	0.086	0.123	0.110	0.143	0.125
$\ln(W/W_*)$	-0.778	-0.814	-0.792	-0.762	-0.628	-0.700	-0.622	-0.572
	0.009	0.015	0.009	0.007	0.008	0.004	0.009	0.011
Avg Adj-R^2	0.909	0.884	0.895	0.895	0.716	0.854	0.719	0.557
Avg # of Obs	6,281	6,281	6,281	6,281	6,281	6,281	6,281	6,281

Table A.11
2001–2014 print sizes: Robustness results.

	Trade-weighted				Volume-weighted			
	Mean	20th	50th	80th	Mean	20th	50th	80th
Panel A: Volatility Estimation #1								
Intercept	-8.729	-9.270	-9.015	-8.361	-7.158	-8.738	-7.834	-6.038
	0.031	0.011	0.031	0.055	0.064	0.053	0.080	0.080
$\ln(W/W_*)$	-0.803	-0.782	-0.795	-0.836	-0.800	-0.821	-0.868	-0.802
	0.002	0.003	0.002	0.003	0.007	0.004	0.008	0.010
Avg Adj-R^2	0.917	0.881	0.900	0.926	0.781	0.903	0.869	0.763
Avg # of Obs	4,457	4,457	4,457	4,457	4,457	4,457	4,457	4,457
Panel B: Volatility Estimation #1								
Intercept	-8.871	-9.445	-9.197	-8.557	-7.372	-8.942	-8.065	-6.268
	0.035	0.014	0.035	0.059	0.067	0.056	0.084	0.082
$\ln(W/W_*)$	-0.777	-0.751	-0.764	-0.806	-0.779	-0.795	-0.846	-0.784
	0.002	0.002	0.002	0.003	0.007	0.004	0.008	0.010
Avg Adj-R^2	0.881	0.872	0.891	0.918	0.770	0.894	0.860	0.753
Avg # of Obs	4,446	4,446	4,446	4,446	4,446	4,446	4,446	4,446
Panel C: Eliminating 10-share Stocks								
Intercept	-9.020	-9.509	-9.259	-8.623	-7.426	-8.999	-8.120	-6.307
	0.034	0.014	0.033	0.057	0.068	0.056	0.084	0.083
$\ln(W/W_*)$	-0.766	-0.743	-0.756	-0.798	-0.768	-0.785	-0.835	-0.778
	0.002	0.003	0.002	0.002	0.006	0.003	0.007	0.010
Avg Adj-R^2	0.905	0.862	0.881	0.912	0.862	0.891	0.860	0.753
Avg # of Obs	4,331	4,331	4,331	4,331	4,331	4,331	4,331	4,331
Panel D: Annual Aggregation								
Intercept	-9.048	-9.500	-9.298	-8.674	-7.457	-9.068	-8.186	-6.375
	0.089	0.037	0.080	0.168	0.195	0.152	0.250	0.250
$\ln(W/W_*)$	-0.801	-0.771	-0.785	-0.842	-0.845	-0.832	-0.899	-0.873
	0.005	0.009	0.007	0.008	0.021	0.011	0.025	0.034
Avg Adj-R^2	0.911	0.861	0.885	0.923	0.799	0.905	0.895	0.800
Avg # of Obs	4,253	4,253	4,253	4,253	4,253	4,253	4,253	4,253

Table A.12
Robustness results: Number of prints with effective price volatility.
This table presents Fama–MacBeth estimates μ_n and a_σ from monthly regressions

$$\ln(N_i) = \mu_n + \frac{2}{3}\cdot\ln\left(\frac{W_i}{W_*}\right) + a_\sigma\cdot\ln\left(\frac{P_i\cdot\sigma_i}{P_*\cdot\sigma_*}\cdot\left(\frac{W_i}{W_*}\right)^{-1/3}\right) + \tilde{\epsilon}_i.$$

