Trading Game Invariance in the TAQ Dataset[*]

Albert S. Kyle
Robert H. Smith School of Business
University of Maryland
akyle@rhsmith.umd.edu

Anna A. Obizhaeva
Robert H. Smith School of Business
University of Maryland
obizhaeva@rhsmith.umd.edu

Tugkan Tuzun
Federal Reserve Board
tugkan.tuzun@frb.gov

Original Version: September 1, 2010
This version: December 22, 2011

Abstract

The trading game invariance hypothesis of Kyle and Obizhaeva (2011a) is tested using the Trades and Quotes ("TAQ") dataset. Over the period 1993-2001, the estimated monthly regression coefficients of the log of trade arrival rate on the log of trading activity has an almost constant value of 0.690, slightly higher than the value of 2/3 predicted by the invariance hypotheses. Over the period 2001-2008, the coefficient estimates rise almost linearly, with an average value of 0.787. Average trade size, normalized for trading activity, falls dramatically over the period 1993-2008. The distribution of trade size adjusted for trading activity resembles a log-normal more closely in 1993 than in 2001 or 2008, with truncation below the 100-share odd-lot boundary becoming a more prominent feature over time. These results suggests that the 2001 reduction in minimum tick size to one cent and the subsequent increase in algorithmic trading have resulted in more intense order shredding in actively traded stocks than inactively traded stocks. The invariance hypothesis explains 91% of the cross-sectional variation in print arrival rates and average print size.

[*] We are grateful to Peter Bossaerts, Elena Asparouhova, and Stewart Mayhew for very helpful comments on an earlier draft of this paper. The views expressed herein are those of the authors and do not necessarily reflect the views of the Board of Governors or the staff of the Federal Reserve System.

1 Introduction

The trading game invariance hypothesis of Kyle and Obizhaeva (2011a) makes precise predictions concerning how arrival rates and the size distributions of intended orders vary across stocks with different levels of trading activity. We use these predictions as a lens to examine the frequency and size of trades printed in the Trades and Quotes ("TAQ") dataset for the period 1993-2008.

The invariance hypothesis is based on the intuition that trading in securities markets can be modeled as games played at different speeds or over different horizons for different stocks. In actively traded stocks, trading takes place at fast speeds over short horizons, perhaps only a few minutes. In inactively traded stocks, trading takes place slowly over longer horizons, perhaps a few months. Letting P denote price per share and V denote expected daily share volume, Kyle and Obizhaeva (2011) define "trading activity," denoted W, as the product of expected daily dollar volume $P \cdot V$ ("dollar volume") and the expected daily percentage standard deviation of returns σ ("volatility"). The product of dollar volume and volatility is a better measure of the risk transfer taking place in the market than the dollar volume alone.

The size of an intended order, which Kyle and Obizhaeva (2011a) call a "bet," is defined as the product of stock price P, signed trade size in shares $\tilde{Q}$, and volatility σ. The expected number of bets per day is denoted by γ. The hypothesis of trading game invariance predicts that when the expected number of bets is normalized by $W^{-2/3}$, the normalized bet arrival rate $\gamma \cdot W^{-2/3}$ is constant across stocks and across time. The invariance hypothesis also predicts that when intended order size as a fraction of daily volume is normalized by $W^{2/3}$, the distribution of normalized intended order size $(\tilde{Q}/V) \cdot W^{2/3}$ is invariant as well. In terms of the log-linear regressions $\ln(|\tilde{Q}|/V) = \mu_Q + a_Q \cdot \ln(W) + \tilde{\epsilon}$ and $\ln(\gamma) = \mu_\gamma + a_\gamma \cdot \ln(W) + \tilde{\epsilon}$, the invariance hypothesis nests several distinct predictions:

- The regression coefficients a_Q and a_γ are predicted to be $-2/3$ and $2/3$, respectively. A one percent increase in trading activity can be decomposed into an increase in the arrival rate of intended orders of $2/3$ of one percent and an increase in the size of intended orders of $1/3$ of one percent, implying that intended order size as a fraction of average daily volume decreases by $2/3$ of one percent.

- The shape of the distribution of the error term in the first regression is the same across stocks and across time and does not vary with other stock characteristics. This implies that the coefficients a_Q and a_γ are equal to $-2/3$ and $2/3$ not only for a mean-regression, but also for any quantile regression. Using portfolio transition data, Kyle and Obizhaeva (2011b) find that the distribution of $\tilde{\epsilon}$ is close to a normal with variance 2.50. This suggests that a one standard deviation change in intended order size is a factor of $exp(2.50^{1/2}) = 4.86$.

- The constant terms μ_Q and μ_γ do not vary across stocks or across time (adjusting for inflation), implying that the log-means and log-percentiles of normalized trade size and normalized number of trades do not vary across stocks or across time.

- If individual stock characteristics—such as stock price or stock volatility—are added to the regression, the R^2 is predicted not to increase much.

The TAQ dataset reports "prints" of tick-by-tick transactions prices and share quantities for all listed U.S. stocks from 1993 to present. For each stock and each month, we construct a frequency distribution of TAQ print sizes (in shares). We describe the distribution of TAQ print sizes by a list of statistical attributes, including the trade-weighted and volume-weighted mean print size and ten equally-spaced trade-weighted and volume-weighted percentiles. The volume-weighted deciles put more weight on larger and more economically significant trades than trade-weighted deciles, thus allowing us to examine the right tail of distributions of TAQ print sizes in more detail. Our goal is to use the invariance hypothesis to interpret time series and cross-sectional variation in TAQ print frequency and print size during the period 1993-2008.

The invariance hypothesis makes predictions about intended orders rather than actual TAQ prints resulting from order execution strategies. As a result of the important distinction between an intended order and a TAQ print, we do not expect estimates based on TAQ prints to match precisely the predictions of the invariance hypothesis for intended orders, even if the invariance hypothesis describes intended orders perfectly. To interpret empirical results, we make the identifying assumption that each intended order is reported as a constant number ξ of TAQ prints, where the order shredding factor ξ does not vary across intended order size, ticker symbol, or time.

Executions of bets are intermediated by market makers, high frequency traders, and other arbitragers. This intermediation results in additional non-bet volume. To interpret our results, we make the identifying assumption that trading volume increases by a factor ζ as a result of intermediation trades, i.e., each intended order generates extra expected non-bet trading of $\zeta - 1$ times the size of the intended order, where the volume multiplier $\zeta \geq 1$ does not vary across intended order size, ticker symbol, or time.

For example, if each intended order is executed against single market maker with no order shredding, we have $\xi = 1$ and $\zeta = 2$.

We do not expect these benchmark assumptions of a constant order-shredding factor ξ and a constant intermediation factor ζ to describe markets correctly. Electronic order handling and regulatory changes have likely encouraged the splitting of intended orders into more TAQ prints with more intermediation over time, relative to our benchmark assumptions. As a result, we expect that our empirical results have the best correspondence with predictions at the beginning of the period, when traditional upstairs markets accounted for a large share of order execution, order shredding was less common, and the intermediation scheme was simpler. We conjecture that during the period 1993-2008, the normalized mean TAQ print size has declined and the normalized mean number of prints has correspondingly increased, reflecting ever increasing order shredding and intermediation. Due to the interaction between order execution strategies and microstructure effects such as minimum tick size and 100-share minimum round lot size, we hypothesize that relatively large intended orders in more active stocks tend to generate more prints per intended order and perhaps more intermediation trades than relatively small intended orders in inactive stocks. We expect therefore the estimated exponents a_Q and a_γ to be higher in absolute values than the benchmark prediction of 2/3 and increase in absolute values over time.

Our empirical results are broadly consistent with these predictions.

A time series of separate month-by-month regression coefficients of the log of TAQ print arrival rates on the log of trading activity shows that the estimated coefficients a_γ remained virtually constant over the subperiod 1993-2001. Over this subperiod, the estimated coefficient of 0.690 is only slightly higher than the benchmark prediction of 2/3, but this difference is statistically significant due to the low standard errors of the estimated coefficient. Under the invariance hypothesis, the fact that 0.690 is only slightly large than 2/3 is consistent with the hypothesis that orders for actively traded stocks generate only slightly more TAQ prints than orders for inactively traded stocks during the period 1993-2001. After 2001, the estimated monthly coefficients increase almost linearly, from about 0.690 in 2001 to about 0.850 in 2008. This increase is statistically highly significant.

Under the invariance hypothesis, the pattern of steadily increasing coefficients is consistent with the interpretation that something happened in 2001 which resulted in intended orders in active stocks being shredded into far more TAQ prints over time than intended orders in inactive stocks. This result is clearly consistent with the interpretation that the reduction in tick size from 1/16 of a dollar to one cent, which occurred in 2001, has led to more and more order shredding over time, especially in large orders for active stocks. The result is also consistent with the interpretation that there has been a larger increase in small intermediation trades as a fraction of total volume in more actively traded stocks than less actively traded stocks, especially after 2001. These interpretations are consistent with our finding that the estimated exponents a_Q in our log-linear regressions for percentiles of trade-weighted and volume-based distributions of TAQ print sizes have increased in absolute values during the period 1993-2008, and these changes are more pronounced for trade-weighted and volume-weighted percentiles representing large prints than small prints.

We examine empirical distributions of logs of normalized TAQ print sizes for ticker symbols sorted into ten dollar volume groups and four price volatility groups in years 1993, 2001, and 2008 (see figures 3-7). In 1993, all forty empirical distributions resemble a bell-shaped normal density function with common mean and variance across the forty subgroups. In 2001 and 2008, the shape of the distributions look much less like a normal than in 1993. Furthermore, average print size decreases during the period 1993-2008 by a factor of about ten. The estimated variances of 1.98 in 1993, 1.75 in 2001, and 1.33 in 2008 for trade-weighted distributions of TAQ print sizes are lower than the variance of 2.50 for distributions of portfolio transition trades, reported in Kyle and Obizhaeva (2011); furthermore, these variances decrease over time, consistent with our hypothesis that larger orders generate relatively more prints than smaller orders over time.

Statistical tests clearly reject the hypothesis that normalized TAQ print sizes are distributed as a common log-normal random variable. We conjecture that the rejection may arise due to microstructure effects such as the one-cent tick size and clustering of trades at round lot sizes such as 100, 1000, and 5000 shares. In particular, the following three issues may be relevant:

First, the distribution of normalized TAQ prints is truncated by a 100-share odd-lot boundary, consistent with the fact that odd-lot trades are not reported on the TAQ tape, as discussed in O'Hara, Yao, and Ye (2011).

Second, TAQ print sizes tend to cluster at "even" quantity levels, as studied in Alexander and Peterson (2007) and Moulton (2005). For example, there are more trades of 5,000 shares than 4,000 or 6,000 shares, and far more than 4,900 and 5,100 shares. As a result of clustering, our empirical distributions of normalized TAQ prints have spikes at "even" quantity levels. The biggest spike corresponds to TAQ prints of 100 shares. The 100-share prints account for about 16% of all trades before 2001 and 50% of trades after 2001; prints of 100 shares made up to 70% of all TAQ prints in 2008. Empirical distributions of TAQ prints for NASDAQ stocks also have another large spike that corresponds to a 1,000-share print size. The frequency of 1,000-share prints is much more pronounced in 1993, than in 2001 or in 2008 (see figure 5). After 1988, NASDAQ dealers often had obligations to make binding bid and ask quotes for 1,000 shares for trades with market participants, as a result of which prints for exactly 1,000 shares were common. This restriction was gradually removed beginning in 1997.

Third, the minimum tick size was reduced from 1/8 of a dollar (12.5 cents) to 1/16 of a dollar (6.25 cents) in 1997, and to one cent in 2001. 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 typical days trading range, and thus there are better opportunities for order shredding and making intermediation trades. Although firms can implement stocks splits to adjust relative tick sizes, these adjustments occur infrequently and with long time lags, as noted in Angel (1997). The invariance hypothesis suggests that the tick size of one cent is effectively greater when the trading game is played faster. This implies that "effective" tick size is lower for stocks with high price and high volatility. Low effective relative tick size is likely to encourage traders to split their orders into smaller trades placed at finer adjacent price points as a strategy to avoid front-running. When effective relative tick size is low, the effective round lot size constraint of 100-shares is more binding, with more odd-lot orders not recorded on the tape and more small trades rounded up to precisely 100 shares.

When we impose restrictions $a_Q = -2/3$ and $a_\gamma = 2/3$, implied by the invariance hypothesis, and estimate only intercepts μ_Q and μ_γ in separate month-by-month regressions, the time series of R-squares fluctuates around 0.91. This implies that the invariance hypothesis explains 91% of the variation across stocks in TAQ print arrival rates and average print sizes. During the two periods 1993-2000 and 2001-2008, we show that an additional 3% and 4% of the variation, respectively, can be attributed to variation across stocks in "effective price volatility," which scales price volatility $P \cdot \sigma$ for the time units implied by trading game invariance. An interesting topic for the further research is to examine the extent to which other microstructure effects explain the remaining variation.

Previous studies have analyzed average print sizes as well. Glosten and Harris (1988) find that average print size (in shares) is negatively related to market depth. Brennan and Subrahmanyam (1998) document that print sizes (in dollars) are also related to return volatility, standard deviation of trading volume, market capitalization, number of analysts following stocks, number of institutional investors holding stocks, and proportion of shares they hold. The fact that an R-square of 0.92 in their cross-sectional regressions is only slightly larger than the average R-square of 0.91 in our restricted regressions suggests that other variables offer limited improvement in explanatory power over the invariance hypothesis.

Our results supplement those of Kyle and Obizhaeva (2011b), which provides evidence in favor of the invariance hypothesis using a sample of portfolio transitions. Transition trades represent a subset of transactions well-suited for testing of the invariance hypothesis because of the correspondence between intended orders and actual trades. While this study of TAQ data involves a much broader sample of trades, this advantage comes at the expense of having to deal with data affected by order shredding and intermediation trades.

The remainder of this paper states the implications of the invariance hypothesis, discusses the design of our empirical tests, and presents our results.

2 Testable Implications of the Invariance Hypothesis.

Kyle and Obizhaeva (2011a) think of trading activity as the outcome of traders playing trading games. Innovations in the order flow, called "bets," are assumed approximately to follow a compound Poisson process with an expected arrival rate of γ bets per day. A bet is a random variable $\tilde{B}$ that measures the risk transferred during one transaction. It is defined as the product of dollar share price P, share quantity $\tilde{Q}$, and daily percentage returns volatility σ,

$$\tilde{B} = \tilde{Q} \cdot P \cdot \sigma. \tag{1}$$

The random variable $\tilde{Q}$ has a zero mean; positive values represent buying; and negative values represent selling. In a similar spirit, "trading activity" is defined as the product of expected daily share volume V, share price P, and daily volatility σ,

$$W = V \cdot P \cdot \sigma. \tag{2}$$

In theory, the product of dollar volume and volatility is a better measure of the risk transfer taking place in the market than dollar volume alone. Empirically, there is so much more cross-sectional variation in dollar volume than volatility across stocks that results are similar if dollar volume alone is used as the measure of trading activity.

Expected daily share volume V is the product of the expected arrival rate of bets and their expected size,

$$V = \frac{\zeta}{2} \cdot \gamma \cdot E|\tilde{Q}|, \tag{3}$$

augmented by non-bet volume. The parameter ζ is a "volume multiplier," which measures the factor by which non-bet "intermediation" trades increase trading volume. The parameter ζ is divided by 2 because each unit of trading volume has both a buy side and a sell side. Non-bet volume includes trading by market makers, high frequency traders, and other arbitragers who intermediate among long-term bets. If each bet is intermediate by a single market maker, similar to an NYSE specialist, this is consistent with $\zeta = 2$. If bets are intermediated by many market makers, who lay off positions trading with each other, similar to NASDAQ in the early 1990's, this is consistent with $\zeta = 3$. Plugging equation (3) into equation (2), we find that trading activity W is the product of the volume multiplier ζ, bet arrival rate γ, and expected bet size $E|\tilde{B}|$:

$$W = \frac{\zeta}{2} \cdot \gamma \cdot E|\tilde{B}|. \tag{4}$$

The invariance hypothesis describes how market microstructure characteristics such as bet arrival rates and bet sizes vary across securities with different levels of trading activity.

