The Journal of FINANCE
The Journal of THE AMERICAN FINANCE ASSOCIATION

THE JOURNAL OF FINANCE • VOL. LXIX, NO. 5 • OCTOBER 2014

What's Not There: Odd Lots and Market Data

MAUREEN O'HARA, CHEN YAO, and MAO YE*

ABSTRACT

We investigate odd-lot trades in equity markets. Odd lots are increasingly used in algorithmic and high-frequency trading, but are not reported to the consolidated tape or in databases such as TAQ. In our sample, the median number of odd-lot trades is 24% but in some stocks odd lots are 60% or more of trading. Odd-lot trades contribute 35% of price discovery, consistent with informed traders using odd lots to avoid detection. Omitting odd-lot trades leads to inaccuracies in order imbalance measures and makes sentiment measures unreliable. Excluding odd lots from the consolidated tape raises important regulatory issues.

ODD LOTS ARE TRADES for less than 100 shares of stock. Traditionally, odd-lot trades and volumes were small, and were thought to originate from retail traders and so they were viewed as having little information content with respect to future price movements—that is, as irrelevant. As a result, the convention followed by all market centers was that only round-lot trades of 100 shares and mixed-lot trades of greater than 100 shares are reported to the consolidated tape.[1]

*Maureen O'Hara is from Johnson Graduate School of Management at Cornell University. Chen Yao is from Warwick Business School at the University of Warwick. Mao Ye is from College of Business at the University of Illinois, Urbana-Champaign. We are grateful to NASDAQ, particularly Frank Hatheway, and the Financial Markets Research Center at Vanderbilt University for providing data, and to Soeren Hvidkjaer for providing computer code. We thank Campbell Harvey (The Editor), an Associate Editor, and an anonymous referee for helpful comments. We also thank Jim Angel; Brad Barber; Shmuel Baruch; Robert Battalio; Jonathan Brogaard; Kee Chung; Robert Connelly; Prachi Deuskar; Engelbert Dockner; Amy Edwards; Larry Glosten; Michael A. Goldstein; Oliver Hansch; Jeff Harris; Joel Hasbrouck; Terry Hendershott; Tim Johnson; Herb Lash; Christian Laux; Jaehoon Lee; Steward Mayhew; Dmitriy Muravyev; Elizabeth Odders-White; Terry Odean; Marios Panayides; Neil Pearson; Patrik Sandas; Jeff Smith; George Sofianos; Neal Stoughton; Sunil Wahal; Chunchi Wu; Hao Zhang; Josef Zechner; and seminar participants at the 2011 NBER market microstructure meeting, the 2013 AFA meetings, Goldman Sachs, University of Illinois, NASDAQ, the State University of New York at Buffalo, Syracuse University, and Vienna Graduate School of Finance for helpful comments and discussions. Mao Ye gratefully acknowledges the support from the National Science Foundation, the Extreme Science and Engineering Discovery Environment (XSEDE), and the Bureau of Economic and Business Research and Research Board at the University of Illinois. We thank Robert Sinkovits of the San Diego Supercomputer Center and Anirban Jana of the Pittsburgh Supercomputer Center for their assistance with supercomputing. Excellent research assistance was provided by Taishan Zhu and Xuefei Xia.

[1] The consolidated 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, use consolidated tape data.

DOI: 10.1111/jofi.12185

However, times have changed. The median trade size on the NASDAQ is 100 shares, with odd lots now a large fraction of trades. 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).[2] The emergence of high-priced stocks such as Google or Apple, where trading a round lot requires an investment of $60,000 or more, results in odd lots constituting a significant fraction of trade for a subset of important stocks in the market. 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.

Yet none of this is apparent to either market watchers or researchers because neither the consolidated tape nor the TAQ data derived from it includes odd-lot trades.[3] That odd-lot trades are now an important fraction of the market is undeniable: in our sample, we find the average ratio of odd-lot trades per stock is 24% but for some stocks odd lots are as high as 60% of total transactions. Perhaps more disquieting is the fact that these trades are not innocuous, as we demonstrate that odd-lot trades have higher information content than round-lot or mixed-lot trades. Moreover, we find that odd lots as a percentage of trades have been growing over time. As we discuss, these findings have important implications for the current regulatory debates regarding market transparency and high-frequency trading, as well as for the design and interpretation of academic studies relying on market data.

Our analysis focuses on a special data set of 120 stocks 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 on NASDAQ for the period 2008 to 2009. Trades are also identified by trader identity (specifically, whether the buyer or seller is a high-frequency trader), 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.[4] We supplement this data set with more recent data on trade executions from 2010 to 2011 to show how odd-lot trading has continued to grow for the stocks in our sample.

Our analysis focuses on three questions. First, how important is odd-lot trading across stocks and what determines its incidence? To address this question, 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

[2] The increased incidence of index trading also leads to increases in odd-lot trades due to rebalancing, as does more extensive use of hedging techniques for option trades.

[3] Even regulators face a blind spot with respect to odd lots in much of the data they collect. For example, the U.S. Securities and Exchange Commission requires each market center to provide on a monthly basis the execution rates of limit orders on those markets (referred to as SEC Rule 605 market quality statistics), but these statistics do not include odd-lot trades.

[4] The sample was constructed by Terrence Hendershott and Ryan Riordan, and details on the data can be found in Brogaard, Hendershott, and Riordan (2013).

identity of odd-lot traders. Second, what are the informational properties of odd-lot trades? Here, we calculate Weighted Price Contribution measures of odd-lot trades and conduct vector autoregressive (VAR) analyses to investigate how the information content differs across trade sizes and trader types. Third, how does the exclusion of odd-lot trades affect researchers? We address this question by showing how these missing odd lot trades influence a variety of measures used by finance researchers.

Why does it matter that 52.9% of trades in Google are not visible to the market? Or that 25% of trades in small stocks (and almost 20% of trades in large stocks) are missing from TAQ data? Or that 85% of price discovery on NASDAQ is now coming from trades of 100 shares or less? Or that odd lots are most frequently used by high-frequency traders? We believe there are some very important reasons to care about these odd-lot trades.

First, odd lots provide an important lens through which to view the new world of high-frequency trading. While odd lots are still used by retail traders, they are more likely to arise from high frequency or algorithmic traders. Our results are consistent with algorithms now slicing and dicing larger orders into odd-lot–sized pieces. The fact that 35% to 39% of price discovery is coming from odd lots is consistent with this being done to "hide" such trades from the market. Our results here contribute to a growing literature on the impact of high frequency and algorithmic trading on markets (see Chaboud et al. (2014), Easley, Lopez de Prado, and O'Hara (2011, 2012a), Hasbrouck and Saar (2011), Hendershott, Jones, and Menkveld (2011), Baron, Brogaard, and Kirilenko (2012)).

Second, the Securities and Exchange Commission (SEC) as well as regulators throughout the globe are increasingly concerned about market transparency. Much of this concern has focused on pretrade transparency in the context of hidden orders or dark pools (see Buti, Rindi, and Werner (2011a, 2011b), Ye (2011), Bloomfield, O'Hara, and Saar (2013)). But posttrade transparency is equally important, and seeing all trades gives traders important information about the current state of the market.[5] Omitting odd lots from the consolidated tape but not from the proprietary data feeds sold by exchanges means that U.S. markets are becoming increasingly opaque (at least to most of us).

Third, and perhaps most pertinent for finance researchers, the exclusion of odd-lot trades can affect a variety of market data–based measures, as well as the interpretation of previous research results. Order imbalance measures, for example, are greatly affected by missing trades, with incorrect classification rates in our sample of 11%.[6] The missing trade problem is also particularly acute for behavioral finance studies that impute retail trading behavior and sentiment (see, for example, Lamont and Frazzini (2007), Hvidkjaer (2008),

[5] It is important to stress that trades involving hidden orders and trades in dark pools are reported to the consolidated tape, so that posttrade transparency issues do not arise in these contexts (unless such trades are odd lots in which case they are also not reported).

[6] 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.

Barber, Odean and Zhu (2009)).[7] We show that, depending upon the time period, up to 15% of all stocks in our sample have zero imputed retail trades because of this missing data problem. Our findings raise red flags in using particular data measures in future research and in interpreting some existing results in the literature.

This missing data issue should concern all researchers using TAQ data. We also believe it raises important regulatory issues for the SEC. While policies surrounding odd lots 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 reveal that odd-lot trades have changed as well, and they now play a new, and far from irrelevant, role in the market. It is time for regulatory policies with respect to odd lots to reflect these new realities.