One observation corresponds to each month and each stock i with trading activity W_i, defined as the product of the average daily dollar volume $V_i\cdot P_i$ and the percentage standard deviation σ_i of daily returns. Effective price volatility is defined as $P_i\cdot\sigma_i\cdot\left(\frac{W_i}{W_*}\right)^{-1/3}$, with the effective price volatility of the benchmark stocks $P_*\cdot\sigma_*$ equal to $40\cdot 0.02$. The value of N_i is the average number of prints per day. The scaling constant $W_* = (40)(10^6)(0.02)$ corresponds to the measure of trading activity for the benchmark stock with a price of \$40 per share, trading volume of 1 million shares per day, and daily volatility of 2%. "Adj-R^2" denotes the adjusted R^2 averaged over monthly regressions. The table also reports the average R^2 from the restricted regressions with $a_\sigma = 0$ as well as the average R^2 from unconstrained regressions

$$\ln(N_i) = \mu_n + \frac{2}{3}\ln\left(\frac{W_i}{W_*}\right) + b_1\cdot\ln\left(\frac{V_i}{(10^6)}\right) + b_2\cdot\ln\left(\frac{P_i}{(40)}\right) + b_3\cdot\ln\left(\frac{\sigma_i}{(0.02)}\right) + \tilde{\epsilon}_i.$$

"# of Obs" is the number of stocks averaged over monthly regressions. The estimates are reported for 1993–2000 and 2001–2014 subperiods.

Panel A

	Volatility estimation #1				Volatility estimation #2			
	1993–2000	1993–2000	2000–2014	2000–2014	1993–2000	1993–2000	2000–2014	2000–2014
Intercept	6.083	6.114	7.805	8.234	6.130	6.148	7.848	8.273
	0.029	0.025	0.039	0.039	0.030	0.034	0.039	0.044
$\ln(EPV/EPV_*)$	-0.438		-0.600		-0.492		-0.594	
	0.004		0.008		0.006		0.007	
$\ln(V/V_*)$		0.120		0.260		0.134		0.266
		0.002		0.004		0.003		0.004
$\ln(P/P_*)$		-0.359		-0.265		-0.380		-0.275
		0.004		0.005		0.005		0.005
$\ln(\sigma/\sigma_*)$		-0.460		-0.441		-0.557		-0.535
		0.020		0.014		0.011		0.009
Avg Adj-R^2	0.915	0.922	0.952	0.966	0.870	0.881	0.929	0.946
Avg # of Obs	6,653	6,653	4,457	4,457	6,605	6,605	4,446	4,446

Panel B

	Eliminating 10-share stocks				Annual aggregation			
	1993–2000	1993–2000	2000–2014	2000–2014	1993–2000	1993–2000	2000–2014	2000–2014
Intercept	6.271	6.242	7.972	8.350	6.197	6.292	7.939	8.419
	0.022	0.028	0.039	0.046	0.095	0.103	0.126	0.138
$\ln(EPV/EPV_*)$	-0.473		-0.590		-0.488		-0.589	
	0.003		0.007		0.007		0.029	
$\ln(V/V_*)$		0.122		0.256		0.151		0.276
		0.002		0.004		0.008		0.014
$\ln(P/P_*)$		-0.372		-0.281		-0.350		-0.228
		0.002		0.003		0.016		0.016
$\ln(\sigma/\sigma_*)$		-0.516		-0.494		-0.451		-0.410
		0.012		0.008		0.063		0.040
Avg Adj-R^2	0.918	0.928	0.953	0.969	0.932	0.939	0.950	0.970
Avg # of Obs	6,618	6,618	4,331	4,331	4,253	4,253	4,253	4,253

Table A.13
Number of prints with effective price volatility: Restriction.

Panel A: All Stocks				
	Restricted Model		Unrestricted Model	
	1993-2000	2000-2014	1993-2000	2000-2014
Intercept	6.270	7.952	6.114	8.216
	0.022	0.038	0.029	0.045
Ln(W/W_bar)	***0.667***	***0.667***	0.631	0.746
			0.002	0.003
Ln(EPV/EPV_bar)	-0.471	-0.608	-0.497	-0.546
	0.003	0.008	0.002	0.007
Adj-R^2	0.918	0.955	0.920	0.965
# of Obs	6,621	4,452	6,621	4,452
Panel B: NYSE/AMEX				
	Restricted Model		Unrestricted Model	
	1993-2000	2000-2014	1993-2000	2000-2014
Intercept	6.316	7.998	6.049	8.167
	0.008	0.049	0.014	0.053
Ln(W/W_bar)	***0.667***	***0.667***	0.597	0.741
			0.001	0.003
Ln(EPV/EPV_bar)	-0.338	-0.475	-0.417	-0.443
	0.004	0.010	0.003	0.010
Adj-R^2	0.924	0.955	0.935	0.965
# of Obs	2,189	1,759	2,189	1,759
Panel C: NASDAQ				
	Restricted Model		Unrestricted Model	
	1993-2000	2000-2014	1993-2000	2000-2014
Intercept	6.247	7.942	6.146	8.293
	0.031	0.032	0.041	0.038
Ln(W/W_bar)	***0.667***	***0.667***	0.646	0.754
			0.002	0.002
Ln(EPV/EPV_bar)	-0.497	-0.676	-0.510	-0.604
	0.006	0.007	0.005	0.006
Adj-R^2	0.917	0.959	0.918	0.969
# of Obs	4,432	2,694	4,432	2,694

Table A.14
Number of prints: Standard-error weighted results.