The invariance hypothesis is based on the intuition that trading games are the same across securities and across time, up to some Modigliani-Miller transformation, except for the speed with which they are being played. In actively traded stocks, trading takes place at fast speeds over short horizons. In inactively traded stocks, trading takes place slowly over long horizons. Trading game invariance is derived from the following invariance assumption:

If units of time are rescaled proportionally to match the arrival rate of bets γ, so that one bet is expected to arrive per tick on the rescaled clock, then the dollar distribution of gains and losses on a bet between one tick on the clock and the next is invariant across different trading games in different stocks. Equivalently, trading game invariance is equivalent to the hypothesis that the random variable $\tilde{I}$, called a "trading game invariant," and defined by

$$\tilde{I} = \frac{\tilde{B}}{\gamma^{1/2}}, \tag{5}$$

has a probability distribution which is invariant across stocks and time. In the above equation, the denominator is $\gamma^{1/2}$ instead of γ because bet size $\tilde{B}$ is proportional to the standard deviation of returns while the time units which are rescaled by bet arrival rate γ are proportional to the variance of returns.

Kyle and Obizhaeva (2011a) show that the trading game invariance hypothesis leads to the following testable predictions concerning how bet arrival rate γ and bet size $\tilde{Q}$ change with trading activity:

$$\gamma = E\Big[\frac{\zeta}{2} \cdot |\tilde{I}|\Big]^{-2/3} \cdot W^{2/3}, \tag{6}$$

$$\frac{|\tilde{Q}|}{V} \approx E\Big[\frac{\zeta}{2} \cdot |\tilde{I}|\Big]^{-1/3} \cdot W^{-2/3} \cdot \tilde{I}. \tag{7}$$

Under the identifying assumption that the volume multiplier ζ is constant, equation (6) and equation (7) imply that the normalized bet arrival rate $W^{-2/3} \cdot \gamma$ and distributions of the normalized bet sizes $W^{2/3} \cdot |\tilde{Q}|/V$ are invariant across trading games in different stocks. Under the identifying assumption that portfolio transition trades are typical bets, Kyle and Obizhaeva (2011) find that the distribution of logs of normalized bet sizes, $2/3 \cdot \ln(W) + \ln(|Q|/V)$, is close to a normal with variance of 2.50.

The invariance hypothesis implies that changes in daily trading activity come from both changes in bet size and changes in bet arrival rate. A one percent increase in trading activity W is associated with an increase of 2/3 of one percent in the bet arrival rate γ and an upward shift by 1/3 of one percent of the entire distribution of bet size $|\tilde{B}|$, which is equivalent to a downward shift by 2/3 of one percent in the distribution of bet sizes $|\tilde{Q}|$ as a fraction of trading volume V.

As discussed by Kyle and Obizhaeva (2011a), market microstructure invariance includes not only the hypothesis of trading game invariance but also the additional hypotheses of market impact invariance and bid-ask spread invariance. This paper focusses only on the predictions of trading game invariance concerning the arrival rate and size distribution of intended orders. Testing the implications market impact invariance and bid-ask spread invariance for TAQ data are left as interesting topics for future research.

Alternative Models. As in Kyle and Obizhaeva (2011a), we consider two alternative models: the model of invariant bet frequency and the model of invariant bet size.

The model of invariant bet frequency assumes that variation in trading activity comes entirely from variation in bet sizes $\tilde{B}$, while the expected number of bets γ over a calendar day is invariant. In this model, the bet arrival rate $\gamma \cdot W^0$ and the distribution of bet sizes

$\frac{|\tilde{Q}|}{V} \cdot W^0$ are invariant across stocks and across time, even if no adjustments for differences in trading activity are made (as represented by zero exponents).

The model of invariant bet size assumes that variation in trading activity comes entirely from variation in the number of bets γ placed over a calendar day, while the distribution of bet size $\tilde{B}$ over a calendar day remains the same across stocks. In this model, the normalized bet arrival rate $\gamma \cdot W^{-1}$ and the distribution of normalized bet size $\frac{|\tilde{Q}|}{V} \cdot W^1$ are invariant across stocks and across time.

All three models make specific nested predictions concerning how the expected number of bets γ and the distribution of bet sizes $\tilde{B}$ per calendar day are related to the measure of trading activity W on the other side. The only differences among these predictions are the implied exponents on trading activity. We discuss next how to test the three models using the TAQ dataset.

Testing Invariance Models Using TAQ Print Data. The TAQ dataset reports "prints" of tick-by-tick transaction prices and share quantities for all listed U.S. stocks from 1993 to present. We test implications of the invariance hypothesis using data on TAQ prints sizes and number of TAQ prints recorded per day. TAQ prints are different from bets. In actual trading, one independent trading decision often generates multiple reports of order executions, since orders may be broken down into smaller pieces for execution and traded with several different counter-parties and at several different prices. Rather than representing bets, some TAQ prints may represent trades by intermediaries who bear risks of bets temporarily, until they find their place in the portfolios of long-term investors.

Let X denote the unsigned number of shares representing a single "print" in the TAQ dataset. We assume that each bet is shredded into ξ trades, and the distribution of TAQ print sizes $\tilde{X}$ differs from the distribution of bets $\tilde{Q}$ by a factor ξ. We can express this assumption as,

$$|\tilde{Q}| \approx \xi^{-1} \cdot X. \tag{8}$$

Let N denote the number of trades printed per day. We assume that each bet $\tilde{Q}$ is shredded into ξ trades and generates $\zeta \cdot \tilde{Q}$ of non-bet volume by intermediaries who endogenously respond to bets. Therefore, we assume that the expected number of bets γ differs from the average number of TAQ prints $\bar{N}$ by a factor of $\xi \cdot \zeta/2$, where division by 2 is necessary because each transaction is a match between a buyer and a seller. We can express thus assumption as,

$$\gamma = (\xi \cdot \zeta/2)^{-1} \cdot \bar{N}. \tag{9}$$

As a benchmark for interpreting our empirical results, we make two identifying assumptions. First, we assume that the order-shredding factor ξ is constant across securities and across time. Second, we assume that the volume multiplier ζ is constant across securities and across time.

For simplicity of exposition, we interpret our results under the identifying assumptions $\xi = 1$ and $\zeta = 2$. This corresponds to the hypothesis that each intended trade is executed as one print against a single intermediary, such as a designated market maker like the NYSE specialist of the mid-1990's. In actual markets, ξ may deviate from the assumed value $\xi = 1$ because order shredding algorithms may vary across securities in a complex and systematic manner, for example, as a function of trade size and stock price (based on tick size). In actual markets, ζ may deviate from the assumed value $\zeta = 2$; the amount of intermediation may potentially depend on various stock characteristics, such as price volatility and trading volume.

Institutional Details Related to the Microstructure of TAQ DAta. We make the benchmark assumptions for the purpose of interpreting our results, but we do not expect these assumptions to describe TAQ prints correctly.

We do not expect that the order-shredding parameter ξ is constant across intended order size, ticker symbol, and time. We hypothesize that large intended orders may result in more TAQ prints than small intended orders. Several factors may have contributed to this pattern:

- Although traders have increasingly shredded orders into "odd lots" of less than 100 shares, some traders probably resist shredding orders into numerous odd lots.
- Under NYSE Rule 411(b), broker-dealer member firms have an obligation to consolidate a customer's odd-lot orders if the share amount of such orders exceeds 100 shares; other exchanges have similar provisions and have brought enforcement cases against member firms that did not comply with those rules.
- During the period under investigation, odd-lot transactions were executed through a separate odd-lot trading system, and these small trades were not reported for dissemination on the Consolidated Tap, as discussed in O'Hara, Yao, and Ye (2011).
- "Tape shredding" affects trading patterns, as suggested by Caglio and Mayhew (2007); large orders can be broken up into more trades than small order to generate additional revenue from sales of consolidated trade and quote data.
- Large intended orders are likely to be matched against multiple intended orders of smaller size, also resulting in more TAQ prints for larger intended orders than for smaller ones. According to the Consolidated Tape Association (CTA), the exchanges are required to collect and report last sale data (CTA Plan (1992) Section VII). At the NYSE, for example, it is the duty of the member representing the seller to ensure that a trade has been reported. Since the rules required reporting of "sales" not "trades," intrinsic order splitting may be more important for intended buy orders than intended sell orders.

Furthermore, we expect technological progress in computing and various regulatory changes to have made the relationship between the predictions of the invariance hypothesis and empirical results for TAQ prints vary during the period 1993-2008. We expect that prints are more similar to intended orders in the first half of the sample rather than in the second one. For example, in the early 1990s, prior to the common use of electronic interfaces, traders more frequently executed intended orders as block trades in the "upstairs" market, in which case at least one side of the reported block trade might correspond precisely to an intended order. Prior to the changes in order handling rules in 1997, NASDAQ dealers often took the other side of entire intended orders because customers could not place their own orders into a central limit order book and NASDAQ dealers might be unhappy if customers "bagged" them by dumping large quantities of stock on many separate dealers at the same time.

Over the entire time period, traders have often broken intended orders into pieces for execution, as documented in Keim and Madhavan (1995), in which case an intended order has shown up as multiple TAQ prints spread out over the time. As the use of the NYSE's Direct Order Transfer system ("DOT" system) became more commonly used by professional traders in the 1990s, the use of electronic order submission strategies and order shredding became more common. For NASDAQ stocks, this practice accelerated after new order handling rules were implemented in the late 1990s.

In 2001, the minimum tick size was cut from 6.25 cents to one cent, as a result of which quoted spreads decreased and fewer shares were shown at the best bid and offer. Traders could use electronic interfaces to place scaled limit orders of small size at adjacent price points separated by one cent, resulting in smaller TAQ print sizes for the same intended order size. Adoption of Regulation NMS in 2005 resulted in market fragmentation, which also fragmented TAQ print sizes over multiple trading venues. Regulation NMS also encouraged competition based on speed and efficiency of electronic interfaces, which further encouraged order shredding. In the last decade, continued improvements in computer technology have widened use of electronic order handling systems, making it practical for traders to shred intended orders for many thousands of shares into tiny pieces of 100 shares or less.

We do not expect that the volume multiplier ζ is constant across intended order size, ticker symbol, and time. In the beginning of our sample, for example, the volume multiplier was probably bigger for orders traded on NASDAQ than orders traded on NYSE. Atkins and Dyl (1997) report that since NASDAQ dealers were either buyers or sellers in almost every trade at the NASDAQ, the NASDAQ trading volume was inflated at least by a factor of two, relative to the number of trades actually occurring between investors. Over time, this pattern may have changed, as dealers' participation rate in facilitation of trading has decreased and trades from other trading systems have begun to be reported on the consolidated tape through the NASDAQ system. Recent technological developments have most likely increased the amount of intermediation in securities markets in the second half of the sample. The number of TAQ prints has soared due to intermediation by high frequency traders, which now accounts for a significant share of volume, as described in Chordia et al. (2011) and Hendershott et al. (2011). For example, Kirilenko et al. (2011) find that high frequency traders account for more than 30% of stock index futures trading volume but hold inventories for only a few minutes. We hypothesize that high frequency trading is more prevalent in stocks with a lower relative tick size, i.e. higher prices and higher volatilities, where these strategies are more profitable.

3 Data.

3.1 Data Description

The NYSE TAQ database contains trades and quotes reported on the consolidated tape by each participant in the Consolidated Tape Association ("CTA") for all stocks listed on exchanges starting from year 1993. Since we examine the distribution of unsigned trades, our analysis employs only data on trades, not quotes. For each trade, the TAQ data records the time, exchange, ticker symbol, number of shares traded, execution price, trade condition, and other parameters. The dataset contains over 19 billion records with the number of data entries steadily increasing over time from over 5 million records per month in 1993 to over 500 million records per month in 2008.

We transform the very large raw TAQ data files into another smaller dataset, convenient for our subsequent analysis. We first remove bad records from the trades data using standard filters. The TAQ database provides information about the quality of recorded trades in their condition and correction codes. We eliminate trades with condition codes of 8, 9, A, C, D, G, L, N, O, R, X, Z or with correction codes greater than 1. The correction code of 8 indicates, for example, that the trade was canceled.

The remaining TAQ prints are aggregated in a specific way to reduce the size of the dataset while preserving information about the distributions of monthly trade sizes. For each ticker symbol and each day, each TAQ print is placed into one of 54 bins constructed based on the number of shares traded. Letting X denote the size of a TAQ print in shares, "even" bins correspond to TAQ prints of the following exact "even" sizes of $X = 100$, $X = 200$, $X = 300$, $X = 400$, $X = 500$, $X = 1000$, $X = 2000$, $X = 3000$, $X = 4000$, $X = 5000$, $X = 10000$, $X = 15000$, $X = 20000$, $X = 25000$, $X = 30000$, $X = 40000$, $X = 50000$, $X = 60000$, $X = 70000$, $X = 75000$, $X = 80000$, $X = 90000$, $X = 100000$, $X = 200000$, $X = 300000$, $X = 400000$, and $X = 500000$ shares. "Odd" bins correspond to TAQ prints with trade sizes X between adjacent even bins, i.e., $100 < X < 200$, $\ldots$, $400000 < X < 500000$, $500000 < X$. Note that the size of bins grows approximately at a log-rate. We consider TAQ prints with even sizes separately, because they are unusually frequent in the data.

For each day and each symbol, we store the number of TAQ prints in each size bin. To simplify the analysis below, once we assign a given print to a bin, we assume its size (in shares) is equal to a midpoint of that bin. If print size is larger than 500,000 shares, we assign it to the 54th bin and assume its size to be 1,000,000 shares. This aggregation allows us to capture the most important properties of TAQ print size distributions and implement our analysis in an efficient way. The convenience comes, however, at the expense of introducing some additional noise into our analysis, which may affect our results.

For each day and each ticker symbol, we also store 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.

We sum the number of prints within each bin over a month for each stock to calculate the frequency of trade sizes in each bin for each stock and each month. Later we average the frequency distributions of normalized print sizes to construct an empirical distribution of TAQ print sizes (in shares) for each stock and each month in the sample. Aggregation by month is done because many stocks do not have enough transactions per day to build a good empirical approximation for theoretical distributions.

In addition to calculating the average number of TAQ prints per day, we calculate several statistics describing the possibly complicated shape of the distribution of TAQ 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 TAQ prints for a given stock in a given month. Note that trade-weighted percentiles effectively put the same weight onto prints of different sizes. This tends to emphasize small trades. For example, if there are 99 prints of 100-shares lot and one print of 100,000 shares, then the distribution of print sizes is mostly concentrated at a 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.

Since 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 previous example, percentile 1-9 are 100 shares and percentiles 10-99 are 100,000 shares.

We report empirical results for both trade-weighted 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. It is a straightforward exercise (involving a change in probability measure) to show that that if the log of trade-weighted print size is distributed $N(\mu, \sigma^2)$, then the log of volume-weighted print size is distributed $N(\mu + \sigma^2, \sigma^2)$. The only difference between the two distributions is the shift in mean.

The monthly data is matched with CRSP data to acquire share and exchange codes for stocks in the sample. Only common stocks listed on the NYSE (New York Stock Exchange), AMEX (American Stock Exchange) and NASDAQ from year 1993 through year 2008 are included in our study. Stocks that had splits in a given month are eliminated from the sample for that month. For each stock and each month, the data is also augmented by adding average daily volume (in dollars and in shares), average price, and the historical volatility. Our final sample includes 1,107,990 stock-month observations. For each 191 months between February 1993 and December 2008, there are, on average, observations for about 5,800 stocks.

3.2 Descriptive Statistics

Table 1 describes the data. Panel A reports statistics for the subperiod 1993-2000. Panel B reports statistics for the subperiod 2001-2008. We report these statistics separately, because the properties of the data have changed substantially following decimalization in 2001. Statistics are calculated for all securities in aggregate as well as separately for ten groups of stocks sorted by average dollar volume. Instead of dividing the securities into ten deciles with the same number of securities, volume break points are set at the 30^{th}, 50^{th}, 60^{th}, 70^{th}, 75^{th}, 80^{th}, 85^{th}, 90^{th} and 95^{th} percentiles of dollar volume for the universe of stocks listed in NYSE with CRSP share codes of 10 and 11. Group 1 contains stocks in the bottom 30^{th} percentile. Group 10 contains stocks in the top 5^{th} percentile. Group 10 approximately corresponds to the universe of S&P100 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 which are economically the most important. For each month, the thresholds are recalculated and stocks are reshuffled across groups.

Panel A of Table 1 reports statistical properties of securities and TAQ prints in the sample before 2001. For the entire sample of stocks, the average trading volume is \$6.28 million per day, ranging from \$0.14 million for the lowest volume decile to \$181.98 million for the highest volume decile. The average volatility for the entire sample is equal 4.1% per day. The volatility tends to be higher for smaller stocks. The volatility is 4.6% for the lowest volume decile and 3.3% for the highest volume group. Thus, the measure of trading activity, equal to the product of dollar volume and volatility, increases from $0.14 \cdot 0.046$ to $181.98 \cdot 0.033$, i.e., by a factor of 913.

