What's Not There: The Odd-Lot Bias in TAQ Data[1]

Maureen O'Hara, Cornell University

Chen Yao, University of Illinois

Mao Ye, University of Illinois

November 2011

Abstract

We investigate the systematic bias that arises from the exclusion of trades for less than 100 shares from TAQ data. In our sample, we find that the median number of missing trades per stock is 19%, but for some stocks missing trades are as high as 66% of total transactions. Missing trades are more pervasive for stocks with higher prices, lower liquidity, higher levels of information asymmetry and when volatility is low. We show that odd lot trades contribute 30 % of price discovery and trades of 100 shares contribute another 50 %, consistent with informed traders splitting orders into odd-lots and smaller trade sizes. The truncation of odd-lot trades leads to a significant bias for empirical measures such as order imbalance, challenges the literature using trade size to proxy individual trades, and biases measures of individual sentiment. Because odd-lot trades are more likely to arise from high frequency traders, we argue their exclusion from TAQ and the consolidated tape raises important regulatory issues.

JLE Code: G10; G14.

Key word: TAQ data, Odd-lots, Price Discovery, Transparency, Order Imbalance, Retail Trading

[1] O'Hara's email is mo19@cornell.edu, Yao's email is chenyao2@illinois.edu and Ye's email is maoye@illinois.edu. Comments are welcome. We are grateful to NASDAQ, particularly Frank Hatheway, and the Financial Markets Research Center at Vanderbilt University for providing data. We also thank Jim Angel, Robert Battalio, Kee Chung, Prachi Deuskar, Engelbert Dockner, Larry Glosten, Oliver Hansch, Jeff Harris, Terry Hendershott , Tim Johnson, Terry Hendershott, Christian Laux,Jaehoon Lee, Steward Mayhew, Dmitriy Muravyev, Neil Pearson, George Sofianos, Neal Stoughton, Sunil Wahal, Chunchi Wu, Hao Zhang, Josef Zechner and seminar participant at Goldman Saches, University of Illinois, the State University of New York at Buffalo, Syracuse University and Vienna Graduate School of Finance Hao Zhang for helpful comments and discussions. Mao Ye thanks the Bureau of Economic and Business Research and Research Board at the University of Illinois and Anirban Jana of Pittsburg Supercomputing Center for research support. We are responsible for remaining errors.

What's Not There: The Odd-Lot Bias in TAQ Data

Odd-lots are trades for less than 100 shares of stock. In the market, such trades were generally viewed as irrelevant: odd lot trades and volumes were small, and they were thought to originate from small retail traders and so would have little information content with respect to future price movements. On the NYSE, odd lots even had their own trading system. The convention followed by all market centers was (and still is) that only round-lot trades of 100 shares and mixed-lot trades of greater than 100 shares are reported to the consolidated tape. [2]

Times have changed. The median trade size on the NASDAQ in 2008 and 2009 is 100 shares, dictating that now a large fraction of trades are odd-lots. Algorithmic trading routinely slices and dices orders into smaller pieces, creating a new clientele of odd-lot traders. Allocation protocols for crossing networks can result in odd-lot fills, as can clearing rules associated with particular order types (such as market-at-close orders). And the fact that odd lots are not reported to the tape provides incentives for informed traders to transact via odd-lots rather than use more visible trade sizes. Indeed, the emergence of high-priced stocks such as Google, where trading a round-lot requires an investment of $50,000 or more, has resulted in odd-lots constituting a significant fraction of trade for a subset of important stocks in the market.

Yet, none of this is apparent to researchers using TAQ data because TAQ does not include odd-lot trades.[3] In this paper, we investigate the systematic bias that arises from the exclusion of trades for less than 100 shares from TAQ data. That the main source of transaction data for researchers is truncated for trades below 100 shares raises a variety of important issues

[2] The consolidate tape was established as part of the national market system in 1975. Currently, there are approximately 2.5 million subscribers and it reaches more than 200 million households. The price updates in financial news TV programs, for example, uses the consolidated tape data.

[3] This bias also affects research using SEC Rule 605 market quality statistics as these also do not include odd-lot data.

both for current researchers using TAQ data and for the interpretation of results from papers using past TAQ data.[4] As we demonstrate, the exclusion of odd lot trades is a substantial omission: cross-sectionally in our sample, we find the median number of missing trades per stock is 19%, but for some stocks missing trades are as high as 66% of total transactions. This omission has important implications for microstructure researchers, for asset pricing studies, and for behavioral finance analyses using measures of retail order flow as sentiment indicators.

Our analysis uses a special data set of 120 stocks that was provided to us by NASDAQ. This data set, which was originally intended to facilitate studies of high frequency trading, includes trades, inside quotes, and the order book for the period 2008-2009 and from 02/22/2010 to 02/26/2010. Trades are also identified by trader identity (specifically, whether the buyer or seller are high frequency traders), by trade type (buy or sell) and by which side of the trade was the maker or taker of liquidity. The 120 stocks in the sample were selected to provide a stratified sample of securities representing different market capitalizations and listing venues.[5] A limitation of the data is that it includes only trades executed on the Nasdaq, but unlike the consolidated tape, these data include odd-lot, round-lot and mixed lot trades in regular trading hours.

Our analysis focuses on three main questions. First, how important is odd lot trading across stocks and what determines its incidence? To address this issue, we analyze the trading patterns of odd-lots, the scale of odd-lot trading across stocks, the types of stocks more frequently traded in odd-lots, and the identity of odd-lot traders. Second, what are the

[4] Since 1993, 182 articles published in the Journal of Finance, Journal of Financial Economics and Review of Financial Studies used the TAQ data. The research covers a wide range of fields such as asset pricing (see, e.g. Sadka(2006), Easley, Hvidkjaer and O'Hara (2002), among others), behavioral finance (see, e.g. Malmendier and Shanthikumar (2007), Barber, Odean and Zhu (2009), among others), corporate finance (see, e.g. Nimalendran Ritter and Zhang (2007), Krigman, Shaw and Womack (1999), Chen, Goldstein and Jiang (2007), among others) and real estate (see, e.g. Gentry, Kemsley and Mayer (2003)).

[5] The sample was constructed by Terrence Hendershott and Ryan Riordan, and details on the data can be found in Brogaard (2010).

informational properties of odd-lot trades? Here our analysis calculates the Weighted Price Contribution measure of odd-lot trades and investigates how this contribution differs across trade sizes. Third, how might the exclusion of odd-lot trades bias the results of research? We examine how the odd-lot bias affects microstructure tools such as order imbalance measures and PIN variables; how it affects behavioral finance studies that use the imbalance of small trades as a proxy for individual sentiment; and the impact of missing trades on several widely-used empirical methodologies that rely on the dollar size of trades as a proxy for retail trading.

Our analysis provides a number of new results. We find that missing trades are a large and significant problem for stocks with higher prices, lower liquidity, higher levels of information asymmetry, and when the volatility is low. Fully 34% of all trades in Google, for example, are missing from TAQ data. Almost 24% of trades in small stocks and 18% of trades in large stocks in our sample are odd-lots, resulting in a substantial missing trade bias in TAQ. Moreover, odd-lots have been increasing, growing from approximately 14% of trades for our sample firms in January 2008 to approximately 22% in December 2009. Traders (or algorithms) appear to be splitting trades into odd-lot pieces, motivated perhaps by such trades' absence from the consolidated tape. We also find that odd-lots trades are more likely to be from high frequency traders, evidence suggestive of the new patterns of trading in the market. Our results here contribute to a growing literature on the impact of high frequency trading on markets (see, Hendershott, Jones and Menkveld (2011); Chaboud, Hjalmarsson, Vega and Chiquoine (2009); Hasbrouck and Saar (2010)).

We find strong evidence that these odd-lot trades have large information content. Odd-lot trades now contribute 30 % of price discovery, suggesting that odd-lots are no longer simply the milieu of small, retail traders. We also find that round-lot trades of 100 shares contribute

approximately another 50 %. That 80% of price discovery on NASDAQ is coming from trades of 100 shares or less is consistent with informed traders splitting orders into odd lots and small trade sizes. These results contrast with Barclay and Warner's (1993) and Chakravarty's (2001) finding that medium size trades (500-9999 shares in their sample) were most informative, and complement more recent studies of which trades move prices (e.g., Choe and Hansch (2005)).[6]

We demonstrate that a variety of biases come from these missing trades. One area directly affected by the odd-lot bias is research using order imbalance measures. The microstructure literature uses order imbalances to impute the existence of asymmetric information and to calibrate liquidity effects; asset pricing research has used order imbalances to investigate stock returns, momentum, volatility, and market efficiency; and behavioral finance has used order imbalances to test for disposition effects in trading. Our results show that order imbalance measures are greatly affected by missing trades, with approximately 11% of imbalances incorrectly classified. This bias particularly affects studies using trade-based measures (see Busse and Green (2002); Chan and Fong (2000), Chordia and Subrahmanyam (2002); Chordia, Goyal, and Jegadeesh (2011)), and it is less of a problem for volume-based measures (Hvidkjaer (2006); Chordia, Roll, and Subrahmanyam (2004)) and for PIN models (Easley, Kiefer, O'Haraand Paperman (1996)).

The missing trade problem is particularly acute for studies imputing retail trading behavior and sentiment. Because TAQ data does not provide trader identities, the literature relies on Lee and Radhakrishna's (2000) suggested $5,000 dollar cut-off to identify individual trades (see, for example, Barber, Odean and Zhu (2009) and Lamont and Frazzini (2007)). For stocks

[6] Hansch and Choe (2005) document a dramatic shift of the distribution of informed trades away from medium-sized and into small-sized trades beginning around 1997. See also Campbell Ramadorai and Schwartz (2007) and Chakravarty, Jain, Upson and Wood (2010])for analyses of the information properties of small trades.

above $50, however, this cutoff will effectively remove all individual trades from the data because only odd-lot trades fall below the cut-off level. We provide evidence that, depending upon the time period, up to 15% of all stocks in our sample have zero imputed retail trades because of this cut-off. These stocks, however, carry up to 70% of market value in value-weighted portfolios. Moreover, trade imbalance measures based on these imputed retail trades will also be biased, with errors of more than 20% for both transaction and volume-based measures. These errors will lead to spurious inferences regarding trader sentiment.

This missing data bias should be of concern to all researchers using TAQ data. We also believe it raises important regulatory issues for the U.S. Securities and Exchange Commission (SEC). Recently, the SEC reaffirmed its policy that odd-lot trades would not be reported to the consolidated tape. While this policy may have been sensible in the past, fragmentation, high frequency trading, and the widespread use of algorithms have changed markets in fundamental ways. Our results suggest that odd-lot trades have changed as well, and they now play a new, and far from irrelevant, role in the market.

The paper is organized as follows. Section 1 provides a short history of odd lot trading. Section 2 describes the data; provides summary statistics; gives results on the composition and cross-sectional properties of odd-lot trading; and examines the relationship between missing trades and price, liquidity, and information asymmetry. Section 3 explores the information content of odd-lot trades and computes price discovery measures for trades of different sizes. Section 4 evaluates qualitatively the potential bias arising from missing trades. Section 5 concludes the paper and discusses its policy implications.

1. A Short History of Odd Lot Trading

Odd-lots have undoubtedly existed since the beginning of trading, but their role in modern markets has generally been of limited importance.[7] Starting in 1976, the NYSE formally allowed trading by specialists in odd-lots but required that odd-lots be handled via a separate odd-lot trading system. The rationale for this separate system was to afford customers "an inexpensive and efficient order execution system compatible with the traditional odd-lot investing practices of small, retail customers".[8] The odd-lot system featured different reporting rules in that odd-lot trades were segregated from round lot volume and were not reported to the consolidated tape. The odd-lot trading system also featured different order handling rules, and it essentially required the specialist to price the odd-lot at the price of the next executed round-lot. The ability to get a "better" price in the odd-lot system created incentives for abuse, and over the years the NYSE instituted disciplinary actions against a number of member firms.[9] For the most part, however, odd-lot trading became increasingly less important, and Figure 1 shows that by 1990 it accounted for less than 1 % of NYSE volume (for discussion of the decline of odd-lot trading, see Wu (1972)).

Insert Figure 1 About Here

Because institutions rarely, if ever, traded odd-lots, researchers often used odd-lots as a proxy for individual trades (see, for example, Francis (1986), Lakonishok and Maberly (1990), Ritter (1988), Rozeff (1985) and Dyl and Maberly (1992)). This individual investor linkage was also the basis for a popular theory in technical analysis called odd-lot theory, which was based on the belief that one could outperform the stock market by identifying the least-informed

[7] Odd lots were important from late 1950s to early 1970s. For a review of history of odd lots from 1958 to 1976, please see the lecture by Paul Miranti and Phil Bradford called "Finance Technology and Organization: Automating Odd-Lot trading at the NYSE, 1958-1976" in American Finance Association (AFA) history of finance thought series.
[8] See NYSE (2007) "Odd Lot Order Requirements", Information Memo 07-60.
[9] See "NYSE Moves to Prevent Abuses in Odd-Lot Trades," *Wall Street Journal*, Nov. 14, 2007.

investors and making investments opposite to them. As Malkiel (1981) notes "the odd-lotter is precisely that person …, and [according to this theory] success is assured by buying when the odd-lotter sells and selling when the odd-lotter buys." While apparently popular in the 1960's and 1970's, this theory found little empirical support and so fell out of common use.