The paper is organized as follows. Section I provides a short history of odd-lot trading. Section II describes the data, provides summary statistics, gives results on the composition and cross-sectional properties of odd-lot trading, investigates how odd lots are used in high-frequency trading, and controls for other factors affecting the growth of odd lots. Section III explores the information content of odd-lot trades and computes price discovery measures for trades of different sizes. Section IV provides a variety of robustness checks, including time aggregation of trades, a VAR analysis of information content, and evidence from more recent data on price discovery. Section V evaluates qualitatively the potential bias for research studies arising from missing trades. Section VI concludes and discusses policy implications. A short Postscript follows. An Internet Appendix provides additional evidence of research biases arising from odd-lot truncation.[8]

I. 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.[9] 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 NYSE (2007)." The odd-lot system featured different reporting rules in that odd-lot trades were segregated from

[7] Behavioral finance studies often rely on dollar trade size cutoffs to determine retail participation and sentiment (see Lee and Radhakrishna (2000)). For higher price stocks, this approach will bias participation rates downward.

[8] The Internet Appendix may be found in the online version of this article.

[9] Odd lots were important from the late 1950s to early 1970s. For a review of the history of odd lots from 1958 to 1976, see the lecture by Paul Miranti and Phil Bradford "Finance Technology and Organization: Automating Odd-Lot Trading at the NYSE, 1958–1976" in the American Finance Association (AFA) history of finance thought video series (available at http://streaming.osu.edu/flash/course/fin694.24wi10/11398-11/).

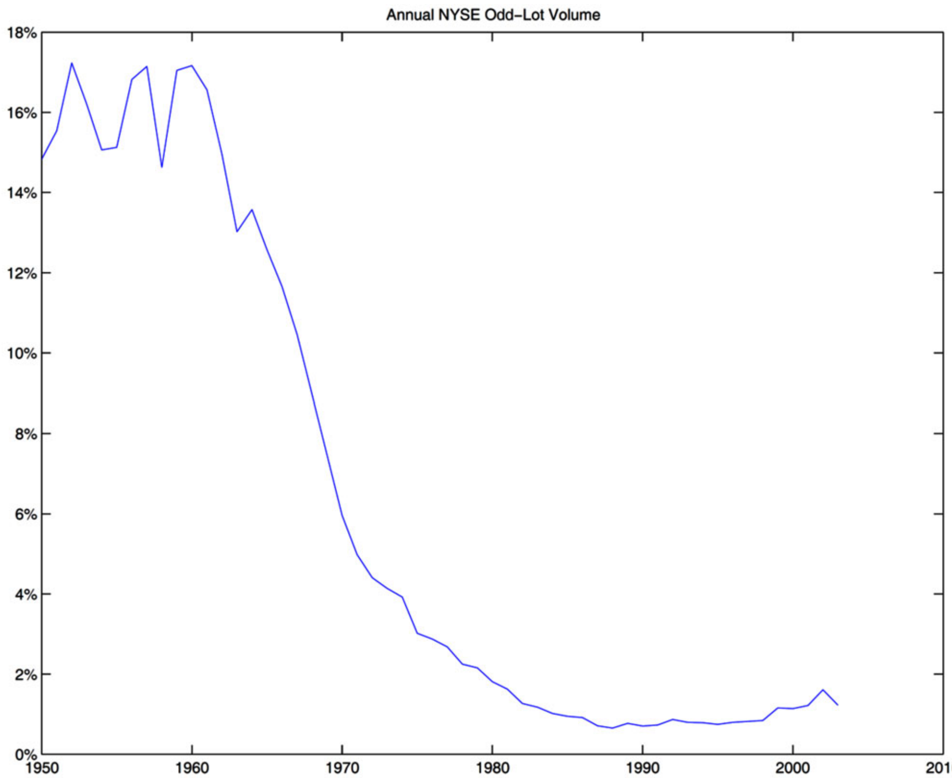

Figure 1. Historical volume of odd lots. This figure shows the historical market shares of NYSE odd lots from 1950 to 2004 in terms of volume. The data are from the NYSE fact book.

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.[10] For the most part, however, odd-lot trading became increasingly less important. Figure 1 shows that by 1990 it accounted for less than 1% of NYSE volume (for discussion on the decline of odd-lot trading, see Wu (1972)).

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

[10] See "NYSE Moves to Prevent Abuses in Odd-Lot Trades," *Wall Street Journal*, November 14, 2007.

stock market by identifying the least informed investors and making investments opposite to them. As Malkiel (1981, p. 140) notes, 'the odd-lotter', according to popular superstitions, is precisely that, kind of person. Thus success is assured by buying when the odd-lotter sells and selling when the odd-lotter buys." While apparently popular in the 1960s and 1970s, this theory found little empirical support and so fell out of common use.

Recent market changes, however, have led to changes in odd-lot trading. In July 2010, the NYSE decommissioned its separate odd-lot trading system, requiring henceforth that odd-lot orders and trades be handled by the same trading system as all other orders and trades.[11] Yet some distinctive features to odd-lot trading remain, especially 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.[12] 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.[13]

II. Data and Analysis

A. Data

The data in this paper come from a variety of sources. Information on price, volume, daily volatility, and market cap are from CRSP. The main data sets we use for transactions data are TAQ, the NASDAQ high-frequency data set (denoted NASDAQ HF), and NASDAQ ITCH data set. The NASDAQ HF data contain trades, inside quotes of the NASDAQ market, and the order book for a sample of 120 U.S. stocks. These stocks are selected to provide a stratified sample of securities representing differing market capitalization levels and listing venues.[14] Table I provides sample statistics on the firms in our study. The trade file for NASDAQ HF data contains each trade conducted on the NASDAQ exchange, excluding trades done in the opening, closing, and intraday crosses, for the sample period 2008 to 2009. To provide evidence on the growth and incidence of odd-lot trading over time, we use data on trade executions

[11] See SEC 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.

[12] As an example, suppose a trader wished to sell 143 shares. If this order were executed in a single trade, then the order to sell 143 shares would be printed to the tape. An order to sell 143 shares that was executed in two trades (say a 100-share trade and a 43-share trade) would appear on the tape as a single trade of 100 shares (the 43-share trade would not appear). If the 143-share order were split into 143 orders for one share each, then none of the trades would appear on the tape.

[13] 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 are reported to the securities information processor (SIP).

[14] Brogaard, Hendershott, and Riordan (2013) show these stocks are representative of the universe of listed stocks trading in U.S. markets.

Table I

Summary Statistics of the 120 Firms in the HF Sample

The table provides summary statistics for the 120 firms in the NASDAQ HF 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 the remaining 40. Spread is the average trade-weighted effective half-spread, which is the absolute difference between the trade price and the quote midpoint; PIN is the probability of informed trading for each stock; Range is the daily high price minus daily low price divided by the daily closing price; Volume is the daily volume; Price is the closing price of the trading day from CRSP; and MarketCap is the market capitalization of the stock on each trading day. Volume and Marketcap are in units of one million. Rankings are based on market caps as of December 31, 2007.

	Mean	StdDev	Max	Min	Type
MarketCap	46,760.98	51,461.22	383,602.92	3,349.12	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	1,554.53	667.15	4,110.46	98.90	medium
Spread	0.04	0.03	0.55	0.01	medium
Range	0.05	0.04	0.87	0.00	medium
Volume	1.00	1.28	23.51	0.02	medium
Price	28.39	18.34	114.17	0.90	medium
PIN	0.15	0.04	0.25	0.02	medium
MarketCap	422.75	248.14	1,797.76	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

(including odd lots) for our sample stocks from the NASDAQ Historical ITCH database for the period January 2, 2010 to November 18, 2011.[15]

The NASDAQ HF data have a number of unique features, three of which are particularly important for our study. First, the HF data include all trades (including odd-lot trades) occurring on the NASDAQ exchange during regular trading hours in 2008 and 2009. This allows us to determine the incidence of odd-lot trading in this period. Second, the data include buy/sell indicators, allowing us to compute trade and imbalance measures without resorting to standard trade classification algorithms.[16] Third, the HF data provide information

[15] Two stocks from the original 120-stock sample were no longer trading in the later sample period. For simplicity, we refer to the ITCH data as covering 2010 to 2011, but note that our data actually end in mid-November and not at year-end. The NASDAQ HF data also carry information for NASDAQ best quote and offer for (1) the first full week of the first month of each quarter during 2008 and 2009 and (2) September 15 to 19, 2008 and February 22 to 26, 2010. We use this quote information to compute the Hasbrouck's (1991a, 1991b) permanent price impact measure as a robustness check.

[16] The buy/sell indicator refers to the liquidity seeking side of the trade.

on whether the traders involved in each trade are high-frequency traders. In particular, trades in the HF data set are categorized into four types: HH for high-frequency traders that take liquidity from high-frequency traders; HN for high-frequency traders that take liquidity from non-high-frequency traders; NH for non-high-frequency traders that take liquidity from high-frequency traders; and NN: non-high-frequency traders take liquidity from non high-frequency traders. These designations allow us to investigate the role and use of odd lots in high-frequency trading strategies. The NASDAQ ITCH data do not include information on high-frequency status or signed trades, and so are not used for analyses needing such inputs.[17]

The NASDAQ data have some limitations. The data include 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 across market centers, but O'Hara and Ye (2011) show 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.

