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
Robert H. Smith School of Business
University of Maryland
stuzun@rhsmith.umd.edu

Preliminary Version

September 1, 2010

Abstract

Our paper tests the theory of trading game invariance using the sample of unsigned trades from the Trades and Quotes dataset from 1993 to 2008. We examine two predictions concerning trading patterns. First, the number of trades should vary across stocks proportionally to their trading activity in 2/3 power. Second, the distribution of trade sizes as a fraction of trading volume should vary across stocks proportionally to their trading activity in -2/3 power. Our results support the invariance theory, though the evidence is distorted by various market frictions such as relative tick size, minimum lot size, clustering of trades, and order shredding. For the number of trades, the power coefficient is equal to 0.69 (with standard errors of 0.002) but, following a reduction in tick size in 2001 and a consequent spread of algorithmic trading, the estimate increases to 0.79 (with standard errors of 0.011). The distribution of trade sizes for individual stocks varies with the trading activity in a manner predicted by the invariance theory as well, i.e., when trade sizes are adjusted for differences in trading activity, their distributions are stable across stocks. These distributions are similar to the distribution of a log-normal variable truncated from below at the 100-share threshold.

1 Introduction

Even a quick look at financial markets reveals a significant variation in how securities are traded. During a month, hundred thousands of large orders are executed for some securities, meanwhile only a few small orders arrive to the marketplace for others. Kyle and Obizhaeva (2010) develop the invariance theory that suggests how the distribution of trade sizes and trade arrival rates should vary across stocks with different levels of trading activity. We examine its predictions using the Trade and Quote (TAQ) dataset for the period from 1993 to 2008.

The main idea of the trading game invariance is that trading games played in different securities are fundamentally the same. The only difference between these games is the time horizon, called a "trading day", over which they are being played. Trading days are short for actively traded stocks, perhaps corresponding to a few minutes, and long for inactively traded stocks, perhaps corresponding to a few months. This invariance theory generates testable predictions concerning how the arrival rate of trades, and the distribution of their sizes vary with trading activity across stocks and across time for the same stock: Increasing trading activity by one percent increases the arrival rate of trades by two-thirds of one percent, increases trade size by one-third of one percent and thus decreases trade size as a fraction of average daily volume by two-thirds of one percent, and leaves unchanged the shape of the distribution of trade size. "Trading activity" is defined as the product of the daily dollar volume and volatility.

The TAQ dataset contains tick-by-tick data recorded on consolidated tape for all listed stocks starting from 1993. For each stock and each month, we calculate the number of trades per month and construct the empirical distributions of trade sizes (in shares) that we describe by a list of attributes. These attributes include the means of trade sizes, the ten equally-spaced percentiles of trade sizes based on the distribution of trade sizes and ten equally-spaced percentiles based on their contribution to the total trading volume. The latter put more weight on larger and economically more significant trades thus allowing us to examine the right tail of distributions in more detail. Our goal is to study whether cross-sectional variation in the constructed variables is consistent with predictions of the invariance theory.

Our tests show that the trading game invariance explains the cross-sectional variation in the arrival rates of trades. We regress the logarithm of number of trades on the logarithm of trading activity and find the coefficient equal to 0.74 for the entire sample from 1993 to 2008, 0.69 for the subsample from 1993 to 2001, and 0.79 for the subsample from 2001 to 2008 with standard errors smaller than 0.015. These estimates are remarkably close to 2/3, predicted by the invariance theory. We suggest that the difference between two subsamples can be due to a structural break related to a reduction in tick size, which unevenly affected small and large stocks. Our results also clearly reject alternative models of invariant bet size and invariant bet frequency. According to these models, the estimated coefficient should be either close to 0 or 1.

Our tests show that the trading game invariance also provides a good explanation for the cross-sectional variation in the distribution of trade sizes. We regress the logarithm of the constructed attributes of the distribution of trade size (as a fraction of trading volume) on the logarithm of the trading activity. When we use the mean of distribution on the left-hand side, our estimates are equal to -0.74 before 2001 and -0.79 after 2001. When we use the trade-based percentiles, the coefficients range from -0.74 to -0.81. For the volume-weighted percentiles, the coefficients range from -0.51 to -0.80. These estimates are close to $-2/3$ predicted by the invariance theory. Alternative theories, predicting these coefficients to be 0 and -1, obviously represent a much less accurate description of actual data.

To further investigate trading patterns, we examine empirical distributions of trade sizes more closely. According to the invariance theory, the distribution of trade sizes for any stock can be transformed, by a proper adjustment for differences in trading activity (i.e. multiplying by trading activity in two-third power), into a single distribution, common across stocks and across time. Can we observe this prediction in the data? For ten volume and four price volatility groups, we plot the distributions of the logarithms of the trades sizes, normalized as suggested by invariance theory. We find that these distributions are indeed quite stable across all subgroups. Furthermore, they closely resemble bell-shaped density function of a normal random variable, suggesting that the normalized trade sizes are distributed as a log-normal random variable. Note that if we adjust trade sizes according to alternative models, then the resulting distributions are much less stable across volume groups, with their means varying significantly across stocks. This is another evidence in favor of the invariance theory.

Our finding that the distribution of order sizes looks similar to log-normal has important economic implications. It suggests that the order flow is dominated by really large orders and that small orders are much smaller than large orders. Almost all trading volume therefore appears in large orders. Almost all variance of price changes comes from the price effects of large orders.

Our plots reveal, however, some systematic differences between empirical distributions for stocks with different levels of trading volume and volatility. The statistical tests on whether normalized trade sizes are distributed as a log-normal random variable are also rejected. We argue that various market frictions, such as a minimum tick size and a minimum lot size, significantly distort trading patterns and do not allow us to observe stronger evidence for the invariance theory.

Market frictions are changing over time. Hendershott et al. (2010) and Chordia et al. (2009) discuss the recent transformation of trading activity after the decimalization of 2001 and a consequent spread of algorithmic trading. With the goal to examine these issues more closely, we plot distributions of normalized trade sizes for stocks sorted into ten volume and four price volatility in years 1993, 2001, and 2008.

Our distributions look similar to bell-shaped distributions being truncated from below at the 100-share threshold. The effect of the minimum lot size is especially pronounced after the reduction in tick size in 2001. For example, 100-share trades accounted for 16% of all trades executed before 2001 and 50% of trades executed after 2001. This number reaches astonishing 70% in 2008. Effectively, most orders, even the largest ones, are now being shredded into sequences of 100-share trades. This practice came along with the introduction of electronic interface and algorithmic trading. Since the minimum lot size restriction and relative tick size certainly influence order shredding algorithms, they have to be taken into account when one tests the invariance theory. It is worth reminding that the invariance theory makes predictions about intended orders rather than actual trades generated by order-shredding algorithms. The extent of order shredding in the markets after 2001 makes us feel that the invariance theory can be hardly tested with the TAQ data on unsigned trades after 2001, unless one makes particular assumptions about the order shredding algorithms and how they vary across stocks.

Market frictions often affect trading patterns in the opposite way. Consider, for example, whether we should expect to observe smaller or larger trades, relative to the level predicted by invariance theory, for volatile stocks. On one hand, volatile stocks are expected to be constrained by the minimum lot size of 100 shares. If trades below the 100-share threshold are never brought into the marketplace, then their actual trades might seem to be "too large." On the other hand, smaller effective tick size makes trader submit smaller orders at finer price levels, therefore trades in volatile stocks might seem to be "too small." Our analysis shows that the first effect is more significant in the data than the second one.

A number of other market frictions can complicate testing the invariance theory using the TAQ dataset. Trades tend to cluster at some levels (e.g., Alexander and Peterson (2007)). Some of this clustering is a result of various regulations. There has been, for instance, the requirement to fix the minimum quotation size at the level of 1,000 shares for trades with large market participants imposed on the Nasdaq market makers from 1988 to 2001. This restriction is clearly reflected in our plots as a disproportionate number of 1,000-share trades for Nasdaq-listed during that period. These types of market frictions are certainly not captured by the invariance theory. An interesting topic for the further research is to design better econometric tests that will account for various market frictions.

Our paper adds to the results in Kyle and Obizhaeva (2010) that documented a strong evidence for the invariance theory using the sample of portfolio transitions. Portfolio transitions represent a special subset of market transactions with unique properties that make them especially valuable for testing invariance theory, since the data allows to observe actual orders intended for execution rather then sequences of executed trades. In contrast, our tests are based on a broader sample of trading data. Its broad coverage comes, however, at the expense of having to deal with less perfect data and thus detecting somewhat less clean evidence.

There has been a literature on what determines trading frequency and trade sizes. Glosten and Harris (1988) found that average trade size (in shares) is negatively related to market depth. Brennan and Subrahmanyam (1998) documented that trade sizes (in dollars) are also related to other stock characteristics such as return volatility, the standard deviation of trading volume, the market capitalization, the number of analysts following stocks, the number of institutional investors holding stocks, and the proportion of shares hold by them. Interestingly, the R-squared in their cross-sectional regressions is about 0.92, which is very similar to R-squared in our regressions even when we restrict the power coefficient to be equal to 2/3, as implied by the invariance theory, leaving only a constant term to estimate. The similarity of R-squared indicates that the inclusion of additional explanatory variables does not help much in explaining the cross-sectional variations of the average trade sizes, in addition to the invariance theory.

The remainder of this paper states the implications of trading game invariance, discusses the issues arising when these implications are tested using the Trades and Quotes dataset, and then describes results of empirical tests.

2 Testable Implications of Trading Game Invariance

In Kyle and Obizhaeva (2010), traders are thought as playing trading games. They arrive to the market and execute orders. Innovations in their order flow follow a compound Poisson process with arrival rate of γ innovations per day. A typical innovation in this order flow, called a "bet," is a random variable $\tilde{Q}$ with a zero mean. A positive value of $\tilde{Q}$ represents buying, and a negative value of $\tilde{Q}$ represents selling.

The invariance theory is formulated in terms of "bet size" that measures risk transferred by the bet. "Bet size" is a random variable defined as the product of dollar bet value (dollar share price P times share quantity $\tilde{Q}$) and volatility σ (percentage standard deviation of returns per day),

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

In a similar spirit, "trading activity" is defined as the product of the arrival rate of bets and the expected absolute value of bet size,

$$W = \gamma \cdot E|\tilde{B}| = \sigma_r \cdot P \cdot V. \tag{2}$$

The last equality follows from the definition of expected trading volume V over a day,

$$V = \gamma \cdot E|\tilde{Q}|. \tag{3}$$

According to this measure of trading activity, active stocks are stocks with high volatility and high dollar trading volume per calendar day. Inactive stocks are stocks with low volatility and low dollar trading volume per calendar day.

The invariance theory describes how market microstructure varies across stocks with different levels of trading activity. Its main idea is that trading games are the same across stocks, up to some Modigliani-Miller transformation, except for the speed with which these games are being played. This speed is related to the level of trading activity W: Trading games are played faster in active stocks and slower in inactive stocks. For this claim to be true, the microstructure invariant

$$\tilde{I} = \frac{\tilde{B}}{\gamma^{1/2}}. \tag{4}$$

needs to have the same distribution across stocks and across time. Trading game invariance generates testable predictions concerning how bet frequency and (unsigned) bet size should vary with trading activity,

$$\gamma = E\{|\tilde{I}|\}^{1/3} \cdot W^{2/3}, \tag{5}$$

$$\frac{|\tilde{Q}|}{V} = \frac{|\tilde{I}|}{E\{|\tilde{I}|\}^{1/3}} \cdot W^{-2/3}. \tag{6}$$

The entire distribution of trade sizes and trading frequency, normalized for differences in trading activity W,

$$\gamma \times W^{-2/3} \quad \text{and} \quad \frac{|\tilde{Q}|}{V} \times W^{2/3}$$

have to be identical across stocks and across time.

This model says that changes in daily trading activity W come from both changes in unsigned bet size and changes in bet frequency per calendar day. In particular, the invariance of trading games requires that a one percent increase in trading activity W is associated with an increase of 2/3 of one percent in bet arrival frequency γ and an upward shift by 1/3 of one percent of the entire distribution of bet size $\tilde{B}$. The latter implication is equivalent to saying that distributions of (unsigned) trade sizes $|\tilde{Q}|$ as a fraction of trading volume V shifts downwards by 2/3 of one percent when trading activity increases by one percent.

Alternative Models. There are two alternative models: Model of Invariant Bet Frequency and Model of Invariant bet Size. The model of Invariant Bet Frequency assumes that the variation in trading activity comes entirely from variation in bet sizes $\tilde{B}$, while the number of bets γ over a calendar day remains invariant across stocks. This model generates testable predictions concerning how (unsigned) bet size should varies with trading activity. The entire distribution of trade sizes and trading frequency,

$$\gamma \cdot W^0 \quad \text{and} \quad \frac{|\tilde{Q}|}{V} \cdot W^0$$

have to be identical across stocks and across time, even if no adjustments for differences in trading activity W are made.

The model of Invariant Bet Size assumes that the variation in trading activity comes entirely from variation in the number of bets γ placed over a calendar day. The distribution of bet size $\tilde{B}$ over a calendar day remains the same across stocks. This model also generates testable predictions concerning how (unsigned) bet size should varies with trading activity. The entire distribution of trade sizes and trading frequency, normalized for differences in trading activity W,

$$\gamma \times W^{-1} \quad \text{and} \quad \frac{|\tilde{Q}|}{V} \times W^1$$

have to be identical across stocks and across time.

All three models imply specific relations between the number of bets γ and the distribution of bet sizes $\tilde{B}$ per a calendar day on one side and the measure of trading activity W on the other side. The only difference between their predictions is the power coefficient by trading activity W. We discuss next how to test which of three models describes the best transactions data in the Trades and Quotes database.

Testing Theories Using the TAQ Dataset. The TAQ database contains a time-stamped record of trades printed for NYSE and NASDAQ stocks. We can thus estimate trade frequency γ looking at the average number of prints per a calendar day as well as the distribution of trade size $|\tilde{Q}|$ looking at the empirical distribution of (unsigned) print sizes in the TAQ database. Equipped with this data, we can test the predictions of the invariance theory and alternatives. There are, however, several important problems that may be a hurdle for our tests.