Panel A						
	All stocks		NYSE/AMEX		NASDAQ	
	1993–2000	2001–2014	1993–2000	2001–2014	1993–2000	2001–2014
Intercept	6.192	8.476	6.125	8.348	6.212	8.586
	0.165	0.053	0.131	0.051	0.182	0.057
$\ln(W/W_*)$	0.666	0.785	0.626	0.753	0.680	0.811
	0.011	0.007	0.009	0.010	0.012	0.009
Avg Adj-R^2	0.873	0.923	0.912	0.933	0.858	0.913
Avg # of Obs	6,621	4,452	2,189	1,759	4,432	2,694

Panel B						
	All stocks		NYSE/AMEX		NASDAQ	
	1993–2000	2000–2014	1993–2000	2000–2014	1993–2000	2000–2014
Intercept	6.290	7.899	6.322	7.948	6.279	7.891
	0.097	0.065	0.077	0.108	0.110	0.067
$\ln(W/W_*)$	0.667	0.667	0.667	0.667	0.667	0.667
$\ln(EPV/EPV_*)$	-0.473	-0.604	-0.340	-0.483	-0.502	-0.669
	0.007	0.007	0.007	0.014	0.010	0.005
Avg Adj-R^2	0.918	0.955	0.924	0.955	0.917	0.959
Avg # of Obs	6,621	4,452	2,189	1,759	4,432	2,694

Table A.15
Print sizes: Standard error weighted results.

	Panel A: January 1993 to December 2001							
	Trade-weighted				Volume-weighted			
	Mean	20th	50th	80th	Mean	20th	50th	80th
Intercept	-7.263	-8.516	-7.320	-6.296	-4.731	-6.415	-4.944	-3.370
	0.201	0.226	0.227	0.186	0.111	0.180	0.125	0.085
$\ln(W/W_*)$	-0.738	-0.776	-0.746	-0.722	-0.562	-0.661	-0.581	-0.483
	0.013	0.013	0.016	0.014	0.008	0.012	0.009	0.007
Avg Adj-R^2	0.896	0.866	0.878	0.880	0.686	0.830	0.679	0.524
Avg # of Obs	6,621	6,621	6,621	6,621	6,621	6,621	6,621	6,621

	Panel B: January 2001 to December 2014							
	Trade-weighted				Volume-weighted			
	Mean	20th	50th	80th	Mean	20th	50th	80th
Intercept	-8.983	-9.508	-9.221	-8.566	-7.267	-8.962	-8.123	-6.271
	0.056	0.063	0.063	0.053	0.134	0.055	0.107	0.096
$\ln(W/W_*)$	-0.778	-0.760	-0.771	-0.805	-0.756	-0.791	-0.834	-0.764
	0.012	0.012	0.013	0.010	0.016	0.008	0.007	0.009
Avg Adj-R^2	0.907	0.866	0.884	0.913	0.778	0.892	0.862	0.760
Avg # of Obs	4,452	4,452	4,452	4,452	4,452	4,452	4,452	4,452

Table A.16
Number-of-prints regressions with stock and month fixed effects.
This table presents the results from the panel regression:

$$\ln\left(N_i\right) = \mu_n + \alpha \cdot \ln\left(\frac{W_i}{W_*}\right) + \tilde{\epsilon}_i.$$

The panel regressions are run with and without stock and month fixed effects. The White standard errors are reported below the coefficient estimates.