Trades do occur off exchange, however, due to practices such as preferencing and internalization.[18] Such trades are reported to Trade Reporting Facilities (TRFs), which in turn report those trades to the consolidated tape, but again odd lot trades are not reported. Because many retail trades are subject to preferencing arrangements, it is likely that odd lots are more common on TRFs, although data to determine this are not available. The SEC (2010) reported that odd lot share volume for the market as a whole was 4% of total share volume, a number that closely tracks what we find in the NASDAQ HF data. It seems reasonable to assume, therefore, that odd-lot behavior in NASDAQ is typical of that in the larger market. However, to the extent that TRF odd-lot trading is larger, our results on the incidence of odd-lot trading will be understated.[19]

B. Odd-Lot Trades and Volume: How Much Is Missing?

Figure 2 demonstrates the time-series pattern of odd-lot trades and volume for the period 2008 to 2011. Panel A shows that in January 2008 about 14% of total trades are odd-lot trades and hence are missing from the consolidated tape and TAQ data, with this number increasing to about 25% by November

[17] NASDAQ TotalView-ITCH is a series of messages that describes orders added to, removed from, and executed on the NASDAQ. This data set provides much less information than the NASDAQ high-frequency data set. For example, it does not provide trader type, nor does the data set directly carry information on the best bid and ask. This restricts the type of tests we can conduct using the data set. Fortunately, ITCH data do include all the trades, which allows us to calculate the market share of odd lots as well as the weighted price contribution.

[18] Trading also takes place in crossing networks, but trades there are batched so odd lots are uncommon. Crossing networks also report trades to TRFs.

[19] The NASDAQ TRF is the largest of the active TRFs, and correspondence with Jeffrey Smith of NASDAQ indicated that TRF odd lot trading there was similar to that found in the HF database.

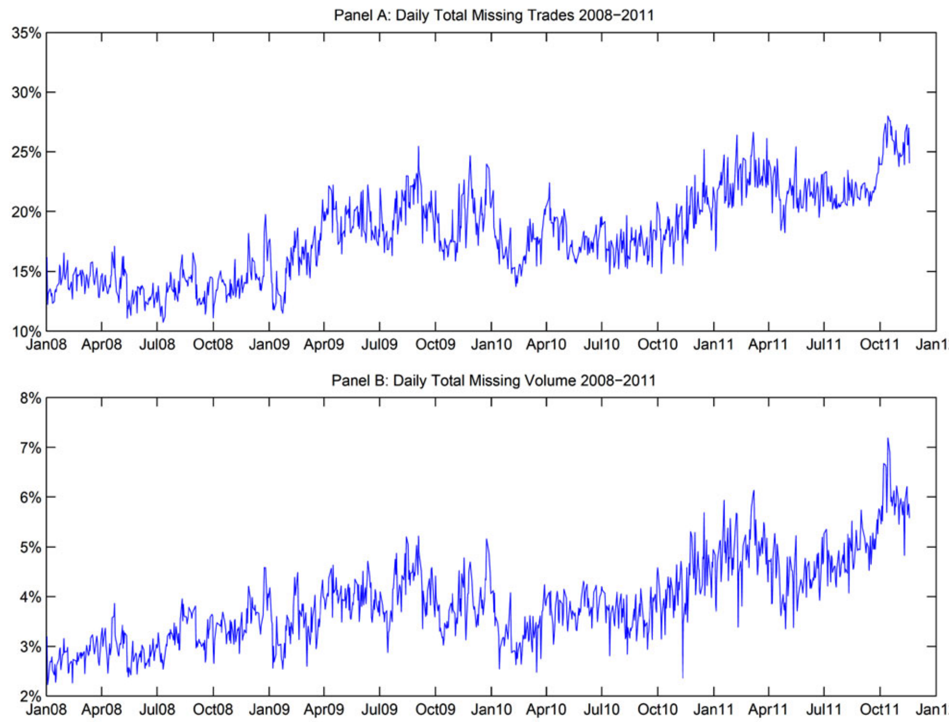

Figure 2. Time-series variation in odd-lot trades. This figure shows the total level of odd-lot trades and volume from January 2008 to November 2011. Panel A depicts the number of odd-lot trades as a percentage of total transactions. Panel B depicts the odd-lot trade volume as a percentage of total trades.

2011. Panel B shows that odd-lot share volume is about 2.25% of total share volume in January 2008, rising to about 6% at the end of 2011. While 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 we investigate these factors later in this section.

Table II gives the level of odd-lot trades and volumes for the 15 largest stocks in our sample (Panel A), and for the 15 stocks with the largest increase in odd-lot trades over the 2008 to 2011 period (Panel B). Figure 3 presents further details on the cross-sectional variation of odd-lot trading.

A number of large, well-known firms have substantial numbers of odd-lot trades, and these appear to be growing over time. For example, Google's fraction of odd-lot trades increased from almost 31% in 2008 to 53% in 2011. Over this same period, Amazon's odd-lot trades increased from approximately 16% to 46%, while Apple's increased from 17% to 38%. Some firms, for example, GE and Cisco, had little change in odd lots over this period. For the largest stocks, odd-lot transactions increased from 12.65% to 20.55% over this period and for

Table II

Sample Stocks and Odd-Lot Trades and Volumes

This table reports odd lots as a percent of all trades and trading volume for the 15 largest stocks as well as for the 15 stocks with the highest growth in odd lots in our 120 stock sample. The results for 2008 to 2009 are based on NASDAQ HF data and the results for 2010 to 2011 are based on NASDAQ ITCH data from January 2, 2010 to November 18, 2011.

	2008		2009		2010		2011	
Stock	Trades (%)	Volume (%)	Trades (%)	Volume (%)	Trades (%)	Volume (%)	Trades (%)	Volume (%)
Panel A: Large Market Cap Stock Odd Lots Percentage								
GE	8.80	1.58	7.73	0.93	8.31	0.84	8.85	1.01
PG	10.93	3.62	17.59	5.88	15.43	5.21	17.23	5.71
AAPL	17.12	5.65	23.88	8.61	26.34	9.50	38.47	13.92
CSCO	9.96	1.26	8.95	1.28	7.60	0.86	8.00	0.66
GOOG	30.92	11.78	38.94	16.99	44.63	19.68	52.95	22.96
PFE	8.57	1.11	9.51	1.36	8.31	0.91	8.22	0.97
INTC	8.95	1.05	9.19	1.16	7.93	0.85	9.56	0.93
HPQ	10.90	3.39	14.21	4.16	11.33	3.26	15.01	4.32
DIS	9.04	2.53	15.40	4.14	12.50	3.73	16.66	5.82
AXP	12.65	3.94	17.34	4.80	15.03	5.61%	21.84	8.18
MMM	16.19	5.75	24.58	8.96	28.76	11.29	29.71	12.34
DELL	10.25	1.79	10.06	1.43	9.62	1.25	10.28	1.39
AMG	14.52	4.13	19.90	6.35	19.85	6.49	26.38	9.39
HON	10.87	3.64	17.52	5.76	17.50	6.49	24.31	10.43
EBAY	10.07	2.27	11.28	2.21	10.09	2.17	20.78	6.51
Avg.	12.65	3.57	16.40	4.93	16.21	5.21	20.55	6.97
Panel B: Stocks with Most Odd-Lot Trade Growth								
ISRG	34.39	13.97	34.96	13.82	49.06	22.46	65.67	30.88
AMZN	16.23	5.02	28.67	8.77	26.44	9.70	46.03	17.97
CTSH	11.72	3.70	18.50	5.44	25.83	9.00	37.60	14.51
SJW	17.63	4.76	42.19	12.44	50.46	17.96	39.89	17.29
GOOG	30.92	11.78	38.94	16.99	44.63	19.68	52.95	22.96
CRVL	30.20	10.69	61.34	24.47	63.76	26.09	51.81	22.16
AAPL	17.12	5.65	23.88	8.61	26.34	9.50	38.47	13.92
LANC	21.45	9.18	28.16	11.17	37.63	15.42	41.54	18.83
CELG	16.96	5.75	21.52	7.15	27.48	9.94	36.58	13.51
GAS	16.59	5.07	38.91	12.63	34.16	14.13	35.59	16.39
BIIB	18.71	6.64	26.39	9.15	23.09	7.79	36.94	15.11
NC	32.97	13.19	49.93	15.84	58.05	29.43	51.01	22.56
AGN	18.63	5.73	26.77	8.58	31.72	11.91	36.61	15.19
AZZ	17.52	6.34	33.74	12.01	37.85	14.53	34.65	14.01
PPD	22.76	7.51	37.92	13.05	39.50	15.16	39.73	16.26
Avg.	21.59	7.67	34.12	12.01	38.40	15.51	43.00	18.10

the stocks with the greatest odd-lot growth, odd lots increased from 21.59% to 43% of trades.