The average print size is equal to \$23,598 before 2001, ranging from \$11,428 for low-volume stocks to \$88,450 for high-volume stocks, corresponding to a decrease from 8% to 0.05% of daily volume from lowest to 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 \$5,706 for low-volume stocks to \$28,440 for high-volume stocks, corresponding to a decrease from 4% to 0.02% of daily volume. Note that the invariance hypothesis predicts that the shape of the distributions of trade sizes as a fraction of daily volume should be similar across stocks, with the only difference that their log-means are shifted downwards by two-thirds of an increase in a log-trading activity. Since from lowest to highest deciles, trading activity increases by a factor of 913, 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 $913^{2/3} \approx 100$. This is more than the observed differences in means of $8\%/0.05\% \approx 16$ and medians of $4\%/0.02\% \approx 20$ between two volume groups.

In the subperiod 1993-2001, the average number of TAQ prints recorded per day is 143 for the entire sample, increasing monotonically from 16 to 2,951 from the first to the tenth volume group. The number of prints increases by a factor of $2951/16 \approx 184$. The invariance hypothesis predicts that the expected number of bets should increase by two-thirds of the increase in trading activity, i.e., $913^{2/3} \approx 100$. While this back-of-the envelope calculation suggests that number of TAQ prints increases more than predicted, potentially reflecting a more intensive order shredding in high-volume groups, further investigation is certainly warranted.

Some print sizes are unusually common in the TAQ data. Before 2001, even-sized trades account for over 60% 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 14% of transactions for low-volume stocks and 25% of transactions for high-volume stocks. There is also a significant number of 1,000-share prints. The large fraction of 1,000-share prints for low-volume stocks relative to high-volume stocks, 18% versus 13%, 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.

Panel B of Table 1 describes statistical properties of the sample after 2001. Daily volume has more than tripled from \$6.28 million before 2001 to over \$19 million after 2001. Volatility has decreased from 4.1% to 3.4% per day. These numbers imply that the trading activity has doubled between the subperiod 1993-2001 and 2001-2008. The average number of TAQ prints has increased by a factor of 12 from 143 to 1761, and the average print size has decreased by a factor of 3 from \$23,598 to \$7,945. Back-of-the-envelope calculations implied by the invariance hypothesis suggests 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 have to be attributed to other factors. One of them is the order shredding that has been becoming 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. During the subperiod 2001-2008, for example, 100-share trades account for about 50% of all transactions and 18% of volume traded, with these numbers reaching 70% and 35 %, respectively, in 2008. The 1,000-share trades have become less important in the latter subperiod. We expect that order shredding will make it difficult to test the invariance hypothesis using TAQ data after 2001.

Frequency and Sizes of TAQ Prints During 1993-2008. For the 16-year period 1993-2008, figure 9 plots the 191 monthly values of the normalized mean of the number of TAQ prints per month, calculated as $\bar{N}_m \cdot W^{-2/3}$, where $\bar{N}_m$ is the number of TAQ prints per month and W is trading activity. The figure also plots the averages of the 20^{th}, 50^{th} and 80^{th} percentiles of the trade-weighted and volume-weighted distributions of logs of normalized TAQ print sizes, calculated as $\ln(|\tilde{X}|/V \cdot W^{2/3})$, where $|\tilde{X}|$ is the TAQ print size. To facilitate comparison across stocks and and across time, all variables are scaled by $W^{-2/3}$ and $W^{2/3}$ as implied by the invariance hypothesis. The left panel shows the normalized variables averaged across low-volume stocks (group 1). The right panel shows the normalized variables averaged across high-volume stocks (groups 9 and 10).

The trading patterns differ significantly during the two subperiods of 1993-2000 and 2001-2008. For high-volume stocks, the percentiles of print sizes and print rates do not change much prior to the beginning of decimalization in 2001; afterwards, percentiles of print size decreases steadily and the average number of prints correspondingly increases. For low-volume stocks, similar changes started to occur even before 2001. Since most of low-volume stocks are NASDAQ stocks, the pre-2001 decrease in print sizes and increase in print arrival rates may be explained by reduction in tick size from 1/8 of a dollar to 1/60 of a dollar announced at NASDAQ in 1997. With an exception of the largest print sizes in high-volume stocks, the downward trend in normalized print size and the upward trend in number of prints seems to stop in 2007. A similar pattern is seen in figures in Hendershot et al. (2011). In the following sections, we will examine these patterns in more detail.

4 Empirical Results

All three models make distinctively different predictions concerning the differences in the distributions of trade sizes and their frequencies across stocks. We run our tests based on both the number of TAQ prints and distributions of their sizes to determine which of the models provides a more reasonable description of the data.

4.1 Tests Based on Trading Frequency

Comparison of Three Models. According to each model, the average number of prints $\bar{N}$ per day is constant across stocks, if normalized by multiplying by a model-dependent power of trading activity W. The three models differ only in the power of W used to normalize the average number of prints. The invariance hypothesis implies that $\ln(\bar{N} \cdot W^{-2/3})$ is constant, the model of invariant bet size implies that $\ln(\bar{N} \cdot W^{0})$ is constant, and the model of invariant bet rate implies that $\ln(\bar{N} \cdot W^{-1})$ is constant.

The three columns of figure 8 contain plots of the log of the average number of TAQ prints per day $\bar{N}$, normalized according to each of the three models, against the log of trading activity W. We present results for April 1993, April 2001, and April 2008, because trading has changed dramatically during the period 1993-2008. For April 1993, we examine separately the NYSE-listed and Nasdaq-listed stocks, because these securities may exhibit different patterns due to a significant difference in regulatory rules across exchanges during that time. We choose the month of April to avoid seasonality, as trades tend to cluster much less before the end of calendar quarter, as showed by Moulten (2005). Each observation corresponds to the average number of TAQ prints per day for a given stock in a given month. There are about 5,800 observations on each subplot. 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, showing that the average number of TAQ prints has increased over time. For April 1993, the average number of TAQ prints is slightly higher for the NYSE stocks than the NASDAQ ones. Some of NASDAQ stocks with low trading activity seem to have too small a number of TAQ prints. A few of the most inactive NASDAQ stocks have too high a number of prints in April 1993.

In subplots for the model of invariant bet frequency, observations are lined up across a line with a positive slope. The model assumes that differences in trading activity come entirely from differences in trade sizes. Since in actual data changes in trading activity are partially explained by changes in trading rates, the model tends to underestimate the number of trades for high-volume stocks and overestimate it for low-volume stocks.

In subplots for the model of invariant bet 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. Since some part of these differences is explained in actual data by differences in trade sizes, the model tends to overestimate the number of trades for high-volume stocks and underestimate it for low-volume stocks.

OLS Estimates of Number of TAQ prints. The three models make distinctly different predictions concerning how arrival rates of TAQ prints vary with the level of trading activity. The predictions of the models can be nested into a simple linear regression,

$$\ln\left[\bar{N}_i\right] = \mu_\gamma + a_\gamma \cdot \ln\left[\frac{W_i}{W_*}\right] + \tilde{\epsilon}. \tag{10}$$

The equation relates the log of the mean number of TAQ prints $\bar{N}_i$ per month for stock i to the level of trading activity W_i. The scaling constant $W_* = (40)(10^6)(0.02)$ measures trading activity for an arbitrary benchmark stock with price \$40 per share, trading volume of one million shares per day, and daily volatility of 2% per day. The model of trading game invariance predicts $a_\gamma = 2/3$, the model of invariant bet frequency predicts $a_\gamma = 0$, and the model of invariant bet size predicts $a_\gamma = 1$. For each month, we estimate of the parameters a and a_γ using an OLS regression in which there is one observation for each stock with TAQ data for that month.

Panel A of figure 9 shows a time series of 191 month-by-month regression coefficients a_γ from regression equations (10) over the period 1993-2008. We superimpose a horizontal line representing $a_\gamma = 2/3$, the value predicted by the invariance hypothesis. The figure shows that there are two distinctive periods, 1993-2000 and 2001-2008. Over the subperiod 1993-2000, all estimated coefficients a_γ remained virtually constant. The average value of 0.690 is only slightly higher than the predicted values of 2/3. Over the subperiod 2001-2008, the estimates clearly begin to diverge from the values implied by the invariance hypothesis, steadily increasing from about 0.690 to about 0.850 by the end of 2008. We suspect that the breakpoint in 2001 most likely results from decimalization and improving electronic interfaces encouraging the use of order shredding strategies and intermediation more strongly for actively traded stocks than inactively traded stocks. The constant term a is scaled to represent the log of expected number of bets for the benchmark stock with trading activity W_*. Both the constant term a and the coefficient a_γ are changing over time and may have linear time trends which are different over the two subperiods 1993-2000 and 2001-2008.

Table 2 presents the results of the regression (10) pooled over time. The six columns show the results for all stocks, the subsets of NYSE/AMEX-listed stocks, and the subset of NASDAQ-listed stock during the two subperiods 1993-2000 and 2001-2008. The table reports Fama-MacBeth estimates of the coefficients. Newey-West standard errors are calculated with three lags relative to a linear time trend estimated by OLS regressions from the estimated coefficients $\hat{\mu}_{\gamma,T}$ and $\hat{a}_{\gamma,T}$ for each month: $\hat{\mu}_{\gamma,T} = \mu_{\gamma,0} + \mu_{\gamma,t} \cdot (T - \bar{T})/12 + \tilde{\epsilon_T}$ and $\hat{a}_{\gamma,T} = a_{\gamma,0} + a_{\gamma,t} \cdot (T - \bar{T})/12 + \tilde{\epsilon_T}$, 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, $T = 95$ for December 2000, and $\bar{T} = 48$. For the subperiod January 2001 to December 2008, $T = 1$ for January 1993, $T = 96$ for December 2008, and $\bar{T} = 48.5$.

The point estimate of $a_{\gamma,0}$ is equal to 0.690 for the subperiod 1993-2001 and 0.787 for the subperiod 2001-2008. The standard errors of these estimates are 0.001 and 0.003, respectively. Clearly, the first estimate is much closer to the predicted value of 2/3 than the second estimate. This is consistent with the interpretation that after 2001, order shredding in high-volume stocks has increased more than in low-volume stocks, most likely because of intensive order shredding in high-volume stocks after 2001. Note also that the alternative models predicting $a_{\gamma,0} = 0$ and $a_{\gamma,0} = 1$ are clearly rejected. For the period 1993-2000, the estimated time trend $a_{\gamma,t}$ of -0.002 per year is statistically insignificant. For the time period 2001-2008, the estimated time trend $a_{\gamma,t}$ of 0.017 per year is statistically significant, consistent with the steady increase in a_γ by 0.16 from about 0.690 to about 0.850 over the eight-year period.

Across the two periods 1993-2000, and 2001-2008, the three estimates of 0.639 and 0.758

for NYSE stocks are closer to the benchmark prediction of 2/3 and than the greater estimates of 0.705 and 0.827 for NASDAQ stocks. This pattern is consistent with figure 8, which reveals that there are fewer trades than predicted in some inactive NASDAQ stocks, even after normalizing for trading activity.

The point estimate of $\mu_{\gamma,0}$ is equal to 6.154 for the subperiod 1993-2000 and 8.043 for the subperiod of 2001-2008. This indicates that the average number of TAQ prints has increased over time. The constant terms $\mu_{\gamma,t}$ of 0.078 and 0.242 also show a statistically significant downward time trend for both periods, especially for the second one. The time trends for the coefficient a_γ and the the constant term μ_γ are further examined in section 4.3 below.

4.2 Tests Based on TAQ Print Sizes

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

To approximate the distribution, we calculate print size $|X|$ based on the mid-point of its print-size bin. For each month and for each volume group, the empirical stock-level distributions of normalized print sizes are combined by averaging across stocks in each volume group the frequency distributions of the number of trades in each bin. The results are plotted on figure 1 and figure 2. For illustrative purposes, we present results only for April 1993, but we will closely examine the data for the entire period 1993-2008 later in the paper.

Figure 1 shows empirical distributions of logs of normalized 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 10, respectively. The first three columns contain plots of the trade-weighted distributions with the density of logs of normalized 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 normalized according to the three models. If one of the three models is correct, then the corresponding distributions should be stable across the three 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-weighted and volume-weighted distributions of normalized print sizes based on the entire sample. As discussed above, if trade-weighted distributions are log-normally distributed, then volume-weighted distributions are log-normally distributed as well. If normalized TAQ print sizes are distributed as a log-normal, then the three plots in each column of plots are expected to coincide with the common superimposed density.

In the first column, the three trade-weighted distributions implied by the invariance hypothesis can be seen by visual inspection to have similar means, variance, and supports. The shapes of empirical distributions bear some resemblance to the superimposed normal density, but the fit is by no means exact. The low-volume group 1 matches the superimposed normal better than the medium- and high-volume groups. In the fourth column, the volume-weighted distributions are more similar to the superimposed common normal density. This suggests that the invariance hypothesis explains a big part of variation in the distribution of TAQ prints sizes and especially in the distribution of economically important large trades. Furthermore, the print sizes are distributed similarly to a log-normal.

For the model of invariant bet frequency, the trade-weighted densities in the second column and the volume-based densities in the fifth column are much less stable across volume groups. In both columns, the distributions shift to the left as trading volume increases. This suggests that the first alternative model understates sizes of TAQ prints for high-volume stocks and overestimates them for low-volume stocks. The models 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 bet size, the trade-weighted densities in the third column and the volume-based densities in the sixth column are unstable across volume groups as well. In both columns, the distributions shift to the right as trading volume increases. The second alternative model overstates sizes of TAQ prints for high-volume stocks and underestimates them for low-volume stocks.

The alternative models clearly provide worse explanations for the observed variations in TAQ print sizes than the invariance hypothesis.

Figure 2 shows results for the NASDAQ stocks. Similarly to the NYSE stocks, the distributions of TAQ print sizes are more stable across volume groups, when print sizes are adjusted according to the invariance hypothesis rather than alternative models. The NASDAQ distributions are less smooth and have more spikes than the NYSE distributions, especially the trade-based ones. 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 of Log-normal Distributions. As discussed above, if the log of trade-weighted normalized print sizes is distributed $N(\mu, \sigma^2)$, then the log of volume-weighted normalized print size is distributed $N(\mu + \sigma^2, \sigma^2)$. It is interesting to examine how closely the means and variances of the distributions superimposed in figure 1 and figure 2 for the invariance hypothesis come to satisfying this constraint.

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 -1.01 and its variance 1.78. Since $-1.01+1.78 = 0.77 \neq 0.97$, we see that the constraint fails to hold by a margin of only about 20%. The log-variance of 2.69 for the volume-based distribution is much larger than the log-variance of 1.78 for the trade-weighted distribution. This is inconsistent with log-normality, which implies that these log-variances should be be the same. For NASDAQ stocks, the volume-weighted mean of 1.19 should be the same as the sum of the trade-weighted mean of -0.18 and its variance of 1.85. Since $-0.18 + 1.85 = 1.67 \neq 1.19$, we see that this constraint fails to hold by a margin of 0.48. The log-variance of 1.87 for the volume-based distribution is similar to the log-variance of 1.85 for the trade-weighted distribution, consistent with the predictions of log-normality. Since 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 trades 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.

OLS Regression Estimates of TAQ Print Sizes, February 1993 - December 2008. We test implications of the invariance hypothesis for TAQ print sizes using OLS regressions in which the left-side variable is either a mean or quantile of the trade-weighted or volume-weighted distribution of logs of TAQ print sizes. For each stock in a given month, we construct the empirical trade-weighted and volume-weighted distributions of logs of TAQ print sizes. Letting $f(.)$ denote an functional which calculates either the mean or the *pth* percentiles (20th, 50th, 80th) of these distributions, we regress these variables on logs of trading activity,

$$f\Big(\ln\Big[\frac{\tilde{X}_i}{V_i}\Big]\Big) = \mu_Q + a_Q \cdot \ln\Big[\frac{W_i}{W_*}\Big] + \tilde{\epsilon}_i, \tag{11}$$

Since expected trading volume V is the product of the expected number of prints $\bar{N}$ and expected print size $E\{\tilde{X}\}$, the left-hand-side variable in equation (11) is similar to the result of reversing the sign on the left-hand-side variable in the regression equation (10). Specifically, $V = \bar{N} \cdot E\{\tilde{X}\}$ implies that the left-side of equation (10) is $\log \bar{N} = -\log(E\{\tilde{X}\}/V)$. This is different in absolute value from the left-hand-side of equation (11) because the concavity of the log function implies by Jensen's inequality that the log of the expectation is less than the expectation of the log: $\log(E\{\tilde{X}/V\}) < E\{\log(\tilde{X}/V)\}$. If $\tilde{X}_i/V_i$ were distributed log-normally with the same variance across stocks, then the coefficient estimates for a_γ and a_Q would be the same in absolute value but opposite in sign in all of the regressions in equations (11) and (10), but the constant terms would be different. In fact, we show below that $\tilde{X}_i/V_i$ deviates from a log-normal sufficiently to make the coefficients a_γ and a_Q vary across the different regression equations using means and quantiles as the left-side variable.