More recently, changes in markets have led to changes in odd-lot trading as well. In July 2010, the NYSE decommissioned its separate odd-lot trading system, requiring now that odd-lot orders and trades be handled by the same trading system as all other orders and trades.[10] Some distinctive features to odd-lot trading remain, however, particularly with respect to reporting rules. In particular, odd-lots trades are not reported to the consolidated tape, meaning that an odd-lot trade remains invisible to the broader market. Odd-lot limit orders are also treated differently in the quote montage. An odd-lot order that would better the existing quote is not included in the quote montage, although an odd-lot that adds depth at an existing displayed quote can be included in the reported depth.[11]

2. Data and Analysis

2.1 Data

The data in this paper are from a variety of sources. Information on price, volume, daily volatility and market cap are from CRSP. The two main datasets we use for transactions data are TAQ and the NASDAQ high frequency dataset. The NASDAQ dataset contains trades, inside quotes, and the order book for a sample of 120 U.S. stocks. These stocks were selected to

[10] See Securities and Exchange Commission Release No. 34-62302; File No. SR-NYSE-2010-43 (June 16, 2010) for details on the new order handling and reporting rules for odd-lot trades.

[11] The history of odd lot trading on Nasdaq differs from that of the NYSE in that until 1997 Nasdaq was a quotation system and not an actual trading platform. Quotes could only be made for 100 shares or above, so by definition odd-lots were not quoted on Nasdaq. After 1997, market makers could post quotes on Nasdaq but again there was a minimum quote size of a round lot. Since 2003, market makers and other firms can post orders to Nasdaq but only round lots were reported to the SIP.

provide a stratified sample of securities representing differing market capitalization levels and listing venues.[12] Table 1 provides sample statistics on these firms.

Insert Table 1 About Here

The Nasdaq dataset has a number of unique features. Of particular importance for our study is that it includes all trades (including odd-lot trades) occurring on the NASDAQ exchange during regular trading hours in 2008, 2009 and 02/22-02/26/2010. In our analysis, we focus only on the data from 2008 and 2009. The dataset also identifies the traders involved in each side of the trades as being either high frequency traders or non-high frequency traders. The dataset also signs trades, allowing us to compute trade imbalance measures without resorting to standard trade classification algorithms.

The Nasdaq data has some limitations. The data do not include trades in opening, closing or intraday crosses. It also includes only trades executing on the Nasdaq and not those executing elsewhere in the market. In the past, this would have raised concerns regarding selection bias, but recent research by O'Hara and Ye (2011) shows that competition between market centers has effectively removed this bias in the current fragmented market structure. In particular, markets now trade stocks irrespective of listing locale, and Nasdaq executes a large fraction of trade in both its listed stocks and stocks listed on the NYSE. A maintained, and we believe reasonable, assumption in our analysis is that odd-lot behavior in Nasdaq is typical of that in the larger market.

As a useful preliminary, we compared the TAQ data and the NASDAQ data for our sample period. Because TAQ data includes trades from all the exchanges and trade reporting facilities (TRFs), we first restrict our analysis to the trades executed in NASDAQ. The subsample of NASDAQ trading in TAQ data and NASDAQ high frequency trading data have

[12] Brogaard (2010) shows these stocks are representative of the universe of listed stocks trading in U.S. markets.

the same number of observations for trades greater than or equal to 100 shares when we exclude the trades during market open and close in TAQ data.[13] Where TAQ and the NASDAQ data sets differ is only with respect to odd-lot trades, which are not reflected in TAQ but are included in the NASDAQ dataset.[14] In our analysis, we will use the terms missing trades and odd lot trades interchangeably depending upon context.

2.2 Odd-lot trades and volume: How much is missing?

Figure 2 demonstrates the time series pattern of missing trades and volume in TAQ data. Panel A of Figure 1 shows that in January 2008 about 14% of total trades were odd lot trades and so are missing from the TAQ data, and this number increases to about 20% by the end of 2009. Panel B shows that total missing odd lot volume is about 2.25% of total volume in January 2008 and it is about 4% at the end of 2009. While both the number and volume of odd lot trades are highly variable, both series show a clear increasing trend over time. Other variables, such as stock price levels and liquidity, may also influence the incidence of odd lots, and so the apparent increase in odd-lots may actually be due to these factors rather than any change in odd-lot usage per se. In section 2.4.1. , we will show that the increasing time trend of odd lot is still strong after we control for price, liquidity, volatility and information asymmetry.

Insert Figure 2 About Here

Table 2 gives the level of missing trades for each sample stock, and Figure 3 demonstrates the cross-sectional pattern of these trades. The median level of missing trades is 19.1%. The lowest percentage of odd lot trades in our sample is 8.67% for BZ (Boise,Inc.), and

[13] We have confirmed with NASDAQ that the trades of 100 shares or more in NASDAQ dataset should be the same as the TAQ data. NASDAQ reports round-lot or mixed lot (trades with more than 100 shares but not in the multiple of 100) trades to the consolidated tape and TAQ data is a mere reflection of the consolidated tape.

[14] From 2008 to 2009, there are 920 odd-lots reported to TAQ data under the market center NASDAQ for our sample of 120 stocks. These odd-lots appears either at market open or market close, 915 of them are under the special trade condition of Q or M. The other 5 trades do not have any special condition, but they appear at the first 5 second of market open. (9:30:00-9:30:05).

the highest is 66.5% for KTII. A number of large, well-known firms have substantial numbers of missing trades. Google, for example, has almost 34% missing trades, Apple has 19.3% and Amazon has approximately 22% odd-lot trades. In volume terms, the median level is 6% and the maximum is 28.2%.

Insert Table 2 and Figure 3 About Here

Table 3 presents further detail on the cross-sectional variation of odd-lot trading based on market capitalization and on price level. Institutions are generally thought to trade larger stocks, so odd-lots may be more prevalent in the smaller stocks favored by retail traders. We divided the 120 stocks into 40 large, 40 medium and 40 small market capitalization groups. Panel A of Table 3 shows that this conjecture is correct: odd-lots in the large firm sample are 18% of trades , and this increases to 21.2% of trades for the medium firm sample, and to 23.8% for small firms. The difference between the small and large samples is strongly statistically significant, but we cannot reject the hypothesis that odd-lot trading in the small and medium samples is the same.

Insert Table 3 About Here

Historically, retail traders used odd-lots to purchase small quantities of high-priced stocks, so we would also expect to find a relationship between missing trades and price levels. We divided the 120 stocks into 40 low, 40 medium and 40 high stock price groups. Panel B of Table 3 shows that that this relationship is not monotonic: high-priced stocks are more likely to have missing trades (24.9%), but the percentage of missing trades in low-priced stocks (19.4%) is higher than it is in medium-priced stocks (18.7%). This result suggests that the motivations for odd-lot trades may be more complex than in times past.

The histogram of odd-lot trades in Figure 4 shows a clear pattern of clustering on particular trade sizes. Two facts are particularly salient. First, trades in a multiple of 10 are more

likely than other trades, with 50 shares being the most frequent trade size. Second, trades at 1 share are the second most frequent trade size. That trade clusters at particular price increments has long been observed in equity markets (see for example Harris (1991); Christie and Schultz (1994)). Our finding here that odd-lot quantities also cluster is consistent with traders generally selecting round numbers of shares in which to trade.[15]

Insert Figure 4 About Here

The popularity of the 1-share trade may reflect a number of different effects. Because markets now feature hidden orders, a common strategy to detect hidden liquidity is to "ping" the market by sending a one-share order. Orders successfully finding liquidity will thus execute for one share. Another explanation is less benign. A round-lot trade can be split into smaller trade sizes to escape reporting requirements. Splitting the order into a 99-share trade and a 1-share trade is consistent with this practice, as of course, is splitting orders into other trade sizes. Interestingly, we find that most odd-lot trades below 50 shares fall into the 1-5 share bin, and most odd lot trades above 50 shares fall into the 95-99 share bin.

2.3 Who's trading odd lots?

Another factor influencing the use of odd-lot trading is the rise of high frequency trading. High frequency trading (HF) is now the norm in equity markets, and using this same data set Brogaard (2010) found that high frequency traders were involved in 73% of all trades. HF traders follow a variety of trading strategies, but virtually all of these strategies involve the use of algorithms to send massive numbers of orders to trading venues. The NASDAQ high frequency

[15] For related work on trade clustering in equities see Alexander and Peterson (2007) and in foreign exchange see Moulton (2005).

dataset differentiates traders into two types: high frequency traders and non-high frequency traders. Accordingly, trades in the dataset are categorized into four types: HH stands for high frequency traders take liquidity from high frequency traders; HN: high frequency trader takes liquidity from non-high frequency traders; NH: non-high frequency trader takes liquidity from high frequency trader; and NN: non-high frequency traders take liquidity from non-high frequency traders.

Figure 5 - Panel A provides the ratio of odd lot trades relative to the total number of trades for each trader type. The figure shows that odd lots are more likely to occur when trades are initiated by high frequency traders. About 20-25% of trades of HH and HN type trades are odd-lots. On the other side, odd-lots are least likely when non-high frequency traders take liquidity from high frequency traders. Less than 15% of NH type trades are odd-lots. Panel B demonstrates a similar pattern for volume and the rankings. Order splitting entails additional trading commissions, and so we would expect odd lots to be more common for HF traders who generally face much lower trading costs than retail traders. The results here are consistent with that hypothesis.

Insert Figure 5 About Here

Focusing on these HF traders, we find that many odd-lots appear in a sequence with no round or mixed-lot trades between them. Table 4 gives an example, which happened on June 20, 2008 for trading in Apple (AAPL). At 13:59:01:107, 111 odd-lot trades happened in the same millisecond with the same direction and price, all of which are HN type trades (high frequency traders taking liquidity from non-high frequency traders). The total volume for all these trades is only 2995 shares. Three milliseconds later, we see another 102 odd-lot trades of the HN type

with the same direction and price, which result in volume of 2576 shares.[16] Such patterns are consistent with sophisticated traders (high frequency traders, in this particular case) who are able to slice and dice their orders and hide from the consolidated tape. This also suggests that odd-lot trades may have information content, an issue we address in more detail in the next section.

Insert Table 4 About Here

2.4.1 Which factors determine odd-lot trading?

As an additional diagnostic, we ran between-effect, random and within-effect (fix-effect) regressions on a panel containing information on the percentage of missing trades and missing volume, and daily price, volume and volatility. Between-effect regression allows us to explore cross sectional variation in missing trades and volume. We regress the cross-sectional mean level of missing trades and volume on the price level and the proportional effective spread, which we use as a proxy for liquidity. Daily price range is included to control for volatility. We also include the Probability of Informed Trade (PIN) to consider whether stocks with more information-based trading are more likely to have greater odd-lot trading.[17] Finally, we include the dummy variable NYSE to control for listing venue effects. We use both time and stock subscripts, but because we run between-effect regressions the coefficient is actually defined over the mean of each variable for each stock.

[16] A sequence of odd-lots may also be generated by mechanical reason. For example, suppose the first order of the day is a 50-share buy. After that sell and buy order of 100 shares appears alternatively. Then, the 50 share buy may result in a trade of 50 shares, and the sell order has 50 shares remaining, which may be matched with the next buy order. Therefore, one odd-lot order can generate a large sequence of odd-lot trades. However, odd-lots generated in this way should follow some mechanical pattern, which is not consistent with the example we give.

[17] PIN can be estimated based on all trades and or based on the trades of 100 shares or more. We estimated both measures, and apply the PIN measure using all trades because it is the true PIN measure. Nerveless, missing trades also proposes a challenge for applying PIN measure to the TAQ data and the authors believe that the measure based on volume, the VPIN measure proposed by Easley, Lopez de Prado and O'Hara (2011) might be a better measure of order flow toxicity using the truncated TAQ data, because volume is less affected by the missing trades.

$$\overline{missingntradepct_i} = \mu + \beta_1 * \overline{logpre_i} + \beta_2 * \overline{spread_i} + \beta_3 \overline{pinall_i} + \beta_4 * \overline{range_i} + \beta_5 \overline{NYSE_i} + \overline{\epsilon_i} \quad (1)$$

$$\overline{missingvolpct_i} = \mu + \beta_1 * \overline{logpre_i} + \beta_2 * \overline{spread_i} + \beta_3 \overline{pinall_i} + \beta_4 * \overline{range_i} + \beta_5 \overline{NYSE_i} + \overline{\epsilon_i} \quad (2)$$

The results are given in Table 5. As expected, high-price stocks have more missing trades and missing volumes. Daily price ranges relative to price, a proxy for volatility as well as stock listing venue, have no explanatory power for cross-sectional variation of missing trades and missing volume. The level of liquidity does, however, affect odd-lot trading. We find that the number of missing trades and volume increases in the proportional spread, suggesting that stocks with lower liquidity have greater odd-lot trading. We also find that stocks with higher PINs have higher levels of missing trades. Because odd-lots are not reported to the tape, this latter result is consistent with informed traders breaking their trades into odd-lots so as to better hide their information. The regression R^2 is 64.6%, meaning that about 2/3 of cross-sectional variation of missing volume is explained by these variables.