Institutions are generally thought to trade larger stocks, so odd lots may be more prevalent in the smaller stocks favored by retail traders. To test this

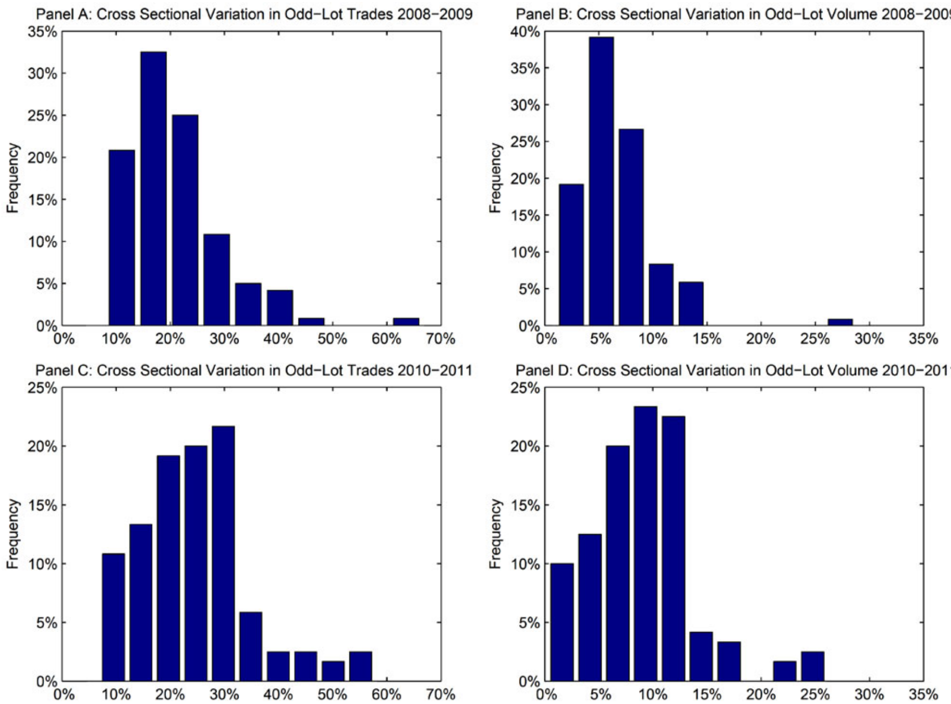

Figure 3. Cross-sectional variation in odd-lot trades. This figure shows the level of missing trades across the 120 stocks. Panel A depicts the percentage of odd-lot trades in the 2008 to 2009 NASDAQ HF sample and Panel B depicts the percentage of odd-lot volume in the 2008 to 2009 NASDAQ HF sample. Panel C depicts the percentage of odd-lot trades in the 2010 to 2011 NASDAQ ITCH sample and Panel D depicts the percentage of odd-lot volume in the 2010 to 2011 NASDAQ ITCH sample.

conjecture, we divide the 120 stocks into 40 large, 40 medium, and 40 small market capitalization groups, and we calculate odd-lot percentages by aggregating the NASDAQ HF sample and the ITCH samples. Panel A of Table III provides supporting evidence: odd lots are 19.6%, 22.2%, and 25% of trades for large firms, medium-size firms, and small firms, respectively. 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.

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 divide the 120 stocks into 40 low, 40 medium, and 40 high stock price groups. Panel B of Table III shows that high-priced stocks are more likely to have odd-lot trades, with 26.9% of transactions in odd lots. The percentage of odd-lot trades in low-priced and medium-priced stocks varies over time, but even in low-priced stocks we find odd lots of more than 19% in

Table III

Odd-Lot Trades by Market Cap and Price

This table presents the odd-lot trades based on market cap and price groups. Panel A divides the 120 stocks into large, medium, and small market cap groups, each of which contains 40 stocks. Panel B divides the 120 stocks into high, medium, and low price groups, each of which has 40 stocks. We aggregate the NASDAQ HF data from 2008 to 2009 and ITCH data from January 2, 2010 to November 18, 2011 in this calculation. The table also tests the hypothesis that the average level of odd lots is equal across different groups. *t*-statistics are presented in parentheses. ***, **, and * indicate significance at the 1%, 5%, and 10% level, respectively.

Panel A: By Market Capitalization

	Large	Medium	Small	Small–Medium	Medium–Large	Small–Large
Ratio of Missing Trades	0.196	0.222	0.250	0.029	0.026	0.055***
				(1.39)	(1.42)	(2.62)
Ratio of Missing Volume	0.062	0.076	0.087	0.011	0.014*	0.025***
				(1.36)	(1.75)	(2.76)

Panel B: By Price

	High	Medium	Low	Low–Medium	Medium–High	Low–High
Ratio of Missing Trades	0.269	0.203	0.196	−0.007	−0.067***	−0.074***
				(−0.40)	(−3.06)	(−3.91)
Ratio of Missing Volume	0.097	0.065	0.064	−0.001	−0.032***	−0.033***
				(−0.16)	(−3.48)	(−4.22)

our sample. This result suggests that the motives for odd-lot trades may be more complex than in times past.

C. Who Is Trading Odd Lots and How?

Understanding current odd-lot usage requires recognizing the new role played in markets by high-frequency trading. High-frequency traders follow a variety of trading strategies, but virtually all of these strategies involve the use of algorithms to slice, dice, and send massive numbers of orders to trading venues. As noted earlier, the NASDAQ HF data set differentiates traders into

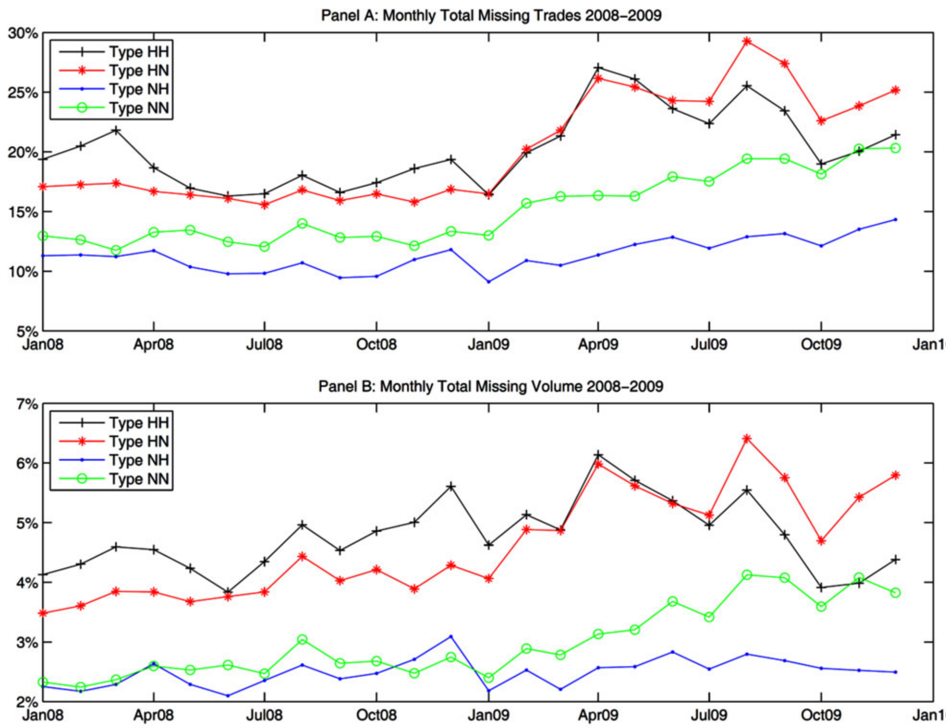

Figure 4. Odd-lot trades by trader type. This figure displays the time-series odd-lot percentage by four different trade types. The first letter indicates the liquidity taker and the second the liquidity maker where H stands for high-frequency trader and N stands for non-high-frequency trader. For example, an HN trade means that a high-frequency trader takes liquidity from a non-high-frequency trader.

high frequency and non-high-frequency categories, and it also distinguishes who was the maker or taker of liquidity in each trade. These data allow us to investigate more carefully the question of who is trading odd lots and how.

Figure 4, Panel A, provides the ratio of odd-lot trades relative to the total number of trades for each trader type (HH, HN, NH, and NN). The figure shows that odd lots are more likely to occur when trades are initiated by high-frequency traders. About 20% to 25% of HH and HN 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 high-frequency traders that generally face much lower trading costs than retail traders. The results here are consistent with that hypothesis.