The assumption that the inventory of traders follows a compound Poisson process implies that their trades are independently distributed. 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 executed against several different counter-parties and prices. The TAQ database gives a time-stamped records of trades as printed for NYSE and NASDAQ stocks. It allow us to observe neither independent intended orders nor the number of prints into which initial order has been split. At the same time, order shredding algorithms may vary across stocks in a complex manner, for example, as a function of stock price (based on tick size) or a stringency of a minimum lot size constraint. Furthermore, some orders can be only partially executed, with the amount of realized cancelations also potentially related to some properties of securities.

If actual trading strategies vary across securities in a systematic way, then our empirical tests may be significantly distorted by cross-sectional differences in these trading strategies. How to modify theories of invariants for a potential autocorrelation of trades and their cancelation is an interesting issue for future research. In this paper, we design our tests assuming that correlated and canceled orders constitute only an insignificant part of total trading volume as, for example, in Kyle (1985).

Dealing with various market frictions is another issue in designing good empirical tests, because market frictions most likely affect trading patterns. For example, there have been various restrictions on trade sizes during a period from 1993 to 2008. These restrictions certainly affect observed trade sizes and trading frequency. If trades can not be smaller than 100 shares, then intended orders with less than 100 shares will be either not submitted at all or rounded up to satisfy a minimum lot size requirement. This restriction is expected to be especially binding for low-volume stocks with small orders traded at high price levels. Another binding restriction could be the fixing of market-maker minimum quotation sizes at the level of 1,000 shares at Nasdaq from 1988 to 2001. As we will see, this restriction has led to clustering of Nasdaq trades a the level of 1,000 shares. Another important market friction to keep in mind is a tick size, or a minimum increment by which price can move. When tick size as a fraction of price volatility is large, traders tend to submit larger orders since fewer price levels are available.

In this study, we provide the evidence on how these frictions influence our results. Exploring these issues in more detail is an interesting issue for future research.

3 Data

3.1 Data Description

The NYSE TAQ database contains trade and quote reported on the consolidated tape by each CTA participants for all stocks listed on exchanges starting from year 1993. Since we examine the distribution of unsigned trades, our analysis employs exclusively data on trades. We leave the interesting issues about signed trades, which may be reconstructed if we use both trades and quotes, for the future research. For each trade, the TAQ data records the time, exchange, security symbol, number of shares traded, execution price, trade condition, and other parameters. This database is of a large size. Its subset from 1993 to 2008 contains over 19 billion records with over 5 million records per month in 1993 and over 500 million records per month in 2008.

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

The remaining trades are aggregated in a specific way. These trades are placed into 55 bins that we construct based on the number of shares traded. "Even" bins correspond to the orders of even sizes, i.e., trades with the following numbers of shares traded: 100, 200, 300, 400, 500, 1000, 2000, 3000, 4000, 5000, 10000, 15000, 20000, 25000, 30000, 40000, 50000, 60000, 70000, 75000, 80000, 90000, 100000, 200000, 300000, 400000, 500000. "Odd" bins correspond to the trades with odd sizes, i.e., when the number of shares is in-between even bins. We chose these size bins so that their size grows at the approximately log-scale. The selection of these bins also reflects our intention to treat separately trades with even sizes, because these trades are especially frequent in the data. For each day and each symbol, we store the number of trades in each size bin. Once we assign a given trade to an appropriate size bin, we assume its size (in shares) is equal to a midpoint of that bin. If trade size is larger than 500,000 shares, it is assigned to 55th bin and its size is assumed to be 1000,000 shares. The aggregation of the data into size bins allows us to capture the main properties of trade size distribution and implement our analysis in a more efficient way. This convenience comes, however, at the expense of introducing additional noise into our analysis, which may affect our results.

For each day and each symbol, we also store other variables such as the open price, the close price, the number of trades per day, the dollar volume per day, the shares volume per day, the close-to-close return, and the volatility defined as the standard deviation of returns over the past 20 trading days.

Since many stocks do not have enough transaction per day, we can not use daily data to build a good empirical approximation for a theoretical distribution of trade sizes for individual stocks. We therefore aggregate our daily data into monthly files. We sum up the number of trades within each bin and construct empirical distributions of trade sizes (in shares) for each stock and each month in the sample.

The theoretical distributions of trade sizes can be of quite a general form. We define several attributes of these distributions with the intention to capture their shapes to the greatest extent. These variables are estimated from the empirically observed distributions. The attributes include the average number of trades per day. We also consider the average trade size and its various percentiles based on trade size distribution. We refer to these percentiles as trade-weighted percentiles. For example, the xth trade-based percentile corresponds to a trade size such that trades with sizes above this threshold constitute x% of all trades for a given stock in a given month. Note that trade-based percentiles effectively put the same weight onto trades of different sizes. This can be misleading. For example, if there are 99 trades executed in 100-share lot and one large trade executed in 100,000-share lot, then the distribution of trade sizes will mostly concentrate at 100-share level. All trade-weighted percentiles below the $99th$ percentile will be equal to 100 shares. The total trading volume, however, is determined by a large trade.

Since large trades are economically more important, we need to investigate the right tail of trade sizes distribution in more detail as well. We therefore also consider the percentiles based on trades' contribution to total volume. The contribution to the total volume by trades from a given trade size bin is calculated based on its midpoint. We refer to these percentiles as volume-weighted percentiles. The xth volume-based percentile corresponds to a trade size such that trades with sizes above this threshold constitute x% of total trading volume. In the previous example, for example, the $99th$ volume-based percentile will correspond to 100,000-share trade. It is worth mentioning that the volume-weighted distributions have an easy interpretation. Essentially, their plots show the actual distribution of trading volume across bins with trades of different sizes.

We report our results for distributions of both types because volume-weighted distributions allow us to focus on economically important trades. Of course, if we know a trade-weighted distribution of trade sizes, then we can easily calculate a volume-weighted distribution as well. For example, we will see that the distribution of trade sizes is close to be log-normal. It can be easily shown that if a random variable $\ln(\tilde{z})$ is normally distributed as $N(\mu, \sigma^2)$ and another random variable $\tilde{y}$ has the $\tilde{z}$-weighted density function of $\tilde{z}$, then the logarithm of this random variable $\ln(\tilde{y})$ is distributed as $N(\mu + \sigma^2, \sigma^2)$. Applying this fact to our situation, we know that if trade size is distributed as a log-normal variable with the density of its logarithm being $N(\mu, \sigma^2)$, then the volume-weighted density is also log-normal with its logarithm being $N(\mu+\sigma^2, \sigma^2)$. Thus, the only difference between these distributions is the shift in the means.

Our monthly data is matched with CRSP data set for the purpose of acquiring share and exchange codes for stocks in the sample. Only common stocks listed on the NYSE (New York Stock Exchange), AMEX (American Stock Exchange) or 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 in that month. Our data is also augmented with the data on the average daily volume (in dollars and in shares), the average price, and historical volatility for each stock and each month. Our final sample includes 1,107,990 stock-month observations. For each 192 months between 1993 and 2008, there are observations on about 5,800 stocks traded.

3.2 Descriptive Statistics

Table 1 provides a description of the data. Panel A reports statistics for the subsample from February 1993 to December 2000. Panel B reports statistics for the subsample from January 2001 to December 2008. We report these statistics separately, because the properties of the data have changed substantially following the decimalization in 2001. Statistics are calculated for all securities in aggregate as well as separately for ten groups of stocks sorted by average daily 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 trading 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 by dollar trading volume. Group 10 contains stocks in the top 5^{th} percentile. It approximately corresponds to the universe of S&P100. 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 the statistical properties of trades and securities 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 deciles to \$181.98 million for the highest volume deciles. 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 volume and volatility, increases from 0.14×0.046 to 181.98×0.033 by about 913.

The average trade size is equal to \$23,598 before 2001, ranging from \$11,428 for low-volume stocks to \$88,450 for for high-volume stocks, corresponding to a decrease from 8% to 0.05%, if considered as a fraction of daily volume. The median is much lower than the mean. This is consistent with the existence of several large orders that make the distribution of trade 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 theory predicts that the distribution of trade size as a fraction of volume, $|Q|/V$, should be similar across stocks. The only difference is that it is shifted downwards by two-third of an increase in a trading activity. Since from lowest to highest decile, 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 traded for low-volume stocks should be shifted downwards relative to high-volume stocks by a factor of $913^{-2/3} = 0.01$. This is roughly similar to the differences in the means and medians between low-volume and high-volume groups.

The average number of trades per day is 143 for the entire sample and monotonically increases from 16 to 2951 over the volume groups. The actual increase in number of trades is a factor of 2951/16=184.43. The invariance theory predicts that average number of trades should increase by two-third of an increase in a trading activity, i.e., $913^{2/3} = 100$. While this back-of-the envelope calculation suggests that number of trades increases more than our model predicts, potentially reflecting a more intensive order shredding in high-volume groups, the further investigation is certainly warranted.

Trading is a subject to various restrictions. The minimum lot size, for example, is equal to 100 shares. Many orders are therefore executed in 100-share lots. These trades correspond to 16% of all trades executed or 2% of volume traded before 2001. The 100-shares trades represent 14% of trades for low-volume stocks and 25% for high-volume stocks. We see that the 100-share restriction is more binding for high-volume stocks. This happens because stocks with high volume usually have high prices, thus making 100-share threshold more significant in dollar terms. Another interesting observations is a large fraction of 1,000-share trades, especially for low-volume stocks, before 2001. This reflects the requirement for Nasdaq market makers to post quotes for at least 1,000 shares prior to 2001. The data also shows that even lots, corresponding to even-share bins with exact number of shares such as 100, 200, 300 shares etc. traded, account for more than 50% of volume traded and about 80% of trades executed. The prevalence of these trades validates our choice of trade size bins with the even-share trades being placed in separate bins.

Panel B of Table 1 reports the statistical properties of trades and securities for our sample after 2001. The difference between the data before and after 2001 is striking. After 2001, the average daily volume is over \$19 million which is three times larger than before 2001. The average number of trades is 1761 increasing by a factor of 12, and the average trade size is only \$7,945 decreasing by a factor of 3, compared to the earlier sample. The back-of-the-envelope calculation based on the invariance theory suggest that cross-sectional differences in trade frequencies and trade sizes after 2001 can be consistent with this theory as well.

The descriptive statistics shows that order shredding and the minimum lot size became very important after 2001. The 100-share trades constitute, on average, 50% of all trades executed and accounts for 18% of volume traded in the latter sample. The migration of trades to smaller trade sizes continues throughout the period from 2001 to 2008. For example, we observe that 100-share trades represent about 70% of trades and accounts for 35 % of volume in 2008 (unreported). This indicates that order shredding can make it difficult for us to test the invariance theory using the data after 2001.

4 Results

All three invariance models make distinctively different predictions concerning the differences in the distributions of trade sizes and their frequencies across stocks. We use the TAQ dataset to determine which of the models is more reasonable in describing the data. We run our tests both based on trading frequencies and on trade sizes.

4.1 Tests Based on Trading Frequency

Comparison of Three Models. According to the theory of trading game invariance as well as the two alternative models, the number of trades will be constant across stocks if normalized appropriately for differences in trading activity W. Three models differ only in the suggested normalization. The theory of trading game invariance says that the number of trades per day γ has to be normalized with the trading activity W in a power of -2/3,

$$\ln(\gamma \times W^{-2/3}). \tag{7}$$

Alternative theories of invariant bet size and bet frequency propose other adjustments,

$$\ln(\gamma \times W^{0}) \quad \text{and} \quad \ln(\gamma \times W^{-1}). \tag{8}$$

Figure 8 plots the logarithm of average number γ of trades per day, normalized according to three models, against the logarithm of the trading activity W, for each stock traded in April 1993, April 2001, and April 2008. Trading activity W is the product of average daily dollar volume and daily volatility for a given stock in a given month. We consider these three years for robustness because, as we have already mentioned, the trading process has changed significantly over the period under consideration. We also consider separately the NYSE-listed and Nasdaq-listed stocks for April 1993, as these two are also significantly different due to differences in market frictions.

Figure 8 clearly shows that the theory of trading game invariance fits the data very well. Especially for the NYSE stocks traded in April 1993, all observations are lined up across a horizontal line. These pattern become slightly less pronounced after 2001.

When the number of trades is normalized according to the theory of invariant bet frequency, all observations are lined up across a line with a positive slope. It is easy to explain. This model assumes that differences in the trading activity come entirely from differences in trade sizes. In real data, however, changes in trading activity are partially explained by changes in trading frequency. Thus, this model tends to underestimate the number of trades for high-volume stocks and overestimate it for low-volume stocks.

When the number of trades is normalized according to the theory of invariant bet size, the results are the opposite. All observations are lined up along the line with a negative slope. Again, it is easy to explain. This model attributes differences in the trading activity entirely to differences in the trading frequencies. Some fraction of these differences comes, however, from differences in trade sizes. This model therefore tends to overestimate the number of trades for high-volume stocks and underestimate it for low-volume stocks.

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

$$\ln\Big[\gamma\Big] = \alpha + a_\gamma \times \ln\Big[\frac{W_i}{W_*}\Big] + \tilde{\epsilon}. \tag{9}$$

The equation relates the average number of trades γ per day to the level of trading activity W, defined as the product of average daily dollar volume $V_i \times P_i$ and the standard deviation σ_i of daily returns. The scaling constant $W_* = (40)(10^6)(0.02)$ corresponds to the measure of trading activity for benchmark stock with price \$40 per share, trading volume of one million shares per day, and daily volatility of 0.02. In the regression, 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 $a_\gamma = 1$. We run monthly regressions and report the Fama-MacBeth estimates with their Newey-West standard errors computed with 3 lags in Table 2. The first three columns of the table report the results for stocks and the entire sample period as well as for the two subsamples, before and after 2001. The next six columns report the results for the NYSE/AMEX-listed stocks and the NASDAQ-listed stocks.

The estimate of a_γ is equal to 0.74 for the entire sample with the standard errors of 0.011. As we see, it is very close to the value of 2/3 predicted by the invariance theory. The formal statistical test, however, rejects this theory with F-test of 44 and p-value of 0.001, due to very small standard errors. Note that alternative models are rejected with overwhelming margins. The magnitude of the F-tests indicates that if we do a Bayesian analysis, then almost regardless of our priors, we will conclude that the theory of trading game invariance provides the most reasonable explanation of the cross-sectional differences in trading frequencies.