We run the regressions for each month between February 1993 and December 2008. Figure 9 shows the time series of 191 month-by-month regression coefficients a_Q from regression equations (11) for the $20th$, $50th$ and $80th$ percentiles of TAQ print sizes over the period 1993-2008. Panel B presents results for trade-weighted distributions. Panel C presents results for volume-weighted distributions. We superimpose horizontal lines representing the levels of -2/3, as benchmarks predicted by the invariance hypothesis.

The figure shows that there are two distinctive subperiods. Over the subperiod 1993-2001, all estimated coefficients remained virtually constant. The estimates of a_Q were slightly lower than the predicted value of -2/3 for the $20th$, $50th$ and $80th$ percentiles of trade-weighted distributions, fluctuating between -0.800 and -0.700 and implying that small print sizes as a fraction of volume decreased faster with trading activity than predicted. For the $50th$ and $80th$ percentiles of volume-weighted distributions, the estimates of a_Q fluctuated between -0.700 and -0.480, somewhat higher than predicted by the invariance hypothesis, while all estimates for the $20th$ volume-weighted percentiles were very close to -2/3. Over the subperiod 2001-2008, the estimates began to diverge from initial levels. For the volume-weighted percentiles, the estimates of a_Q decreased from about -0.500 to -1.000 for the $80th$ percentile, from about -0.650 to -1.000 for the $50th$ percentile, and from -0.670 to -0.870 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 are gradually decreasing from about -0.670 to -0.850. Most changes occur in the distributions of large TAQ print sizes (right tails) rather than small print sizes (left tails).

Table 3 reports the estimates from regressions (11) pooled over the period 1993-2000. 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. Since the monthly estimates of $\hat{a}_T$ and $\hat{a}_{Q,T}$ for each month T are changing over time, we choose to 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 $\hat{\mu}_{Q,T}$ and $\hat{a}_{Q,T}$ for each month: $\hat{\mu}_{Q,T} = \mu_{Q,0} + \mu_{Q,t} \cdot (T - \bar{T})/12 + \tilde{\epsilon_T}$ and $\hat{a}_{Q,T} = a_{Q,0} + a_{Q,t} \cdot (T - \bar{T})/12 + \tilde{\epsilon_T}$, 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 $a_{Q,0}$ is equal to -0.759 for the means. This estimate is larger in absolute value by 0.069 than the estimate of 0.690 for the number of TAQ prints in table 2. For the trade-weighted percentiles, the estimated coefficients range from -0.797 for the 20^{th} percentile to -0.742 for the 80^{th} percentile. All of these estimates are slightly bigger in absolute value than the value of -2/3 predicted by the invariance hypothesis. This implies that TAQ print sizes as a faction of volume tend to decrease with trading activity faster than the invariance hypothesis implies. For the volume-weighted distributions, the estimated coefficient a_Q is equal to -0.591 for the means; the estimates range from -0.688 for the 20^{th} percentile to -0.514 for the 80^{th} percentile. During the subperiod 1993-2000, the volume-weighted print size as a fraction of volume does not decrease with trading activity as fast as trade-weighted print sizes. Across means and percentiles, the standard errors of estimates $a_{Q,0}$ range from 0.001 to 0.003; these values are similar in magnitude to the averages of standard errors of a_Q from the cross-sectional monthly regressions (11), whose values range from 0.003 to 0.005. The data suggests that the invariance hypothesis $a_{Q,0} = -2/3$ explains the data much better than the alternatives $a_{Q,0} = 0$ and $a_{Q,0} = -1$.

The estimated time trend $a_{Q,t}$ ranging from 0.006 to 0.012 for the trade-weighted distributions per year is statistically significant. The time trend is either statistically insignificant or negative for the volume-weighted distributions. This implies that a_Q has increased for small prints and increased for large prints over that subperiod.

The estimated intercept $\mu_{Q,0}$ of -7.219 in the regression based on the trade-weighted means implies that the median print size for the benchmark stock is equal to $exp(-7.219)$, or 0.07 percent of daily volume. The estimated intercepts of -8.470 and -6.240 in the regressions for the 20^{th} and 80^{th} percentiles suggest that the average 20^{th} and 80^{th} print-size percentiles are equal to 0.02 percent and 0.20 percent of daily volume for the benchmark stock, respectively. Under the assumption of log-normality, Kyle and Obizhaeva (2011a) 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 $1 - N(z - \sigma)$, where σ is the standard deviation for the distribution of the log of trade sizes. Based on the trade-weighted variance of 1.78 in figure 1, 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 $\mu_{Q,0}$ in cross-sectional regressions are similar in magnitude and range between 0.013 to 0.030, across means and percentiles. The negative and statistically significant estimates of time trend $\mu_{Q,t}$ indicate that the TAQ print sizes have been gradually decreasing in size during the subperiod of 1993-2000, with a greater downward drift for the right tails of distributions than the left tails.

The R-square is lower in regressions based on trade-weighted distributions than regressions based on volume-weighted distributions. For the means, the R-squares are 0.93 and 0.86, respectively. The difference in R-square increases monotonically from a difference between 0.90 and 0.91 for the 20^{th} percentiles to a difference between 0.93 and 0.77 for the 80^{th} percentiles. These numbers show that there is more unexplained variation in large print sizes than in small print sizes. 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 (11) for the subperiod 2001-2008. For the means, the estimates of $a_{Q,0}$ are -0.793 for trade-weighted distributions and -0.743 for volume-weighted distributions. These estimates are larger in absolute value than the corresponding estimates of -0.759 and -0.591 for the subperiod 1993-2001 in table 3. All but one estimates of -0.787, -0.793, and -0.805 for the 20^{th}, 50^{th}, and 80^{th} trade-weighted percentiles and -0.799, -0.800, -0.720 for the 20^{th}, 50^{th}, and 80^{th} volume-weighted percentiles are also higher in absolute values than the estimates of -0.797, -0.765, -0.742, -0.688, -0.611, and -0.514 for the earlier subperiod. The biggest changes occur in the estimates for the 80^{th} percentile of trade-weighted distributions and the 50^{th} and 80^{th} percentiles of volume-weighted distributions. This suggests that recent technological and regulatory changes had the largest effect on the the right tail of TAQ print-size distributions. The standard errors of $a_{Q,0}$ are between 0.003 and 0.004, as the averages of standard errors of a_Q in monthly regressions (11). This validates the adjustment for time trend in the Fama-McBeth procedure, because without inclusion of a time trend, the standard errors would range from 0.004 to 0.030.

The estimates of the intercept $\mu_{Q,0}$ for the subperiod 1993-2001 are lower than for the subperiod 2001-2008, for the means and all percentiles. For the pooled sample, for example, the estimate of -8.692 in table 4 is lower than the corresponding estimate of -7.219 in table 3, i.e., the typical print size for the benchmark stock fell from 0.07 to 0.02 percent of daily volume over the period 1993-2001. The estimated time trend $\mu_{Q,t}$ is negative and statistically significant in all columns, except for $20th$ percentile of trade-weighted distributions, also implying that the distributions of TAQ print sizes has been shifting downwards.

4.3 Other Microstructure Effects Over Time

In this section, we focus exclusively on the relationship between the invariance hypothesis and microstructure effects related to the minimum tick size of one cent and the minimum round lot size of 100 shares. Our goal is to better understand how various market frictions and regulatory rules affect the observed number and size distribution of TAQ prints.

Effective Price Volatility, Effective Relative Tick Size, and Effective Relative Lot Size. For the purpose of comparing price volatility with minimum tick size and the round lot size of 100 shares, it is better to begin the analysis with daily dollar price volatility $P \cdot \sigma$ than daily percentage returns volatility σ. The invariance hypothesis suggests that it is better to measure volatility defined over a period proportional to the average arrival time between intended orders rather than over a period of one calendar day. Since $(W/W_*)^{-2/3}$ is proportional to the average arrival time between intended orders, we use its square root to scale daily dollar volatility. This results in a definition of effective price volatility given by

$$\text{Effective Price Volatility} := P \cdot \sigma \cdot \left(\frac{W}{W_*}\right)^{-1/3}, \tag{12}$$

In comparison with calendar day volatility, this definition makes effective price volatility lower for active stocks and higher for inactive stocks. We expect this variable to be related to several microstructure effects associated with minimum tick size and minimum lot size.

Traders often measure daily price volatility in units of minimum tick size. For example, if a \$40 stock has a volatility of 2% per day, the minimum tick size is 1/80 of the daily dollar price volatility of \$0.80 when the minimum tick size is one cent. Rather than defining effective relative tick size as a fraction of daily dollar price volatility, the invariance hypothesis makes it natural to measure effective relative tick size as a fraction of effective price volatility. Assuming a minimum tick size of one cent, the definition is

$$\text{Effective Relative Tick Size} := \frac{\$0.01}{\frac{P \cdot \sigma}{P_* \cdot \sigma_*} \cdot \left(\frac{W}{W_*}\right)^{-1/3}}. \tag{13}$$

The presence of $P_* \cdot \sigma_*$ in the denominator scales the definition of relative tick size so that relative tick size is one cent for the benchmark stock. For traders who trade different securities over different horizons in trading games operating at different speeds, defining effective relative tick size using effective volatility better reflects the manner in which minimum tick size restricts trading than a definition based on calendar day price volatility. Higher effective price volatility makes the effective relative tick size lower. We conjecture that lower effective relative tick size encourages traders to shred intended orders into a larger number of smaller pieces for execution and may also induce a larger number of intermediation trades.

Median bet size is proportional to $V \cdot W^{-2/3}$. We define effective relative lot size as the ratio of the 100-share round-lot size to median bet size:

$$\text{Effective Relative Lot Size} := \frac{100}{\frac{V}{V_*} \cdot \left(\frac{W}{W_*}\right)^{-2/3}}. \tag{14}$$

The presence of V_* in the denominator scales the definition of effective relative lot size so that effective relative lot size is 100 shares for the benchmark stock. Since $W = V \cdot P \cdot \sigma$, the product of effective relative tick size and effective relative lot size is a constant equal to exactly one dollar, the dollar value of one tick on a trade of minimum round-lot size.

Effective relative lot size is proportional to effective price volatility and inversely proportional to effective relative tick size. Higher effective volatility therefore makes the 100-share round-lot boundary more binding, in the sense that more intended orders begin to fall below that boundary. Some of these small intended orders may be executed as odd lots which are not recorded on the consolidated tape, some may not be executed at all, and others may be rounded to precisely 100 shares. We expect the amount of such censoring and rounding to be more pronounced for stocks with large effective relative lot size.

To summarize, increases in effective price volatility are expected to have two opposite kinds of effects on the number of TAQ prints and the distribution of their sizes. The effect operating through effective relative tick size encourages more order shredding and perhaps more intermediation trades. The effect operating through effective relative lot size encourages more censoring of odd lots and more rounding of odd lot trades to precisely 100 shares.

Trade-Weighted Distributions for NYSE-listed Stocks, April 1993. After sorting stocks into ten volume groups and four effective price volatility groups, we analyze distributions of the logs of normalized print sizes $\ln(\frac{X}{V} \cdot W^{2/3})$ with the scaling factor $W^{2/3}$ implied by the invariance hypothesis. The invariance hypothesis predicts these 40 distributions to be the same. To examine how these distributions change over the period 1993-2008, we examine the shape of the distributions for three months: April 1993, April 2001, and April 2008.

Figure 3 shows the trade-weighted distributions of logs of normalized print sizes for 5 of the 10 volume groups and all 4 effective prince volatility groups for the NYSE-listed stocks in April 1993. We highlight 100-share trades in light grey and 1,000-share trades in dark grey. We also report the number of stocks in each subgroup and the average number of trades per day for these stocks. On each subplot, we superimpose a density of a normal distribution with the mean of -1.01 and variance of 1.78, 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 forty subplots and coincide with a superimposed density of a normal. For most subgroups, distributions are indeed close to a superimposed normal.

There is, however, a clear truncation below the 100-share odd-lot boundary with clustering of 100-share trades, shown in light grey, 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 empirical distributions have spikes due to clustering of trades at round-lot levels. There are visible spikes at the 100-share level and the 1,000-share level, marked by clustering of light grey and dark grey columns, as well as two spikes in-between, corresponding to the 200-share and 500-share levels. For the subsample of stocks with low volume and low price volatility, large variation in trading activity smoothes 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 the 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 shredding. The first effect seems to dominate, because the average number of TAQ prints decreases with price volatility. For high-volume stocks, for example, the average number of prints decreases monotonically from 938 prints recorded per day for low-volatility stocks to only 307 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 4 shows the volume-weighted distributions of logs of normalized print sizes for the NYSE stocks in April 1993. In comparison with the trade-weighted distributions in figure 3, the volume-weighted distributions put more weight onto larger trades and allow us to see more clearly the distribution of large print sizes.

Compared with the trade-weighted distributions in figure 3, the volume-weighted distributions are more stable across subgroups and more closely resemble the superimposed normal distribution. On most plots, the space below a 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 represent intended orders shredded into smaller trades. The strong visual resemblance of the graphs to a log-normal, as in Kyle and Obizhaeve (2011b), are consistent with the interpretation that most of the largest orders in 1993 appear to have been executed in the upstairs market, 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. There are also a few spikes in far-right tails of several distributions, suggesting that a few very large prints occur in the data more often than explained by log-normality.

Trade-Weighted Distributions for NASDAQ-listed Stocks, April 1993. Figure 5 shows the trade-weighted distributions of logs of normalized print sizes for the 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 grey spikes, typically in the middle of the NASDAQ distributions. These spikes can be explained by the 1988 Securities and Exchange Commission requirement for 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 additional 104 stocks in November 1997, and finally for all others. 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 2008. Figure 6 and figure 7 show the trade-weighted distributions of logs of normalized print sizes for all stocks traded in April 2001 and April 2008, respectively. In 2001, decimalization and more widespread use of electronic interfaces led to a significant increase in order shredding, the effect of which is clearly seen in both figures.

The frequency of trades has increased significantly over time. For high-volume and low-volatility stocks, for example, there were, on average, only 938 trades per day in April 1993, increasing to 11,154 trades per day in April 2001 and 74,420 trades per day in April 2008. The distributions of normalized print sizes has shifted substantially to the left during the period 1993-2008. Based on the means of superimposed normals, for example, we conclude that median print size drops from 0.36% of daily volume for the NYSE stocks in April 1993 and 0.84% for the NASDAQ stocks to 0.27% in 2001, and only 0.07% in 2008. The block-order market seems almost to have disappeared, and the trading is now dominated by transactions of 100 shares. Trades of 100 shares constitute about 70% of trades executed and 35% of volume traded in 2008.

Regressions With Effective Price Volatility. Table 5 presents the Fama-MacBeth estimates μ_γ and a_σ from monthly regressions

$$\ln\left[\bar{N}_i\right] = \mu_\gamma + \frac{2}{3}\cdot\ln\left[\frac{W_i}{W_*}\right] + a_\sigma\cdot\ln\left[\frac{P_i\cdot\sigma_i}{P_*\cdot\sigma_*}\cdot\left(\frac{W_i}{W_*}\right)^{-1/3}\right] + \tilde{\epsilon}_i. \tag{15}$$

The regression effectively imposes the invariance restriction of $a_Q = 2/3$ in regression (10) and adds effective price volatility as an additional explanatory variable. 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 $\hat{\mu}_{\gamma,T}$ and $\hat{a}_{\sigma,T}$ for each month: $\hat{\mu}_{\gamma,T} = \mu_{\gamma,0} + \mu_{\gamma,t}\cdot(T-\bar{T})/12 + \tilde{\epsilon_T}$ 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 all sample, the subsets of NYSE/AMEX-listed stocks and the NASDAQ-listed stock during the two subperiods 1993-2000 and 2001-2008.