Insert Table 5 About Here

We also ran equations (3) - (4) using the random effect model. The regression takes the following form:

$$missingntradepct_{i,t} = \mu + \alpha_i + \gamma_t + \beta_1 * logpre_{i,t} + \beta_2 * spread_{i,t} + \beta_3 * pinall_i + \beta_4 * range_{i,t} + \beta_5 NYSE_i + \epsilon_{i,t} \quad (3)$$

$$missingvolpct_{i,t} = \mu + \alpha_i + \gamma_t + \beta_1 * logpre_{i,t} + \beta_2 * spread_{i,t} + \beta_3 * pinall_i + \beta_4 * range_{i,t} + \beta_5 NYSE_i + \epsilon_{i,t} \quad (4)$$

The results are very similar, except that we now find that higher volatility as measured by daily price range results in lower odd-lots trades and volume. Engle, Ferstenberg and Russell (2007)

model the decision to split orders as the trade-off between execution cost and the volatility of execution cost. Breaking trades into small pieces may lead to a lower transaction cost, however, splitting trades across time leads to execution risk because it is hard to predict future transaction costs. This risk is certainly higher when volatility is high, so our results here are consistent with their result.

Finally, we ran the following two regressions using a fixed effect model. Since PIN and listing venue do not vary across time and are captured by the dummy coefficients, they are not included in the following regressions.

$$missingntradepct_{i,t} = \mu + \alpha_i + \gamma_t + \beta_1 * logpre_{i,t} + \beta_2 * spread_{i,t} + \beta_3 * range_{i,t} + \epsilon_{i,t} \quad (5)$$

$$missingvolpct_{i,t} = \mu + \alpha_i + \gamma_t + \beta_1 * logpre_{i,t} + \beta_2 * spread_{i,t} + \beta_3 * range_{i,t} + \epsilon_{i,t} \quad (6)$$

Columns (5) and (6) of Table 5 report the fixed-effect regression of missing volume and trades. The findings are similar: higher price, lower liquidity and low volatility lead to more missing trades and volume.

The other application of fixed effect regression is to see whether there is an increasing trend for odd lots after we control for price, liquidity, and volatility and information asymmetry. Figure 2 demonstrates that odd lot trading increases over time, but the increase in odd lots could be an artifact of other variables. For example, we know that stock prices were low in late 2008 and early 2009 and then bounced back in late 2009. As odd lot trading increases in price, the increase in the level of odd lots may simply be a consequence of an increase in price. To address this issue, we plot the time effect variable in the fixed effect regression. Figure 6 demonstrates a clear trend of an increase in odd lot trading after we control for price, liquidity, and volatility and information asymmetry.

Insert Figure 6 About Here

3. Do Odd-Lot Trades Move Prices?

The results of the previous section suggest that odd-lot trades are now part of a variety of trading strategies, and in particular may have information content for future price movements. There is a large literature looking at the informativeness of stock trades. In general, microstructure research suggests that trades from informed traders will permanently move prices, while trades from uninformed traders will generally have transient price effects. To investigate the informativeness of odd-lot trades, we follow the literature using weighted price contribution (WPC), which measures how much of a stock's cumulative price change or return change is attributable to trades in particular trade-size categories (see, e.g., Barclay and Warner (1993), Chakravarty (2001) Choe and Hansch (2005) and Alexander and Peterson(2007)).

3.1 Weighted Price Contribution

Suppose there are *N* trades for a stock *s* on day *t*, and each trade falls in one of the *J* size categories. Price contribution of the trade belonging to category *j* for stock s on day *t* is defined as:

$$PC_j^{s,t} = \frac{\sum_{n=1}^{N} \delta_{n,j} r_n^{s,t}}{\sum_{n=1}^{N} r_n^{s,t}} \tag{7}$$

$\delta_{n,j}$ is an indicator variable which takes the value of 1 if the *n*-th trade belongs to size category *j*, and zero otherwise. Barclay and Warner (1993) define $r_n^{s,t}$ as the difference between the price of trade *n* and *n-1*, while Choe and Hansch (2005) defines $r_n^{s,t}$ as the log return between the price of trade *n* and *n-1*.

The weighted cross-sectional average price contributions following Barclay and Warner (1993) (henceforth "$WPC_{price\ change}$"), and following Choe and Hansch (2005) (henceforth

"$WPC_{\text{return change}}$") are calculated as follows. The weight for stock *s* on day *t* for the $WPC_{\text{price change}}$ measure is the ratio of its absolute cumulative price change to the sum of all stocks' absolute cumulative price changes on day *t*; the weight for stock *s* on day *t* for $WPC_{\text{return change}}$ measure is the ratio of its absolute cumulative return to sum of all stocks' absolute cumulative returns on day *t.*[18] We weigh each stock's price contribution to mitigate the problem of heteroskedasticity, which may be severe for firms with small cumulative changes. Suppose there are *N* trades for a stock *s* on day *t*, the weight for stock *s* on day *t* is defined as

$$w^{s,t} = \frac{|\sum_{n=1}^{N} r_n^{s,t}|}{\sum_{s=1}^{S}|\sum_{n=1}^{N} r_n^{s,t}|} \tag{8}$$

The WPC of trades in size category *j* on day *t* is defined as

$$WPC_j^t = \sum_{s=1}^{S}\left(w^{s,t} PC_j^{s,t}\right) \tag{9}$$

Suppose there are *T* days in total, the WPC of trades in size category *j* is defined as

$$WPC_j = \sum_{t=1}^{T} WPC_j^t / T \tag{10}$$

Table 6 presents results on price discovery by trade size.[19] Two results are striking. First, more than 80% of price discovery is accounted for by trades of 100 shares or less. Thus, price discovery has shifted away from larger trades and into the smaller trade categories. Indeed, our results show that trades greater than 500 shares now contribute only 0.041 of price discovery. Second, the less-than-100-share trade category is responsible for 26.5% and 31.8% weighted price contribution. Thus, odd-lot trades are clearly informative.

Insert Table 6 About Here

[18] One difference between our WPC measure and the WPC measures by Barclay and Warner (1993), and Choe and Hansch (2005) is that we first find the daily WPC for each size category and then take the arithmetic averages across all days. The difference in approaches arises because our data lacks daily opening and closing trades while they have continuous datasets. Our WPC measure resembles Barclay and Hendershott (2003) in that they measure WPC from close-to-open while we measure WPC from open-to-close.

[19] Market opens are often viewed as times of high information content so we ran our analysis both including and excluding the first 15 minutes of trading. The results are virtually identical we so we report results from the entire trading day.

An interesting question is whether odd lots from particular trader types are more important for price discovery. High frequency traders may use odd lots to hide their information, but HF traders operate on a very short time horizon. Therefore, the information they want to hide is often measured in minutes if not seconds or parts thereof, and so these odd lot trades may not be expected to contribute much to daily price discovery. Moreover, many HF traders act as market maker, providing liquidity to position takers in the market. HH trades are more likely to be inter-dealer trades and so might expect these trades to be less important for daily price discovery. Table 7 presents results on price discovery by trader type. As expected HH trades play the smallest role, while trades involving non-HF participants play the largest role.

Insert Table 7 About Here

3.2 Sources of Cumulative Price Changes: Formal Tests

The stealth trading hypothesis by Barclay and Warner (1993) states that informed traders are concentrated in particular size categories and that price movements are due mainly to informed trader's private information. We follow Barclay and Warner (1993) in testing this stealth trading hypothesis in the context of odd-lot trading. In particular, we investigate whether odd-lot trades have information content, and whether they contribute to the price discovery process. Two alternative hypotheses, the public information hypothesis and the trading volume hypothesis, also address the relation between price contribution and percentage of transactions or total trading volume in each trade-size category.

The public information hypothesis claims the release of public information causes most stock price change. The testable implication of this theory is that the price contribution in a trade size category is proportional to the percentage of trades in that category. The stealth trading hypothesis implies the price contributions would not be proportional.

Regression (1) in Table 8 reports weighted-least-squares regression of the price contribution on two trade-size category dummies and the percentage of transactions in that category. The regression equation is as follows:

$$PC_j^{s,t} = d_{<100}\delta_{<100}^{s,t} + d_{\geq 100}\delta_{\geq 100}^{s,t} + \beta * pct_transcation_j^{s,t} + \epsilon_j^{s,t} \quad (11)$$

$PC_j^{s,t}$ is the price contribution for stock *s* on day *t* of trade size category j. Trades are classified into two categories: less than 100 shares, and equal or greater than 100 shares. $\delta_{<100}^{s,t}$ and $\delta_{\geq 100}^{s,t}$ denote the two indicator variables that take the value one if $PC_j^{s,t}$ falls into their trade categories, and zero otherwise; $d_{<100}$ and $d_{\geq 100}$ represents coefficients for two indicator variables. β is the coefficient for percentage of transactions for stock *s* on day *t* of trade size category j. Regression weight is the ratio of stock *s* absolute cumulative price change on day *t* to the sum of all stocks' absolute cumulative price changes on day *t*.

Insert Table 8 About Here

If the public information hypothesis holds, the coefficient percentage of transactions or percentage of trading volume in that category should equal one and the coefficient of the dummy variable should equal 0. The t-statistics for $\beta = 1$ of 1.98 means that the public information hypothesis can be rejected at 0.047 level of significance. The results also show that the coefficient of less-than-100 trade size is positive significantly differently from zero, while the indicator coefficient of equal-or-greater-to-100 trade size is insignificant. This indicates that missing trades contribute disproportionally to the price discovery process. The hypothesis that the coefficients for the two indicator variables are equal can be rejected at 0.001 level of significance. These transactions-based results are consistence with stealth trading hypothesis.

An alternative trading volume hypothesis states that large trades move stock prices more than small trades. The price contribution in a trade size category is proportional to the percentage

of trading volume in that category. Regression (2) in Table 8 reports weighted-least-squares regression of the price contribution on two trade-size category dummies and the percentage of trading volume in that category. The regression equation is as follows:

$$PC_j^{s,t} = d_{<100}\delta_{<100}^{s,t} + d_{\geq 100}\delta_{\geq 100}^{s,t} + \beta * pct_trdqty_j^{s,t} + \epsilon_j^{s,t} \quad (12)$$

where $PC_j^{s,t}$, $d_{<100}$, $\delta_{<100}^{s,t}$, $d_{\geq 100}$ follow the definitions in the previous regression. β is the coefficient for percentage of trading volume for stock s on day t of trade size category j.

Table 8 indicates that the hypothesis for coefficient of the percentage of trading volume in that category should equal to one can be rejected at 0.001 level of significance. The hypothesis that the coefficients for the two indicator variables are equal can be rejected at 0.0001 level of significance. The volume-based results suggest that odd lot trades are embedded with more private information, again consistent with the stealth trading hypothesis.

4. Why Does It Matter? The Impact of Missing Trades on Empirical Measures

4.1 Order Imbalance

One important application of TAQ data is to calculate order imbalances. The literature uses buy and sell imbalance as a proxy for information asymmetry, price pressure and sentiment of investors. The measure has been used to explain stock returns (Chordia, Roll and Subrahmanyam (2002), Chordia, and Subrahmanyam (2004)), momentum (Hvidkjaer, 2006), herding (Jame and Tong (2010) and Christoffersen and Tang (2009)), disposition effect, (Chordia, Goyal, and Jegadeesh, 2011) and volatility (Chan and Fong, 2000). Busse and Green (2002) use order imbalance to test market efficiency, and Barber, Odean and Zhu (2009) use order imbalance to study whether retail trades move price.

Order imbalance can be measured in three ways. Busse and Green (2002) and Chan and Fong (2000) use the number of buyer-initiated trades minus the number of seller-initiated trades. Hvidkjaer (2006) and Sias (1997) use the volume of trades to define order imbalance. Chordia, Roll and Subrahmanyam (2002) and Chordia and Subrahmanyam (2004) use the dollar volume in addition to the first two definitions.

Missing trades not only affect order imbalance measures quantitatively, but also affect these measures qualitatively. Because of missing trades, we may falsely identify a buy imbalance as a sell imbalance and conversely. If order imbalance is then used as an independent variable in regression analysis, the sign of the coefficient may be reversed.

Table 9 demonstrates the degree of misclassification of order imbalance based on the number of trades (OIBNUM), the number of shares (OIBSH) and the dollar volume (OIBDOL).

Insert Table 9 About Here

We consider a trading day for each stock as one observation. Using our complete data series, we denote the true order imbalance of all trades as true buy imbalance, true balance and true sell imbalance. If we use TAQ data, we only observe trades of 100 shares or more, so we define these measures as observed buy imbalance, observed balance and observed sell imbalance. Because the NASDAQ high frequency dataset provides buy and sell indicators for all trades, we do not need to sign trades using the Lee and Ready (1991) algorithm, which may lead to additional noise in calculating order imbalance. We omit the examination of bias arising from the Lee and Ready (1991) algorithm for this paper and focus instead on the bias of signing order imbalance arising from missing trades.