When we break the sample into two subsamples before and after 2001, we observe that the results are better in the first part of the sample. The point estimate of α_γ is equal to 0.69 before 2001 and 0.79 after 2001. This suggests that order shredding has become more significant in the later sample, generating a large number of prints in the transactions data. The increase in coefficients can be most likely attributed to the fact that order shredding is more intensive for high-volume stocks.

Separate Coefficients for Price, Volume, and Volatility. Table 3 estimates the monthly regressions

$$\ln\Big[\gamma\Big] = \alpha + \frac{2}{3}\ln\Big[\frac{W_i}{W_*}\Big] + b_1 \times \ln\Big[\frac{V_i}{(10^6)}\Big] + b_2 \times \ln\Big[\frac{P_i}{(40)}\Big] + b_3 \times \ln\Big[\frac{\sigma_{r,i}}{(0.02)}\Big] + \tilde{\epsilon}. \qquad (10)$$

This regression imposes the restriction that the coefficient $a_0 = 2/3$ as predicted by the model of trading game invariance. It then allows the coefficient on the three components of W_i to vary freely. Thus, the model of trading game invariance predicts $b_1 = b_2 = b_3 = 0$. The model of invariant bet frequency predicts $b_1 = b_2 = b_3 = -2/3$, and the model of invariant bet size predicts $b_1 = b_2 = b_3 = 1/3$. Table 3 reports the Fama-MacBeth estimates and Newey-West standard errors based on monthly regressions.

The table reports that the estimate of $\hat{b}_1 = 0.18$ for the coefficient on volume V_i, the estimate $\hat{b}_2 = -0.3$ for the coefficient on price P_i, and the estimate of $\hat{b}_3 = -0.41$ for the coefficient on volatility $\sigma_{r,i}$. The corresponding standard errors imply that these estimates are significantly different from zero. Thus, the number of trades increases faster with trading volume and slower with dollar volatility than suggested by the invariance theory. Note that although we have here three explanatory variables, the increase in the R-square relative to the univariate regressions has been insignificant.

4.2 Tests Based on Trade Size

Comparison of Three Models. Figure 1 and Figure 2 show the trade-weighted and volume-weighted distributions of normalized trade sizes for the NYSE-listed and the NASDAQ-listed stocks traded in April 1993. Our choice of month April guarantees that trade size distribution figures are not influenced by this seasonality in the market, because trade sizes tend to cluster less before the end of calendar quarter, as showed by Moulten (2005). We present the distributions from year 1993 only for illustrative purposes and closely examine data from 1993 to 2008 later.

Trade sizes are normalized as implied by the three models. These models predict that if trade sizes are adjusted appropriately for differences in the trading activity W, then their distributions will be similar across stocks. The models differ only in the adjustment they suggest. The theory of trading game invariance says that trade size $|Q|/V$ should be normalized with the trading activity W in a power of 2/3,

$$\ln(\frac{\tilde{|Q|}}{V} \times W^{2/3}). \qquad (11)$$

Alternative theories of invariant bet size and bet frequency propose other adjustments,

$$\ln(\frac{\tilde{|Q|}}{V} \times W^{1}) \quad \text{and} \quad \ln(\frac{\tilde{|Q|}}{V} \times W^{0}). \qquad (12)$$

Note that we calculate the trade size $|Q|$ based on the mid-point of its trade size bin. The empirical stock-level distributions of normalized trade sizes are pooled together for April 1993, averaged across stocks in each volume group, and plotted on the figures. The subplots of the trade-weighted distributions have the frequency of normalized trades on the vertical axis. The subplots with the volume-weighted distributions have the volume contribution of these trades on the vertical axis. We also superimpose the normal distribution with the same mean and the same variance equal to the mean and the variance of normalized trade sizes for the entire sample. We superimpose different distributions for trade-based and volume-based distributions. These superimposed distributions make it easy to see the results. Indeed, if the theories are correct, then distributions of normalized trade sizes should be identical across volume groups. They should also coincide with the superimposed normal distributions, if trade sizes are distributed as a log-normal random variable.

Our results reveal that both the trade-weighted and volume-weighted distributions of trades, normalized according to the theory of trading game invariance, seem to be stable across volume groups. The support of these distributions is clearly similar for both low-volume and high-volume stocks. This suggests that the data fit the theory of trading game invariance quite well. These empirical distributions are also quite similar to the superimposed normal distributions, implying that the normalized trade sizes are indeed distributed similarly to a log-normal random variable.

Note that the distributions of trade sizes for the NASDAQ-listed stocks in Figure 2 are less smooth than those for the NYSE-listed stocks in Figure 1. We will see that this happens because of a particular market regulation existing at NASDAQ in the 90s. NASDAQ dealers were restricted to quote prices for at least 1,000 shares. This led to a disproportionably large fraction of 1,000-share trades at NASDAQ. These 1,000-share trades are responsible for the spikes on the graphs for the NASDAQ-listed stocks.

Figure 1 and 2 show that the distributions of trade sizes normalized as suggested by alternative models are not stable across volume groups. The theory of invariant bet frequency, for example, understates the magnitude of trade sizes for high-volume stocks. This happens because the high volume partially stems from high trading frequency, while this theory assumes that all variation in trading activity comes entirely from the variation in trade sizes. The theory of invariant bet size, in contrast, overstates the magnitude of trade sizes for high-volume stocks. It assumes that all variations in trading activity come entirely from variations in trading frequency. We know, however, that it partially comes from variations in trade sizes. To summarize, alternative theories provide much worse explanations for observed variations in trade sizes comparing to the invariance theory.

It is also interesting to examine more closely the parameters of the superimposed normal distributions. For the NYSE-listed stocks, the distribution superimposed on the trade-weighted distributions has the mean of -1.01 and the variance of 1.78; the distribution superimposed on the volume-weighted distributions has the mean of 0.97 and the variance of 2.69. As mentioned in the previous section, when if trade sizes are distributed as log-normal random variables, then both the trade-weighted and volume-weighted distributions should be similar to that of normal random variables with the same variances but different means. These means should be such that the mean of volume-weighted distribution is equal to the mean of the trade-weighted distribution plus its variance. So, if normalized trade sizes are indeed distributed as log-normal random variables, then the volume-weighted distributions should be similar to the trade-weighted ones but shifted upwards. The magnitude of this upward shift should be equal to the variance. We can easily check whether these relations hold in our data. The volume-weighted mean of 0.97 is only slightly higher than the trade-weighted mean of -1.01 plus the variance of 1.78. Also, the variance of 2.69

for the volume-weighted distributions is somewhat higher than the variance of 1.78 for the trade-weighted distributions. We conclude that although the normalized trade sizes do seem to be distributed similarly to log-normal random variables, there are some deviations. These deviations might be consistent with the existence of some market frictions. For example, the log-normal distribution might be truncated at some level from below. When we examine distributions of normalized trade sizes in more detail, we will see that they are, indeed, truncated from below by a minimum lot size restricted to be 100 shares. Truncated trades are small in size and economically insignificant. This truncation thus will significantly affect the trade-weighted distributions but not the volume-weighted distributions, potentially resulting in a higher variance of the latter.

Next, we proceed to the tests how the data on trade sizes fit the model of trading game invariance using a regression analysis framework.

OLS Estimates of Trade Size, February 1993 - December 2000. Table 4 shows the estimates based on the data between February 1993 and December 2000 from the following regression,

$$\ln\left[\frac{\tilde{Q}_i}{V_i}\right] = \ln\left[\bar{q}\right] + a_Q \times \ln\left[\frac{W_i}{W_*}\right] + \tilde{\epsilon}_i, \tag{13}$$

where the left-hand side $\ln\left[\frac{\tilde{Q}_i}{V_i}\right]$ is the mean or the *pth* (20th, 50th and 80th) percentiles of the stock-level distributions of the logarithms of trade sizes as a fraction of daily volume. In particular, for each stock in a given month, we construct the empirical distribution of the logarithms of trade sizes over the logarithms of our trade-size bins. We then calculate various attributes of these empirical distributions. The means and the percentiles are calculated both for the trade-weighted distributions and the volume-weighted distributions. We run these regressions for each month and report the Fama-MacBeth estimates with their Newey-West standard errors. In the regression, the model of trading game invariance predicts that $a_Q = -2/3$, the model of invariant bet frequency predicts that $a_Q = 0$, and the model of invariant bet size predicts that $a_Q = -1$.

Table 4 shows that the estimates of α_Q range from -0.80 for the 20^{th} percentile to -0.74 for the 80^{th} percentile of trade-weighted distributions. Its estimate based on the mean is equal to -0.76. These coefficients are close to -2/3 predicted by the theory of trading game invariance. Although the hypothesis a_Q=-2/3 is rejected, the F-tests suggest that the theory of trading game invariance has a better fit than other two theories for all the percentiles and the means of the trade size distributions. If we apply a Bayesian analysis, we will conclude that, almost regardless of our priors, the theory of trading game invariance is the most probable one.

The estimates of α_Q from the volume-weighted distributions are lower than those from the trade-weighted distributions, ranging between -0.69 to -0.51. We know that the volume-weighted distributions better describe the behavior of the large trades, while the trade-weighted distributions focus more on small trades. It seems that the size of smaller (larger) trades, as a fraction of volume, decrease with the trading activity at a slower (faster) rate than the invariance theory predicts.

There is another interesting property. Although we use only one explanatory variable $\ln W$, the R-square of our regressions is fairly large, ranging between 0.90 and 0.93 for trade-weighted distributions. Similar R-square has been documented, for example, in Brennan and Subrahmanyam (1998) in regressions with multiple explanatory variables such as the market depth, the return volatility, the standard deviation of trading volume, the market capitalization, the number of analysts following stocks, the number of institutional investors holding stocks, the proportion of shares hold by them and others. The similarity of R-squares suggests that these additional variables do not add too much of explanatory power to the regression relative to our measure of trading activity W. Note also that the R-square is much lower for the volume-weighted distributions. One possible explanation for these lower R-squares is that our trade-size bins are much wider for larger trades, for instance, our thresholds are 100, 200, 300, 400, 500 shares for small trades and 100000, 200000, 300000, 400000, 500000 shares for large trades. In our analysis, each trade is assigned with a trade size equal to a mid-point of a corresponding bin. This procedure mechanically adds more noise into the regressions based on the volume-weighted distributions. Also, we expect that only a few number of large trades per month do not allow us to reconstruct a smooth distribution of trades in the right tail.

OLS Estimates of Trade Size, January 2001 - December 2008. Table 5 reports the estimates from the same regressions as in Table 4 but for the sample from January 2001 to December 2008. As before, the estimates of α_Q are close to -2/3 as predicted by the model of trading game invariance. These estimates, however, are somewhat lower than -2/3, indicating that the trade size, as a fraction of trading volume, decreases with increasing trading activity at a faster rate than the invariance theory suggests.

There is one noticeable difference between our results before and after 2001. After 2001, the estimates for the trade-weighted and volume-weighted percentiles are quite similar to each other. We believe that that the decimalization in 2001 and technological improvements in trading that allowed algorithmic traders to execute a large number of trades over a fairly short period of time, have resulted in a substantial amount of order shredding. Large trades became very infrequent because most of them are shredded now into sequences of 100-share trades. This made trade-weighted and volume-traded distributions more similar to each others after 2001 than before that. Since we expect order shredding to be more significant among high-volume stocks, the realized trade sizes will be lower for high-volume stocks than predicted by the invariance theory. This tilt might be reflected in lower estimates of α_Q after 2001.

Nevertheless, the data on the distributions of trade sizes support the model of trading game invariance and soundly reject assumptions made in alternative models. The reason is that the variation in trading activity in real markets is associated with both variations in trading frequency and trade sizes; neither remains constant as trading activity changes. Our analysis so far also suggests that a number of market frictions might have distorted our results and that the further investigation is warranted.

4.3 Market Frictions

Trade sizes and frequencies are certainly affected by various market frictions such as the minimum lot size, the clustering of trade sizes, the relative tick size, and the propensity for order shredding. These frictions might have distorted our results so far. In this section, we focus exclusively on the theory of trading game invariance. Our goal is to better understand how various market frictions affect observed trade sizes and trade frequencies. With this goal in mind, we examine the distributions of the logarithms of the trade sizes normalized according to the invariance theory as $\ln(\frac{Q}{V} \times W^{2/3})$ over various subsets of stocks.

As before, we consider ten volume groups. We also split our sample into four price volatility groups. In particular, each month we divide all observations into four equally-sized groups based on the variable $P\sigma_r/W^{1/3}$. This variable is equal to the price volatility $P\sigma_r$ normalized according to the theory of trading game invariance for the difference in trading activity W. The normalized price volatility is an important sorting variable because the effects of at least two market frictions are expected to depend on its levels.

First, the normalized price volatility $P\sigma_r/W^{1/3}$ is inversely related to a concept of the relative tick size. Usually, the relative tick size is defined as the tick size (in cents) divided by the price volatility $P\sigma_r$. According to the invariance theory, however, it is more reasonable to define the relative tick size as the tick size (in cents) divided by the price volatility over a trading day equal to H calendar days, i.e.,

$$\text{Relative Tick Size} := \frac{\text{Tick Size}}{P\sigma_r\sqrt{H}} \sim \Big(\frac{P\sigma_r}{W^{1/3}}\Big)^{-1}, \tag{14}$$

where H is related to a speed with which trading games are being played. According to the invariance theory, it is proportional to $W^{2/3}$. The relative tick size is therefore inversely proportional to normalized price volatility.

This clearly shows that the normalized relative tick size is inversely related to normalized price volatility. When the relative tick size is low, stocks can be traded at finer price levels. Their trade sizes are expected to be small and their trading frequencies are expected to be high. Indeed, as an order walks up the limit order book with limit orders placed at a finer grid, more prints of smaller sizes will be generated. Thus, trade size is expected to decrease and trading frequency is expected to increase with the normalized price volatility.