Panel A: All stocks: February 1993 to December 2014				
Intercept	7.828	6.295	7.479	5.543
	0.002	0.008	0.003	0.007
$\ln(W/W_*)$	0.837	0.741	0.746	0.595
	0.000	0.000	0.001	0.000
Stock Fixed Effects	No	No	Yes	Yes
Month Fixed Effects	No	Yes	No	Yes
Avg Adj-R^2	0.78	0.94	0.87	0.97
Avg # of Obs	1,383,857	1,383,857	1,383,857	1,383,857
$\ln(W/W_*)$=2/3	$p < 0.001$	$p < 0.001$	$p < 0.001$	$p < 0.001$

Panel B: All stocks: February 1993 to December 2000				
Intercept	6.221	6.146	5.714	5.173
	0.002	0.002	0.003	0.006
$\ln(W/W_*)$	0.683	0.666	0.567	0.517
	0.000	0.002	0.001	0.001
Stock Fixed Effects	No	No	Yes	Yes
Month Fixed Effects	No	Yes	No	Yes
Avg Adj-R^2	0.87	0.89	0.93	0.95
Avg # of Obs	634,322	634,322	634,322	634,322
$\ln(W/W_*)$=2/3	$p < 0.001$	$p = 0.0242$	$p < 0.001$	$p < 0.001$

Panel C: All stocks: January 2001 to December 2014				
Intercept	8.551	8.439	8.050	6.226
	0.002	0.001	0.003	0.007
$\ln(W/W_*)$	0.818	0.785	0.669	0.587
	0.000	0.000	0.001	0.001
Stock Fixed Effects	No	No	Yes	Yes
Month Fixed Effects	No	Yes	No	Yes
Avg Adj-R^2	0.88	0.93	0.92	0.97
Avg # of Obs	749,535	749,535	749,535	749,535
$\ln(W/W_*)$=2/3	$p < 0.001$	$p < 0.001$	$p = 0.0052$	$p < 0.001$

Table A.17
Volume comparison statistics.

Panel A: 1993–2000							
	5th	25th	50th	75th	95th	Mean	Std Dev
Dollar Volume ($ 1000)	9.01	55.35	275.40	1,749.73	21,001.95	6,186.11	47,661.55
Dollar Volume from Bins ($ 1000)	8.95	55.68	278.68	1,781.85	21,571.51	6,368.07	49,229.03
Panel B: 2001–2014							
	5th	25th	50th	75th	95th	Mean	Std Dev
Dollar Volume ($ 1000)	11.96	123.96	1,365.52	10,692.92	108,488.27	24,407.87	122,187.42
Dollar Volume from Bins ($ 1000)	12.11	125.85	1,389.41	10,902.08	110,867.82	24,914.80	124,924.79

Table A.18
Volume comparison regression.

	Dollar Volume	
	1993–2000	2001–2014
Intercept	68,111	220,509
	55,022	162,649
Dollar Volume From Bins	0.9682	0.9715
	0.0099	0.0071
# of obs.	634,322	749,535
Adj-R^2	0.9795	0.9867

Table A.19
Regression after tick-size adjustment.

	1993–2000	2001–2014
Intercept	6.1390	8.5132
	0.0244	0.0513
$\ln(W\_Tick\_Adj/W^*)$	0.6619	0.7900
	0.0010	0.0033
# of Stocks	6502	4448
Adj-R^2	0.867	0.923