Order imbalance based on number of trades suffers the most from missing odd-lot trades. Altogether, we observe about 11% misspecification due to missing odd-lot trades. This error

arises from 5.42% of imbalances classified as buys when they are actually sell imbalances or no imbalance. We also find 5.52% of imbalances classified as sells when they are buy imbalances or no imbalance. Finally, there are also days classified as no imbalance when they are actually buy or sell imbalance days (approximately .23%). Chordia, Roll, and Subrahmanyam (2002) recommended using the number of trade imbalance measure for empirical work, but as our results show this is not advisable: the OBINUM measure is seriously biased by missing trades.

Table 9 also shows that using volume-based order imbalance or dollar-volume based order imbalance greatly reduces the misclassification problem. This improvement occurs because while the number of missing trades can be large, the amount of missing volume is often small. Altogether, only 3.33% of order imbalances are misclassified when volume measured are used, which suggests the superiority of volume or dollar–volume based measure for order imbalance measurement. As we show shortly, however, volume-based measures can be biased for other applications.

We also investigate whether the PIN measure, another measure of order imbalance, is affected by the missing trades. The PIN measure uses the daily number of buys and sells to impute the level of information-based trading in a stock. Surprisingly, the last panel of Table 9 shows that PIN estimated through all trades and PIN estimated through trades of 100 shares and above are not statistically different. The reason is because PIN measures order imbalance without direction, where buy imbalance and sell imbalance have equal impact on the PIN measure. Buy and sell imbalance may have different impact on the estimation of positive or negative news for the trading day, but the parameter estimation of positive or negative news does not enter the final formula to calculate PIN. Nevertheless, while the bias may be small, the volume-based PIN, VPIN, (Easley, de Prado and O'Hara (2011)) may be a better measure of order imbalance using

the truncated TAQ data because volume is less affected by missing trades.

4.2 Identification of Retail Traders and Investor Sentiment

Because TAQ data do not reveal a trader's identity, Lee and Radhakrishna (2000) propose a $5000 cut-off value to identify individual (or retail) trades. The method is used extensively in the literature to study individual trader's behavior (see, e.g. Shanthikumar (2004); Barber, Odean and Zhu (2009); Frazzini and Lamont (2006); Jame and Tong (2010); and Christoffersen and Tang (2009)). Although it is easy to see that an increase of price affects this proxy of individual trades, a more subtle problem is that price change affects this proxy in a highly nonlinear way. First, this cut-off value implies that any stock with a price over 50 dollars will have zero individual trades because odd-lots will not be reported. Second, if the stock price fluctuates around 50 dollars, individual trades identified through the $5,000 cut-off fluctuate between 0 and a large positive number. For example, suppose that there are 100 individuals, each trading 100 shares. When the stock price is 50.01, zero individual trades are observed, but when the price falls to 49.99, $499,900 in individual trades are reported. This truncation creates outliers that can severely bias regression results. More importantly, the truncation does not directly depend on the market share of odd lots: as long as the price of a stock rises above 50, it is truncated. .

To get a sense of the severity of this problem, we first consider CRSP data. We start from the CRSP daily data file from January 1983 to December 2010 for all common shares issued by U.S. corporations. We apply the usual filters to remove stocks with price lower than 5 dollars and also Berkshire Hathaway. Based on stock price level alone, Figure 7 Panel A shows that on average about 10% of stocks will have zero individual trading volume using the 5,000-dollar cut-

off. The percentage fluctuates with the overall market level, and there are two peaks (the tech bubble and before the financial crisis) where about 15% of stocks would have registered zero individual trading. Because higher-priced stocks often have high trading volume, the bias may in fact be much larger. Panel B of Figure 7 weights each firm by their daily trading volume and Panel C provides the value-weighted average. During the tech-bubble period, we find that more than 70% of individual trading is missing if the $5000 cut-off rule is used.

Insert Figure 7 About Here

Turning to the NASDAQ data, we can specify this bias more directly for our sample firms. Figure 8 shows the total percentage of missing trades for all trades with 5,000 dollars or less. Over time, about 20-30% of trades and 8% to 12% of volume are missing. It is worth noting that the sample period covered by the NASDAQ data will partially mitigate this bias due to the low market price level during 2008-2009. As market prices recovered in 2009, the percentage of missing trades and volumes rose as well, suggesting that this bias is likely to be an even larger problem in today's markets.

Insert Figure 8 About Here

The literature also uses the order imbalances of small trades as an indicator of retail traders' sentiments. This measure can be misleading for a variety of reasons. One is that due to algorithmic and other trading practices small trades now increasingly come from institutional or high frequency traders. Second there are two technical biases that arise because TAQ data cannot include odd-lots. One is that a true buy imbalance can be falsely identified as a sell imbalance and vice versa. Second, and more importantly, for stocks with prices higher than 50 dollars, we infer zero order imbalances from TAQ data but in reality it should be a buy or sell imbalance. Table 10 presents these two biases. The first type of error is of the same magnitude as the error

for the whole sample without the $5,000 cut-off. Based on order numbers, 9.61% of imbalances are mis-classified, with 4.77% of buy imbalances are classified as sell imbalances and 4.58% of sell imbalances are classified as buy imbalances. 0.11% of stock day are classified as buy imbalance although there is a balance of trades. 0.15% of stock day are misclassified as sell imbalance though there is a true balance. Again, the problem is less severe for volume and dollar volume-based imbalance measures where in total about 4% of orders are misclassified.

Insert Table 10 About Here

The problem is much more severe when we observe a zero trade imbalance. This error is less affected by the way to define order imbalance because it is a truncation based on price. Across all the three measures, we observe 17% of balanced trades that are actually buy or sell imbalances. If order imbalances from individual traders are used to explain other variables such as stock return, this can cause either one of two problems. If order imbalance is treated as missing because there are no observed trades, it leads to a 17% truncation of the regression sample. If order imbalance is treated as zero because zero buy and zero sell implied zero order imbalance, it results in 17% sample with zero values in individual trading.

Summing the two types of errors together, about 27% of imbalance is misclassified in terms of transaction and 21% in terms of volume or dollar volume. These errors are significant, because randomly assigning buy as sell order imbalances has a 50% chance of being correct.

5. Conclusions

In this research we investigated the odd-lot bias in TAQ data. We have demonstrated that missing trades are a large and pervasive problem in TAQ data. That trade sizes are truncated below 100 shares means there is a censored sample problem for all stocks. For some stocks, however, this problem is acute, with as much as 40% or more of trades missing from the data.

Moreover, these missing trades are highly informative, meaning that analyses of issues related to market efficiency are also subject to error. Equally important, measures such as order imbalance or imputed trader identity and sentiment measures can be severely biased.

Our analysis shows that odd-lot trades are now far from unusual, and market practices such as algorithmic trading and high frequency trading are only increasing their incidence. For researchers using TAQ data, these trends highlight the need to choose empirical measures carefully. Trade-based measures of order imbalance, for example, are more affected by this bias than are volume-based measures, suggesting a preferred approach for such research. Standard imputations regarding retail trades, or trader sentiment, however, appear to be fatally flawed, and researchers should eschew using TAQ data for such purposes. The development of new, more complete data bases such as the consolidated audit trail may be needed for continued research in this area.

We believe our results also have important policy and regulatory implications. The recent SEC Concept Release (2010) raised a number of questions regarding odd lot trades. In particular, the SEC queried:

> *"Why is the volume of odd lots so high? Should the Commission be concerned about this level of activity not appearing in the consolidated trade data? Has there been an increase in the volume of odd lots recently? If so, why? Do market participants have incentives to strategically trade in odd lots to circumvent the trade disclosure or other regulatory requirements? Would these trades be important for price discovery if they were included in the consolidated trade data? Should these transactions be required to be reported in the consolidated trade data? Why?"*

To our knowledge, our paper is the first one to address these issues.

As we have demonstrated, TAQ data is biased because the consolidated tape is biased: odd-lot trades are not recorded to either data source. When odd-lots were a trivial fraction of market activity, this omission was of little consequence. But new market practices mean that these missing trades are both numerous and informationally important. Moreover, while these trades are invisible to the 2.5 million subscribers to the consolidated tape, they are not invisible to all market participants. Nasdaq ITCH data contains odd-lots, and other market venues also now sell proprietary data that allow purchasers to see all market activity (see Easley, O'Hara and Yang (2010) for an analysis of the detrimental effects of differential access to market information).

The SEC also does not include odd-lots in other market reporting requirements. Rule 605, for example, requires market centers to report market quality statistics on a monthly basis, but these reports are based on trades of various size categories starting at 100 shares and above. Our results suggest that odd-lot trades now play a new, and far from irrelevant, role in the market. The SEC should recognize this new role and change the reporting rules regarding odd-lot trades for the consolidated tape and other regulatory data.

References

Alexander, G. J., and M. A. Peterson, 2007, "An Analysis of Trade-Size Clustering and Its Relation to Stealth Trading," *Journal of Financial Economics*, 84, 435-471.

Barber, B. M., T. Odean, and N. Zhu, 2009, "Do Noise Traders Move Markets?" *Review of Financial Studies*, 22, 151–186.

Barclay, M. J., and J. B. Warner, 1993, "Stealth Trading and Volatility: Which Trades Move Prices?" *Journal of Financial Economics*, 34, 281-305.

Brogaard, J. 2010, "High Frequency Trading and Its Impact on Market Quality," working paper, Northwestern University.

Busse, J. A., and T. C. Green, 2002, "Market Efficiency in Real Time," *Journal of Financial Economics*, 65, 415-437.

Campbell, J. Y., T. Ramadorai, and A. Schwartz, 2009, "Caught On Tape: Institutional Trading, Stock Returns, and Earnings Announcements," *Journal of Financial Economics*, 92, 66-91.

Chaboud, A., B. Chiquoine, E. Hjalmarsson, and C. Vega, 2009, "Rise of the machines: Algorithmic trading in the foreign exchange market," working paper, Board of Governors of the Federal Reserve System.

Chakravarty, S. 2001, "Stealth-Trading: Which Traders' Trades Move Stock Prices?" *Journal of Financial Economics*, 61, 289-307.

Chakravarty, S., P. K. Jain, J. Upson, and R. Wood, 2011, "Clean Sweep: Informed Trading through Intermarket Sweep Orders," working paper, Purdue University and University of Memphis.

Chan, K., and W. M. Fong, 2000, "Trade Size, Order Imbalance, and the Volatility-Volume Relation," *Journal of Financial Economics*, 57, 247-273.

Chen, Q., I. Goldstein, and W. Jiang, 2007, "Price Informativeness and Investment Sensitivity To Stock Price," *Review of Financial Studies*, 20, 619-650.

Choe, H., and O. Hansch, 2005, "Which Trades Move Stock Prices in the Internet Age?" working paper, Pennsylvania State University and Seoul National University.

Chordia, T., A. Goyal, and N. Jegadeesh, 2011, "Buyers Versus Sellers: Who Initiates Trades and When?" working paper, Emory University and University of Lausanne.

Chordia, T., R. Roll and A. Subrahmanyam, 2002, "Order Imbalance, Liquidity, and Market Returns," *Journal of Financial Economics*, 65, 111-130.

Chordia, T., and A. Subrahmanyam, 2004, "Order Imbalance and Individual Stock Returns: Theory and Evidence," *Journal of Financial Economics*, 72, 485-518.

Christie, W. and P. Schulz, 1994, "Why Do NASDAQ Market Makers Avoid Odd-Eighth Quotes?" *Journal of Finance*, 49, 1813-1840.

Christoffersen, S. K., and Y. Tang, 2009, "Institutional Herding and Information Cascades: Evidence from Daily Trades," working paper, McGill University.

Dyl, E. A., and E. D. Maberly, 1992, "Odd-Lot Transactions around the Turn of the Year and the January Effect," *Journal of Financial and Quantitative Analysis*, 27, 591-604.

Easley, D., M. M. L. de Prado, and M. O'Hara, 2011, "The Microstructure of the Flash Crash: Flow Toxicity, Liquidity Crashes and the Probability of Informed Trading," *Journal of Portfolio Management*, 37, 118-128.

Easley, D., S. Hvidkjaer, and M. O'Hara, 2002, "Is Information Risk A Determinant of Asset Returns?" *Journal of Finance*, 57, 2185-2221.

Easely, D., M. O'Hara, and J. Paperman, 1996, "Liquidity, Information and Infrequently Traded Stocks," *Journal of Finance*, 51, 1405-1436

Easley, D., M. O'Hara, and L. Yang, 2010, "Differential Access to Price Information in Financial Markets," work paper, Cornell University.

Engle, R., R. Furstenberg, and J. Russell, 2006, "Measuring and Modeling Execution Cost and Risk," working paper, New York University.

Francis, J. C. 1986. Management of Investments, McGraw-Hill, New York.