Second, the normalized price volatility $P\sigma_r/W^{1/3}$ is also related to a concept of minimum lot size of 100 shares. It is easy to see, if we calculate the normalized 100-share trade as a fraction of trading volume. Effectively, we can normalize 100-share trade size as suggested by the invariance theory and get,

$$\text{Normalized 100-share Trade} := \frac{100}{V} \times W^{2/3} \sim \Big(\frac{P\sigma_r}{W^{1/3}}\Big). \tag{15}$$

The minimum lot size is proportionally related to the normalizes volatility. When the normalized price volatility is high, the 100-share constraint is expected to be more binding. In other words, for high-volatility stocks, the realized trade sizes will be "too high" and realized trading frequency will be "too low", as many small trades will not be even submitted to the system. As we see, the effect of the 100-share threshold is opposite to the effect of the relative ticks size. We will see that the 100-share effect is more pronounced in the data.

It is interesting to examine the distributions of normalized trade sizes over different volume and volatility groups as well as over different time periods.

Trade-Weighted Distributions for NYSE-listed Stocks, April 1993 Figure 3 shows the trade-weighted distributions of the logarithms of the normalized trade sizes for 10×4

groups sorted by volume and price volatility for the NYSE listed stocks in April 1993. Trade sizes are normalized as suggested by the theory of trading game invariance. To show the composition of trades, we highlight 100-share trades in light grey and 1,000-share trades in dark grey. We also superimpose the normal distribution with the same mean and the same variance, calculated based on the entire sample. If the invariance theory holds and trade sizes are distributed log-normally, then distributions should be identical across stocks and they should coincide with the superimposed normal distribution.

For most subgroups, the distributions are indeed close to the normal one. There is also a clear truncation from below at the 100-share threshold. A visual inspection suggests that holding the price volatility constant, the support of the distributions stays reasonably constant across volume groups. Holding the volume constant, however, the distributions change across volatility groups. When price volatility increases, the 100-share trades, shown in light grey, shift to the right and the number of average trades decreases. This indicated that the 100-share constraint is becoming more binding and small orders are not even being placed into the system. As for the tick size effect, when volatility increases, the relative tick size decreases making trade sizes smaller and trade frequency greater. We do not observe, however, these patterns. Obviously, the 100-share effect dominates the effect of the relative tick size. Note also that 100-share trades are placed close to each other on the charts, usually located in one or two adjacent columns. It means that there is not a lot of variation in the measure of trading activity W within the groups. The only exception is the first volume group, where the variation in W is quite significant and the 100-share trades are spread over more than four columns.

There is another issue which requires further investigation. The trade sizes seem to deviate from the log-normal distribution when price volatility is low. In this case, there are too many 100-share trades and trade sizes are lower than suggested by the invariance theory. Otherwise, the distribution of the logarithms of the normalized trade sizes closely fit the truncated normal distribution.

Volume-Weighted Distributions for NYSE-listed Stocks, April 1993 Figure 4 is similar to Figure 3 but it shows the volume-weighted distributions of the logarithms of the normalized trade sizes for the NYSE-listed stocks in April 1993. The volume-weighted distributions put more weight onto large trades and allow us to examine in more detail the right tail of the distributions. These charts have an intuitive interpretation. They represent the distribution of trading volume across trade size bins.

We see that for most subgroups, the distributions of trade sizes are much closer to the imposed normal distribution for the volume-weighted distribution rather than for the trade-weighted distributions. There is a simple explanation. The distortions related to the minimum lot size have almost no effect on the volume-weighted distributions. Numerous 100-share trades contribute a little to the overall volume traded. The 100-share trades almost disappear from the charts and the 100-share constraint thus becomes invisible.

The volume-weighted distributions, by and large, provide a supportive evidence for the invariance theory, but there are a number of caveats. For low-volume stocks, large trades are somewhat smaller than they should be according to the invariance theory. There is also a few of really large orders in the right tail of these distributions.

Trade-Weighted Distributions for NASDAQ-listed Stocks, April 1993 Figure 5 is similar to Figure 3 but it shows the trade-weighted distributions of the logarithms of the normalized trade sizes for the NASDAQ-listed stocks in April 1993.

The first thing to notice is a large fraction of 1,000-share trades, shown in dark grey. The clustering of these trades certainly distorts the distribution of trade sizes for NASDAQ-listed stocks, creating spikes and potentially contaminating the results of our analysis. A number of 1,000-share trades is especially significant in high-volume groups. Note that a disproportionably large number of 1,000-share trades is not observed after 2001 (unreported). This distortion most likely corresponds to the restriction on the minimum quotation sizes existing at NASDAQ since mid-1988. The Securities and Exchange Commission made it mandatory for market makers to have the quotation size of at least 1,000 shares. This rule affected mostly large stocks. For small stocks, the rule was slightly different. After 1996, this restriction has been gradually removed, first for a subset of securities and then for all securities traded at NASDAQ. We expect that this artificial market friction is the reason why our results for the NASDAQ stocks are worse than those for the NYSE stocks before 2001.

Apart of the 1,000-share trades, the empirical distributions of trade sizes are reasonably close to the superimposed normal distribution. This is true especially for low-volume stocks with many stocks placed in those groups. For high-volume groups, the number of stocks decreases significantly and the empirical distributions are not very smooth.

Trade-Weighted Distributions for All Stocks, April 2001 and 2008. Figure 6 shows the trade-weighted distributions of the logarithms of the normalized trade sizes for 10×4 groups sorted by volume and price volatility for stocks traded in April 2001. Figure 7 shows the same distributions for stocks traded in April 2008. One striking pattern clearly appears in the figures. Starting 2001, the distributions of trade sizes is becoming dominated by 100-share trades, shown in light grey on the plots. The block-order market disappears.

The NYSE implemented a new pricing scheme on January 29, 2001, reducing the tick size from 1/16 to 1/100. The NASDAQ started to use the decimal pricing on April 9, 2001. After decimalization and a consequent introduction of electronic interfaces, order shredding has become prevailing. Large trades are now broken into numerous small trades. It is not infrequent to see a million-share trade being shredded into a sequence of 100-share trades. In 2008, for example, the 100-share trades constitute about 70% of all trades executed and 35% of the volume traded. Both figures clearly show these changes.

In our sample, the trade size has decreased significantly over time. The distributions of normalized trades were centered around -1.01 and -0.18 for the NYSE and the NASDAQ-listed stocks in April 1993 (Figures 3 and 5). The average trade size decreases from -1.31 in April 2001 to -2.66 in April 2008. The trading frequency, in contrast, has exploded. For high-volume low-volatility stocks, for example, the average number of trades has increased from 938 trades per month in April 1993 to 74,420 trades per month in April 2008.

The extent of order shredding makes it difficult to test the invariance theories using the TAQ dataset after the decimalization of 2001, unless one makes particular assumptions about the order shredding algorithms and how they depend on the trading activity W. Note that the invariance theory is formulated for intended orders or "ideas", independently arriving to the market. The TAQ dataset, however, contains prints. Thus, we effectively make an assumption that ideas generate the same number of prints, regardless of trading activity W. This assumption may not hold after 2001. Orders in actively-traded stocks may be shredded into more traders comparing to orders in actively-traded stocks.

Dynamics of Trading Patterns for Small and Large Stocks, February 1993 - December 2008. Figure 10 shows how the trading process changed over time from a different angle. It shows the dynamics of the normalized number of trades per month and the distribution of normalized trade sizes over time between 1993 and 2008. As before, we normalize number of trades and their size according to the theory of trading game invariance. The figure shows the number of trades as well as the 20^{th}, 50^{th} and 80^{th} percentiles for trade-weighted and volume-weighted distributions. High-volume and low-volume stocks are examined separately.

We see that the normalized numbers of trades have increased and the normalized trade sizes have decreased significantly over time, even after adjustment for the invariance theory. For the high-volume stocks stocks, the normalized trade size distributions were stable until 2001 and then shifted downwards. For the low-volume stocks, the normalized trade sizes are decreasing gradually after 1997. These changes probably represent the impact of the reduction in tick sizes. If we recollect the history, both NYSE and NASDAQ went from 1/16 quotations to the decimal pricing in 2001. Earlier in 1997, NASDAQ had also announced the change from 1/8 to 1/16. Most of NASDAQ-listed stocks are placed into the low-volume group. These earlier changes at NASDAQ might therefore explain the decrease in trade sizes for low-volume group prior to 2001. Interestingly, with the exception of the volume-weighted distributions for large stocks, the effects of changes seem to start stabilizing in 2007 and 2008.

The effect of order shredding is certainly not uniform across stocks. Figure 10 shows that the number of trades has increased more significantly for high-volume stocks than for low-volume stocks. The distributions of trade sizes have changed in a different manner as well. For high-volume stocks, the distribution of trade sizes is somewhat tighter after 2001. For low-volume stocks, the distribution is stable and even widen for larger trades, as shown by the volume-weighted distributions. Comparing the trade size percentiles of low-volume and high-volume stocks, we observe that the latter have slightly larger trade sizes in trade-weighted distributions but smaller trade sizes in volume-weighted distributions. Order shredding might influence more significantly trading in high-volume stocks.

Dynamics of OLS Estimates, February 1993 - December 2008. Figure 10 shows the dynamics of coefficients from monthly regressions for trading frequency (9) and for trade sizes (13). The coefficients predicted by the theory of trading game invariance are superimposed on the plots.

Panel A shows that the estimates of coefficients from (9) are not only close to the predicted $a_\gamma = 2/3$ on average but also for each month between 1993 and 2008. These coefficients are especially close to the predicted value before 2001 and they are slightly higher after 2001. This increase indicates that order shredding, which started to prevail after decimalization in 2001, is more intensive in high-volume stocks, probably because algorithmic trading is concentrated primarily in these stocks.

Panel B and C show the estimates of coefficients from (13) for the percentiles and means of the stock-level distributions of trade sizes. Again, we see that each month the estimates are quite close to $a_Q = -2/3$ predicted by the invariance theory. Before 2001, the coefficients tend to be closer to the predicted ones. After 2001, order shredding seems to contaminate the transaction data making the coefficients deviate from the values predicted by the invariance theory. In the recent period, the volume-weighted percentiles start behaving more similar to the trade-weighted percentiles, as the market for block traders has been disappearing.

5 Conclusions

We employ the TAQ dataset to test the theory of trading game invariance introduced by Kyle and Obizhaeva (2010). The theory puts forward the idea that financial markets have a particular structure. Securities are traded in such a way that trading games played by traders are the same across stocks. The only difference between these games is the speed with which these games are being played. In other words, the time may run at a different pace for different stocks. When trading games, equivalent in a trading time, are considered in a calender time, several specific cross-sectional patterns naturally appear. For example, if one stock has the trading activity that is one percent higher than another stock, then the invariance theory makes two predictions. Its trade sizes, as a fraction of daily volume, should be two-third of one percent smaller and its number of trades per day should be two-third of one percent larger.

We test these predictions using the data on (unsigned) trades in the TAQ dataset from 1993 to 2008. Our tests based on trading frequencies show that the estimated coefficient 0.69 is remarkable close to the predicted value of 2/3, especially before 2001. After decimalization of 2001 and a consequent spread of algorithmic trading, the coefficient is slightly higher, possibly reflecting a more intensive order shredding in high-volume securities. Our tests based on the distribution of trade sizes also provide the evidence in favor of the invariance theory. We find, for example, that the distributions of trade sizes, normalized according to this theory, are quite stable across stocks and that these distributions are similar to a log-normal distribution truncated from below at the level of 100 shares.

Why order sizes are log-normally distributed is an interesting question for the future research. There may be several explanations. Traders may be symmetric and each of them may be drawing quantities to trade from the same log-normal distribution. Alternatively, each trader may have a natural trade size, related to his own size, if for example large hedge funds submit large orders and small retail investors submit small orders. Although the distribution of trades may look normal if conditioning on the type of a trader, the variation in sizes of traders themselves may turn this distribution into a log-normal one.

There are several other issues that require further investigation. The invariance theory is formulated in terms of “bets” or “ideas” arriving to the market independently. We do not observe these ideas in the TAQ dataset, rather we observe realized prints. These prints are influenced by various market frictions such as minimum lot size or relative tick size as well as by order shredding, which became particularly prevailing after recent technological changes in the trading process. The interesting topic for further research, therefore, is how to design better econometric tests dealing with these issues.

Our tests in this paper are based solely on unsigned trades. Thus, the next logical step is to test the invariance theory using the trade that are signed as buys or sells, for example, according to the Lee and Ready (1991) algorithm.

So far, the predictions of the invariance theory about the cross-sectional pattern in trade frequencies and trade sizes have found support in the samples of trades from the portfolio transition dataset and the TAQ dataset. It would be interesting to see whether the predictions concerning quantities and frequencies also hold in the datasets containing changes in holdings of mutual funds or other reporting institutional traders as well as transactions from other markets.

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.

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

Chordia, Tarun, Richard Roll, and Avanidhar Subrahmanyam, 2008, "Recent trends in trading activity", *Working Paper*.

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

Hendershott, Terrence, Charles Jones, and Albert Menkveld, 2010, "Does Algorithmic Trading Improve Liquidity?", *Journal of Finance*, forthcoming.

Kyle, Albert, and Anna Obizhaeva, 2010, "Market microstructure invariants", *Working Paper*, University of Maryland.

Lee, Charles, and Mark Ready, 1991, Inferring trade direction from intraday data, *Journal of Finance* 46, 733–746.

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.

Table 1: Descriptive Statistics.

Volume Groups:	All	1	2	3	4	5	6	7	8	9	10
	Panel A: Before 2001										
Avg. Trade Size, $\lvert\tilde{Q}\rvert$	23,598	11,428	27,348	36,370	43,387	49,167	53,390	60,404	67,649	77,789	88,450
Med. (TW) Trade 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) Trade 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 Trades, 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: %Trades/%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: %Trades/%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: %Trades/%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: After 2001										
Avg. Trade Size, $\lvert\tilde{Q}\rvert$	7,945	4,337	8,323	10,961	13,235	15,666	17,794	19,966	23,218	27,049	34,743
Med. (TW) Trade 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) Trade Size	34,358	29,124	27,884	30,489	36,134	41,724	48,484	56,585	66,787	82,316	118,826
Avg. Number of Trades, 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: %Trades/%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: %Trades/%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: %Trades/%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

Table reports the properties of securities and trades in the two subsamples, before and after 2001. 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 trade size, the trade-weighted median trade size, the volume-weighted median trade size, the average number of trades per day, the daily dollar volume (in thousands of $), the average volatility of daily returns, the average price, the percent of trades in the 100-share lot, the percent of volume in the 100-share lot, the percent of trades in the 1000-share lot, the percent of volume in the 1000-share lot, the percent of trades in the even lots, and the percent of volume 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 common NYSE-listed stocks. Volume group 1 (group 10) has stocks with the lowest (highest) trading volume.