The point estimates for $a_{\sigma,0}$ are negative and statistically significant for all subsamples. The estimates of -0.447, -0.315, and -0.472 for the subperiod 1993-2000 are smaller in absolute terms than the corresponding estimates of -0.546, -0.378, and -0.647 for the subperiod 2001-2008. The standard errors range from 0.003 to 0.008. The point estimates of $a_{\sigma,t}$ are equal to -0.009, -0.038, -0.008, -0.036, -0.026, and -0.039, with standard errors between 0.001 and 0.004. There is an inverse relation between effective price volatility and the number of TAQ prints. The higher effective volatility implies fewer TAQ prints in the context of the invariance hypothesis. The estimates of $\mu_{\gamma,0}$ and $\mu_{\gamma,t}$ are not too different from the corresponding estimates in table 2.

The comparison of R-squares in regressions (15) with R-squares in regression (10) constrained with $a_\gamma = 2/3$ shows what variations in the number of TAQ prints, unexplained by the invariance hypothesis, can be attributed to differences in effective price volatility. Addition of effective price volatility as an explanatory variable significantly improves the R-square. For the entire sample, adding effective price volatility as an explanatory variable increases the R-square from 0.906 to 0.938 and from 0.915 to 0.956 for the subperiods 1993-2000 and 2001-2008, respectively. For NYSE stocks, the R-square increases from 0.924 and 0.937 to 0.936 and 0.955; for NASDAQ stocks, the R-square increases from 0.901 and 0.899 to 0.941 and 0.963 during the two subperiods.

Finally, we analyze the R-square in the regression which imposes the invariance restriction $a_Q = 2/3$ in regression (10), 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,

$$\ln\left[\bar{N}_i\right] = \mu_\gamma + \frac{2}{3}\ln\left[\frac{W_i}{W_*}\right] + b_1 \cdot \ln\left[\frac{V_i}{(10^6)}\right] + b_2 \cdot \ln\left[\frac{P_i}{(40)}\right] + b_3 \cdot \ln\left[\frac{\sigma_i}{(0.02)}\right] + \tilde{\epsilon}. \qquad (16)$$

For the entire period 1993-2008, the estimates are $\hat{b}_1 = 0.18$ for the coefficient on volume V_i, $\hat{b}_2 = -0.3$ for the coefficient on price P_i, and $\hat{b}_3 = -0.41$ for the coefficient on volatility σ_i (not reported). All coefficients are statistically different from zero. Similar patters are observed for other subperiods and subsamples of the NYSE stocks and the NASDAQ stocks. The exponents for volatility behave similarly to the exponents for price and differently from the exponents for volume. 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 increased R-square from 0.906 to 0.943 for the subperiod 1993-2000 and from 0.915 to 0.965 for the subperiod 2001-2008. The R-square of 0.943 and 0.965 are only slightly bigger than the R-square of 0.938 and 0.956 in regressions (15), respectively. While statistically significant, the addition of two extra degrees of freedom beyond the effective price volatility improves the R-square by only a small amount.

5 Conclusion.

We find that the distributions of print sizes (adjusted for trading activity) resemble a log-normal, with truncation below the 100-share odd-lot boundary. The resemblance was stronger during the earlier period 1993-2001 than the later period 2001-2008, and it shows up more clearly in volume-weighted distributions than trade-weighted distributions.

The invariance hypothesis explains about 91% of variation across stocks in number of TAQ prints. The other unexplained 9% can be most likely attributed to other microstructure effects such as order shredding, intermediation activity, and various market frictions like minimum lot size and tick size. For example, subtle effects of effective price volatility explain additional 3% and 4% of variations during periods 1993-2000 and 2001-2008, respectively. An interesting topic for the future research is to analyze these effects at a deeper level by designing econometric tests that model these issues more carefully.

6 Literature

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

Angel J., James, 1997, "Tick size, share prices, and stock splits," *Journal of Finance* 52(2), 655–681.

Atkins, Allen B., and Edward A. Dyl, 1997, "Market structure and reported trading volume: Nasdaq versus the NYSE," *Journal of Financial Research* 20, 291-304.

Brennan, Michael, and Avanidhar Subrahmanyam, 1998, "The determinants of average trade size," *Journal of Business* 71(1), 1–25.

Caglio, Cecilia, and Stewart Mayhew, 2007, "Equity trading and the allocation of market data revenue," *Working Paper.*

Chordia, Tarun, Richard Roll, and Avanidhar Subrahmanyam, 2011, "Recent trends in trading activity and market quality," *Journal of Financial Economics*, 101, 243–263.

Glosten, Lawrence, and Lawrence Harris, 1988, "Estimating the components of the bid-ask spread," *Journal of Financial Economics* 21, 123-142.

Harris, Lawrence, 1994, "Minimum price variations, discrete bid-ask spreads, and quotation sizes," *Review of Financial Studies* 7(1), 149–178.

Hendershott, Terrence, Charles Jones, and Albert Menkveld, 2011, "Does Algorithmic Trading Improve Liquidity?" *Journal of Finance* 66 (1), 1–33.

Keim, Donald B., and Anand Madhavan, 1995, "Anatomy of the trading process: Empirical evidence on the behavior of institutional trades," *Journal of Financial Economics* 37(3), 371–398.

Kirilenko, Andrei A., Albert S. Kyle, Mehrdad Samadi, and Tugkan Tuzun, 2011, "The flash crash: The impact of high frequency trading on an electronic market," *Working Paper.*

Kyle, Albert S., and Anna A. Obizhaeva, 2011a, "Market microstructure invariants: Theory and Implications of Calibration," *Working Paper*, University of Maryland.

Kyle, Albert S., and Anna A. Obizhaeva, 2011b, "Market microstructure invariants: Empirical Evidence from Portfolio Transitions," *Working Paper*, University of Maryland.

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

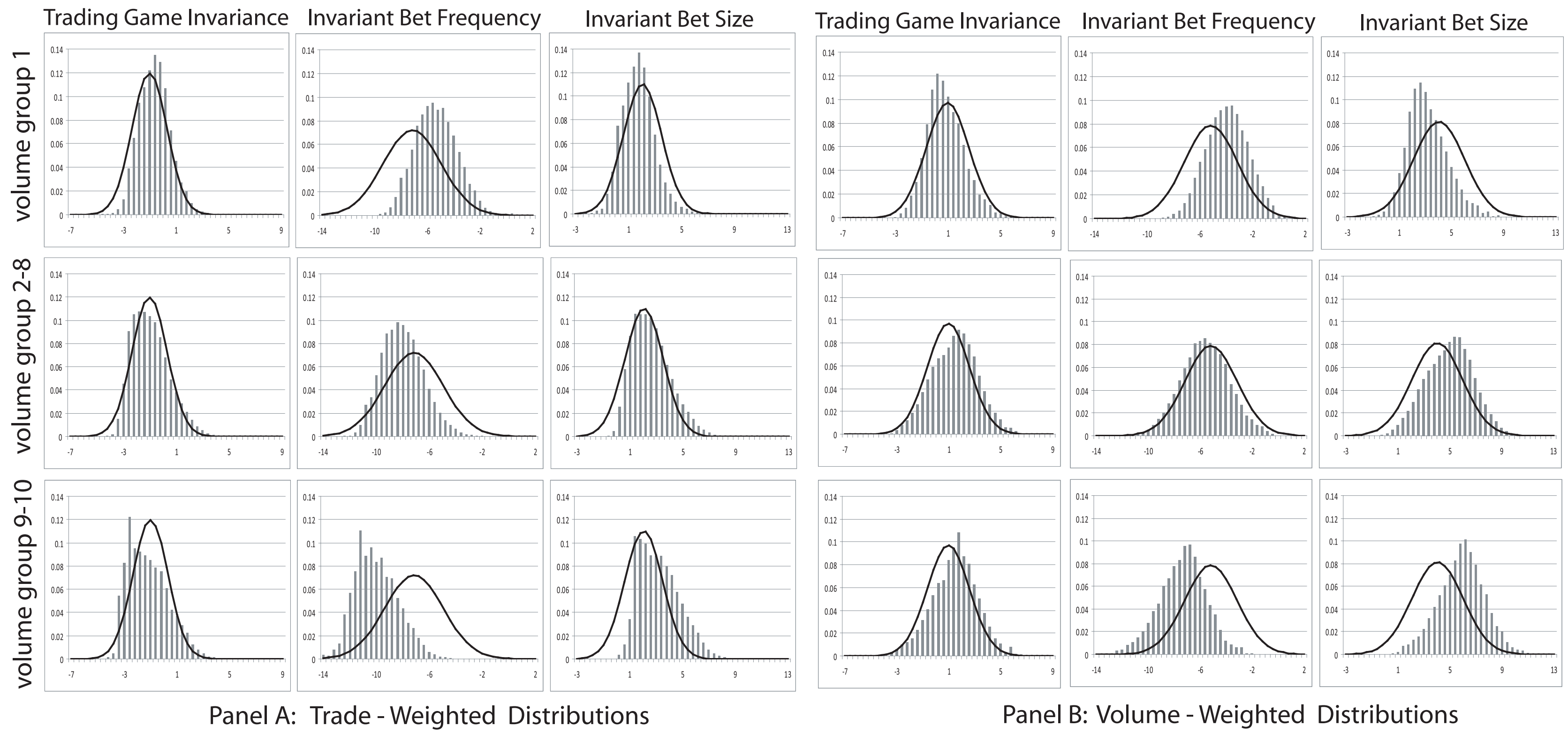
O'Hara, Maureen, Chen Yao, and Mao Ye, 2011, "Whats not there: The odd-lot bias in TAQ data," *Working Paper.*

Table 1: Descriptive Statistics.

Volume Groups:	All	1	2	3	4	5	6	7	8	9	10
					Panel A: 1993-2000						
Avg. Print Size ($)	23,598	11,428	27,348	36,370	43,387	49,167	53,390	60,404	67,649	77,789	88,450
Med. (TW) Print Size ($)	9,386	5,706	10,864	13,445	15,452	16,946	18,115	20,140	21,944	24,598	28,440
Med. (VW) Print Size ($)	105,819	48,176	153,785	178,670	193,535	207,665	230,830	242,800	273,980	308,606	337,984
Avg. # of Prints, γ	143	16	72	124	182	251	327	398	536	836	2951
Avg. Daily Volume($1000)	6,286	143	1,176	2,744	4,947	7,883	11,130	15,657	23,744	42,295	181,987
Avg. Volatility	0.041	0.046	0.034	0.033	0.033	0.033	0.033	0.032	0.032	0.033	0.033
Avg. Price	17.64	10.29	20.46	24.53	27.99	31.67	34.82	38.52	43.10	50.12	64.60
100-Shares: %Prints/%Vol	16/2	14/2	17/2	19/2	19/2	20/2	21/2	21/2	21/2	22/2	25/3
1000-Shares: %Prints/%Vol	18/14	18/15	18/13	17/12	16/12	15/11	15/11	14/10	14/10	13/10	13/11
Even Lots: %Prints/%Vol	80/61	79/63	80/60	80/59	80/58	80/57	79/57	79/56	79/55	79/56	82/58
# Obs	636,271	392,812	93,722	37,288	33,309	14,987	14,234	13,074	12,381	11,893	12,571
					Panel B: 2001-2008						
Avg. Print Size ($)	7,945	4,337	8,323	10,961	13,235	15,666	17,794	19,966	23,218	27,049	34,743
Med. (TW) Print Size ($)	3,170	1,812	3,416	4,447	5,256	6,252	7,027	7,723	8,900	10,025	12,056
Med. (VW) Print Size ($)	34,358	29,124	27,884	30,489	36,134	41,724	48,484	56,585	66,787	82,316	118,826
Avg. # of Prints, γ	1,761	267	1,370	2,143	2,951	3,720	4,611	5,663	7,309	10,318	24,090
Avg. Daily Volume($1000)	19,317	796	6,834	13,563	21,912	31,999	42,816	57,954	83,025	136,204	421,325
Avg. Volatility	0.034	0.038	0.031	0.029	0.028	0.028	0.027	0.026	0.026	0.026	0.027
Avg. Price	19.44	12.26	24.80	28.94	31.52	35.70	38.23	39.87	43.96	46.78	52.23
100-Shares: %Prints/%Vol	50/18	50/18	55/21	52/19	49/17	47/16	46/15	44/15	43/14	42/13	39/11
1000-Shares: %Prints/%Vol	5/7	5/8	3/6	3/5	3/6	3/6	4/6	4/5	4/5	4/5	5/6
Even Lots: %Prints/%Vol	86/61	86/61	90/64	89/63	88/61	87/60	86/59	85/58	85/57	84/56	82/54
# Obs	471,719	306,524	58,794	24,525	22,775	10,687	10,025	9,545	9,328	9,424	10,092

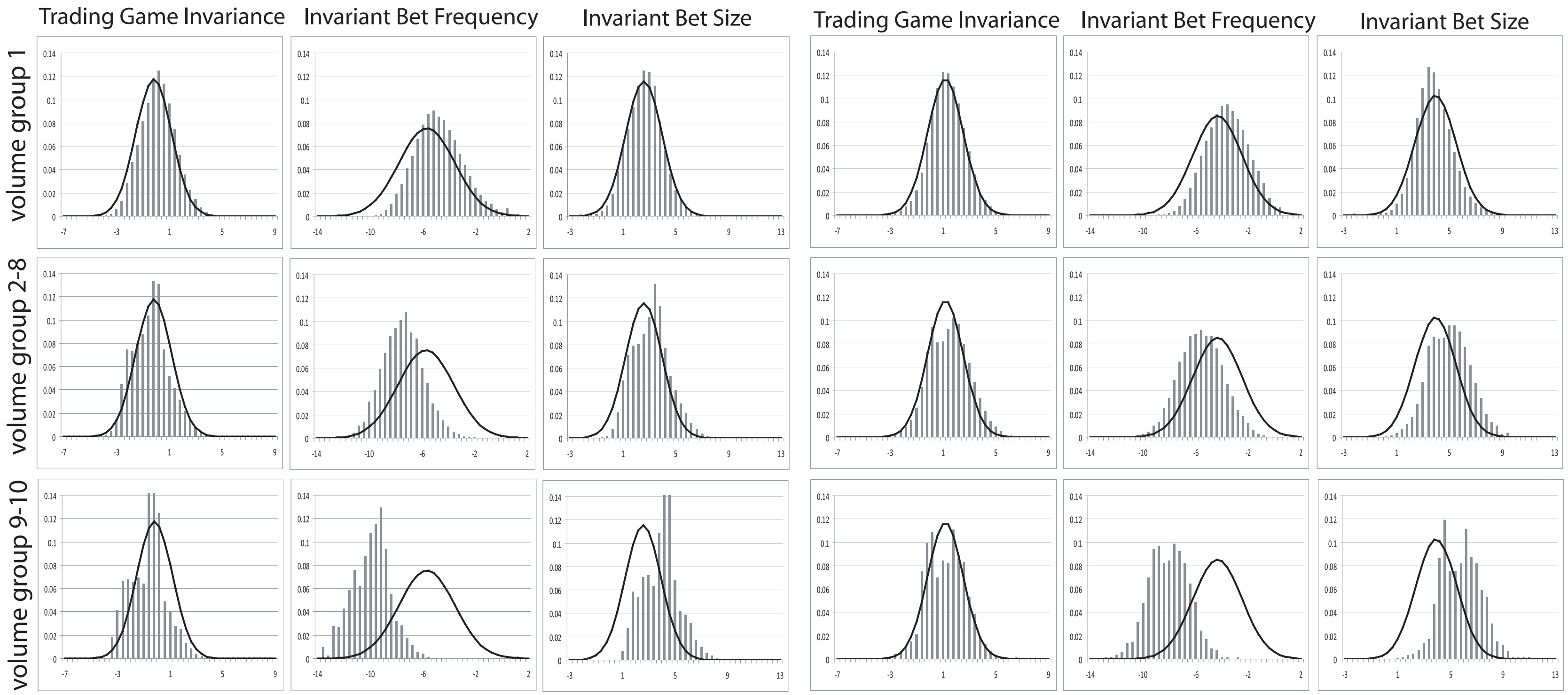
This table reports descriptive statistics for securities and TAQ prints. Each observation represents averages for one security over one month. Panel A reports statistics for data from February 1993 to December 2000. Panel B reports statistics for data from January 2001 to December 2008. Both panels show the average of print size, the trade-weighted median print size, the volume-weighted median print size (in $), the average number of TAQ print per day, the daily dollar volume (in thousands of $), the average volatility of daily returns, the average price, the percent of trades and the percent of volume in the 100-share lot, in the 1000-share lot, and in the even lots for all sample as well as for ten volume groups. Volume groups are based on average dollar trading volume with thresholds corresponding to 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.