The histograms of odd-lot trades in Figure 5 show 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

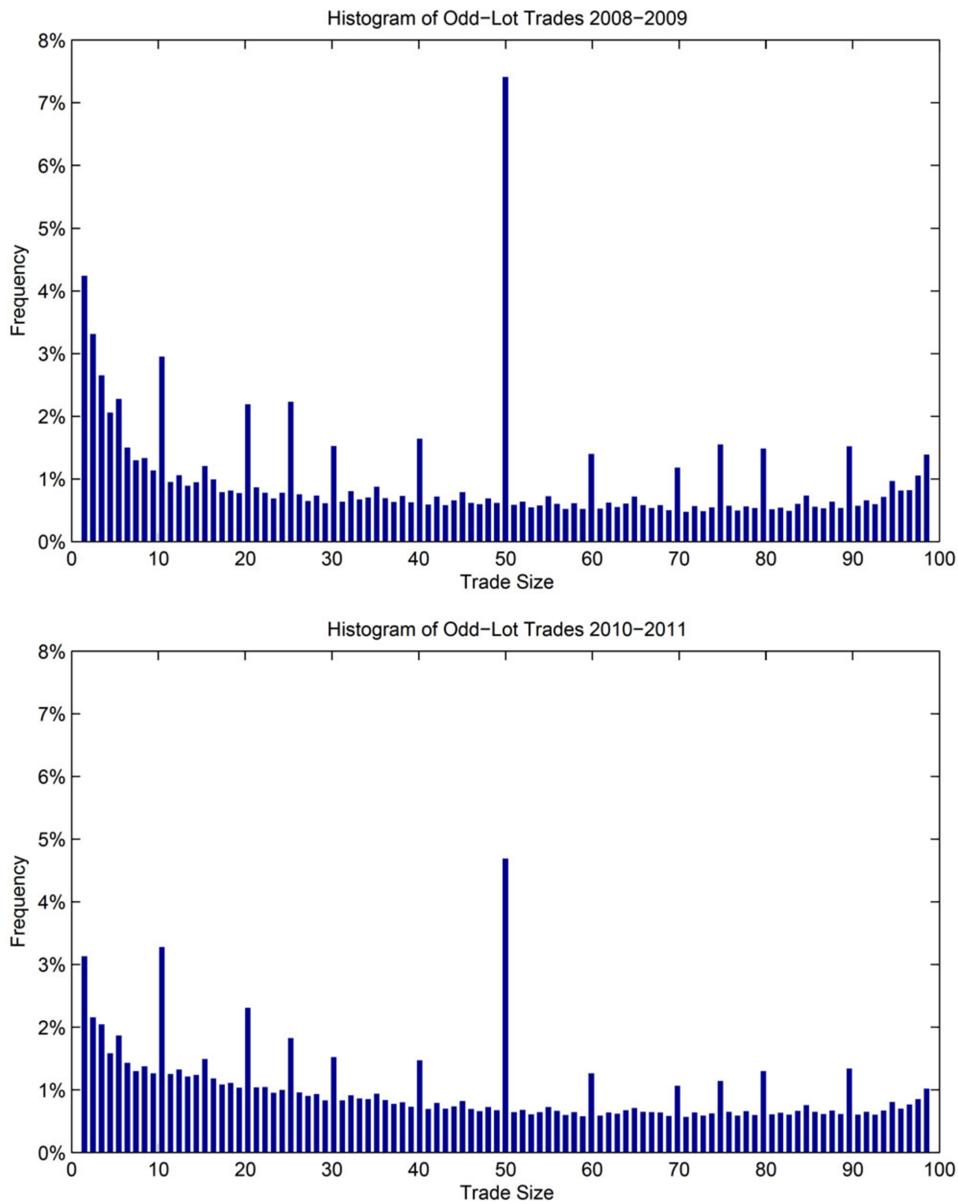

Figure 5. Histogram of odd-lot trades.

most frequent trade size. Second, trades of one share are the second most frequent trade size in the 2008 to 2009 sample and third largest in the 2010 to 2011 sample. That trade clusters at particular price increments has long been observed in equity markets (see, for example, Harris (1991), Christie and

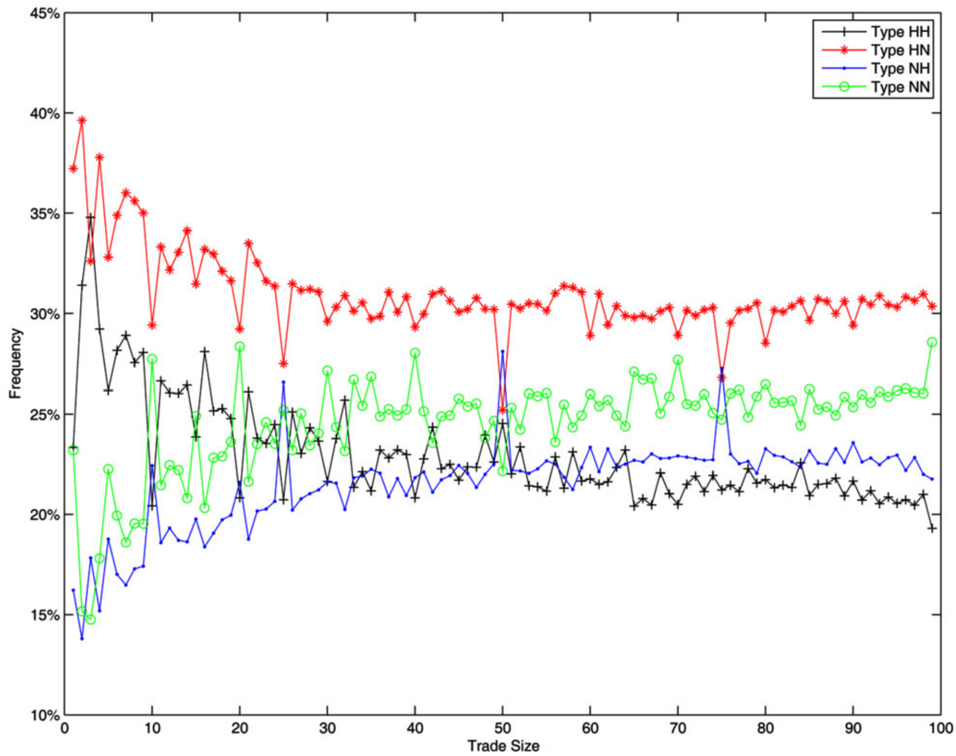

Figure 6. Odd-lot trades by trader type. This figure displays the usage of specific odd-lot sizes by four different trade types. The first letter indicates the liquidity taker and the second the liquidity maker where H stands for high-frequency trader and N stands for non-high-frequency trader. For example, an HN trade means that a high-frequency trader takes liquidity from a non-high-frequency trader. The sample period is 2008 to 2009.

Schultz (1994)). Our finding that odd-lot quantities cluster raises a variety of questions as to how odd lots are being used in markets and by whom.[20]

We can get more insights into these strategies by determining who is trading various odd-lot sizes, which we do in Figure 6. Focusing on trades initiated by high-frequency traders, the data show two interesting patterns. First, the market share of HN and HH traders decreases in odd-lot size, implying that high-frequency traders are more likely to initiate very small odd-lot trades. Indeed, almost 60% of one-share odd-lot trades are initiated by high-frequency traders. Second, the market share of high-frequency traders dips down for round-lot numbers such as 10, 25, 50, while it jumps up for non-high-frequency usage. This pattern reflects the new reality in markets that "silicon traders" (i.e., machines) are not predisposed to favor one number over another, unlike human traders, who prefer to trade in round numbers. This greater tendency

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

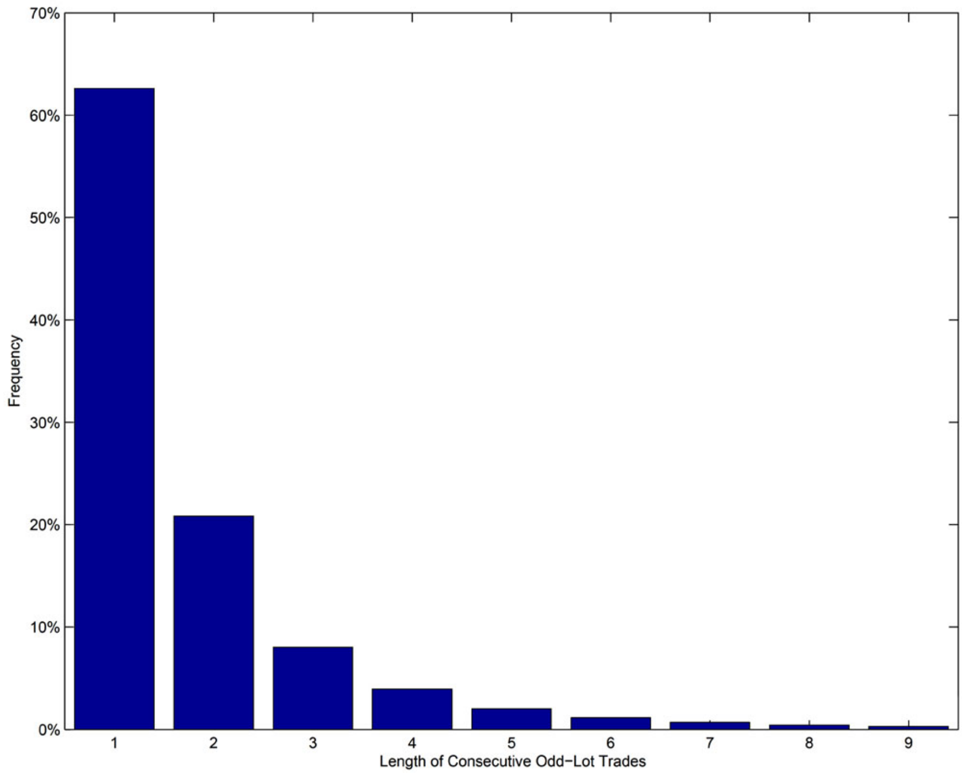

Figure 7. Length of consecutive odd-lot sequences. This figure displays a histogram of the number of consecutive odd lots in each odd-lot sequence, which is defined as a sequence of odd lots without round or mixed lots between them.

of humans to use round numbers also means that silicon traders can exploit the predictable tendencies of their human counterparties (see Easley, Lopez de Prado, and O'Hara (2012b) for more discussion).