Figure 1: Comparison of Three Models based on Trade Size, NYSE-listed Stocks, April 1993

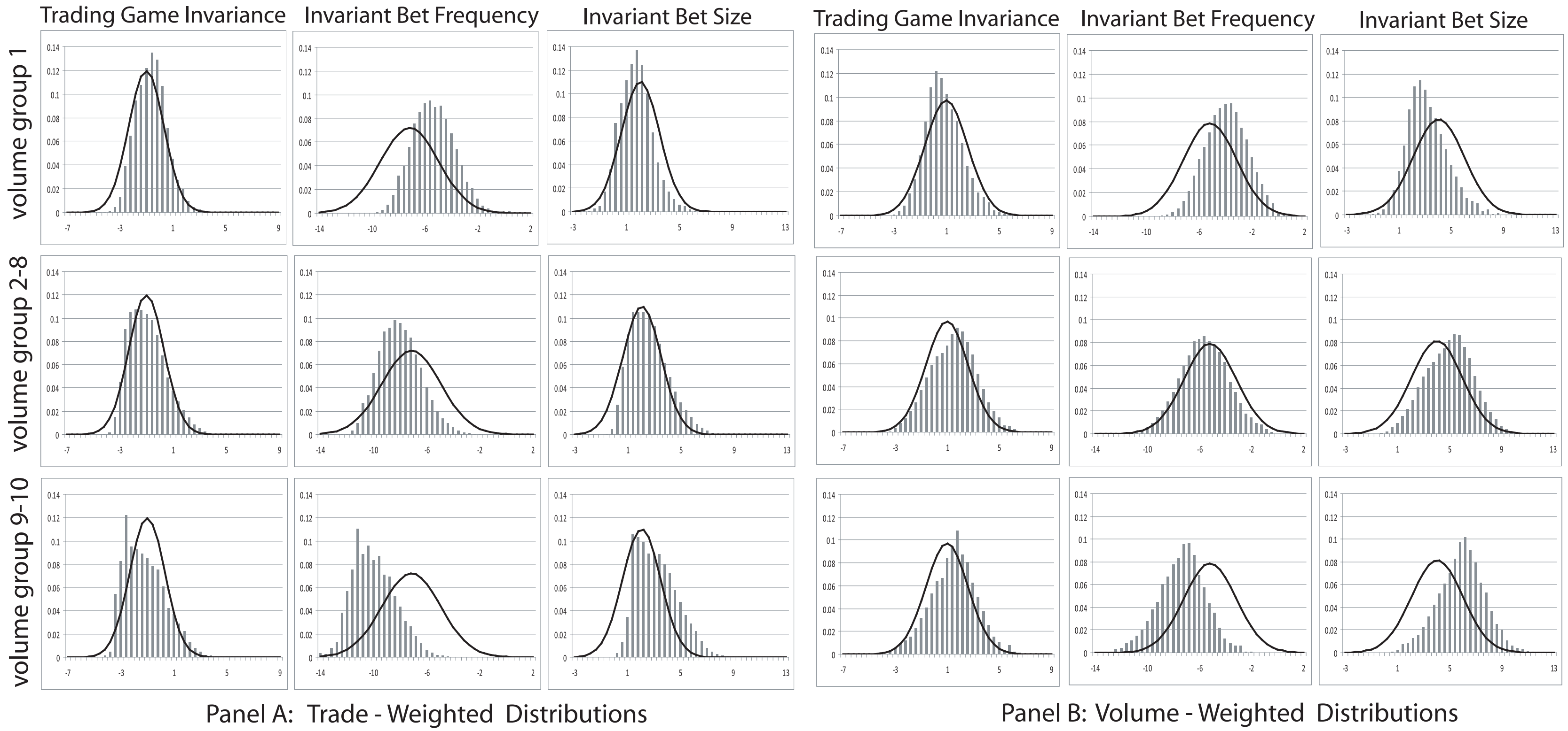

Figure shows the distribution of the logarithm of normalized trade sizes for three different models for the NYSE-listed stocks traded in April 1993. For the model of trading game invariance, the trade sizes are normalized as $|Q|/V \times W^{2/3}$. For the model of invariant bet frequency, the trade sizes are plotted as $|Q|/V$, without any adjustment. For the model of invariant bet size, the trade sizes are normalized as $|Q|/V \times W^1$. Trading activity W is calculated as the product of dollar volume and returns standard deviation. Panel A shows the trade-weighted distributions. Panel B shows the volume-weighted distributions. The stock-level distributions, averaged across stocks for volume groups, are shown on subplots. 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. Volume group 1 (group 10) has stocks with the lowest (highest) trading volume.

Figure 2: Comparison of Three Models based on Trade Size, NASDAQ-listed Stocks, April 1993

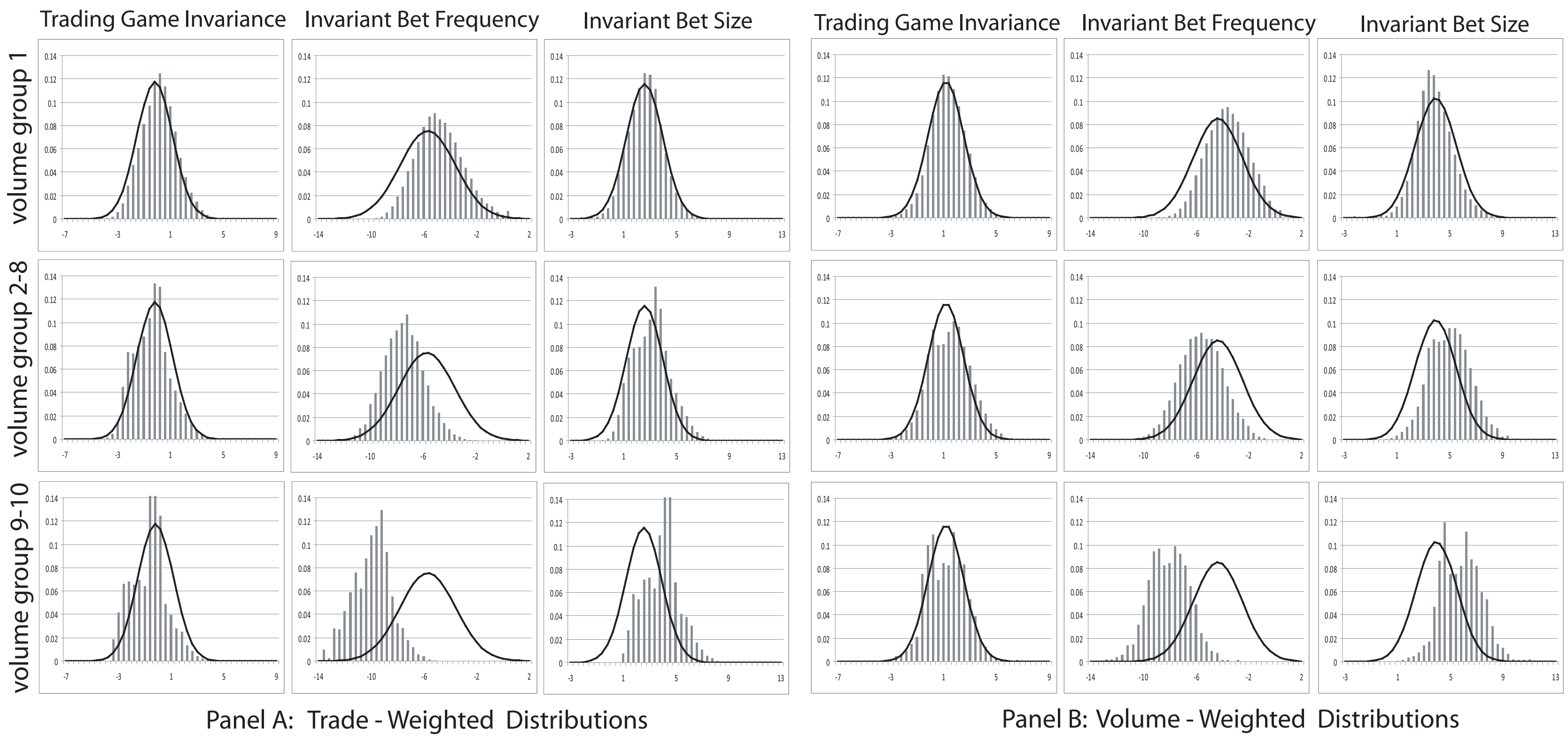

Figure shows the distribution of the logarithm of normalized trade sizes for three different models for the NASDAQ-listed stocks traded in April 1993. For the model of trading game invariance, the trade sizes are normalized as $|Q|/V \times W^{2/3}$. For the model of invariant bet frequency, the trade sizes are plotted as $|Q|/V$, without any adjustment. For the model of invariant bet size, the trade sizes are normalized as $|Q|/V \times W^1$. Trading activity W is calculated as the product of dollar volume and returns standard deviation. Panel A shows the trade-weighted distributions. Panel B shows the volume-weighted distributions. The stock-level distributions, averaged across stocks for volume groups, are shown on subplots. 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. Volume group 1 (group 10) has stocks with the lowest (highest) trading volume.

Figure 3: Trade-Weighted Distributions of Trade Sizes across Volume and Price Groups, NYSE-listed Stocks, April 1993

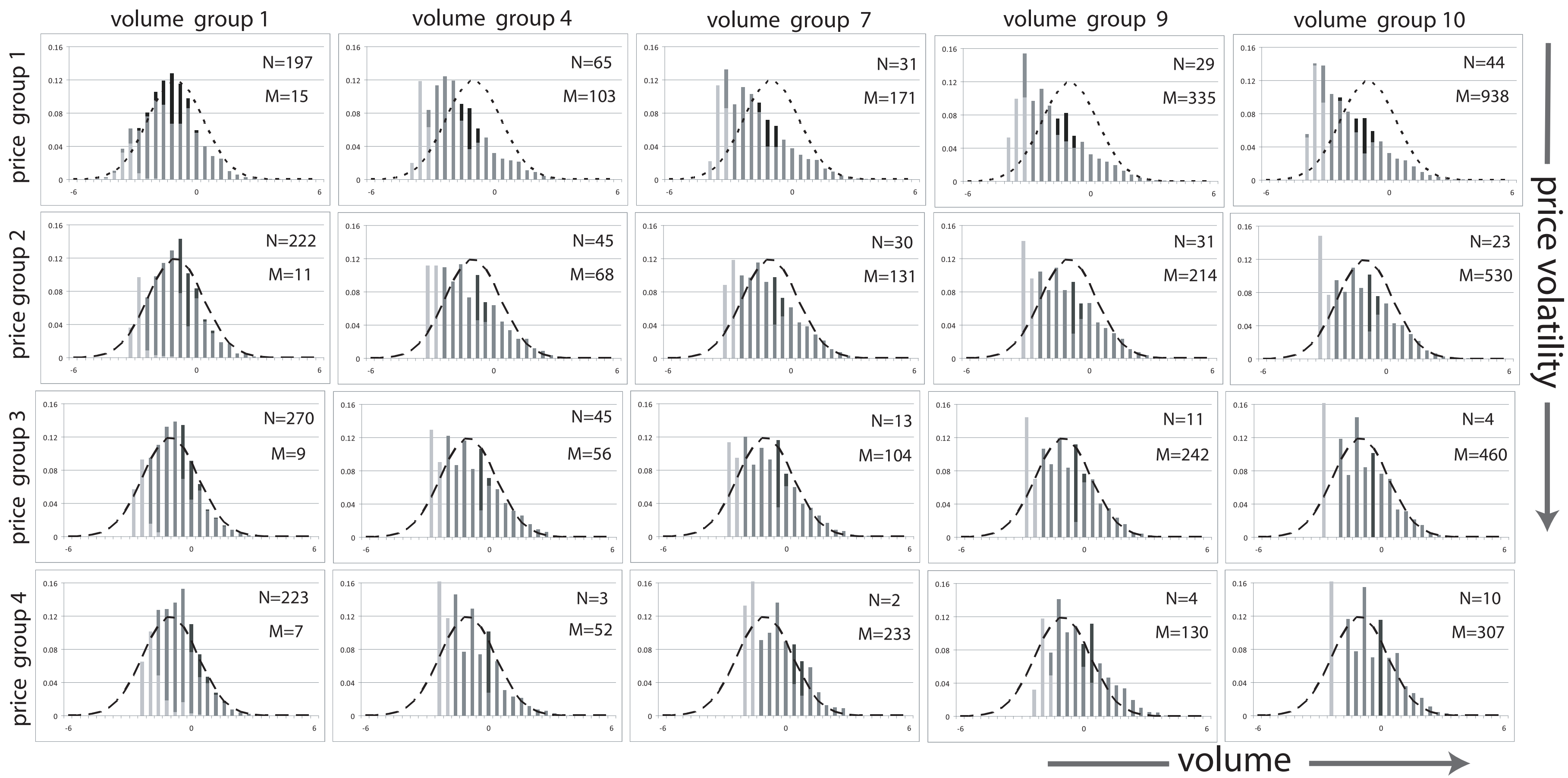

Figure shows distributions of the logarithms of normalized trade sizes for the NYSE stocks in April 1993. For each trade, the normalized trade size is calculated according to the model of trading game invariance, i.e. $\ln(\frac{|Q|}{V} \times W^{2/3})$, where $|Q|$ is a midpoint of a trade size bin in shares, in which a trade locates, 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 stock-level distributions, averaged across stocks for 10 volume and 4 price groups, are shown on subplots. 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. Volume group 1 (group 10) has stocks with the lowest (highest) trading volume. The equally-spaced price volume groups are based $P\sigma_r/W^{1/3}$, which is related to the stringency of minimum 100-share lot size restriction and the relative tick size. Price volume group 1 (group 4) contains the least (most) volatile stocks. The 100-share trades are highlighted in light grey, the 1000-share trades are highlighted in dark grey. On each subplot, the normal distribution with the average trade size mean of -1.01 and averaged standard deviation of 1.33 is imposed. N is the number of stocks and M is the average number of trades per day for these stocks in a given subgroup.