Figure 1: Trade-Weighted and Volume-Weighted Distributions of Normalized TAQ Print Size for Three Models, NYSE-listed Stocks, April 1993



This figure shows the distribution of the logarithm of normalized TAQ print sizes for three different models for NYSE-listed stocks traded in April 1993. The print sizes are normalized as $W^{\alpha} \cdot |X|/V$, with $\alpha = 2/3$ for the model of invariant bet frequency, $\alpha = 0$ for the model of invariant bet frequency, and $\alpha = 1$ for the model of invariant bet size. Trading activity W is calculated as the product of dollar volume $P \cdot V$ and the daily percentage standard deviation of returns σ. 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 (high volume), volume groups 2-9, and volume group 10 (low volume). Volume groups are based on average dollar trading volume with thresholds corresponding to 30th, 50th, 60th, 70th, 75th, 80th, 85th, 90th, and 95th percentiles of the dollar volume for NYSE-listed common stocks.

Figure 2: Trade-Weighted and Volume-Weighted Distributions of Normalized TAQ Print Size for Three Models, NASDAQ-listed Stocks, April 1993

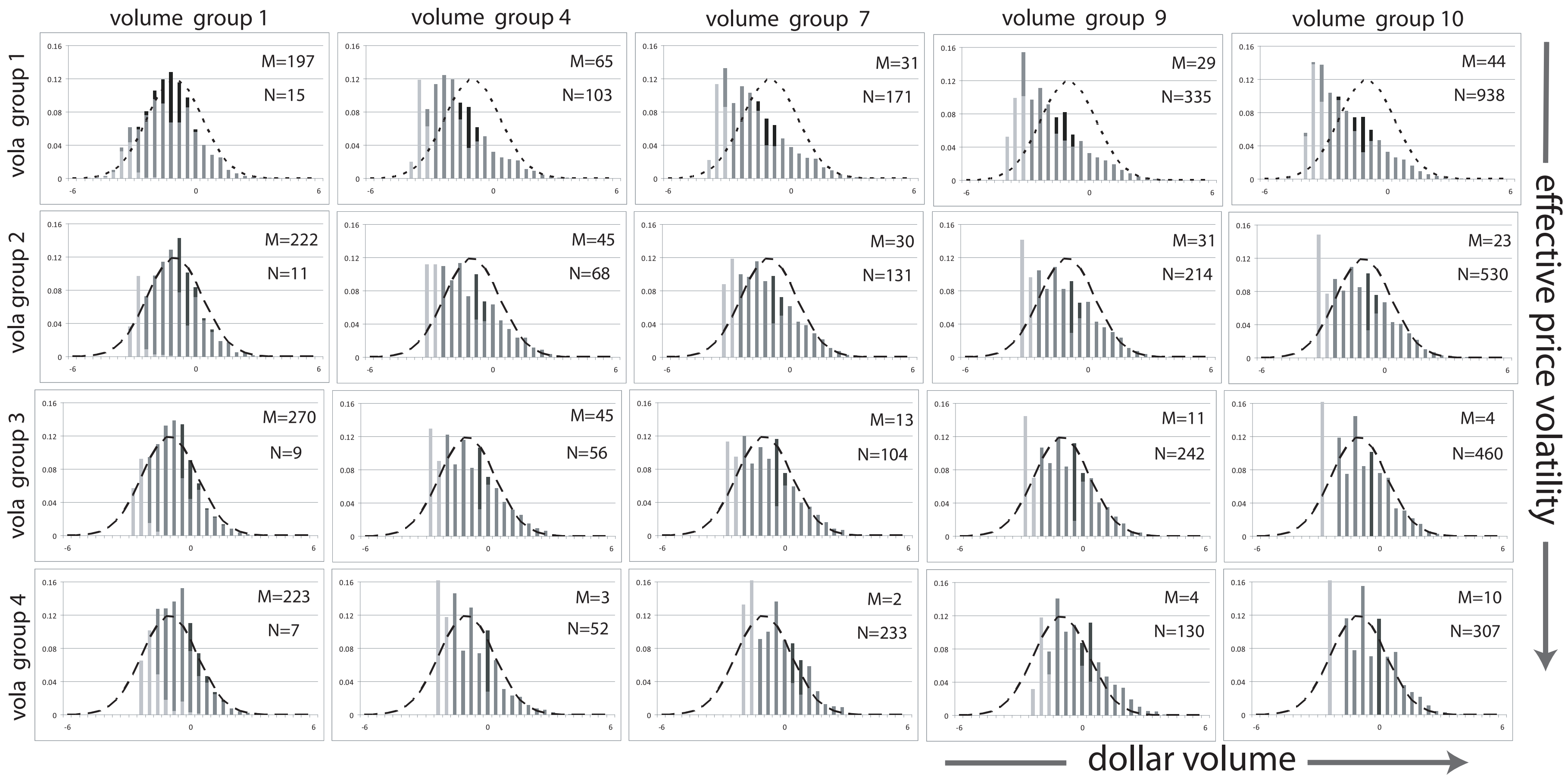


Panel A: Trade - Weighted Distributions

Panel B: Volume - Weighted Distributions

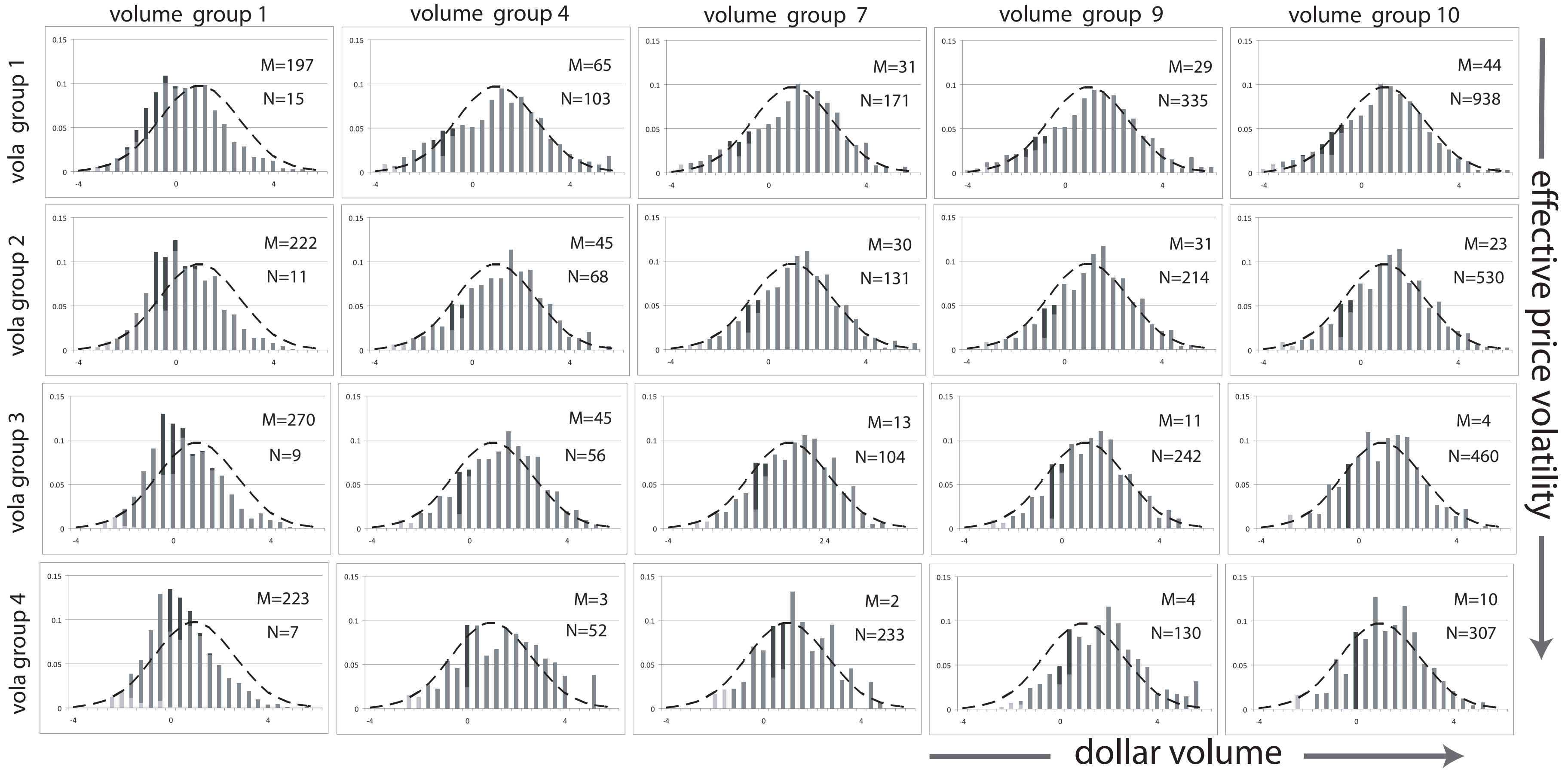
This figure shows the distribution of the logarithm of normalized TAQ print sizes for three different models for NASDAQ-listed stocks traded in April 1993. The print sizes are normalized as $W^{\alpha} \cdot |X|/V$, with $\alpha = 2/3$ for the model of invariant bet frequency, $\alpha = 0$ for the model of invariant bet frequency, and $\alpha = 1$ for the model of invariant bet size. Trading activity W is calculated as the product of dollar volume $P \cdot V$ and the daily percentage standard deviation of returns σ. 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 (high volume), volume groups 2-9, and volume group 10 (low volume). Volume groups are based on average dollar trading volume with thresholds corresponding to 30th, 50th, 60th, 70th, 75th, 80th, 85th, 90th, and 95th percentiles of the dollar volume for NASDAQ-listed common stocks.

Figure 3: Trade-Weighted Distributions of Normalized TAQ Print Sizes, NYSE-listed Stocks, April 1993



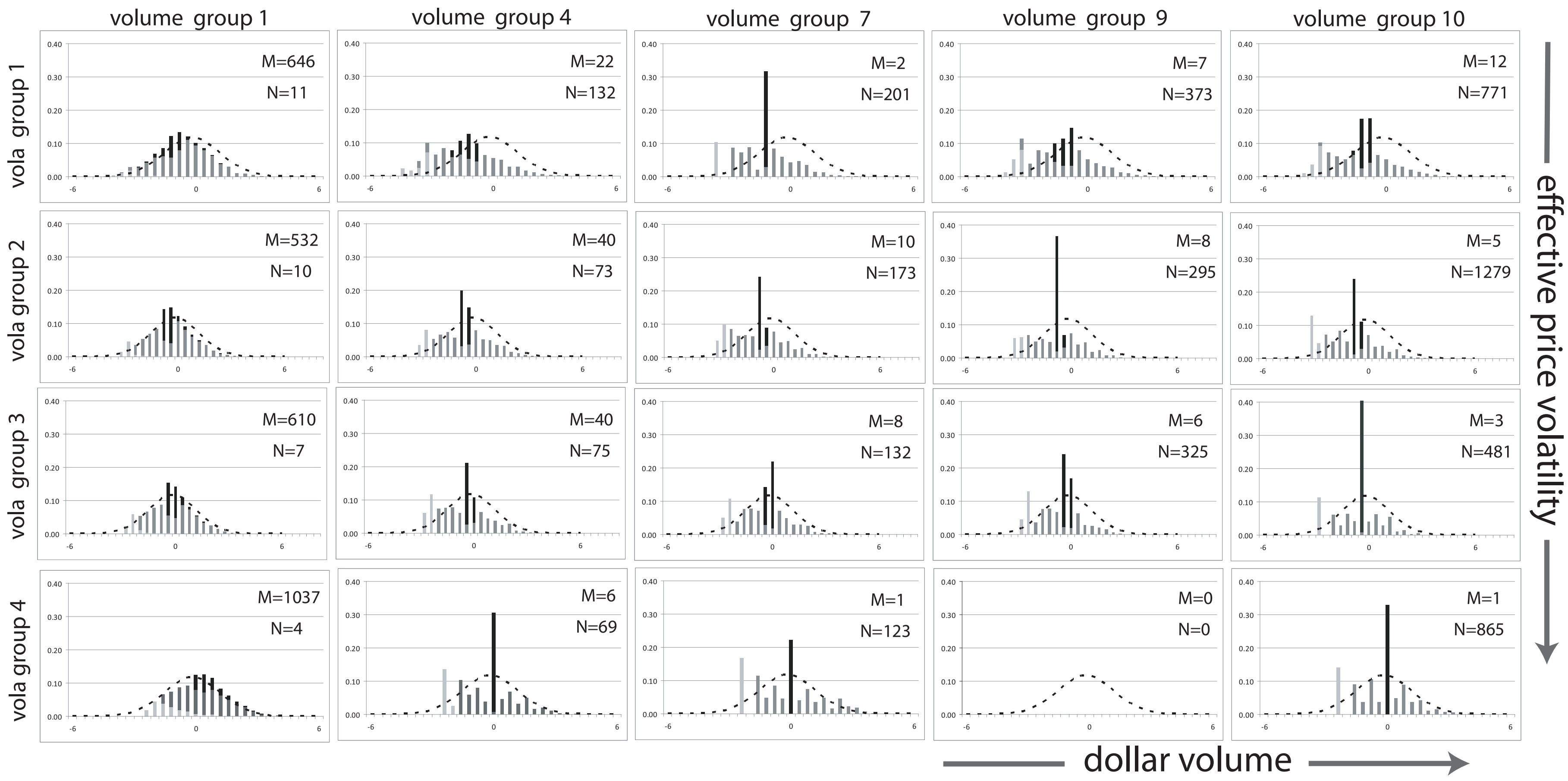
This figure shows distributions of the logarithms of normalized TAQ print sizes for NYSE stocks in April 1993. For each trade, the normalized print size is calculated as $\ln(W^{2/3} \cdot |X|/V)$ based on the invariance hypothesis, where $|X|$ is the midpoint of the print size bin in shares, V is the average daily volume in shares, and W 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 30th, 50th, 60th, 70th, 75th, 80th, 85th, 90th, and 95th percentiles of the dollar volume for common NYSE-listed stocks. The four equally-spaced effective price volatility on the definition $P \cdot \sigma \cdot (W/W_*)^{-1/3}$. The subplots show stock-level distributions averaged across stocks for volume groups 1 (low volume), 4, 7, and 10 (high volume) and for all four price volatility groups 1 (low price volatility), 2, 3, 4 (high price volatility). The 100-share trades are highlighted in light grey; the 1000-share trades are highlighted in dark grey. Each subplot also shows a normal distribution with the pooled average print size mean of -1.01 and pooled averaged standard deviation of 1.33. M is the number of stocks, and N is the average number of TAQ prints per day for the stocks in a given subgroup.