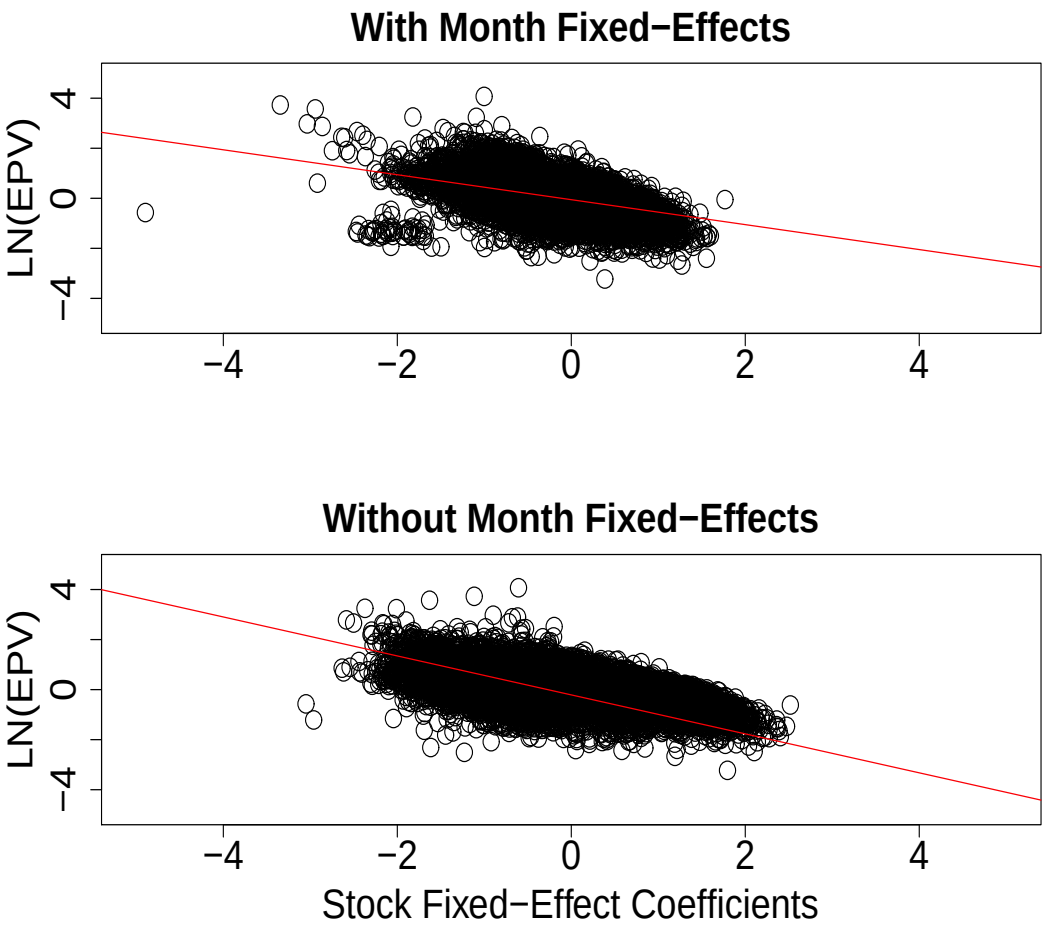

Figure A.12: Stock fixed effects.

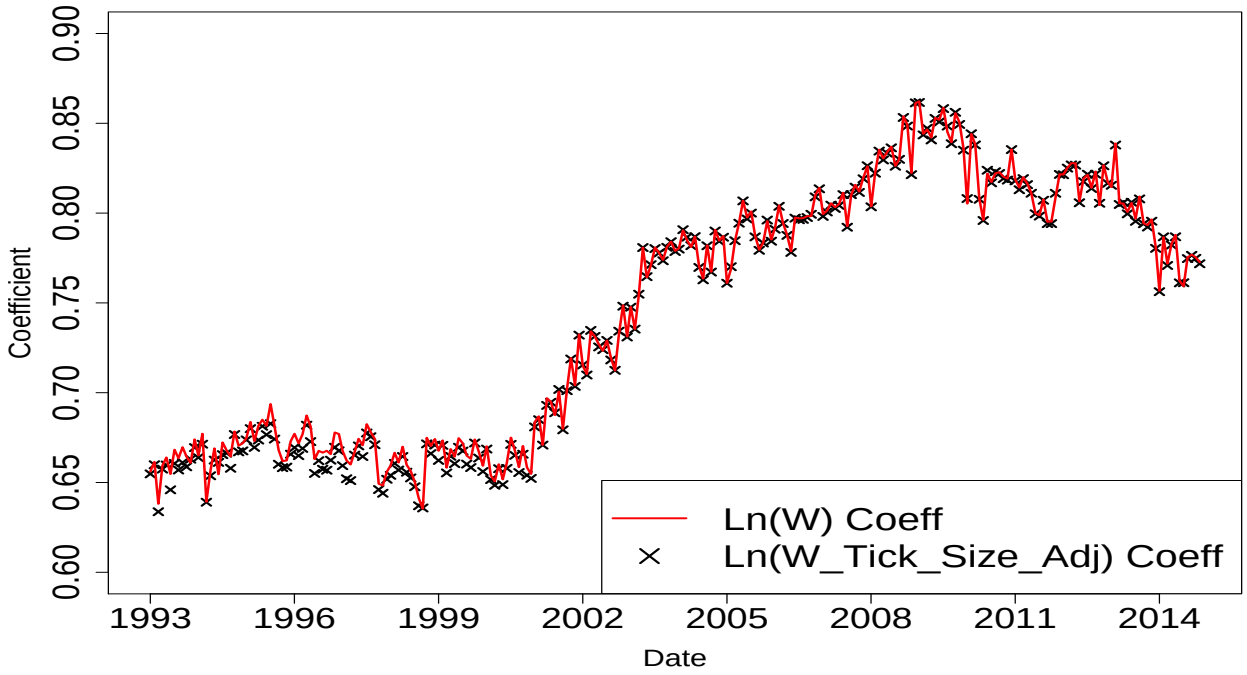

Figure A.13: Number-of-prints regression coefficients with tick-size adjustment.