Figure 4: Volume-Weighted Distributions of Trade Sizes across Volume and Price Groups, NYSE-listed Stocks, April 1993

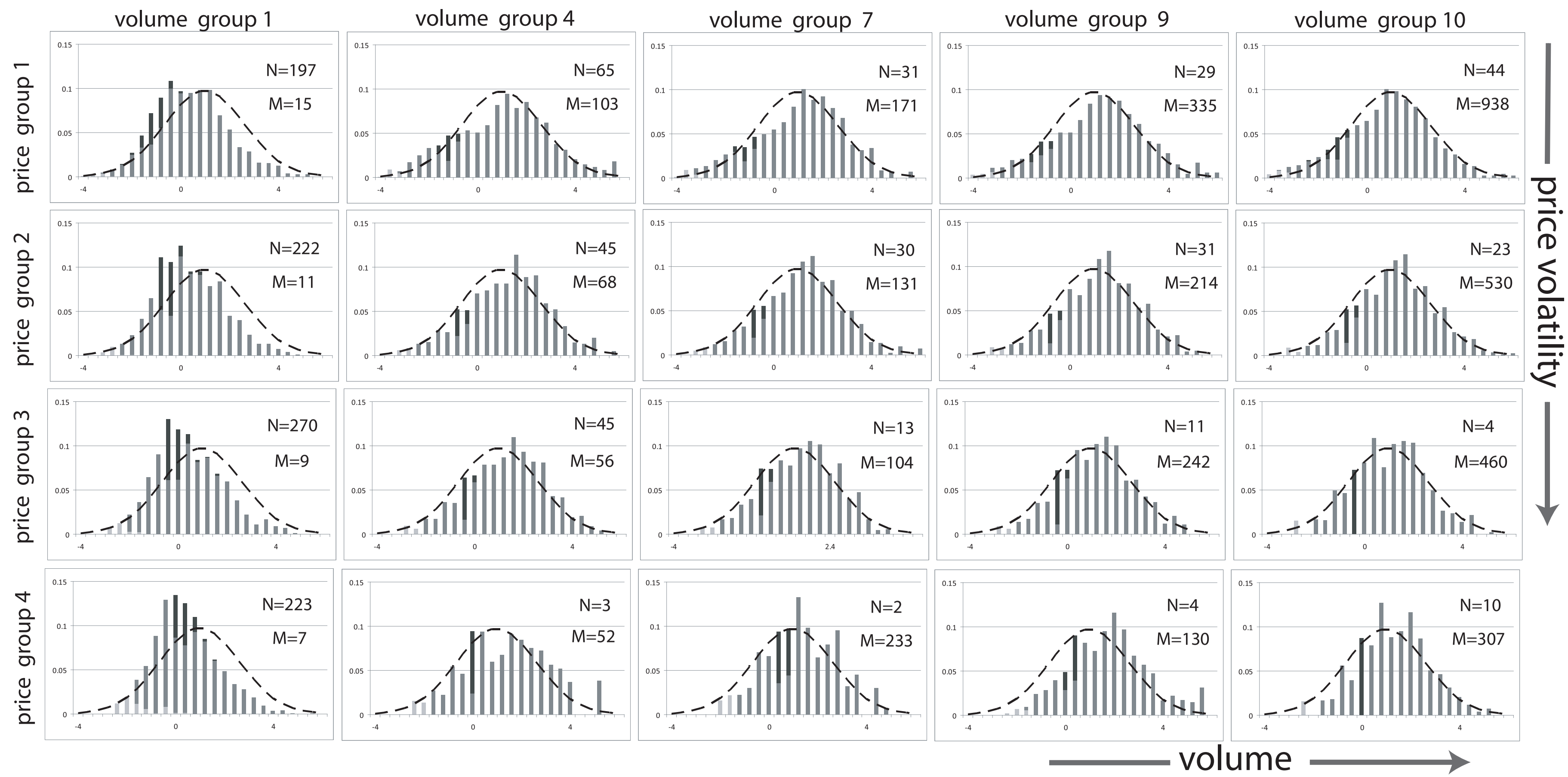

Figure shows distributions of total volume across different normalized trade size bins for the NYSE stocks in April 1993. For each stock, the volume distribution is calculated as the contribution to the total volume by trades from a given trade size bin. On x-axis, the logs of normalized trade sizes are depicted. The adjustment is done according to the model of trading game invariance, i.e. $\ln(\frac{|Q|}{V} \times W^{2/3})$, where $|Q|$ is a trade 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. The stock-level distributions, averaged across stocks for 10 volume and 4 price groups, are shown on subplots. 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. Volume group 1 (group 10) has stocks with the lowest (highest) trading volume. The equally-spaced price volume groups are based $P\sigma_r/W^{1/3}$, which is related to the stringency of minimum 100-share lot size restriction and the relative tick size. Price volume group 1 (group 4) contains the least (most) volatile stocks. The 100-share trades are highlighted in light grey, the 1000-share trades are highlighted in dark grey. On each subplot, the normal distribution with the average trade size mean of 0.97 and averaged standard deviation of 1.64 is imposed. N is the number of stocks and M is the average number of trades per day for these stocks in a given subgroup.

Figure 5: Trade-Weighted Distributions of Trade Sizes across Volume and Price Groups, NASDAQ-listed Stocks, April 1993

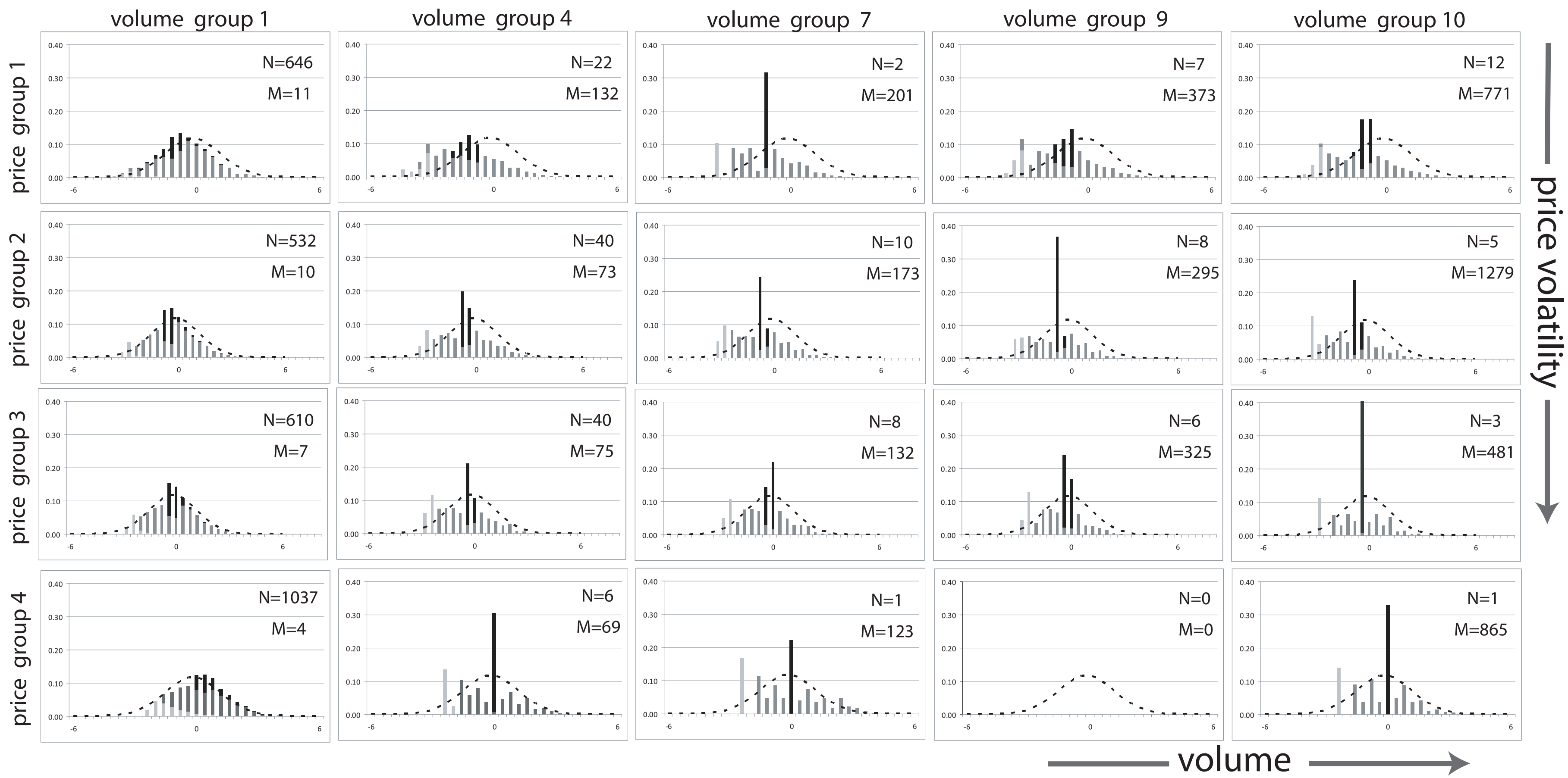

Figure shows distributions of the logarithms of normalized trade sizes for the Nasdaq stocks in April 1993. For each trade, the normalized trade size is calculated according to the model of trading game invariance, i.e. $\ln(\frac{|Q|}{V} \times W^{2/3})$, where $|Q|$ is a midpoint of a trade size bin in shares, in which a trade locates, 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 stock-level distributions, averaged across stocks for 10 volume and 4 price groups, are shown on subplots. 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. Volume group 1 (group 10) has stocks with the lowest (highest) trading volume. The equally-spaced price volume groups are based $P\sigma_r/W^{1/3}$, which is related to the stringency of minimum 100-share lot size restriction and the relative tick size. Price volume group 1 (group 4) contains the least (most) volatile stocks. The 100-share trades are highlighted in light grey, the 1000-share trades are highlighted in dark grey. On each subplot, the normal distribution with the average trade size mean of -0.18 and averaged standard deviation of 1.36 is imposed. N is the number of stocks and M is the average number of trades per day for these stocks in a given subgroup.

Figure 6: Trade-Weighted Distributions of Trade Sizes across Volume and Price Groups, NYSE-listed Stocks, April 2001

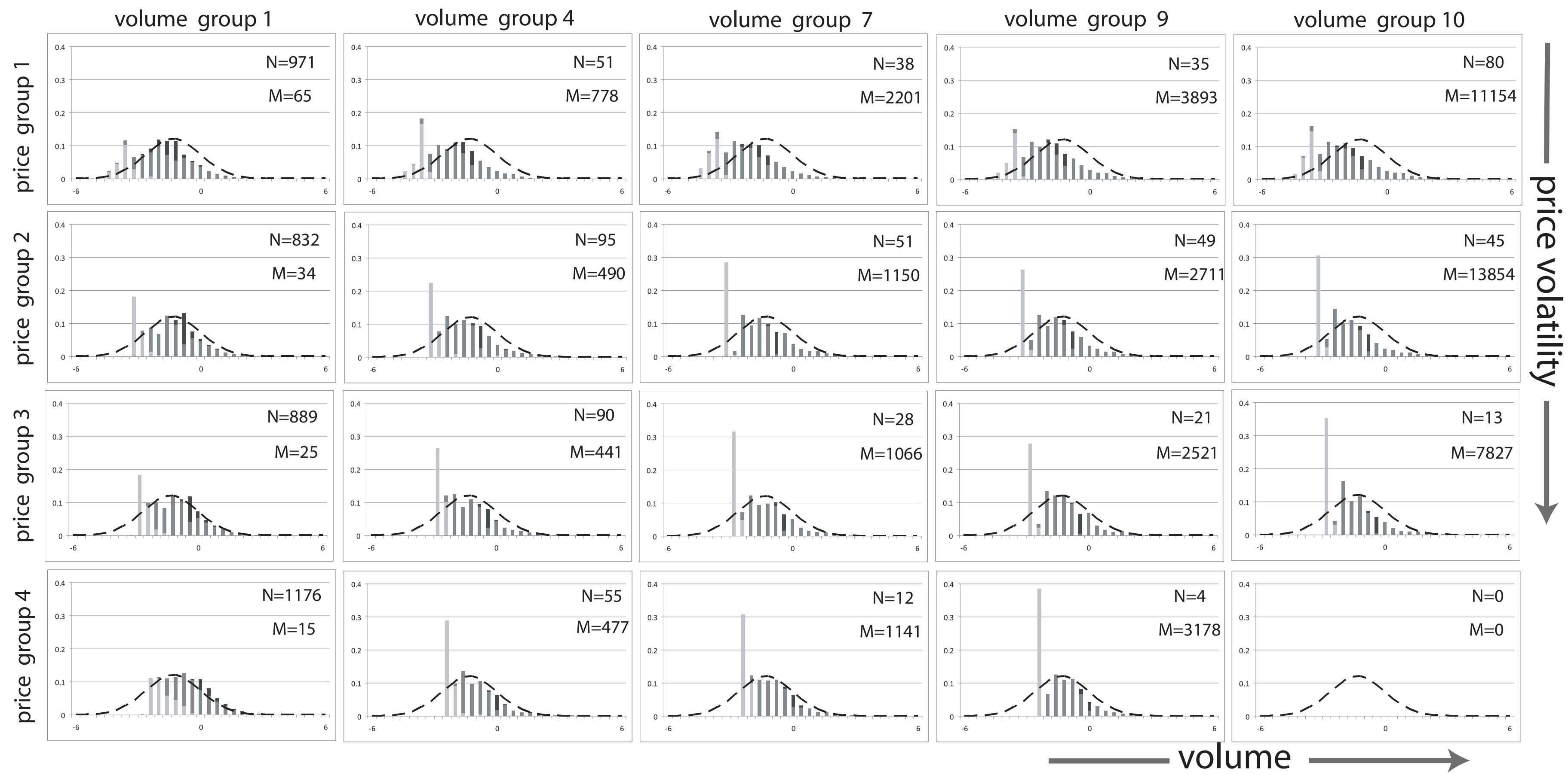

Figure shows distributions of the logarithms of adjusted trade sizes for stocks in April 2001. For each trade, the normalized trade size is calculated according to the model of trading game invariance, i.e. $\ln(\frac{|Q|}{V} \times W^{2/3})$, where $|Q|$ is a midpoint of a trade size bin in shares, in which a trade locates, 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 stock-level distributions, averaged across stocks for 10 volume and 4 price groups, are shown on subplots. 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. Volume group 1 (group 10) has stocks with the lowest (highest) trading volume. The equally-spaced price volume groups are based $P\sigma_r/W^{1/3}$, which is related to the stringency of minimum 100-share lot size restriction and the relative tick size. Price volume group 1 (group 4) contains the least (most) volatile stocks. The 100-share trades are highlighted in light grey, the 1000-share trades are highlighted in dark grey. On each subplot, the normal distribution with the average trade size mean of -1.31 and averaged standard deviation of 1.32 is imposed. N is the number of stocks and M is the average number of trades per day for these stocks in a given subgroup.