Getry, W. M., D. Kemsley, and C. J. Mayer, 2003, "Dividend Taxes and Share Prices: Evidence From Real Estate Investment Trusts," *Journal of Finance*, 58, 261-282.

Harris, L. 1991, "Stock Price Clustering and Discreteness," *Review of Financial Studies*, 4, 389-415.

Hasbrouck, J., and G. Saar, 2010, "Low-Latency Trading," working paper, Cornell University.

Hendershott, T., C. M. Jones, and A. J. Menkveld, 2011, "Does Algorithmic Trading Improve Liquidity?" *Journal of Finance*, 66, 1-33.

Hendershott, T., and R. Riordan, 2009, "Algorithmic Trading and Information," working paper, University of California, Berkeley and Karlsruhe Institute of Technology.

Hvidkjaer, S. 2006, "A Trade-Based Analysis of Momentum," *Review of Financial Studies*, 19, 457.

Jame, R., and Q. Tong, 2009, "Retail Investor Industry Herding," working paper, Emory University .

Kelley, E. K., & Tetlock, P. C. 2011. How wise are crowds? Insights from retail orders and stock returns. Working Paper

Krigman, L., W. H. Shaw, and K. L. Womack, 1999, "The Persistence of IPO Mispricing and the Predictive Power of Flipping," Journal of Finance, 54, 1015-1044.

Kyle, A. S., A. A Obizhaeva, and T. Tuzun, 2010, Trading Game Invariance in the TAQ Dataset Working Paper, The University of Maryland

Lakonishok, J., and E. Maberly, 1990, "The Weekend Effect: Trading Patterns of Individual and Institutional Investors," Journal of Finance, 49, 231-243.

Lee, C., and B. Radhakrishna, 2000, "Inferring Investor Behavior: Evidence From TORQ Data," Journal of Financial Markets, 3, 83-111.

Malkiel, B. 1981, A Random Walk Down Wall Street, Norton, New York.

Malmendier, U., and D. Shanthikumar, 2007, "Are Small Investors Naive About Incentives?" *Journal of Financial Economics*, 85, 457-489.

Moulton, P. C. 2005, "You Can't Always Get What You Want: Trade-Size Clustering and Quantity Choice In Liquidity," *Journal of Financial Economics*, 78, 89-119.

Nimalendran, M., J. R. Ritter, and D. Zhang, 2007, "Do Today's Trades Affect Tomorrow's IPO Allocations?" *Journal of Financial Economics*, 84, 87-109.

O'Hara, M., and M. Ye, 2011, "Is Market Fragmentation Harming Market Quality?" *Journal of Financial Economics*, 3, 459-474.

Ritter, J. R., 1988, "The Buying and Selling Behavior of Individual Investors At the Turn of the Year," *Journal of Finance*, 43, 701-717.

Rozeff, M. S., 1985, "*The Tax-Loss Selling Hypothesis: New Evidence from Share Shifts*," working paper, University of Iowa.

Sadka, R., 2006, "Momentum and Post-Earnings-Announcement Drift Anomalies: The Role of Liquidity Risk," *Journal of Financial Economics*, 80, 309-349.

Sias, R. W., 1997, "Price Pressure and the Role of Institutional Investors in Closed-End Funds," *Journal of Financial Research*, 20, 211-229.

Wooldridge, J. M. (2002). *Econometric analysis of cross section and panel data* The MIT press.

Wu, K-W., 1972, "Odd Lot Trading in the Stock Market and Its Market Impact," *Journal of Financial and Quantitative Analysis*, 7, 1321-1344.

Table 1: Summary Statistics of the Sample Firms.

Table 1 provides summary statistics for the 120 firms ranging from January 2008 to December 2009 in the NASDAQ High Frequency data set. Large firms contain the 40 firms with the largest market cap. Small firms contain the 40 firms with the smallest market cap. Medium firms are firms between large and small firms. Spread is the average trade weighted relative spread; Pin is the probability of informed trading for each stock; Range is the daily price range; Volume is the daily volume; Price is the closing price of the trading day from CRSP; MarketCap is the market capitalization of the stock of the trading day. Volume and Marketcap are in the unit of one million. Rankings are based on market capitation of December 31, 2007.

Variable	Mean	StdDev	Max	Min	Type
MarketCap	46760.98	51461.22	383602.92	3349.12	large
Spread	0.04	0.07	0.87	0.01	large
Range	0.04	0.03	0.68	0.00	large
Volume	16.61	24.34	752.91	0.17	large
Price	56.72	76.76	685.33	5.22	large
Pin	0.07	0.02	0.12	0.02	large
MarketCap	1554.53	667.15	4110.46	98.90	median
Spread	0.04	0.03	0.55	0.01	median
Range	0.05	0.04	0.87	0.00	median
Volume	1.00	1.28	23.51	0.02	median
Price	28.39	18.34	114.17	0.90	median
Pin	0.15	0.04	0.25	0.02	median
MarketCap	422.75	248.14	1797.76	19.13	small
Spread	0.10	0.24	4.40	0.01	small
Range	0.06	0.05	1.59	0.00	small
Volume	0.28	0.36	15.37	0.00	small
Price	19.44	19.84	169.00	0.24	small
Pin	0.18	0.05	0.33	0.02	small

Table 2: Sample stocks and missing odd lot trades and volumes

symbol	missingntr adepct	missing volpct	symbol	missingntr adepct	missing volpct	symbol	missingntr adepct	missing volpct
AA	9.64%	2.15%	CPWR	14.08%	2.99%	JKHY	14.65%	5.61%
AAPL	19.33%	6.55%	CR	18.34%	6.09%	KMB	20.03%	7.00%
ABD	15.29%	5.08%	CRI	10.67%	3.72%	KNOL	22.69%	7.94%
ADBE	15.40%	4.53%	CRVL	40.37%	13.93%	KR	23.51%	6.81%
AGN	22.13%	6.86%	CSCO	9.55%	1.26%	KTII	66.51%	28.70%
AINV	10.90%	3.09%	CSE	13.79%	3.14%	LANC	24.84%	10.17%
AMAT	10.93%	1.79%	CSL	19.52%	6.28%	LECO	23.14%	8.76%
AMED	19.95%	7.40%	CTRN	16.99%	5.97%	LPNT	15.70%	5.14%
AMGN	16.85%	5.01%	CTSH	14.88%	4.48%	LSTR	28.53%	10.81%
AMZN	21.91%	6.57%	DCOM	29.14%	9.24%	MAKO	24.14%	5.54%
ANGO	22.67%	7.82%	DELL	10.17%	1.63%	MANT	22.91%	8.84%
APOG	19.23%	7.07%	DIS	11.66%	3.19%	MDCO	18.28%	6.51%
ARCC	10.72%	3.45%	DK	15.37%	4.34%	MELI	16.11%	5.56%
AXP	14.92%	4.37%	DOW	11.57%	3.25%	MFB	28.39%	9.34%
AYI	12.69%	4.71%	EBAY	10.59%	2.24%	MIG	18.31%	5.16%
AZZ	24.57%	8.48%	EBF	28.09%	9.70%	MMM	19.93%	7.12%
BARE	10.51%	3.17%	ERIE	43.60%	13.89%	MOD	13.44%	3.92%
BAS	19.74%	5.65%	ESRX	22.80%	8.03%	MOS	18.50%	6.53%
BHI	25.83%	9.12%	EWBC	22.38%	6.02%	MRTN	23.13%	8.28%
BIIB	22.23%	7.75%	FCN	16.43%	5.96%	MXWL	23.42%	7.43%
BRCM	14.08%	3.61%	FFIC	25.53%	8.86%	NC	42.66%	14.51%
BRE	40.85%	12.99%	FL	30.53%	8.91%	NSR	17.59%	4.86%
BW	17.21%	5.64%	FMER	25.73%	8.73%	NUS	14.97%	5.41%
BXS	42.83%	13.55%	FPO	25.38%	7.01%	NXTM	17.24%	4.99%
BZ	8.67%	1.75%	FRED	16.93%	5.88%	PBH	23.90%	7.53%
CB	35.25%	11.76%	FULT	22.33%	6.06%	PFE	9.00%	1.22%
CBEY	18.99%	6.16%	GAS	26.50%	7.86%	PG	13.61%	4.49%
CBT	28.46%	8.15%	GE	8.36%	1.26%	PNC	33.51%	10.86%
CBZ	17.73%	5.58%	GENZ	23.13%	8.26%	PNY	16.65%	6.16%
CCO	18.80%	5.17%	GILD	18.99%	5.88%	PPD	31.11%	10.23%
CDR	19.32%	5.12%	GLW	8.14%	1.82%	PTP	19.97%	6.02%
CELG	19.07%	6.39%	GOOG	33.95%	13.56%	RIGL	17.23%	5.55%
CETV	15.35%	4.93%	GPS	19.72%	5.02%	ROC	19.92%	5.79%
CHTT	23.16%	9.03%	HON	13.90%	4.59%	ROCK	27.04%	8.76%
CKH	40.34%	11.90%	HPQ	12.19%	3.70%	ROG	25.10%	6.96%
CMCSA	12.53%	2.50%	IMGN	23.67%	6.91%	RVI	10.97%	2.96%
CNQR	14.74%	5.37%	INTC	9.05%	1.10%	SF	36.78%	11.72%
COO	16.58%	5.30%	IPAR	25.37%	8.46%	SFG	32.21%	9.47%
COST	28.68%	8.85%	ISIL	10.49%	2.88%	SJW	25.26%	6.30%
CPSI	28.12%	10.20%	ISRG	34.67%	13.90%	SWN	19.31%	5.98%

Table 3: Odd-lot trades by market cap and price

This table presents the odd-lot trades based on market cap and price groups. Panel A divide the 120 stocks into large, median and small market cap group, each of which contains 40 stocks. Panel B divide the 120 stocks into high, median and low price group, each of which has 40 stocks. The table also tests the hypothesis that the average level of odd-lots is equal across different group. The t-statistics of the test are presented in the parentheses.

Panel A : By Market Capitalization						
	Large	Median	Small	Small - Median	Median - Large	Small - Large
missingntradepct	0.180	0.212	0.238	0.026	0.032*	0.058***
				(1.24)	(1.74)	(2.93)
missingvolpct	0.055	0.068	0.078	0.010	0.013*	0.023***
				(1.26)	(1.81)	(2.68)

Panel B: By Price						
	High	Median	Low	Low - Median	Median - High	Low - High
missingntradepct	0.249	0.187	0.194	0.008	-0.063***	-0.055***
				(0.44)	(-2.83)	(-2.82)
missingvolpct	0.085	0.057	0.059	0.003	-0.029***	-0.026***
				(0.45)	(-3.32)	(-3.35)

t statistics in parentheses. ***, ** and * means the significance at 1%, 5% and 10% level.

Table 4: Example of odd-lots pattern 1

This table demonstrates an example of a sequence of odd-lots trading happened in June 20, 2008. The patterns are generated by high frequency traders take liquidity from non high frequency traders. There are 111 odd lot sells happened in 13:59:01:107, which has a total of 2995 shares. Another 102 odd lot sells happened 3 milliseconds later, which has a total of 2576 shares.

Sequence	Symbol	Hour	Minute	Second	Millisecond	Shares	BuySell	Price	Type
1	AAPL	13	59	1	107	20	S	125	HN
2	AAPL	13	59	1	107	10	S	125	HN
3	AAPL	13	59	1	107	50	S	125	HN
4	AAPL	13	59	1	107	25	S	125	HN
5	AAPL	13	59	1	107	12	S	125	HN
6	AAPL	13	59	1	107	35	S	125	HN
7	AAPL	13	59	1	107	10	S	125	HN
8	AAPL	13	59	1	107	12	S	125	HN
9	AAPL	13	59	1	107	24	S	125	HN
10	AAPL	13	59	1	107	6	S	125	HN
11	AAPL	13	59	1	107	4	S	125	HN
12	AAPL	13	59	1	107	75	S	125	HN
13	AAPL	13	59	1	107	1	S	125	HN
14	AAPL	13	59	1	107	15	S	125	HN
15	AAPL	13	59	1	107	50	S	125	HN
.....									
108	AAPL	13	59	1	107	50	S	125	HN
109	AAPL	13	59	1	107	50	S	125	HN
110	AAPL	13	59	1	107	30	S	125	HN
111	AAPL	13	59	1	107	3	S	125	HN
112	AAPL	13	59	1	110	47	S	125	HN
113	AAPL	13	59	1	110	80	S	125	HN
114	AAPL	13	59	1	110	80	S	125	HN
115	AAPL	13	59	1	110	8	S	125	HN
116	AAPL	13	59	1	110	8	S	125	HN
117	AAPL	13	59	1	110	60	S	125	HN
118	AAPL	13	59	1	110	8	S	125	HN
119	AAPL	13	59	1	110	32	S	125	HN
120	AAPL	13	59	1	110	30	S	125	HN
.....									
210	AAPL	13	59	1	110	5	S	125	HN
211	AAPL	13	59	1	110	25	S	125	HN
212	AAPL	13	59	1	110	50	S	125	HN
213	AAPL	13	59	1	110	12	S	125	HN

Table 5: Variation of Missing Trades and Volume

This table explains the variation of missing trades and volume. We run the between, random and fixed effect regression on the panel of miss trades and volume for each stocks on each day. *Missingntradepct* and *missingvolpct* are percentage of missing trades and volume; *logprc* is the price level; *spread* is the bid-ask spread; *pin* is the probability of informed trading for each stock; *range* is daily price range; *NYSE* equals to 1 if the stock is listed in NYSE and 0 if it list in NASDAQ. The sample period is 504 trading days from 2008-2009.