Figure 4: Volume-Weighted Distributions of Normalized TAQ Print Sizes, NYSE-listed Stocks, April 1993



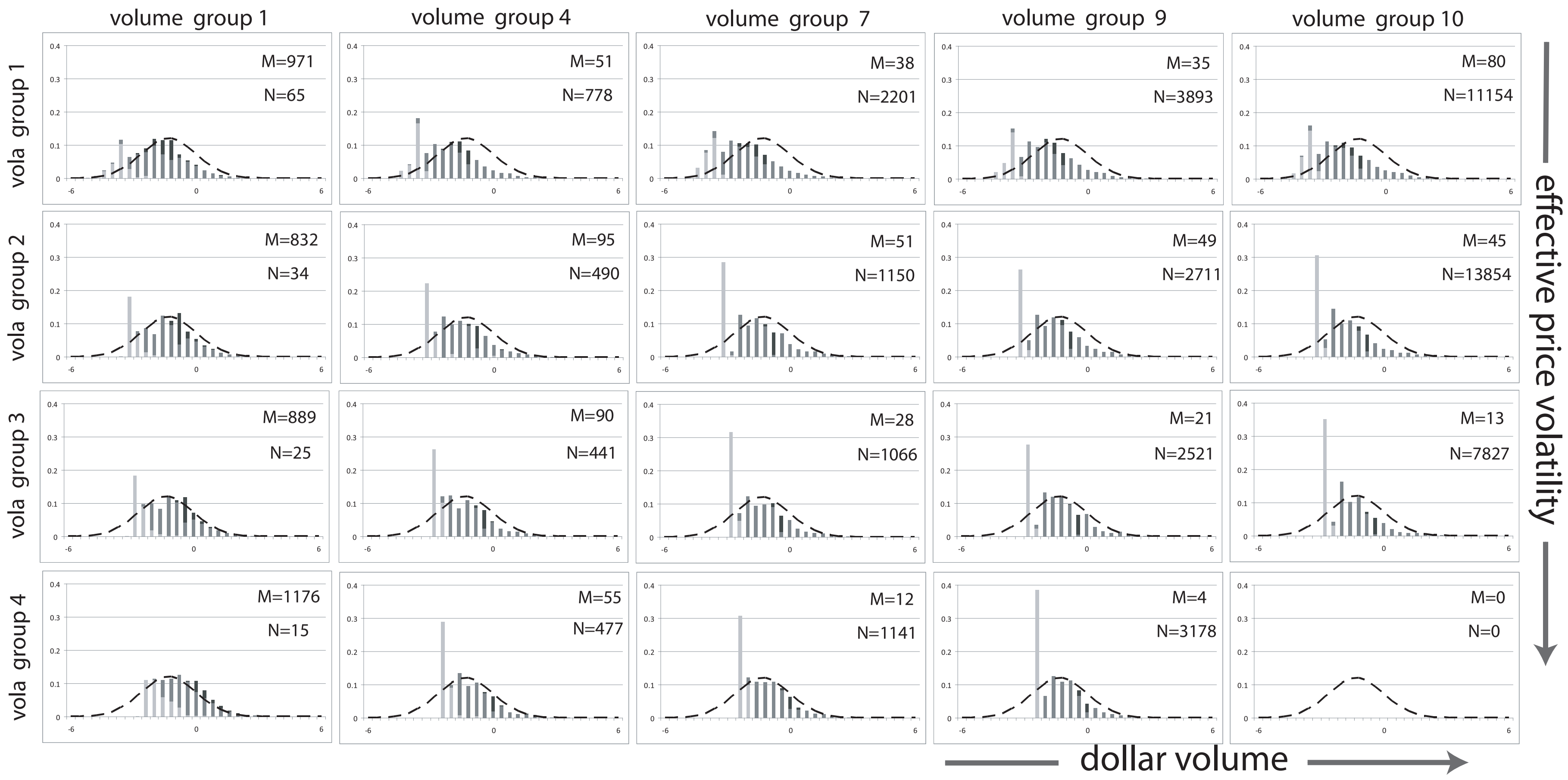
This figure shows distributions of total volume across different normalized TAQ print size bins for the NYSE stocks in April 1993. For each stock, the volume distribution is calculated as the contribution to the total volume by trades from a given trade size bin. The x-axis is the log of normalized print sizes, defined by $\ln(W^{2/3} \cdot \frac{|X|}{V})$ according to the invariance hypothesis, where $|X|$ is a print size in shares (midpoint of a bin), V is the average daily volume in shares, and W 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 30th, 50th, 60th, 70th, 75th, 80th, 85th, 90th, and 95th percentiles of the dollar volume for NYSE-listed common stocks. The four equally-spaced effective price volatility are based on the definition $P \cdot \sigma \cdot (W/W_*)^{-1/3}$. The subplots show stock-level distributions averaged across stocks for volume groups 1 (low volume), 4, 7, and 10 (high volume) and for all four price volatility groups 1 (low price volatility), 2, 3, 4 (high price volatility). The 100-share trades are highlighted in light grey, and the 1000-share trades are highlighted in dark grey. Each subplot also shows a normal distribution with the pooled average print size mean of 0.97 and pooled averaged standard deviation of 1.64. M is the number of stocks, and N is the average number of TAQ prints per day for the stocks in a given subgroup.

Figure 5: Trade-Weighted Distributions of Normalized TAQ Print Sizes, NASDAQ-listed Stocks, April 1993



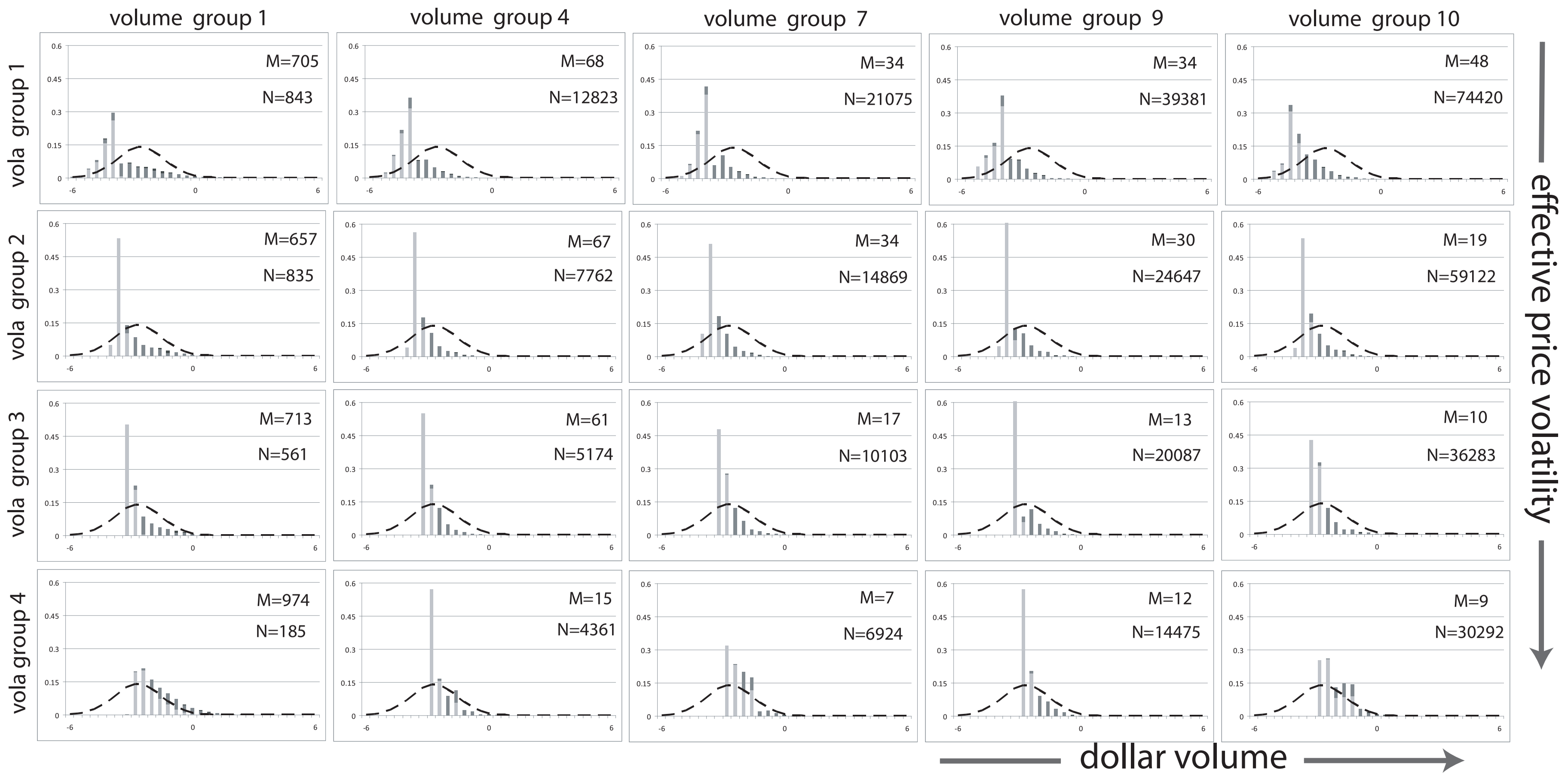
This figure shows distributions of the logarithms of normalized TAQ print sizes for Nasdaq stocks in April 1993. For each trade, the normalized print size is calculated as $\ln(W^{2/3} \cdot |X|/V)$ based on the invariance hypothesis, where $|X|$ is the midpoint of the print size bin in shares, V is the average daily volume in shares, and W 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 30th, 50th, 60th, 70th, 75th, 80th, 85th, 90th, and 95th percentiles of the dollar volume for common NYSE-listed stocks. The four equally-spaced effective price volatility on the definition $P \cdot \sigma \cdot (W/W_*)^{-1/3}$. The subplots show stock-level distributions averaged across stocks for volume groups 1 (low volume), 4, 7, and 10 (high volume) and for all four price volatility groups 1 (low price volatility), 2, 3, 4 (high price volatility). The 100-share trades are highlighted in light grey; the 1000-share trades are highlighted in dark grey. Each subplot also shows a normal distribution with the pooled average print size mean of -0.18 and pooled averaged standard deviation of 1.36. M is the number of stocks, and N is the average number of TAQ prints per day for the stocks in a given subgroup.

Figure 6: Trade-Weighted Distributions of Normalized TAQ Print Sizes, All Stocks, April 2001



This figure shows distributions of the logarithms of normalized TAQ print sizes for NYSE and Nasdaq stocks in April 2001. For each trade, the normalized print size is calculated as $\ln(W^{2/3} \cdot |X|/V)$ based on the invariance hypothesis, where $|X|$ is the midpoint of the print size bin in shares, V is the average daily volume in shares, and W 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 30th, 50th, 60th, 70th, 75th, 80th, 85th, 90th, and 95th percentiles of the dollar volume for common NYSE-listed stocks. The four equally-spaced effective price volatility on the definition $P \cdot \sigma \cdot (W/W_*)^{-1/3}$. The subplots show stock-level distributions averaged across stocks for volume groups 1 (low volume), 4, 7, and 10 (high volume) and for all four price volatility groups 1 (low price volatility), 2, 3, 4 (high price volatility). The 100-share trades are highlighted in light grey; the 1000-share trades are highlighted in dark grey. Each subplot also shows a normal distribution with the pooled average print size mean of -1.31 and pooled averaged standard deviation of 1.32. M is the number of stocks, and N is the average number of TAQ prints per day for the stocks in a given subgroup.

Figure 7: Trade-Weighted Distributions of Normalized TAQ Print Sizes, All Stocks, April 2008



This figure shows distributions of the logarithms of normalized TAQ print sizes for NYSE and Nasdaq stocks in April 2008. For each trade, the normalized print size is calculated as $\ln(W^{2/3} \cdot |X|/V)$ based on the invariance hypothesis, where $|X|$ is the midpoint of the print size bin in shares, V is the average daily volume in shares, and W 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 30th, 50th, 60th, 70th, 75th, 80th, 85th, 90th, and 95th percentiles of the dollar volume for common NYSE-listed stocks. The four equally-spaced effective price volatility on the definition $P \cdot \sigma \cdot (W/W_*)^{-1/3}$. The subplots show stock-level distributions averaged across stocks for volume groups 1 (low volume), 4, 7, and 10 (high volume) and for all four price volatility groups 1 (low price volatility), 2, 3, 4 (high price volatility). The 100-share trades are highlighted in light grey; the 1000-share trades are highlighted in dark grey. Each subplot also shows a normal distribution with the pooled average print size mean of -2.66 and pooled averaged standard deviation of 1.15. M is the number of stocks, and N is the average number of TAQ prints per day for the stocks in a given subgroup.

Table 2: OLS Estimates of Number of TAQ Prints.

	All Stocks		NYSE/AMEX		NASDAQ	
	1993–2000	2001–2008	1993–2000	2001–2008	1993–2000	2001–2008
$\mu_{\gamma,0}$	6.154	8.043	6.067	7.814	6.170	8.300
	(0.017)	(0.034)	(0.012)	(0.027)	(0.019)	(0.050)
$\mu_{\gamma,t}$	0.078	0.242	0.037	0.268	0.123	0.225
	(0.007)	(0.019)	(0.005)	(0.014)	(0.007)	(0.026)
$a_{\gamma,0}$	0.690	0.787	0.639	0.758	0.705	0.827
	(0.001)	(0.003)	(0.001)	(0.003)	(0.002)	(0.005)
$a_{\gamma,t}$	-0.002	0.017	0.002	0.018	0.000	0.016
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.002)
Adj-R^2	0.91	0.94	0.93	0.95	0.90	0.94
# Obs	6,698	4,914	2,199	1,904	4,499	3,010

This table presents Fama-MacBeth estimates μ_γ and a_γ from monthly regressions

$$\ln\left[\bar{N}_i\right] = \mu_\gamma + a_\gamma \cdot \ln\left[\frac{W_i}{W_*}\right] + \tilde{\epsilon_i}.$$

For each month, there is one observation for each stock i. The value of $\bar{N}_i$ is the average number of TAQ 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 price \$40 per share, trading volume of one 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 $\hat{\mu}_{\gamma,T}$ and $\hat{a}_{\gamma,T}$ for each month: $\hat{\mu}_{\gamma,T} = \mu_{\gamma,0} + \mu_{\gamma,t} \cdot (T - \bar{T})/12 + \tilde{\epsilon_T}$ and $\hat{a}_{\gamma,T} = a_{\gamma,0} + a_{\gamma,t} \cdot (T - \bar{T})/12 + \tilde{\epsilon_T t}$, where T is the number of months from the beginning of the sample and $\bar{T}$ is the mean month. "Adj-R^2" denotes the adjusted R^2 averaged over monthly regressions, and "# Obs" denotes the number of stocks averaged over monthly regressions. The estimates are reported for two subperiods 1993-2000 and 2001-2008.

Table 3: Regression Estimates of TAQ Print Sizes, February 1993 - December 2000.

	Trade-Weighted Distribution				Volume-Weighted Distribution			
	Mean	20*th*	50*th*	80*th*	Mean	20*th*	50*th*	80*th*
$\mu_{Q,0}$	-7.219	-8.470	-7.258	-6.240	-4.717	-6.388	-4.933	-3.376
	(0.020)	(0.027)	(0.030)	(0.013)	(0.018)	(0.015)	(0.017)	(0.023)
$\mu_{Q,t}$	-0.037	-0.031	-0.039	-0.054	-0.114	-0.081	-0.134	-0.134
	(0.008)	(0.011)	(0.012)	(0.006)	(0.007)	(0.006)	(0.007)	(0.010)
$a_{Q,0}$	-0.759	-0.797	-0.765	-0.742	-0.591	-0.688	-0.611	-0.514
	(0.002)	(0.003)	(0.002)	(0.003)	(0.002)	(0.001)	(0.002)	(0.003)
$a_{Q,t}$	0.009	0.012	0.008	0.006	-0.003	0.002	-0.005	-0.008
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Adj-R^2	0.93	0.90	0.91	0.91	0.75	0.87	0.75	0.61
#Obs	6,698	6,698	6,698	6,698	6,698	6,698	6,698	6,698

This table presents Fama-MacBeth estimates μ_Q and a_Q from the monthly regressions of the mean and percentiles of TAQ 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

$$\ln\left[\frac{|X_i|}{V_i}\right] = \mu_Q + a_Q \cdot \ln\left[\frac{W_i}{W_*}\right] + \tilde{\epsilon_i},$$

where the left-hand side is either the mean or the *pth* (20th, 50th and 80th) percentile of the distribution of logarithms of (unsigned) TAQ 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. For each month, there is one observation for 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 price \$40 per share, trading volume of one 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 $\hat{\mu}_{Q,T}$ and $\hat{a}_{Q,T}$ for each month: $\hat{\mu}_{Q,T} = \mu_{Q,0} + \mu_{Q,t} \cdot (T - \bar{T})/12 + \widetilde{\epsilon_T}$ and $\hat{a}_{Q,T} = a_{Q,0} + a_{Q,t} \cdot (T - \bar{T})/12 + \widetilde{\epsilon_T}t$, where T is the number of months from the beginning of the sample and $\bar{T}$ is the mean month. "Adj-R^2" denotes the adjusted R^2 averaged over monthly regressions, and "#Obs" denotes the number of stocks averaged over monthly regressions.

Table 4: Regression Estimates of TAQ Print Sizes, January 2001 - December 2008.