Odd lots can also be generated by mechanical reasons due to order mismatching. For example, suppose the first order of the day is a 50-share buy and that subsequently sell and buy orders of 100 shares appear alternatingly. Then the 50-share buy may result in a trade of 50 shares, in which case the sell order has 50 shares remaining, which may then execute against half of the next buy order, and so on. To investigate the importance of this effect, we calculate the incidence of sequences of odd-lot trades in the data. Figure 7 presents the histogram of these sequences, where each sequence is defined as a group of odd lots without a round-lot trade between them.[21] More than 60% of odd-lot sequences have only one odd-lot, while another 20% have two odd lots. More than 99% of odd-lot sequences have less than nine odd lots. The data suggest

[21] Figure 7 displays the histogram up to the 99th percentile of the observations, since the graph has a very long right-hand tail.

that the "odd-lot cascade" is not strong enough to explain the large number of odd lots in the data.

Finally, odd lots can originate for less benign reasons. 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 is splitting orders into other trade sizes. Interestingly, we find that most odd-lot trades below 50 shares fall into the 1- to 5-share bin, and most odd-lot trades above 50 shares fall into the 95- to 99-share bin.

Table IV gives an example for Apple (AAPL) trades on June 20, 2008. At 13:59:01:107, 111 odd-lot trades occurred in the same millisecond with the same direction and price, all of which are HN trades (high-frequency traders taking liquidity from non-high-frequency traders). The total volume for all these trades is only 2,995 shares. Three milliseconds later, we see another 102 odd-lot trades of the HN type with the same direction and price, which result in a volume of 2,576 shares. Such patterns are consistent with sophisticated traders (high-frequency traders, in this particular case) slicing and dicing their orders and hiding 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.

D. Odd-Lot Regression Results

As an additional diagnostic to understand the incidence of odd-lot trades, we run between-effect, random-effect, and within-effect (fixed-effect) regressions on a panel containing information on the percentage of odd-lot trades and volume, as well as daily price, volume, and volatility. The between-effect regression allows us to explore cross-sectional variation in odd-lot trades and volume. We regress the cross-sectional mean level of odd-lot trades and volume on the price level and the proportional effective spread, which we use as a proxy for liquidity. We include daily price range 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.[22] Finally, we include an NYSE dummy variable 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. Our estimating equations are given by

$$\overline{OLtrade_i} = \mu + \beta_1 * \overline{\log prc_i} + \beta_2 * \overline{spread_i} + \beta_3 pinge100_i + \beta_4 * \overline{range_i} + \beta_5 NYSE_i + \overline{\varepsilon_i}, \tag{1}$$

$$\overline{OLvol_i} = \mu + \beta_1 * \overline{\log prc_i} + \beta_2 * \overline{spread_i} + \beta_3 pinge100_i + \beta_4 * \overline{range_i} + \beta_5 NYSE_i + \overline{\varepsilon_i}. \tag{2}$$

[22] PIN can be estimated based on all trades or on the trades of 100 shares or more in NASDAQ. We employed both measures, and report the PIN measure based on trades greater than 100 shares (results are very similar with either calculation). Nevertheless, missing trades also pose a challenge for estimating PIN measures with TAQ data.

Table IV

Example of Odd-Lot Pattern

This table provides an example of a sequence of odd-lot trades on February 6, 2008. The patterns are generated by high-frequency traders taking liquidity from non-high-frequency traders. There are 111 odd-lot sells at 13:59:01:107, for a total of 2,995 shares. Another 102 odd-lot sells happened three milliseconds later, for a total of 2,576 shares. The first letter in the type variable refers to the liquidity taker and the second one the liquidity maker; letter H designates high-frequency traders and N designates non-high-frequency traders.

Sequence	Symbol	Hour	Minute	Second	Millisecond	Shares	BuySell	Price	Type
1	AAPL	13	59	1	107	20	S	125.00	HN
2	AAPL	13	59	1	107	10	S	125.00	HN
3	AAPL	13	59	1	107	50	S	125.00	HN
4	AAPL	13	59	1	107	25	S	125.00	HN
5	AAPL	13	59	1	107	12	S	125.00	HN
6	AAPL	13	59	1	107	35	S	125.00	HN
7	AAPL	13	59	1	107	10	S	125.00	HN
8	AAPL	13	59	1	107	12	S	125.00	HN
9	AAPL	13	59	1	107	24	S	125.00	HN
10	AAPL	13	59	1	107	6	S	125.00	HN
11	AAPL	13	59	1	107	4	S	125.00	HN
12	AAPL	13	59	1	107	75	S	125.00	HN
13	AAPL	13	59	1	107	1	S	125.00	HN
14	AAPL	13	59	1	107	15	S	125.00	HN
15	AAPL	13	59	1	107	50	S	125.00	HN
				....					
108	AAPL	13	59	1	107	50	S	125.00	HN
109	AAPL	13	59	1	107	50	S	125.00	HN
110	AAPL	13	59	1	107	30	S	125.00	HN
111	AAPL	13	59	1	107	3	S	125.00	HN
112	AAPL	13	59	1	110	47	S	125.00	HN
113	AAPL	13	59	1	110	80	S	125.00	HN
114	AAPL	13	59	1	110	80	S	125.00	HN
115	AAPL	13	59	1	110	8	S	125.00	HN
116	AAPL	13	59	1	110	8	S	125.00	HN
117	AAPL	13	59	1	110	60	S	125.00	HN
118	AAPL	13	59	1	110	8	S	125.00	HN
119	AAPL	13	59	1	110	32	S	125.00	HN
120	AAPL	13	59	1	110	30	S	125.00	HN
				....					
210	AAPL	13	59	1	110	5	S	125.00	HN
211	AAPL	13	59	1	110	25	S	125.00	HN
212	AAPL	13	59	1	110	50	S	125.00	HN
213	AAPL	13	59	1	110	12	S	125.00	HN

The results are given in Table V. As expected, high-price stocks have more odd-lot trades and volumes. Neither the daily price range relative to price nor stock listing venue have explanatory power for the cross-sectional variation in odd-lot trades and volume. The level of liquidity, however, does affect odd-lot trading. We find that the number of odd-lot trades and odd-lot volume increase in the proportional spread, suggesting that stocks with lower liquidity have greater odd-lot trading. We also find that stocks with a higher PIN have higher levels of

Table V

Variation of Odd-Lot Trades and Volume

This table reports results of the variation in missing trades and volume. We run between-, random-, and fixed-effect regressions on the panel of missing trades and volume for each stock on each day. *OLTrade%* and *OLVol%* are the percentage of missing trades and volume; log*prc* is the price level; *spread* is the bid-ask spread; *pinge*100 is the probability of informed trading for each stock for trades greater than 100 shares; and *range* is daily price range; *NYSE* is equal to one if the stock is listed on NYSE and zero if it is listed on NASDAQ. The sample period is 504 trading days from 2008 to 2009 *t*-statistics are presented in parentheses. ***, **, and * indicate significance at the 1%, 5%, and 10% level, respectively.

	(1)	(2)	(3)	(4)	(5)	(6)
	OLTrade%	*OLVol%*	*OLTrade%*	*OLVol%*	*OLTrade%*	*OLVol%*
log*prc*	0.068***	0.035***	0.008***	0.007***	0.012***	0.009***
	(7.26)	(8.48)	(5.20)	(9.34)	(6.90)	(11.24)
*pinge*100	0.555***	0.267***	0.543***	0.283***		
	(5.75)	(6.29)	(4.87)	(5.40)		
spread	12.918***	7.79***	0.221***	0.074***	0.062***	0.027***
	(3.21)	(4.41)	(3.39)	(2.42)	(10.59)	(9.91)
range	0.039	−0.017	−0.248***	−0.123***	−0.287***	−0.140***
	(0.07)	(−0.07)	(−18.24)	(−19.29)	(−18.81)	(−19.56)
NYSE	−0.002	−0.004	−0.008	−0.007		
	(−0.16)	(−0.67)	(−0.52)	(−0.95)		
Constant	−0.098**	−0.082***	0.136***	0.028***		
	(−2.00)	(−3.79)	(7.00)	(3.06)		
Effect	Between	Between	Random	Random	Fixed	Fixed
Observations	60,412	60,412	60,412	60,412	60,412	60,412
R^2	0.482	0.561	0.008	0.011	0.583	0.563
No. tickers	120	120	120	120	120	120

odd-lot trades. This latter result is consistent with informed traders breaking trades into odd lots so as to better hide their information. The regression R^2 is 56.1%, meaning that about two-thirds of cross-sectional variation in odd-lot volume is explained by these variables.