	(1)	(2)	(3)	(4)	(5)	(6)
VARIABLES	missingntradepct	missingvolpct	missingntradepct	missingvolpct	missingntradepct	missingvolpct
logprc	0.044***	0.020***	0.011***	0.008***	0.012***	0.009***
	(4.16)	(4.76)	(6.43)	(10.62)	(6.90)	(11.24)
pinall	0.469***	0.253***	0.475***	0.283***		
	(3.79)	(5.14)	(3.82)	(4.89)		
spread	0.273***	0.164***	0.063***	0.028***	0.062***	0.027***
	(5.08)	(7.65)	(10.91)	(10.34)	(10.59)	(9.91)
range	0.372	0.059	-0.299***	-0.148***	-0.287***	-0.140***
	(0.68)	(0.27)	(-19.77)	(-20.8)	(-18.81)	(-19.56)
NYSE	0.003	-0.002	-0.005	-0.006		
	(0.22)	(-0.33)	(-0.35)	(-0.89)		
constant	-0.018	-0.031	0.137***	0.023***	0.220***	0.061***
	(-0.33)	(-1.42)	(6.75)	(2.46)	(19.67)	(11.66)
Effect	Between	Between	Random	Random	Fixed	Fixed
Observations	60,412	60,412	60,412	60,412	60,412	60,412
R-squared	0.4913	0.6458	0.0111	0.0139	0.5833	0.5627
Number of tickers	120	120	120	120	120	120

Table 6: Price Discovery, Share on Number of Trades and Volume for each Size Category

This table demonstrates the weighted price contribution for each order size category. Panel A is based on the return from 9:45-16:00 and Panel B is based on the return from 9:30-16:00.

Panel A: 9:45-16:00

Trade size category	$WPC_{return\ change}$	$WPC_{price\ change}$	Shares of Trades	Shares of Volume
<100	0.265	0.318	0.156	0.034
100	0.544	0.533	0.541	0.281
200	0.058	0.045	0.117	0.121
300	0.022	0.01	0.042	0.065
400	0.014	0.016	0.026	0.053
500	0.012	0.005	0.024	0.061
100-500	0.694	0.651	0.793	0.633
501-900	0.018	0.012	0.027	0.099
901-1900	0.016	0.011	0.017	0.109
1901-4900	0.006	0.008	0.005	0.078
4901-9999	0.001	0.000	0.001	0.028
501-9999	0.041	0.031	0.051	0.314
>=10000	0.000	0.000	0.000	0.019

Panel B 9:30-16:00

Trade size category	$WPC_{return\ change}$	$WPC_{price\ change}$	Shares of Trades	Shares of Volume
<100	0.306	0.354	0.158	0.034
100	0.504	0.497	0.54	0.281
200	0.053	0.041	0.117	0.121
300	0.022	0.01	0.041	0.065
400	0.014	0.016	0.025	0.053
500	0.013	0.006	0.024	0.062
100-500	0.652	0.615	0.791	0.633
501-900	0.018	0.012	0.027	0.099
901-1900	0.016	0.011	0.017	0.109
1901-4900	0.006	0.008	0.005	0.078
4901-9999	0.001	0.000	0.001	0.028
501-9999	0.042	0.031	0.051	0.313
>=10000	0.001	0.000	0.000	0.019

Table 7. Weighted Price Contribution by Trade Types

This table shows the weighted price contribution for four types of odd lot trades where HH is a high frequency trader providing liquidity to a high frequency trader; HN is a high frequency trader providing liquidity to a non-high frequency trader; NH is a non-high frequency trader providing liquidity to a high frequency trader; and NN is a non-high frequency trader providing liquidity to a non-high frequency trader.

Odd Lots				
Trade Type	$WPC_{return\ change}$	$WPC_{price\ change}$	Shares of Trades	Shares of Volume
HH	0.017	0.018	0.037	0.008
HN	0.067	0.09	0.049	0.01
NH	0.043	0.04	0.034	0.008
NN	0.178	0.206	0.038	0.009

Table 8: Test for price discovery

This table reports the weighted least square regressions of price contribution on dummy of less-than-100-share category, dummy of equal-or-greater-than-100-share category, and percentage of transactions or percentage of trading volume in that category. The dependent variable, price contribution for stock *s* on day *t* of category *j*, is the sum of stock *s* price changes belonging to category *j* on day *t* divided by the total cumulative stock *s* price changes on day *t*. The regression is weighted by the ratio of stock *s* absolute cumulative price change to the sum of all stocks' absolute cumulative price changes on day *t*. The null hypothesis is the coefficients of dummies in each category equal to zero and the coefficient of percentage of transactions or percentage of trading volume in that category equal to one. T-statistics are given in parentheses.

	regresssion	
	(1)	(2)
Trade Size		
< 100 shares	0.120***	0.175***
	(7.31)	(12.39)
>= 100 Shares	-0.023	-0.997***
	(-0.60)	(-11.15)
Percent of Transactions	0.903**	
	(1.98)	
Percent of Volume		1.821***
		(8.34)
Adj R^2	0.043	0.043
Tests on Dummy Variables		
	p-value	p-value
Dummy<100 shares = Dummy of >= 100 Shares	<.0001	<.0001

t-statistics in parentheses *** p<0.01, ** p<0.05, *p<0.1

Table 9. Correctly Signed Order Imbalance and PIN measure

This table demonstrates the percentage of correctly signed buy and sell imbalance and the PIN estimated through all trades and trades greater or equal to 100 shares. The table provides a conservative estimation because it is based on the assumption that Lee and Ready (1991) makes no mistakes in assigning buy and sell trades. True Buy Imbalance, True Balance and True Sell Imbalance are the true daily order imbalances. Observed Buy Imbalance, Observed Balance, and Observed Sell Imbalance are daily imbalances we would observe through the TAQ data, if all the buy and sells are correctly signed. OIBNUM is the defined as the number of buy trades minus the number of sell trades. OIBSH is defined as the number of buy volume minus sell volume. OIBDOL is defined as the buy dollar volume minus sell dollar volume. PINall are PIN estimated using all trades and PINge100 is PIN estimated through trades greater or equal to 100 shares. The sample period is from 2008-2009, where each observation is the imbalance of each 120 stocks on each day.

OIBNUM	Total incorrectly assigned imbalance: 11.37%			
	Observed Buy	Observed Balance	Observed Sell	Sum
True Buy Imbalance	43.60%	0.23%	5.34%	49.16%
True Balance	0.13%	0.02%	0.18%	0.33%
True Sell Imbalance	5.29%	0.00%	45.02%	50.31%
Sum	49.02%	0.25%	50.54%	100%

OIBSH	Total incorrectly assigned imbalance: 3.33%			
	Observed Buy	Observed Balance	Observed Sell	Sum
True Buy Imbalance	47.84%	0.04%	1.62%	49.50%
True Balance	0.00%	0.00%	0.00%	0.00%
True Sell Imbalance	1.64%	0.02%	48.84%	50.50%
Sum	49.49%	0.06%	50.46%	100%

OIBDOL	Total incorrectly assigned imbalance: 3.27%			
	Observed Buy	Observed Balance	Observed Sell	Sum
True Buy Imbalance	47.95%	0.00%	1.64%	49.59%
True Balance	0.00%	0.00%	0.00%	0.00%
True Sell Imbalance	1.62%	0.00%	48.79%	50.41%
Sum	49.57%	0.00%	50.43%	100%

PIN				
	PINall	PINge100	PINall-PINge100	p-value
Mean	0.1364	0.1415	-0.0051	0.314

Table 10. The Percentage of Correctly Signed Order Imbalance for Individual Trades

This table demonstrates the percentage of correctly signed buy and sell imbalance based on the Lee and Radhakrishna's 5,000 dollars cut-off for individual trades. True Buy Imbalance, True Balance and True Sell Imbalance are the true daily order imbalances. Observed Buy Imbalance, Observed Balance, and Observed Sell Imbalance are daily imbalances we would observe through the TAQ data, if all the buy and sells are correctly signed. OIBNUM is the defined as the number of buy trades minus the number of sell trades. OIBSH is defined as the number of buy volume minus sell volume. OIBDOL is defined as the buy dollar volume minus sell dollar volume. The sample period is from 2008-2009, where each observation is the imbalance of each 120 stocks on each day.

OIBNUM	Total incorrectly assigned imbalance: 26.82%			
	Observed Buy	Observed Balance	Observed Sell	Sum
True Buy Imbalance	35.71%	8.11%	4.77%	48.59%
True Balance	0.11%	0.12%	0.15%	0.38%
True Sell Imbalance	4.58%	9.11%	37.34%	51.03%
Sum	40.39%	17.35%	42.26%	100%

OIBSH	Total incorrectly assigned imbalance: 20.72%			
	Observed Buy	Observed Balance	Observed Sell	Sum
True Buy Imbalance	38.45%	7.96%	1.82%	48.23%
True Balance	0.00%	0.01%	0.00%	0.01%
True Sell Imbalance	1.86%	9.07%	40.83%	51.76%
Sum	40.31%	17.04%	42.65%	100%

OIBDOL	Total incorrectly assigned imbalance: 20.70%			
	Observed Buy	Observed Balance	Observed Sell	Sum
True Buy Imbalance	38.55%	7.93%	1.86%	48.34%
True Balance	0.00%	0.00%	0.00%	0.00%
True Sell Imbalance	1.89%	9.02%	40.76%	51.66%
Sum	40.44%	16.95%	42.62%	100%

Figure 1: Historical volume of odd-lots

This graph shows the historical market shares of NYSE odd-lots from 1950-2004. The data are from NYSE fact book.

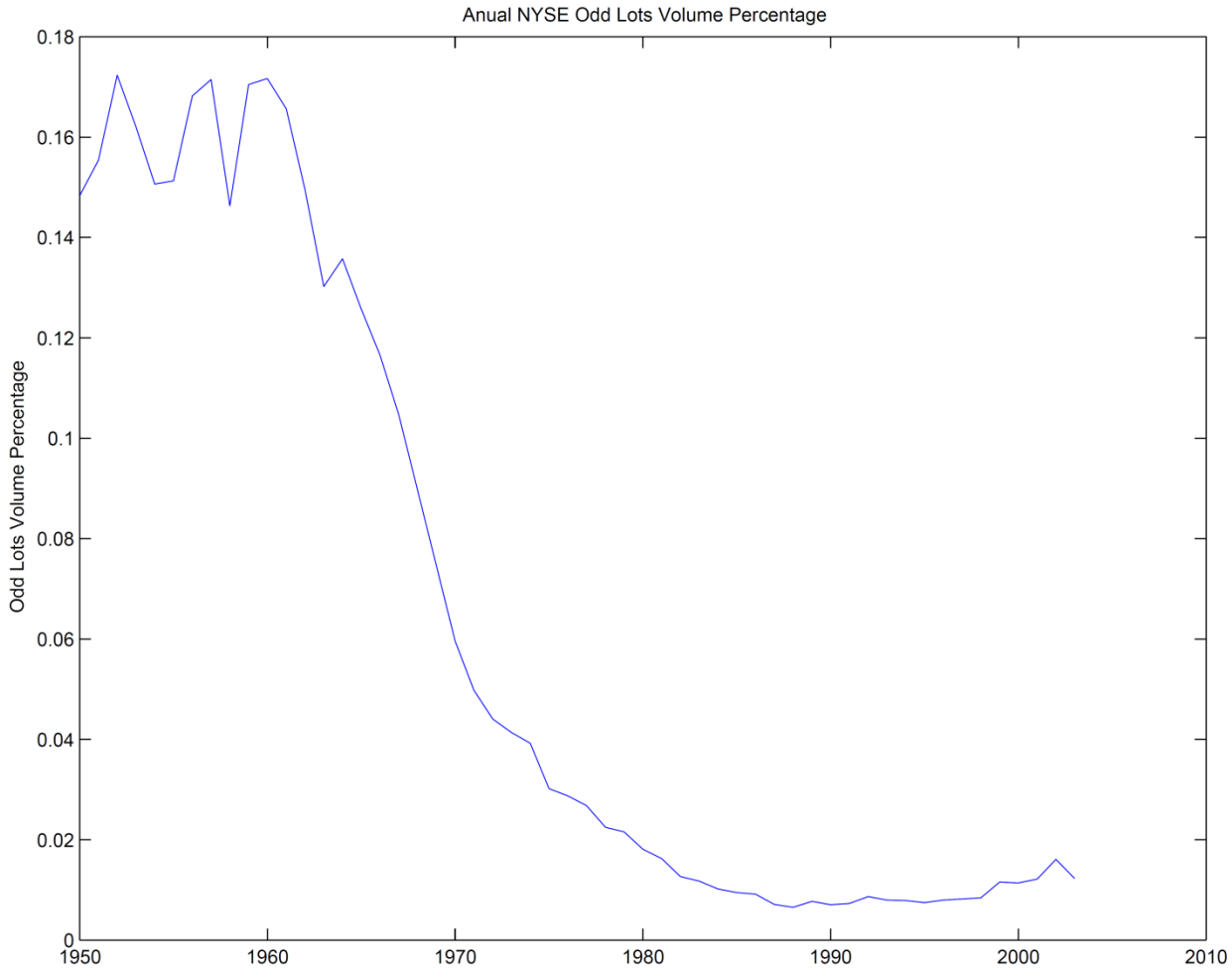


Figure 2: Time Series Variation of Missing Trades

This figure illustrates total level of missing trades and volume from 2008 to 2009. Panel A demonstrates the volume not reported to TAQ data as a percentage of total volume. Panel B demonstrate the trades not reported to TAQ data as a percentage of total trades.