	Trade-Weighted Distribution				Volume-Weighted Distribution			
	Mean	$20th$	$50th$	$80th$	Mean	$20th$	$50th$	$80th$
$\mu_{Q,0}$	-8.692	-9.417	-8.977	-8.072	-6.824	-8.549	-7.355	-5.558
	(0.035)	(0.032)	(0.038)	(0.032)	(0.032)	(0.047)	(0.035)	(0.032)
$\mu_{Q,t}$	-0.149	-0.025	-0.162	-0.255	-0.326	-0.289	-0.388	-0.382
	(0.021)	(0.017)	(0.022)	(0.018)	(0.015)	(0.026)	(0.018)	(0.015)
$a_{Q,0}$	-0.793	-0.787	-0.793	-0.805	-0.743	-0.799	-0.800	-0.720
	(0.003)	(0.005)	(0.003)	(0.002)	(0.003)	(0.002)	(0.003)	(0.004)
$a_{Q,t}$	-0.003	0.005	0.001	-0.012	-0.034	-0.020	-0.041	-0.050
	(0.002)	(0.002)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.002)
Adj-R^2	0.93	0.90	0.92	0.93	0.86	0.91	0.87	0.77
# Obs	4,914	4,914	4,914	4,914	4,914	4,914	4,914	4,914

This table presents Fama-MacBeth estimates μ_Q and a_Q from the monthly regressions of the mean and percentiles of TAQ print size on trading activity W for the sample from January 2001 to December 2008. The coefficients μ_Q and a_Q are based on monthly regressions

$$\ln\left[\frac{|X_i|}{V_i}\right] = \mu_Q + a_Q \cdot \ln\left[\frac{W_i}{W_*}\right] + \tilde{\epsilon_i},$$

where the left-hand side is either the mean or the pth (20th, 50th and 80th) percentile of the distribution of logarithms of (unsigned) TAQ 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. For each month, there is one observation for 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 price \$40 per share, trading volume of one 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 $\hat{\mu}_{Q,T}$ and $\hat{a}_{Q,T}$ for each month: $\hat{\mu}_{Q,T} = \mu_{Q,0} + \mu_{Q,t} \cdot (T - \bar{T})/12 + \tilde{\epsilon_T}$ and $\hat{a}_{Q,T} = a_{Q,0} + a_{Q,t} \cdot (T - \bar{T})/12 + \tilde{\epsilon_T t}$, where T is the number of months from the beginning of the sample and $\bar{T}$ is the mean month. "Adj-R^2" denotes the adjusted R^2 averaged over monthly regressions, and "#Obs" denotes the number of stocks averaged over monthly regressions.

Table 5: OLS Estimates of Number of TAQ Prints with Effective Volatility.

	All Stocks		NYSE/AMEX		NASDAQ	
	1993–2000	2001–2008	1993–2000	2001–2008	1993–2000	2001–2008
$\mu_{\gamma,0}$	6.212	7.595	6.247	7.508	6.190	7.673
	(0.016)	(0.018)	(0.009)	(0.016)	(0.019)	(0.027)
$\mu_{\gamma,t}$	0.088	0.184	0.024	0.219	0.129	0.167
	(0.007)	(0.010)	(0.004)	(0.007)	(0.009)	(0.014)
$a_{\sigma,0}$	-0.447	-0.546	-0.315	-0.378	-0.472	-0.647
	(0.003)	(0.007)	(0.005)	(0.008)	(0.003)	(0.008)
$a_{\sigma,t}$	-0.009	-0.038	-0.008	-0.036	-0.026	-0.039
	(0.001)	(0.003)	(0.002)	(0.004)	(0.001)	(0.003)
Adj-R^2	0.938	0.956	0.936	0.955	0.941	0.963
# Obs	6,698	4,914	2,199	1,904	4,499	3,010
	Regression With Coefficient on Effective Price Volatility $a_\sigma = 0$.					
Adj-R^2	0.906	0.915	0.924	0.937	0.901	0.899
	Regression With Separate Coefficients for Price, Volume, and Volatility.					
Adj-R^2	0.943	0.965	0.946	0.971	0.942	0.975

This table presents Fama-MacBeth estimates μ_γ and a_σ from monthly regressions

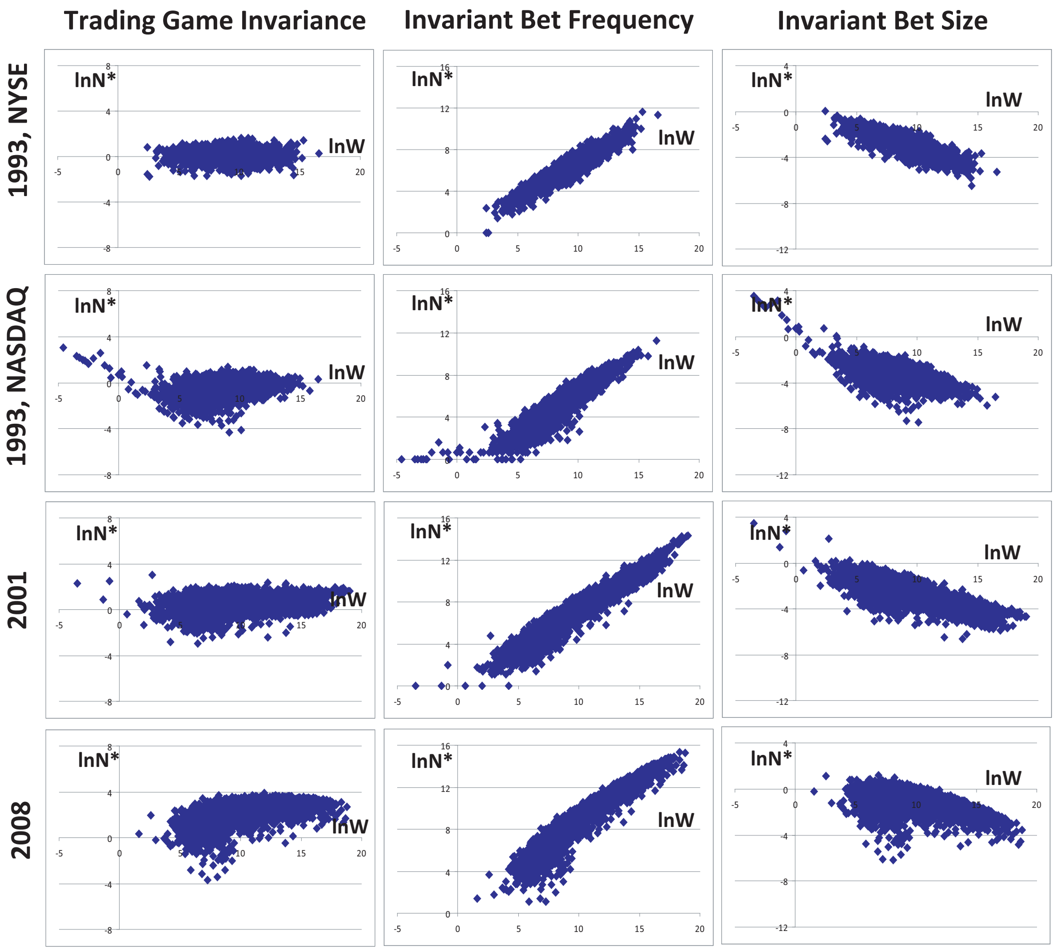
$$\ln\left[\bar{N}_i\right] = \mu_\gamma + \frac{2}{3}\cdot\ln\left[\frac{W_i}{W_*}\right] + a_\sigma\cdot\ln\left[\frac{P_i\cdot\sigma_i}{P_*\cdot\sigma_*}\cdot\left(\frac{W_i}{W_*}\right)^{-1/3}\right] + \tilde{\epsilon}_i.$$

For each month, there is one observation for 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 equal to $40 \cdot 0.02$. The value of $\bar{N}_i$ is the average number of TAQ prints per day. The scaling constant $W_* = (40)(10^6)(0.02)$ corresponds to the measure of trading activity for the benchmark stock with price \$40 per share, trading volume of one 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 $\hat{\mu}_{\gamma,T}$ and $\hat{a}_{\sigma,T}$ for each month: $\hat{\mu}_{\gamma,T} = \mu_{\gamma,0} + \mu_{\gamma,t}\cdot(T-\bar{T})/12 + \tilde{\epsilon_T}$ and $\hat{a}_{\sigma,T} = a_{\sigma,0} + a_{\sigma,t}\cdot(T-\bar{T})/12 + \epsilon_{\tilde{T}}t$, where T is the number of months from the beginning of the sample and $\bar{T}$ is the mean month. "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

$$\ln\left[\bar{N}\right] = \mu_\gamma + \frac{2}{3}\ln\left[\frac{W_i}{W_*}\right] + b_1\cdot\ln\left[\frac{V_i}{(10^6)}\right] + b_2\cdot\ln\left[\frac{P_i}{(40)}\right] + b_3\cdot\ln\left[\frac{\sigma_i}{(0.02)}\right] + \tilde{\epsilon}. \quad (17)$$

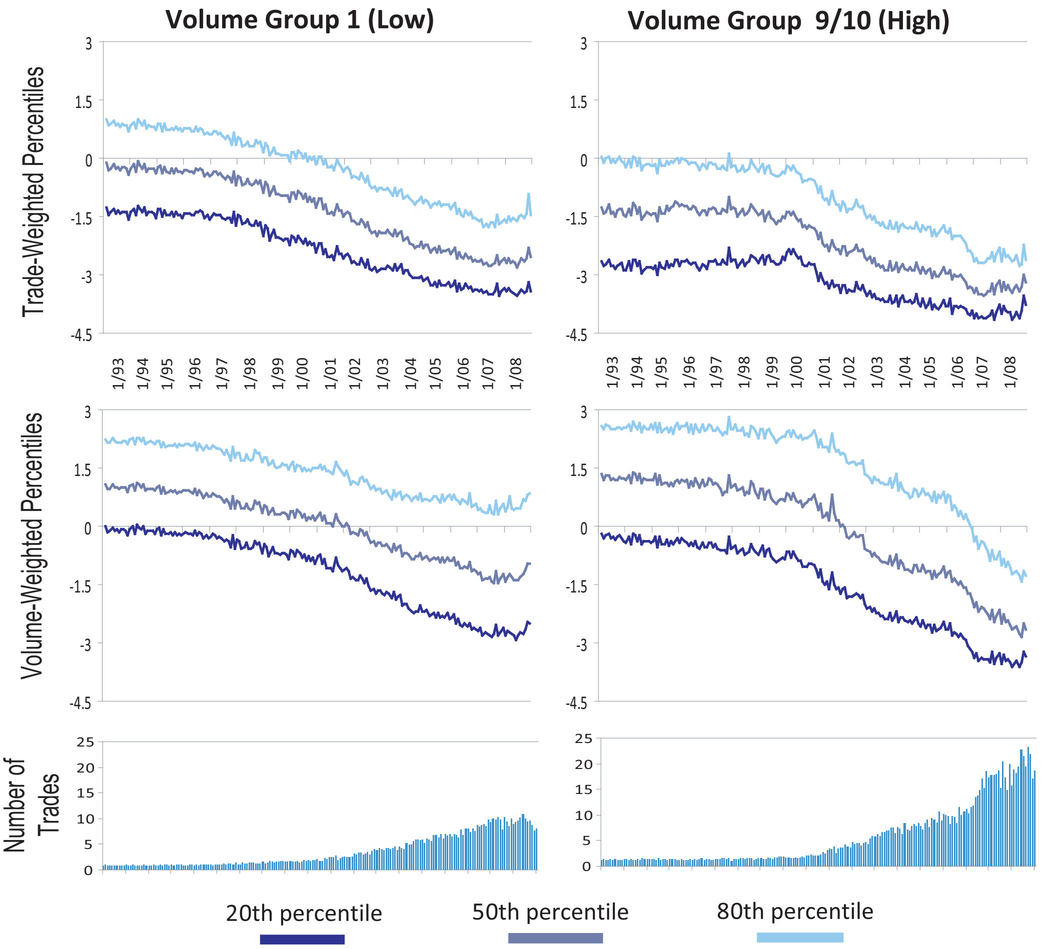
"# Obs" is the number of stocks averaged over monthly regressions. The estimates are reported for two subperiods 1993-2000 and 2001-2008.

Figure 8: The Normalized Number of TAQ Prints Relative to Trading Activity for Three Models.



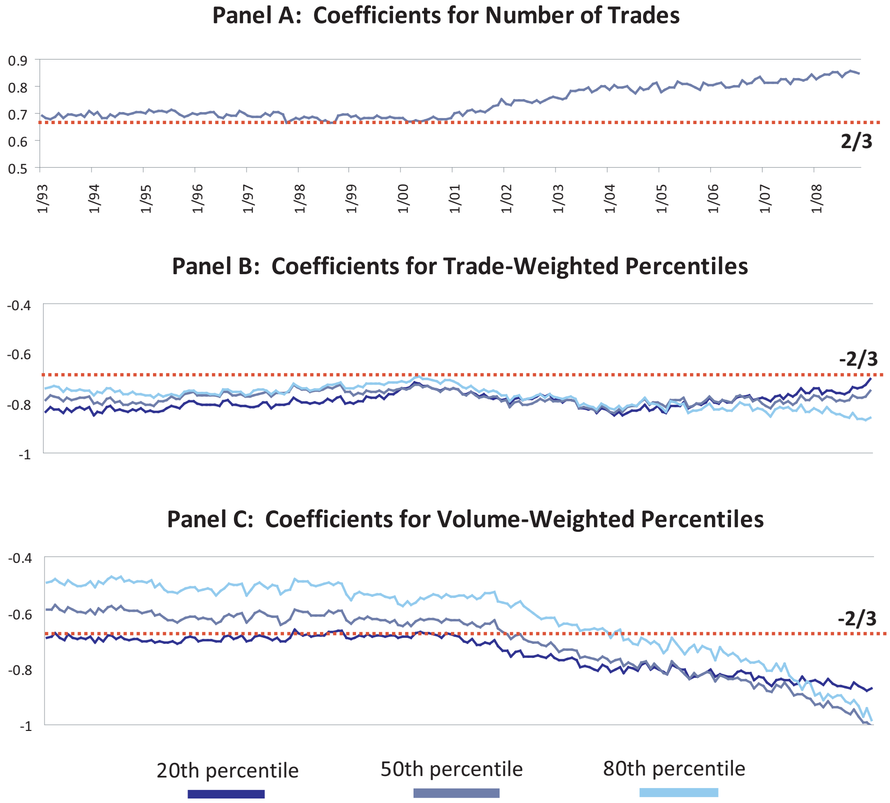
The figure shows the logarithm of the normalized number of TAQ prints across different levels of the logarithm of trading activity W. The normalize number of TAQ prints is defined by $\bar{N}/W^{\alpha}$, with $\alpha = 2/3$ for the model of trading game invariance, $\alpha = 0$ for the model of invariant bet frequency, and $\alpha = 1$ for the model of invariant bet size. Four subsamples are considered: NYSE-listed stocks in April of 1993, Nasdaq-listed stocks in April of 1993, both NYSE and Nasdaq stocks in April of 2001 and both Nasdaq and NYSE stocks in April of 2008. Trading activity W is calculated as the product of average daily dollar $P \cdot V$ volume and the percentage standard deviation of daily returns for a given month.

Figure 9: Time Series of Average Percentiles of Normalized TAQ Print Size and Mean Number of Prints, 1993-2008.



The figure shows the dynamics of the sample means of the 20th, 50th and 80th percentiles for logarithms of the normalized print sizes as well as the means of the normalized number of prints per month from 1993 to 2008. Volume groups are based on average dollar trading volume with thresholds corresponding to 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 normalized print size is calculated based on the midpoint of the print size bin, normalized according to the model of trading game invariance, i.e. $\ln(W^{2/3} \cdot |X|/V)$, where $|X|$ is a midpoint of a print size bin in shares, V is the average daily volume in shares, and W is the measure of trading activity equal to the product of dollar volume and returns standard deviation. The normalized number of TAQ prints per month is calculated as $\bar{N}_m \cdot W^{-2/3}$, where $\bar{N}_m$ is the number of trades per month. The stock-level distributions of normalized print sizes are averaged across stocks for volume groups 1 and 9-10 in a given month. The trade-weighted and volume-weighted percentiles are plotted on this figure.

Figure 10: Time Series of Monthly OLS Coefficient Estimates for Number of Trades, Trade-Weighted Percentiles, and Volume-Weighted Percentiles, 2003-2008.



The figure shows the dynamics of coefficients from regressions of number of prints and various percentiles on the measure of trading activity W from 1993 to 2008. Panel A shows the coefficient a_γ from monthly regressions

$$\ln\left[\bar{N}_i\right] = a + a_\gamma \cdot \ln\left[\frac{W_i}{W_*}\right] + \tilde{\epsilon}_i,$$

where $\bar{N}_i$ is the average number of TAQ prints per day in a given month. The model of trading game invariance predicts $a_\gamma = 2/3$ and alternative models predict that $a_\gamma = 0$ or $a_\gamma = 1$. Panel B shows the coefficient a_Q from monthly regressions

$$\ln\left[\frac{\tilde{X}_i}{V_i}\right] = a + a_Q \cdot \ln\left[\frac{W_i}{W_*}\right] + \tilde{\epsilon}_i,$$

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 $a_Q = -2/3$ and alternative models predict that $a_Q = 0$ or $a_Q = -1$. Panel C shows the coefficient a_Q 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 $a_Q = -2/3$ and alternative models predict that $a_Q = 0$ or $a_Q = -1$. Trading activity W is defined as the product of dollar volume and daily percentage standard deviation of returns, and W_* measures of trading activity of the benchmark stock.