Figure 7: Trade-Weighted Distributions of Trade Sizes across Volume and Price Groups, NYSE-listed Stocks, April 2008

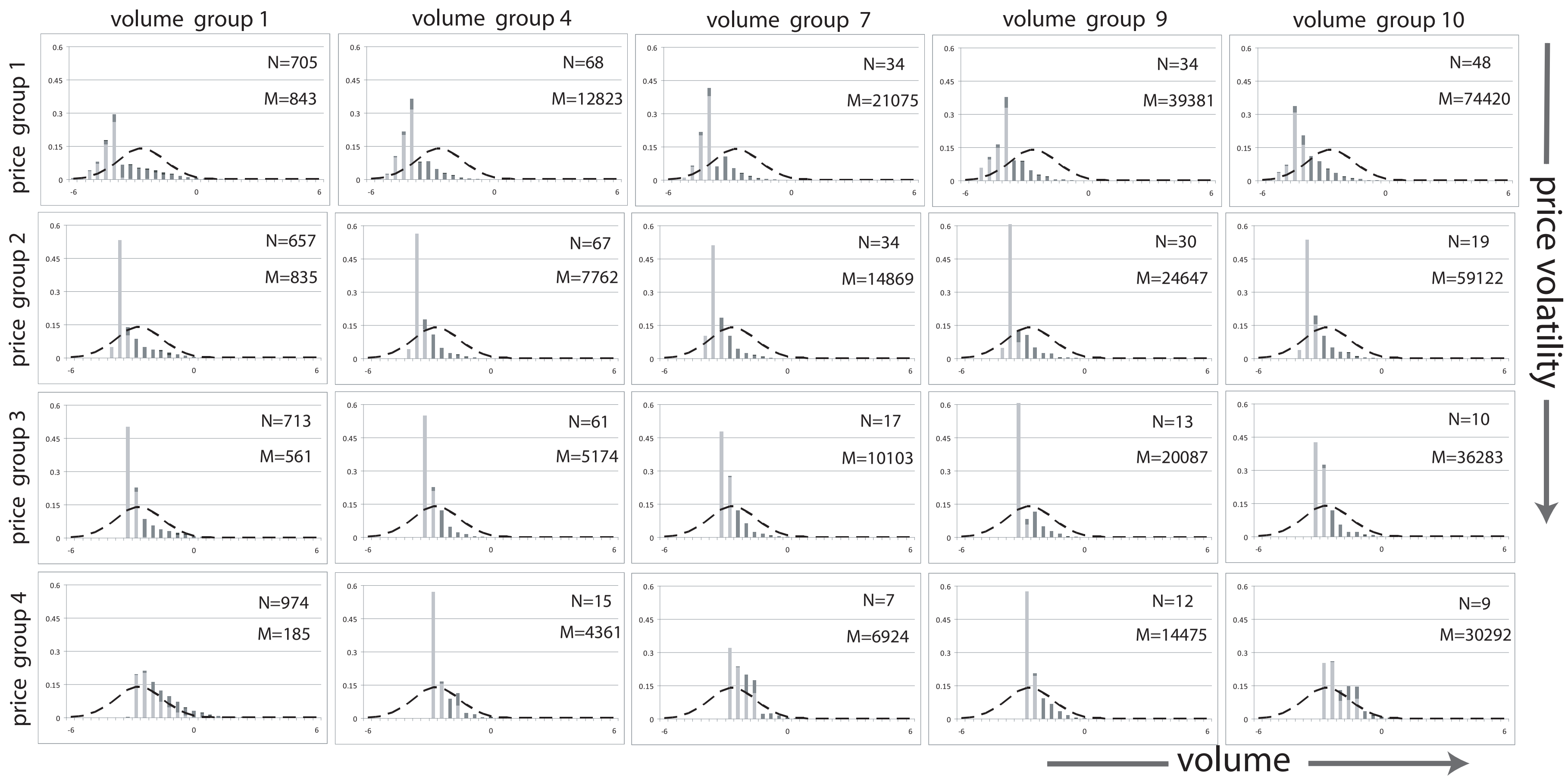

Figure shows distributions of the logarithms of normalized trade sizes for stocks in April 2001. For each trade, the adjusted trade size is calculated according to the model of trading game invariance, i.e. $\ln(\frac{|Q|}{V} \times W^{2/3})$, where $|Q|$ is a midpoint of a trade size bin in shares, in which a trade locates, 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 stock-level distributions, averaged across stocks for 10 volume and 4 price groups, are shown on subplots. 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. Volume group 1 (group 10) has stocks with the lowest (highest) trading volume. The equally-spaced price volume groups are based $P\sigma_r/W^{1/3}$, which is related to the stringency of minimum 100-share lot size restriction and the relative tick size. Price volume group 1 (group 4) contains the least (most) volatile stocks. The 100-share trades are highlighted in light grey, the 1000-share trades are highlighted in dark grey. On each subplot, the normal distribution with the average trade size mean of -2.66 and averaged standard deviation of 1.15 is imposed. N is the number of stocks and M is the average number of trades per day for these stocks in a given subgroup.

Table 2: OLS Estimates of Number of Trades.

	All Stocks			NYSE/AMEX			NASDAQ		
	93/08	93/00	01/08	93/08	93/00	01/08	93/08	93/00	01/08
α	7.10	6.15	8.04	6.95	6.07	7.81	7.24	6.17	8.30
	(0.199)	(0.052)	(0.155)	(0.188)	(0.027)	(0.168)	(0.222)	(0.078)	(0.152)
a_γ	0.74	0.69	0.79	0.70	0.64	0.76	0.77	0.71	0.83
	(0.011)	(0.002)	(0.011)	(0.013)	(0.002)	(0.012)	(0.013)	(0.002)	(0.012)
Adj-R^2	0.92	0.91	0.94	0.94	0.93	0.95	0.92	0.90	0.94
# Obs	5,801	6,698	4,914	2,051	2,199	1,904	3,750	4,499	3,010
	Model of Trading Game Invariance: H_0: $a_\gamma = 2/3$								
F-Test	44.2	92.9	122.4	6.3	260.9	62.4	58.5	336.2	193.0
p-Value	<0.001	<0.001	<0.001	0.0129	<0.001	<0.001	<0.001	<0.001	<0.001
	Model of Invariant Bet Frequency: H_0: $a_\gamma = 0$								
F-Test	4646.2	81391.5	5245.6	2994.2	141943.9	4317.6	3465.8	112044.3	5160.2
p-Value	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001
	Model of Invariant Bet Size: H_0: $a_\gamma = 1$								
F-Test	581.5	16432.1	384.9	556.1	45201.6	441.2	322.6	19561.5	227.4
p-Value	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001

Table presents the Fama-MacBeth estimates α and a_γ from monthly regressions

$$\ln\left[\gamma_{1,i}\right] = \alpha + a_\gamma \times \ln\left[\frac{W_i}{W_*}\right] + \tilde{\epsilon}_i.$$

Each observation corresponds to the stock i with $\gamma_{1,i}$ being the average number of trades per day and the trading activity W_i being the product of the average daily dollar volume $V_i \times P_i$ and the 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 the benchmark stock with price \$40 per share, trading volume of one million shares per day, and daily volatility of 0.02. Adj-R^2 is the adjusted R^2 averaged over monthly regressions. # Obs is the number of stocks averaged over monthly regressions. The Newey-West standard errors computed with 3 lags from the Fama-MacBeth regressions are in parentheses. F-statistics and p-values are calculated from the Fama-MacBeth regressions with Newey-West correction for three different models. The estimates are reported for the entire sample from February 1993 to December 2008 as well as for two subsamples, before and after 2001.

Table 3: OLS Estimates of Number of Trades: Robustness Check.

	All Stocks			NYSE/AMEX			NASDAQ		
	93/08	93/00	01/08	93/08	93/00	01/08	93/08	93/00	01/08
α	2.07	2.51	1.64	3.30	3.94	2.67	1.25	1.98	0.52
	(0.157)	(0.057)	(0.262)	(0.196)	(0.057)	(0.302)	(0.188)	(0.116)	(0.225)
b_1	0.18	0.14	0.23	0.13	0.08	0.18	0.22	0.16	0.28
	(0.011)	(0.003)	(0.014)	(0.013)	(0.003)	(0.016)	(0.014)	(0.007)	(0.013)
b_2	-0.30	-0.35	-0.24	-0.26	-0.34	-0.17	-0.31	-0.34	-0.28
	(0.012)	(0.006)	(0.010)	(0.017)	(0.006)	(0.010)	(0.008)	(0.006)	(0.010)
b_3	-0.41	-0.46	-0.36	-0.45	-0.49	-0.41	-0.44	-0.41	-0.47
	(0.022)	(0.036)	(0.016)	(0.023)	(0.013)	(0.042)	(0.015)	(0.016)	(0.022)
Adj-R^2	0.95	0.94	0.96	0.96	0.95	0.97	0.96	0.94	0.97
# Obs	5,801	6,698	4,914	2,051	2,199	1,904	3,750	4,499	3,010
	Model of Trading Game Invariance: H_0: $b_1 = b_2 = b_3 = 0$								
F-Test	872.8	3072.4	324.1	1448.6	2692.0	504.8	1015.5	5496.5	299.7
p-Value	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001
	Model of Invariant Bet Frequency: H_0: $b_1 = b_2 = b_3 = -2/3$								
F-Test	1287.2	14521.5	1475.1	1245.5	14023.5	1108.9	3324.7	6546.2	3465.7
p-Value	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001
	Model of Invariant Bet Size: H_0: $b_1 = b_2 = b_3 = 1/3$								
F-Test	2013.4	14062.9	1429.0	1226.0	13580.6	1074.2	3272.5	6339.5	3357.3
p-Value	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001

Table presents the Fama-MacBeth estimates α, b_1, b_2 and b_3 from monthly regression

$$\ln\left[\gamma_{1,i}\right] = \alpha + a_\gamma \times \ln\left[\frac{W_i}{W_*}\right] + b_1 \times \ln\left[\frac{V_i}{10^6}\right] + b_2 \times \ln\left[\frac{P_i}{40}\right] + b_3 \times \ln\left[\frac{\sigma_i}{0.02}\right] + \tilde{\epsilon}_i.$$

Each observation corresponds to the stock i with $\gamma_{1,i}$ being the average number of trades per day and the trading activity W_i being the product of the average daily dollar volume $V_i \times P_i$ and the standard deviation σ_i of daily returns in a given month. 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 0.02. Variables V_i, P_i and σ_i are the average trading volume (in shares), average price, and average daily volatility. Adj-R^2 is the adjusted R^2 averaged over monthly regressions. # Obs is the number of stocks averaged over monthly regressions. The Newey-West standard errors computed with 3 lags from the Fama-MacBeth regressions are in parentheses. F-statistics and p-values are calculated from the Fama-MacBeth regressions with Newey-West correction for three different models. The estimates are reported for the entire sample from February 1993 to December 2008 as well as for two subsamples, before and after 2001.

Table 4: OLS Estimates of Trade Sizes, February 1993 - December 2000.

	Trade-Weighted Distribution				Volume-Weighted Distribution			
	Mean	$20th$	$50th$	$80th$	Mean	$20th$	$50th$	$80th$
α	-7.22	-8.47	-7.26	-6.24	-4.72	-6.39	-4.93	-3.38
	(0.033)	(0.039)	(0.045)	(0.036)	(0.073)	(0.053)	(0.084)	(0.087)
a_Q	-0.76	-0.80	-0.76	-0.74	-0.59	-0.69	-0.61	-0.51
	(0.006)	(0.008)	(0.006)	(0.005)	(0.003)	(0.002)	(0.004)	(0.006)
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
	Model of Trading Game Invariance : H_0: $a_Q = -2/3$							
F-Test	254	269	319	226	503	93	180	644
p-Value	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001
	Model of Invariant Bet Frequency : H_0: $a_Q = 0$							
F-Test	17074	10030	19319	21827	30332	97083	21820	7261
p-Value	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001
	Model of Invariant Bet Size: H_0: $a_Q = -1$							
F-Test	91674	50968	102846	120269	220002	584391	151664	63056
p-Value	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001

Table presents the Fama-MacBeth estimates α and a_Q from the monthly regressions of the mean trade size of its percentiles on the measure of trading activity W for the sample from February 1993 to December 2000. The coefficients α and a_Q are based on monthly regressions

$$\ln\Big[\frac{|Q_i|}{V_i}\Big] = \ln\big[\bar{q}\big] + a_Q \times ln\Big[\frac{W_i}{W_*}\Big] + \tilde{\epsilon}_i,$$

where the left-hand side is either the mean or the pth (20th, 50th and 80th) percentile of the distribution of (unsigned) trade sizes $|Q_i|$, as a fraction of daily volume V_i in a given month. The means and percentiles are calculated both based on the distributions of trade size themselves (trade-weighted distribution) and based on the contribution to total trading volume (volume-weighted distribution). Each observation corresponds to the stock i with γ_i being the average daily number of trades and the trading activity W_i being the product of the average daily dollar volume $V_i \times P_i$ and the 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 0.02. Variables V_i, P_i and σ_i are the average trading volume (in shares), average price, and average daily volatility. Adj-R^2 is the adjusted R^2 averaged over monthly regressions. #Obs is the number of stocks averaged over monthly regressions. The Newey-West standard errors computed with 3 lags from the Fama-MacBeth regressions are in parentheses. F-statistics and p-values are calculated from the Fama-MacBeth regressions with Newey-West correction for three different models.