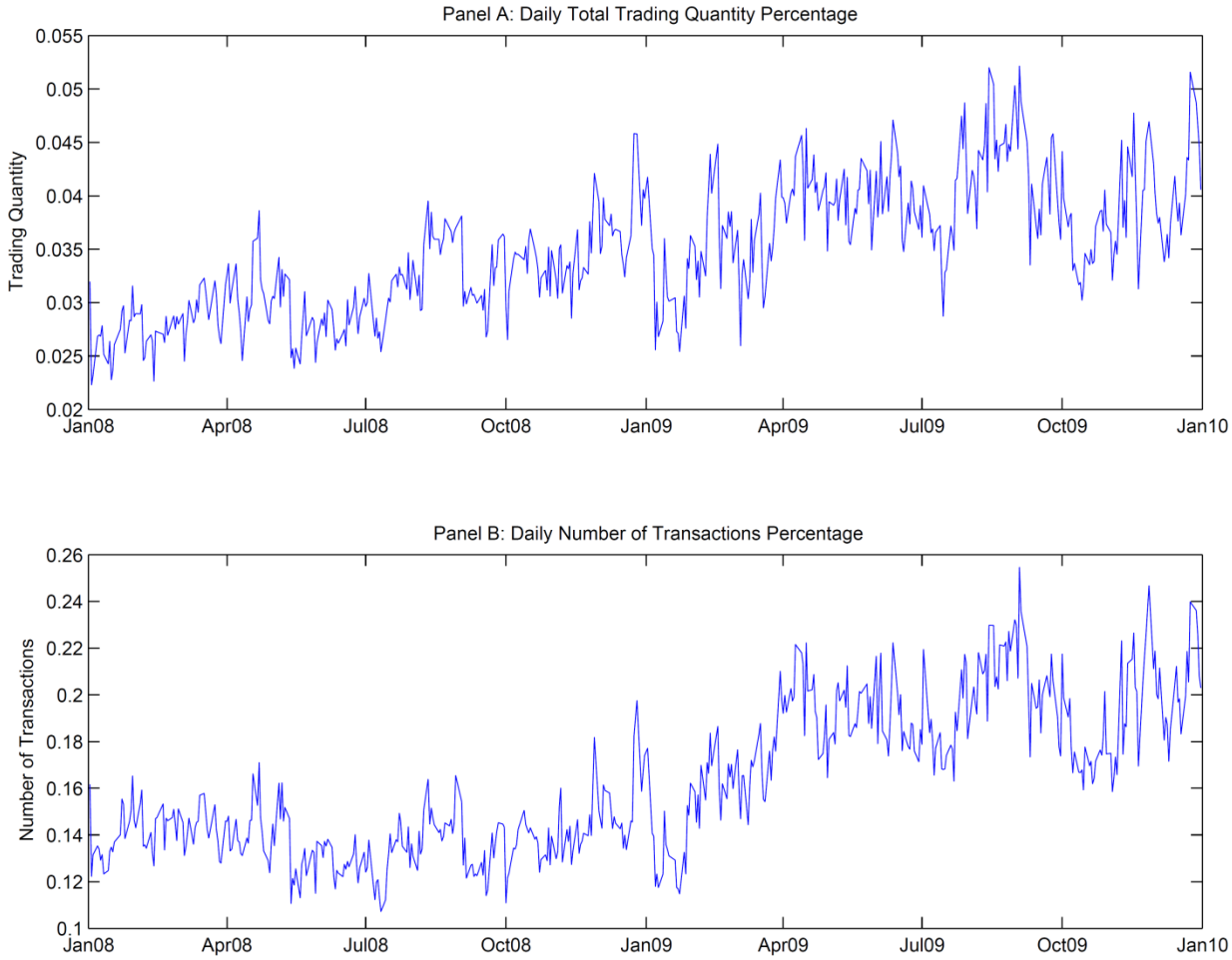
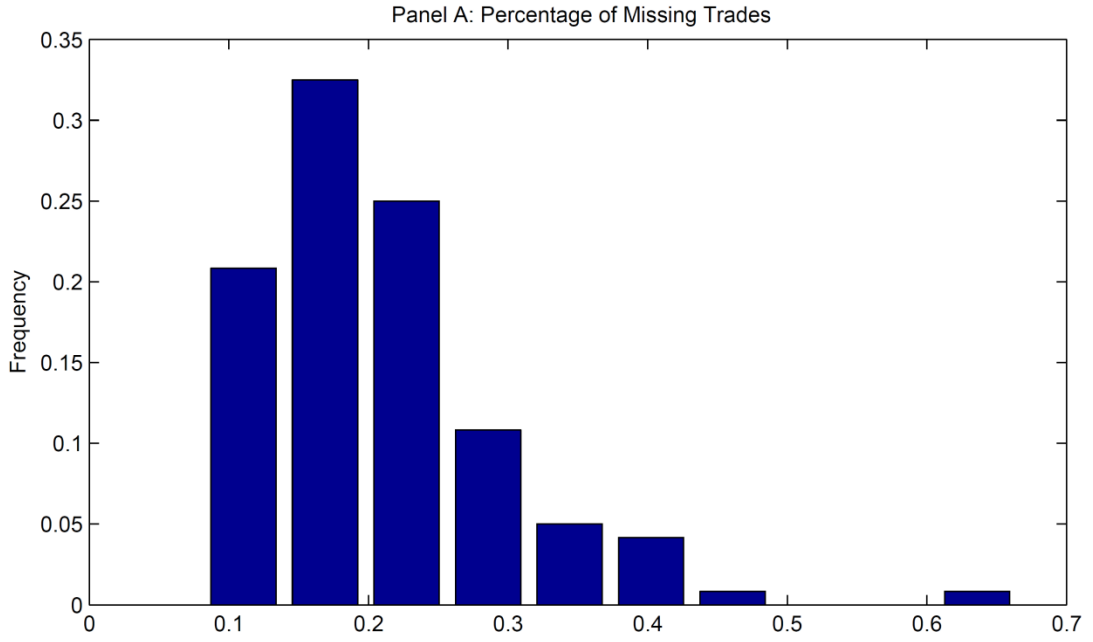


Figure 3: Cross-sectional variation of Missing Trades

This figure illustrates the level of missing trades across the 120 stocks in our sample. Panel A demonstrates the percentage the number of trades which are not reported to TAQ and Panel B demonstrates the percentage of missing volume.



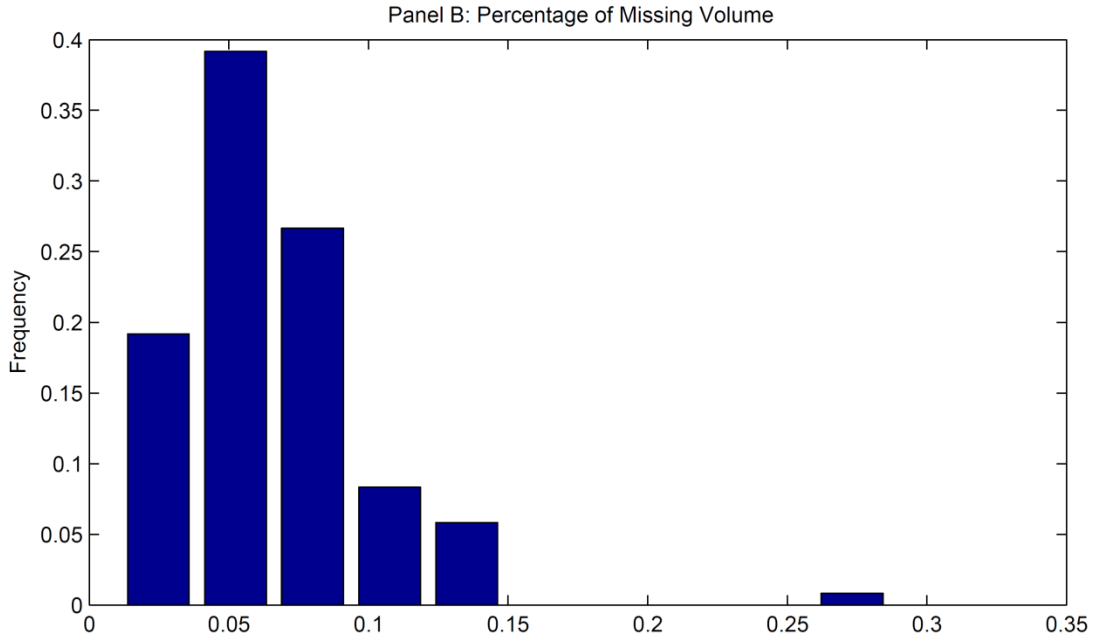


Figure 4: Histogram of Odd-lot Trades

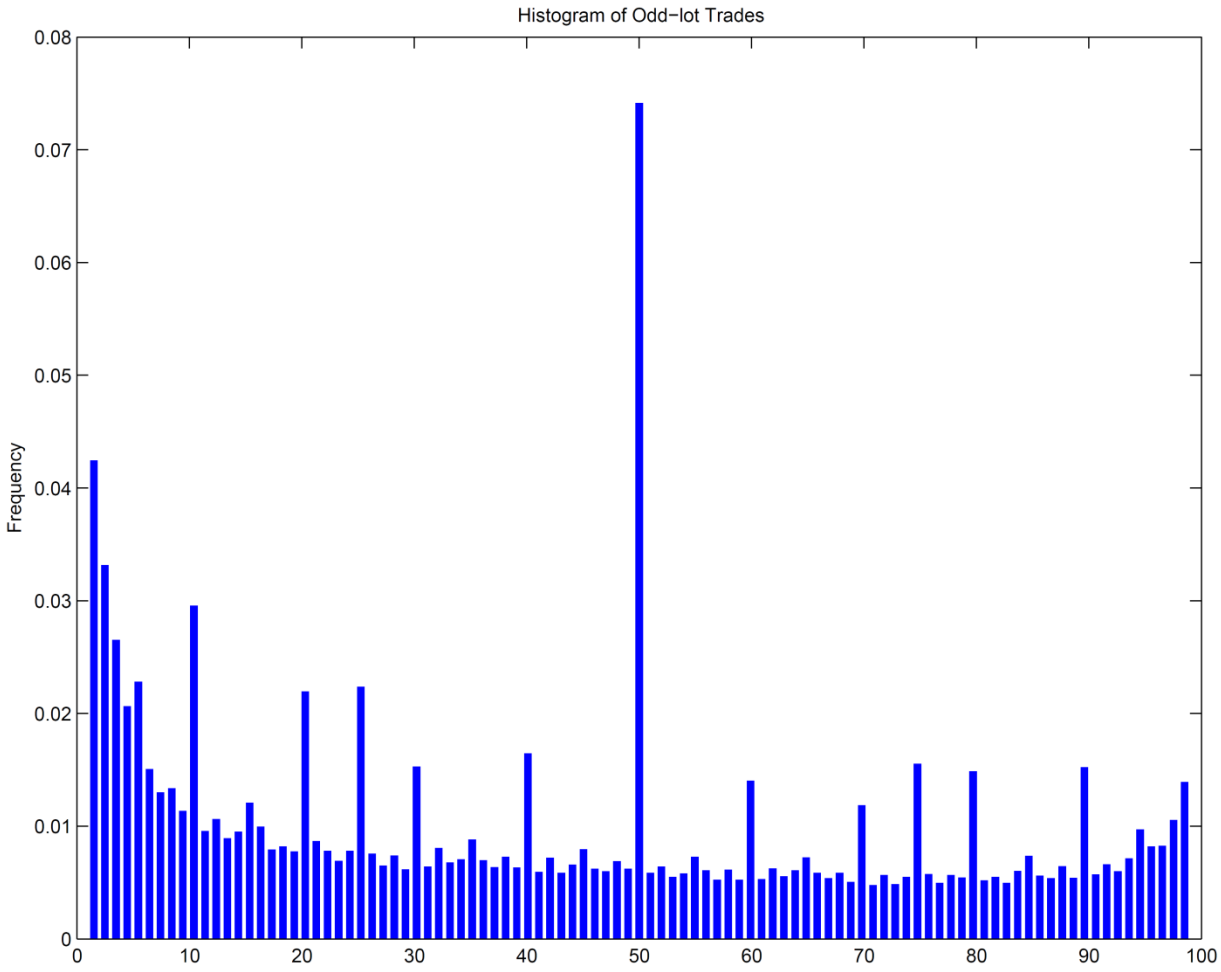
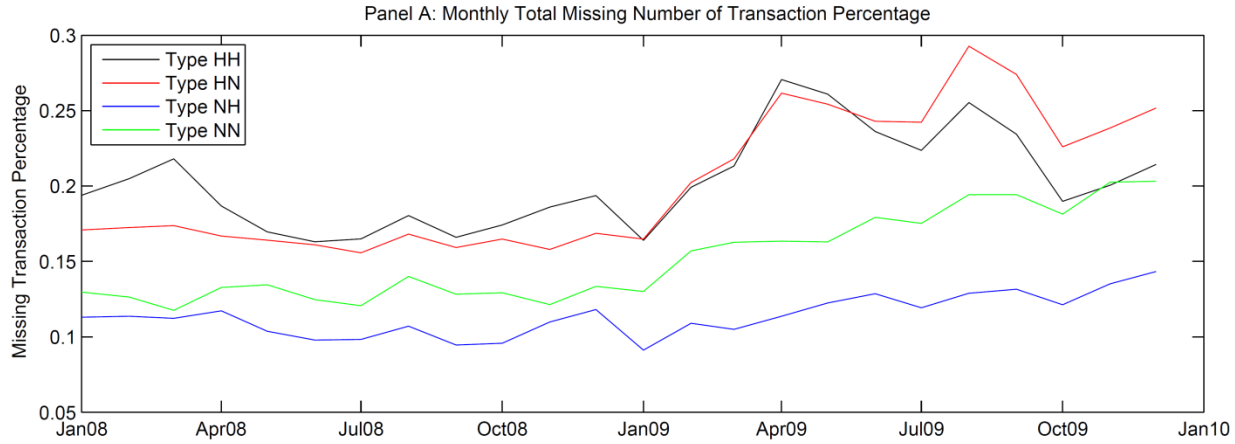