We next run equations (3) and (4) using random effects. The regressions take the form

$$OLtrade_{i,t} = \mu + \alpha_i + \gamma_t + \beta_1 * \log prc_{i,t} + \beta_2 * spread_{i,t} + \beta_3 * pinge100_i + \beta_4 * range_{i,t} + \beta_5 NYSE_i + \varepsilon_{i,t}, \tag{3}$$

$$OLvol_{i,t} = \mu + \alpha_i + \gamma_t + \beta_1 * \log prc_{i,t} + \beta_2 * spread_{i,t} + \beta_3 * pinge100_i + \beta_4 * range_{i,t} + \beta_5 NYSE_i + \varepsilon_{i,t}. \tag{4}$$

The results are very similar, except we now find that higher volatility as measured by the daily price range results in lower odd-lot trades and volume. Engle, Ferstenberg, and Russell (2006) 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, but splitting trades over time leads to execution risk because it is hard to predict future price changes. This risk is certainly higher when volatility is high, so our results here are consistent with their result.

Finally, we run the following two regressions using a fixed effect model:

$$OLtrade_{i,t} = \alpha_i + \gamma_t + \beta_1 * \log prc_{i,t} + \beta_2 * spread_{i,t} + \beta_3 * range_{i,t} + \varepsilon_{i,t}, \quad (5)$$

$$OLvol_{i,t} = \alpha_i + \gamma_t + \beta_1 * \log prc_{i,t} + \beta_2 * spread_{i,t} + \beta_3 * range_{i,t} + \varepsilon_{i,t}. \quad (6)$$

Since PIN and listing venue do not vary over time and are captured by the dummy coefficients, they are not included in these regressions. The findings reported in columns (5) and (6) are again similar: higher price, lower liquidity, and lower volatility lead to more odd-lot trades and volume.

III. Do Odd-Lot Trades Move Prices?

The results of the previous section show that odd-lot trades are now part of a variety of trading strategies used by high frequency traders. Such trades may well have information content for future price movements.[23] To investigate the informativeness of odd-lot trades, we follow the literature using weighted price contribution, which measures how much of a stock's cumulative price change or return change is attributable to trades in particular trade-size categories (see, for example, Barclay and Warner (1993), Chakravarty (2001), Choe and Hansch (2005), and Alexander and Peterson (2007)). In this section, we provide results using the HF data set. We provide robustness checks using the ITCH data as well as an alternative price informativeness measure proposed in Hasbrouck (1991b) in the next section.

A. Weighted Price Contribution

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

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

where $\delta_{n,j}$ is an indicator variable that takes the value of one if the nth trade belongs to size category j, and zero otherwise.

[23] A large literature in microstructure looks at the informativeness of stock trades, with the general conclusion being that trades from informed traders permanently move prices, while trades from uninformed traders have more transient price effects (see Hasbrouck (1991a)).

Barclay and Warner (1993) define $r_n^{s,t}$ as the difference between the price of trades n and $n-1$.[24] The weighted cross-sectional average price contribution following Barclay and Warner (1993) is calculated as follows. The weight for stock s on day t for the weighted price contribution measure is the ratio of the stock's absolute cumulative price change to the sum of all stocks' absolute cumulative price changes on day t.[25] 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 stock s on day t. The weight for stock s on day t is defined as

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

The weighted price contribution 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. Then, the weighted price contribution of trades in size category j is defined as

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

Table VI presents results on price discovery by trade size.[26] Several results are striking. First, approximately 80% to 85% of price discovery is accounted for by trades of 100 shares or less. Barclay and Warner (1993) find that medium-sized trades are most informative, but this is clearly no longer the case—it is the smaller trades that are moving the markets. Second, the less-than-100-share trade category is responsible for 35% of weighted price contribution in the 2008 to 2009 period. Since odd lots over this period comprise in aggregate only 16% of trades and 3.4% of volume, the information content of odd lots far exceeds their incidence, consistent with odd-lot trades being used by informed traders.

[24] Choe and Hansch (2005) define $r_n^{s,t}$ as the log return between the price of trades n and $n-1$. We also calculate the weighted price contribution based on Choe and Hansch (2005), and the result is similar.

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

[26] 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.

Table VI
Price Discovery, Share of Trades, and Share of Volume for Each Size Category

This table reports the weighted price contribution for each order size category using individual trades. WPC is the weighted price contribution using price changes. Share of trades (volume) gives the percentage of trades (volume) in each size category. The data are from the NASDAQ HF database for the 2008 to 2009 sample period.

Trade Size Category	WPC	Share of Trades	Share of Volume
<100	0.354	0.158	0.034
100	0.497	0.54	0.281
200	0.041	0.117	0.121
300	0.01	0.041	0.065
400	0.016	0.025	0.053
500	0.006	0.024	0.062
100–500	0.615	0.791	0.633
501–900	0.012	0.027	0.099
901–1,900	0.011	0.017	0.109
1,901–4,900	0.008	0.005	0.078
4,901–9,999	0.000	0.001	0.028
501–9,999	0.031	0.051	0.313
≥10,000	0.000	0.000	0.019

B. Sources of Cumulative Price Changes: Formal Tests

The stealth trading hypothesis by Barclay and Warner (1993) holds that informed traders concentrate in particular size categories and that price movements are due mainly to informed traders' private information. 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 posits that the release of public information causes most stock price changes. 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, in contrast, implies that the price contributions would not be proportional.

Following Barclay and Warner (1993), we run weighted least squares regressions of the price contribution on two trade-size category dummies and the percentage of transactions in that category. The regression equation is as follows:

$$PC_j^{s,t} = d_{<100}\delta_{<100}^{s,t} + d_{\geq 100}\delta_{\geq 100}^{s,t} + \beta * pct\_transcation_j^{s,t} + \varepsilon_j^{s,t}, \tag{11}$$

where $PC_j^{s,t}$ 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 greater than or equal to 100 shares. The terms $\delta_{<100}^{s,t}$ and $\delta_{\geq 100}^{s,t}$ are indicator variables that take the value of one if $PC_j^{s,t}$ falls into their trade category, and zero otherwise;

Table VII

Test for Price Discovery

This table reports weighted least square regressions of price contribution on a dummy for the less-than-100-share category, a dummy for the greater-than-or-equal-to-100-share category, and the percentage of transactions or percentage of trading volume in that category. The dependent variable, price contribution for stock s on day t in 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 weight is the ratio of the stock s absolute cumulative price change to the sum of all stocks' absolute cumulative price changes on day t. The null hypothesis is that the coefficients on the dummies in each category are equal to zero and the coefficient on the percentage of transactions or percentage of trading volume in that category is equal to one. t-statistics are given in parentheses. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% level, respectively.

	(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	<0.0001	<0.0001

$d_{<100}$ and $d_{\geq 100}$ are coefficients on these indicator variables. The term β is the coefficient on the percentage of transactions for stock s on day t of trade size category j. The regression weight is the ratio of the absolute cumulative price change of stock s on day t to the sum of all stocks' absolute cumulative price changes on day t. Regression (1) in Table VII reports the results.

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

An alternative trading volume hypothesis holds 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 VII reports a 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:

$$PC_j^{s,t} = d_{<100}\delta_{<100}^{s,t} + d_{\geq 100}\delta_{\geq 100}^{s,t} + \beta * pct\_trdqty_j^{s,t} + \varepsilon_j^{s,t}, \tag{12}$$

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

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

IV. Robustness Checks

In this section, we provide a number of robustness checks on our results on the incidence and informational content of odd-lot trades. First, we aggregate trades occurring in the same millisecond with the same active side, the same direction (buy or sell), and the type (HH, HN, NH, and NN) into one large trade. This aggregation addresses the concern that odd-lot trades may come from a single active order interacting with multiple passive orders on the book. Second, we examine an alternative measure of price impact following Hasbrouck (1991a, 1991b). This analysis is designed to address concerns that the weighted price contribution methodology is not appropriate for current high-frequency markets. Finally, we investigate the price informativeness of trades in the ITCH data, which allows us to examine how this measure changes over time. These analyses show that our main results—that odd lots are now a substantial fraction of market activity and that odd-lot trades are informative of future price movements—are robust.