Table 5: OLS Estimates of Trade Sizes, January 2001 - December 2008.

	Trade-Weighted Distribution				Volume-Weighted Distribution			
	Mean	$20th$	$50th$	$80th$	Mean	$20th$	$50th$	$80th$
α	-8.69	-9.42	-8.98	-8.07	-6.82	-8.55	-7.35	-5.56
	(0.103)	(0.044)	(0.111)	(0.162)	(0.079)	(0.188)	(0.242)	(0.237)
a_Q	-0.79	-0.79	-0.79	-0.81	-0.74	-0.80	-0.80	-0.72
	(0.005)	(0.007)	(0.004)	(0.008)	(0.008)	(0.012)	(0.025)	(0.030)
Adj-R^2	0.93	0.90	0.92	0.93	0.86	0.91	0.87	0.77
# bs	4,914	4,914	4,914	4,914	4,914	4,914	4,914	4,914
	Model of Trading Game Invariance : H_0: $a_Q = -2/3$							
F-Test	636	266	935	314	85	118	29	3
p-Value	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	0.086
	Model of Invariant Bet Frequency : H_0: $a_Q = 0$							
F-Test	25116	11432	37000	10574	7994	4286	1056	565
p-Value	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001
	Model of Invariant Bet Size: H_0: $a_Q = -1$							
F-Test	128423	58963	189234	53135	43978	21713	5347	3226
p-Value	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001

Table presents the Fama-MacBeth estimates α and a_Q from the monthly regressions of the mean trade size o its percentiles on the measure of trading activity W for the sample from January 2001 to December 2008. The coefficients α and a_Q are based on monthly regressions

$$\ln\left[\frac{|Q_i|}{V_i}\right] = \ln\left[\bar{q}\right] + a_Q \times \ln\left[\frac{W_i}{W_*}\right] + \tilde{\epsilon}_i,$$

where the left-hand side is either the mean or the pth (20th, 50th and 80th) percentile of the distribution of (unsigned) trade sizes $|Q_i|$, as a fraction of daily volume V_i in a given month. The means and percentiles are calculated both based on the distributions of trade size themselves (trade-weighted distribution) and based on the contribution to total trading volume (volume-weighted distribution). Each observation corresponds to the stock i with γ_i being the average daily number of trades and the trading activity W_i being the product of the average daily dollar volume $V_i \times P_i$ and the 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 0.02. Variables V_i, P_i and σ_i are the average trading volume (in shares), average price, and average daily volatility. Adj-R^2 is the adjusted R^2 averaged over monthly regressions. #Obs is the number of stocks averaged over monthly regressions. The Newey-West standard errors computed with 3 lags from the Fama-MacBeth regressions are in parentheses. F-statistics and p-values are calculated from the Fama-MacBeth regressions with Newey-West correction for three different models.

Figure 8: Comparison of Three Models based on Number of Trades

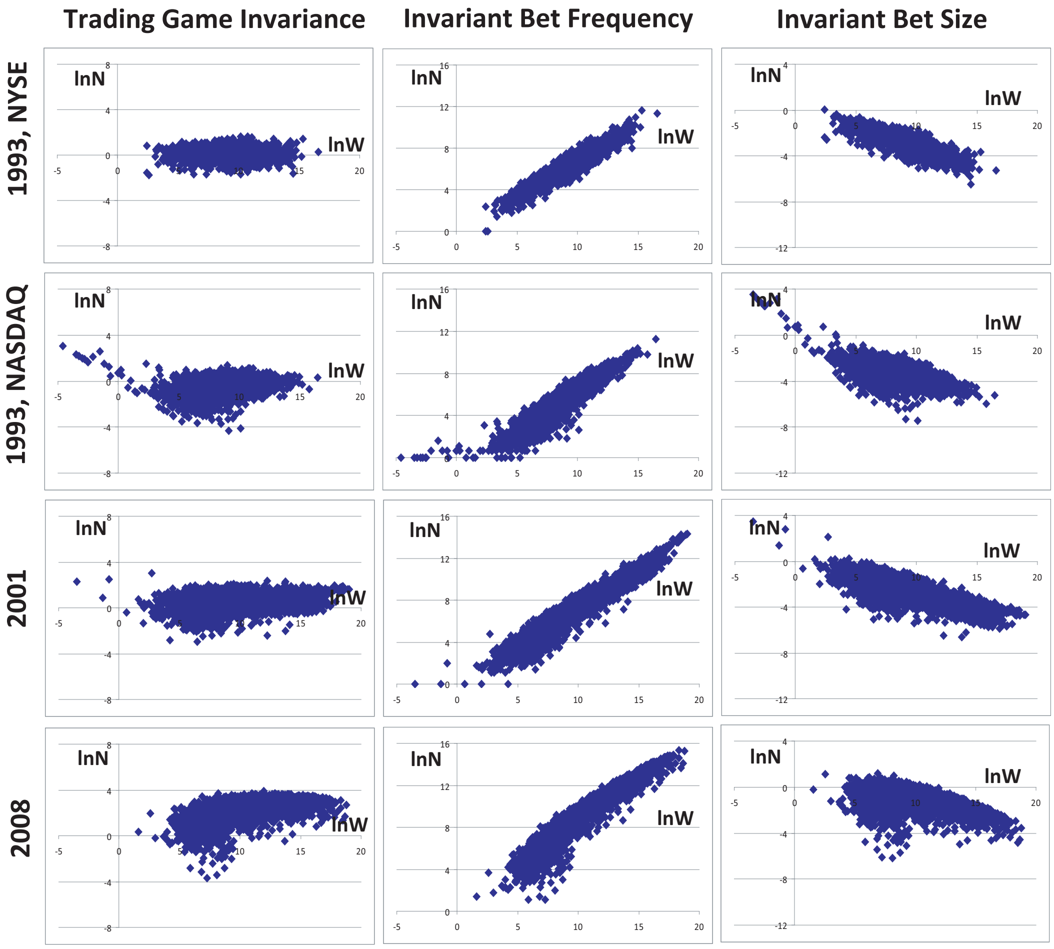

Figure shows the logarithm of normalized number N of trades across different levels of the logarithm of trading activity W. For the model of trading game invariance, the number γ of trades is normalized as $N = \gamma/W^{2/3}$. For the model of invariant bet frequency, the number of trades is shown as $N = \gamma$, without any adjustment. For the model of invariant bet size, the number of trades is normalized as $N = \gamma/W$. Four subsamples are considered: NYSE-listed stocks in April of 1993, Nasdaq-listed stocks in April of 1993, stocks in April of 2001 and in April of 2008. Trading activity W is calculated as the product of average daily dollar volume and daily returns standard deviation in that month.

Figure 9: Trading Patterns for Small and Large Stocks, February 1993 - December 2008.

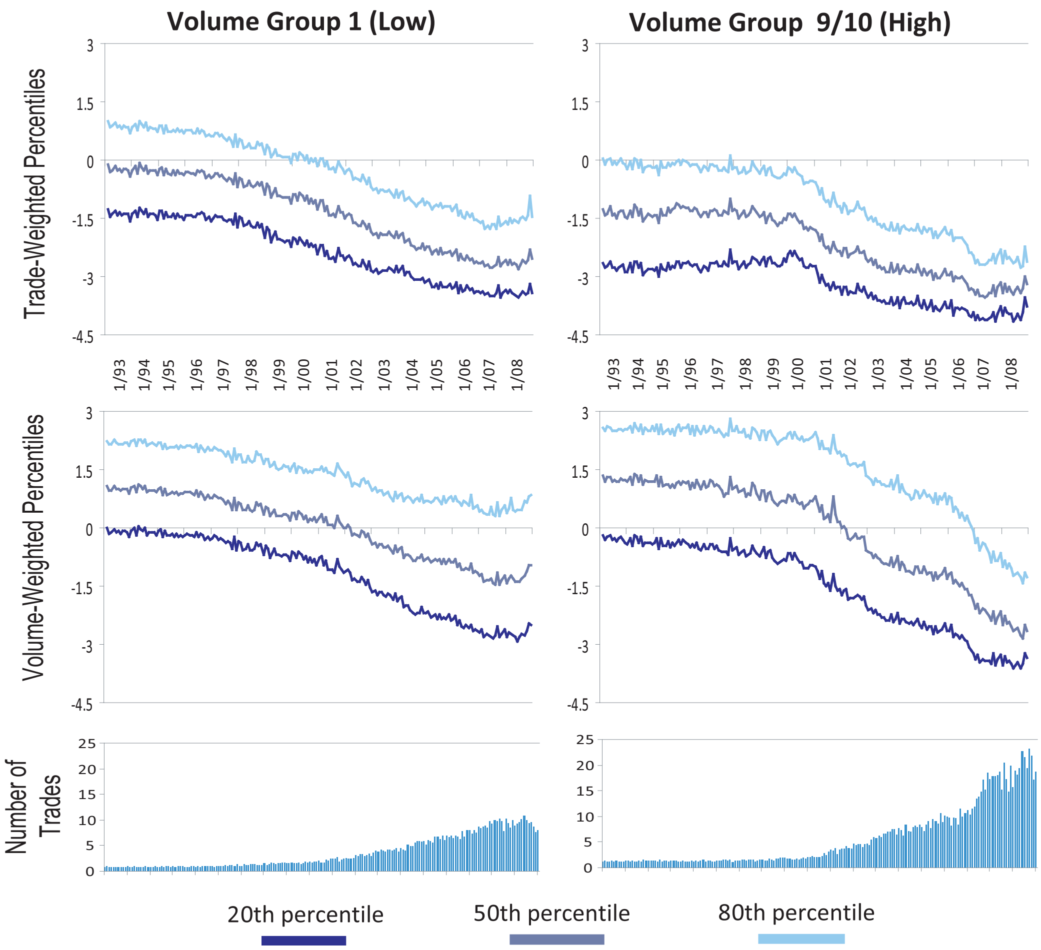

Figure shows the dynamics of average number of trades per month and the 20th, 50th and 80th percentiles for normalized trade size from 1993 to 2008. Trade-weighted percentiles and volume-weighted percentiles are shown for stocks in volume group 1 and volume groups 9 and 10. Volume groups are based on 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. Volume group 1 (group 10) has stocks with the lowest (highest) trading volume. For each trade, the normalized trade size is calculated based on the midpoint of a trade size bin, in which a trade locates, and normalized according to the model of trading game invariance, i.e. $\ln(\frac{|Q|}{V} \times W^{2/3})$, where $|Q|$ is a midpoint of a trade 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 stock-level distributions of normalized trade 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. W is calculated as the product of dollar volume and returns standard deviation. W_* is the measure of trading activity of the benchmark stock.

Figure 10: Dynamics of OLS Estimates, February 1993 - December 2008.

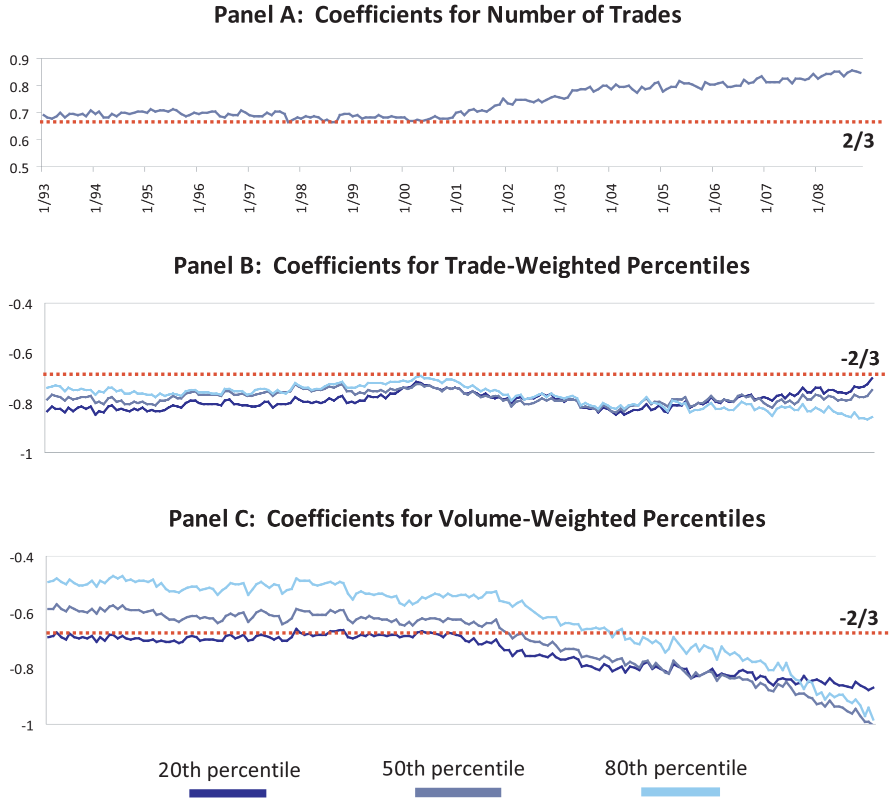

Figure shows the dynamics of coefficients from regressions of number of trades and various percentiles on the measure of trading activity W from 1993 to 2008. Panel A shows the coefficient a_γ from monthly regressions

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

where γ is the number of trades per 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

$$\ln\left[\frac{\tilde{Q}_i}{V_i}\right] = \ln\left[\bar{q}\right] + a_Q \times \ln\left[\frac{W_i}{W_*}\right] + \tilde{\epsilon}_i,$$

where the left-hand side is the *pth* (20th, 50th and 80th) percentiles of the distribution of trade sizes $\tilde{Q}_i$. 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 $a_Q = -1$. W is calculated as the product of dollar volume and returns standard deviation. W_* is the measure of trading activity of the benchmark stock.