Figure 5: Odd-Lot Trades by Trader Type

This figure displays the time series odd-lots percentage breaking down to four different trade types. The first letter symbolizes the liquidity taker and the second one is the liquidity maker. Letter H stands for higher liquidity traders and N stands for non liquidity traders. For example, an HN trade means that a high frequency trader takes liquidity from a non-high frequency trader.



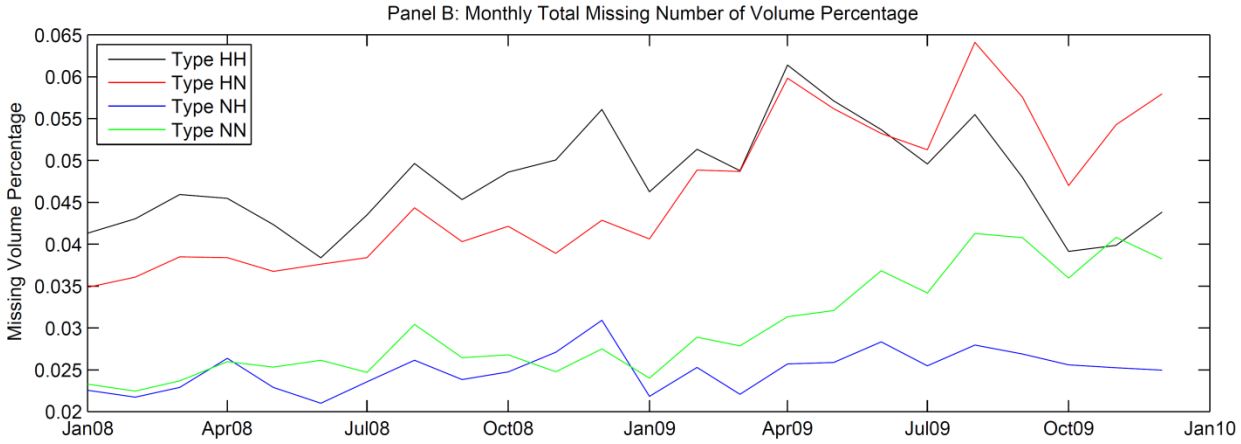


Figure 6: Time Series Variation of Missing Trades with Fixed Effects

This figure illustrates total level of missing trades and volume from 2008 to 2009 controlling for fixed effects. Panel A plots the time dummy coefficient for trades not reported to TAQ data as a percentage of total trades. Panel B reports the time dummy coefficient for volume not reported to TAQ data as a percentage of total volume.

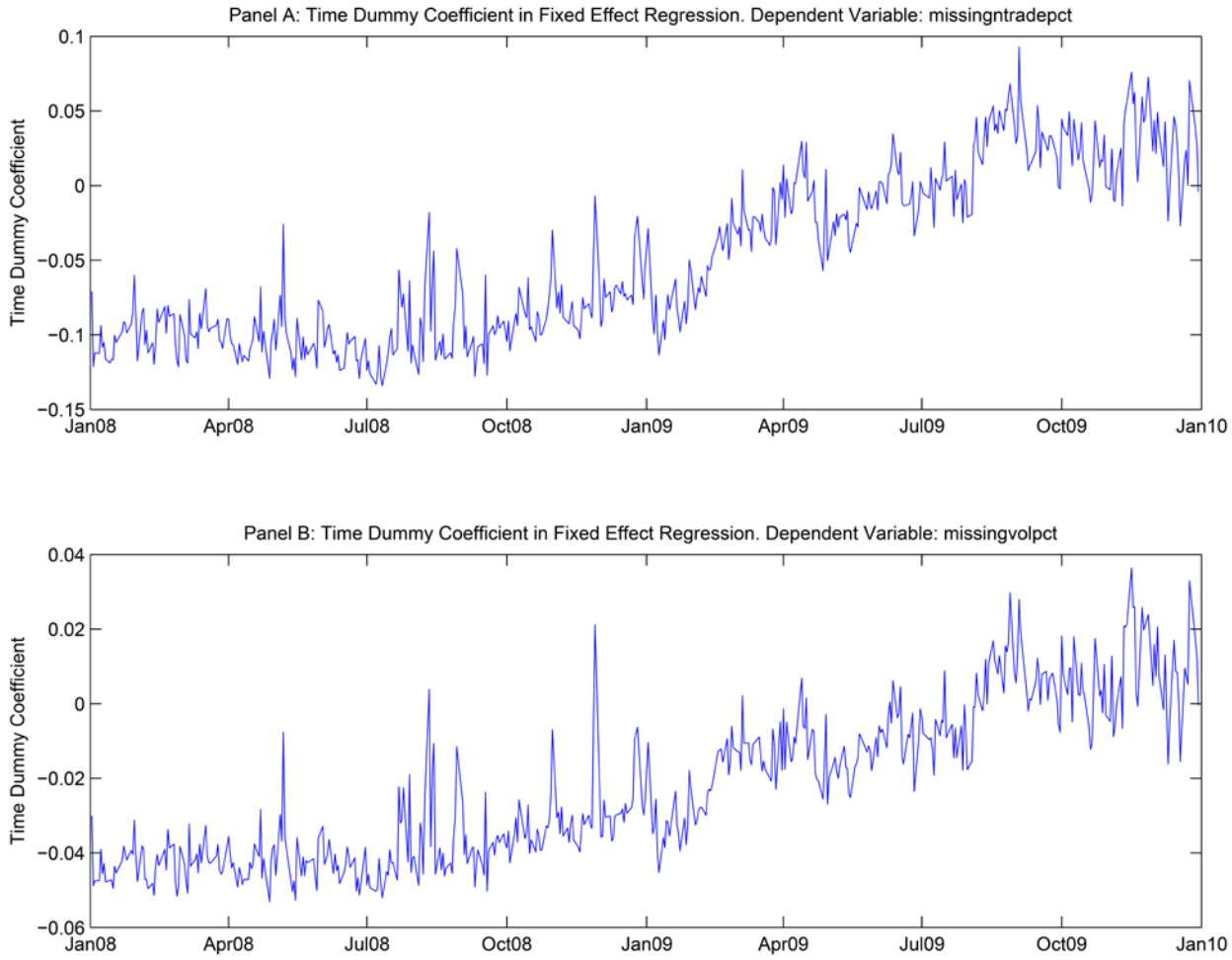


Figure 7: Stocks Shows Zero Individual Trading as A Result of 5000-dollar Cut-off Value

This figure demonstrate the percentage of stocks shows 0 individual trading by applying Lee and Radhakrishna's 5,000 dollars dollar cut-off to the TAQ data. Because TAQ does not report trades less than 100 shares, we observe 0 trading for individual trades for stocks with a price higher than 50. The graph is computed through CRSP. Panel A is the percentage of stocks with 0 individual trades. Panel B weights each stock by their dollar volume and Panel C provides value weighted average.

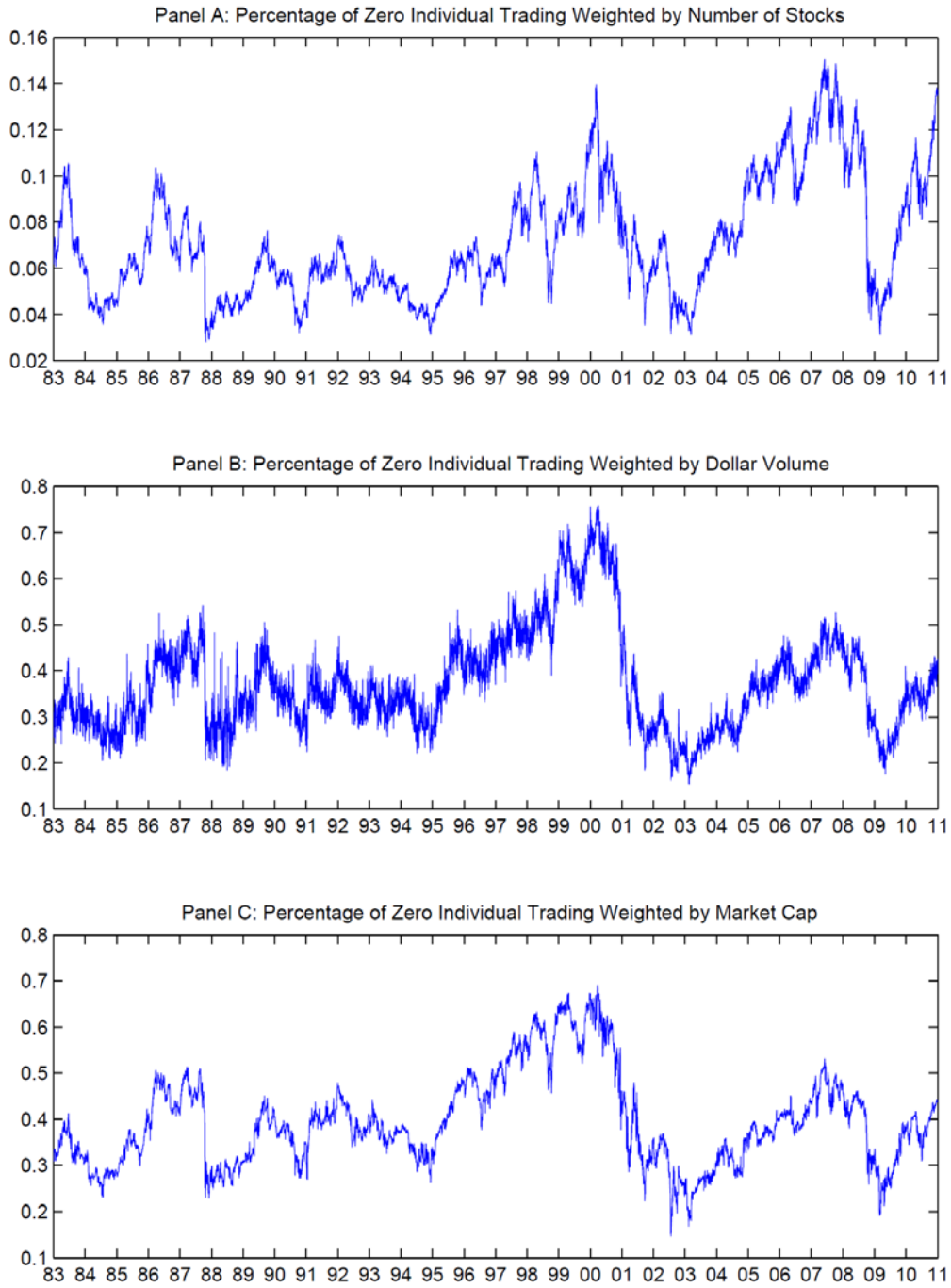


Figure 8: Time Series Variation of Missing Individual Trades using the 5,000 dollars cut-off

This figure illustrates total level of missing trades and volume using the 5,000 dollars cut-off value for individual trades from 2008 to 2009. Panel A demonstrates the volume not reported to TAQ data as a percentage of total volume. Panel B demonstrate the trades not reported to TAQ data as a percentage of total trades.