A. Aggregate Trades in the Same Millisecond

If a large active order interacts with multiple passive orders on the book, then the resulting trade prints may overstate the actual incidence of odd lots in the market. To address this concern, we combine reported trades within the same millisecond and with the same active side, same direction (buy or sell), and same type (HH, HN, NH, and NN) into one large trade. Note that this aggregation, while allaying concerns regarding overestimation of odd-lot trades, is also likely to underestimate both the incidence and price impact of odd lots. To see why, recall the example in Table IV where 111 odd-lot trades executed at 13:59:01:107 on June 20, 2008 are followed by another 102 odd lots executed at 13:59:01:110. After aggregation, these trades will be treated

Table VIII
Weighted Price Contribution by Aggregating Trades within the Same Millisecond

This table reports the weighted price contribution of odd lots where trades within the same millisecond with the same trade direction (buy or sell) and same type (HH, HN, NH, and NN) are aggregated into a single trade. The sample period is from 2008 to 2009.

Trade Size Category	WPC (%)	Share of Trades (%)	Share of Volume (%)
<100	20.36	5.94	0.14
100	35.26	24.53	1.48
200	8.63	12.36	1.49
300	6.46	7.00	1.27
400	3.63	5.12	1.24
500	2.22	4.11	1.24
100–500	66.79	58.71	7.57
501–900	6.75	9.82	4.23
901–1,900	3.16	9.70	7.66
1,901–4,900	2.04	7.56	14.03
4,901–9,999	0.46	4.07	16.91
501–9,999	12.42	31.15	42.82
≥10,000	0.44	4.20	49.47

as two large trades instead of a number of small trades. If algorithms slice the original parent order into odd-lot child orders to reduce the price impact of trading, then aggregation will obliterate this effect. Moreover, if traders slice orders into odd lots at the submillisecond level to escape trade reporting requirements, then aggregation will similarly underestimate the true price impact of the odd lots. Gai, Yao, and Ye (2012) find that high-frequency traders could cancel their orders at two to six microseconds in January 2010, so it seems sensible that they could slice trades in less than one millisecond in our sample period.

As expected, aggregation leads to a dramatic decrease in odd-lot trades and volume. Table VIII shows that, while odd lots fall to 5.94% of trades and 0.14% of volume, they still contribute 20.36% in weighted price contribution. It may seem strange that so few trades can have such a large price contribution, but it arises because the weighted price contribution is a signed measure in which individual trades can have a positive and negative weighted price contribution. The reference is the open-to-close return. Therefore, trades moving the price in the same direction as the daily return contribute positively to weighted price contribution, whereas trades moving the price in the opposite direction of the daily price movement contribute negatively to the weighted price contribution. As a result, price impact for a trade-size category can be zero, if buy and sell trades are equal in number and they move the price by the same magnitude. The outsized effects of odd-lot trades arise because these trades are more likely to be on the correct side of the price movement.

B. The Hasbrouck Price Impact Measure

The original stealth trading hypothesis in Barclay and Warner (1993) only uses weighted price contribution to measure the informativeness of the trade. Here, we support our results using an alternative measure of price impact: Hasbrouck's (1991a, 1991b) permanent price measure. Using this approach, we can measure whether executed odd lots or round and mixed lots have a more permanent impact on prices. To this end, we estimate the return and executed order dynamics in a structural VAR framework. We follow the method of Barclay, Hendershott, and McCormick (2003) and Chaboud et al. (2014) and estimate the impulse response function.

Specifically, we estimate the following system of equations, where r_t is the midpoint return during the one-minute interval, x_t^{odd} is the sum of the signed odd-lot volume (buy-initiated volume minus sell-initiated volume) during the one-minute interval, and $x_t^{rd,mix}$ is the sum of the signed round and mixed lots volume during the one-minute interval, and where we follow Hasbrouck (1996) to calculate the price impact for half an hour, that is, we estimate the VAR system with 30 lags:

$$\begin{aligned} r_t &= \sum_{i=1}^{30} \alpha_i r_{t-i} + \sum_{i=0}^{30} \beta_i x_{t-i}^{odd} + \sum_{i=0}^{30} \gamma_i x_{i,t-i}^{rd,mix} + \varepsilon_{1,t} \\ x_t^{odd} &= \sum_{i=1}^{30} \delta_i r_{t-i} + \sum_{i=1}^{30} \zeta_i x_{t-i}^{odd} + \sum_{i=1}^{30} \eta_i x_{i,t-i}^{rd,mix} + \varepsilon_{2,t} \\ x_t^{rd,mix} &= \sum_{i=1}^{30} \kappa_i r_{t-i} + \sum_{i=0}^{30} \lambda_i x_{t-i}^{odd} + \sum_{i=1}^{30} \mu_i x_{i,t-i}^{rd,mix} + \varepsilon_{3,t}. \end{aligned} \tag{13}$$

In this specification, the contemporaneous odd-lot trading variable, x_t^{odd}, appears in the quote and round- and mixed-lot trade equations. Thus, we assume that odd-lot volume causes contemporaneous quote changes and volume of round and mixed lots. We then reverse the assumption by removing the contemporaneous odd-lot volume and add the contemporaneous round- or mix-lot volume. These two specifications provide upper and lower bounds for the price impact of odd lots.

We estimate the three equations for each stock day, and then determine the arithmetic average of the impulse coefficients, which are given in basis points. Statistical inference is conducted using each stock date as an observation. We calculate the impulse response function to a 100-share shock to odd-lot volume and round- and mixed-lot volume. We calculate the cumulative long-run response of minute-by-minute returns, which is the cumulative impact of the shock after 30 minutes. Panel A of Table IX shows that the lower bound of an odd-lot shock is 3.57 basis points, which is about three times higher than the upper bound of round lots (1.13 basis points). The difference between these two price impacts is 2.44 basis points, with a t-statistic of 7.00. The upper bound for an odd-lot shock (5.20% basis points) is about five times as large as the lower bound of mixed and round lots (1.05 basis points). The difference is 4.15 basis

Table IX

Permanent Price Impact by Odd Lots and Mixed and Round Lots

This table reports the impulse response function for returns for odd-lot trades and round- and mixed-lot trades. Panel A presents the results based on a 100-share shock, and Panel B presents the results for a one-trade shock. We calculate the cumulative long-run response of minute-by-minute returns, which is the cumulative impact of the shock after 30 minutes. Odd lots upper bound and mixed- and round-lots lower bound assume that odd lots cause contemporaneous mixed and round lots, and vice versa. The coefficients (in basis points) are the average price impact across each stock for each day. t-statistics for the differences are also presented. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% level, respectively.

	Odd Lots	Mixed and Round Lots	Difference	t-statistic for the Difference
Panel A: 100-Share Shock				
Odd lots lower, Mixed and round lots upper				
	3.57	1.13	2.44***	7.00
Odd lots upper, Mixed and round lots lower				
	5.20	1.05	4.15***	11.24
Panel B: One-Trade Shock				
	Odd Lots	Mixed and Round Lots	Difference	t-statistic for the Difference
Odd lots lower, Mixed and round lots upper				
	2.16	2.02	0.14	0.30
Odd lots upper, Mixed and round lots lower				
	3.19	1.74	1.45***	3.01

points, with a t-statistic of 11.24. These data provide supportive evidence that odd lots are more informative than mixed and round lots.

As a robustness check, we also compute the result for a one-trade shock using the Hasbrouck method. Therefore, we estimate equation (12) again, except that x_t^{odd} is the sum of the signed odd-lot trades (buy initiated trades minus sell

Table X
Weighted Price Contribution in ITCH Data

This table reports the weighted price contribution for each order size category using individual trades. WPC is the weighted price contribution using price changes. Share of trades (volume) gives the percentage of trades (volume) in each size category. The data are from the NASDAQ ITCH database for the 2010 to 2011 sample period.

Trade Size Category	WPC (%)	Share of Trades (%)	Share of Volume (%)
<100	39.02	19.96	4.29
100	47.45	53.27	27.59
200	4.76	9.25	9.59
300	1.49	3.61	5.61
400	0.79	2.28	4.73
500	0.70	1.79	4.63
100–500	58.77	74.46	57.03
501–900	1.04	2.63	9.20
901–1,900	0.71	2.09	13.30
1,901–4,900	0.24	0.72	10.31
4,901–9,999	0.04	0.11	3.60
501–9,999	2.20	5.92	38.49
≥10,000	0.01	0.03	2.27

initiated trades) during the one-minute interval, and $x_t^{rd,mix}$ is the sum of the signed round- and mixed-lot trades during the one-minute interval. Compared to the results using volume, the price impact per trade is smaller for odd lots and larger for round and mixed lots, but Panel B of Table IX shows that the price impact of odd lots is still higher than that of the round and mixed lots. The lower bound of one-trade odd-lot shock is 2.16 basis points, which is higher than the upper bound of one-trade round- or mixed-lot shock (2.02 basis points). The upper bound for an odd-lot shock (3.19 basis points) is higher than the lower bound of mixed and round lots (1.74 basis points). The difference is 1.45 basis points, with a t-statistic of 3.01.

C. Price Informativeness over Time

As a final robustness check, we estimate weighted price contribution measures for the 2010 to 2011 period covered in the NASDAQ ITCH data. The estimation process is as described in the previous section, and the results are given in Table X. As in our earlier period, we find that odd lots are clearly informative, but now we also find that this informativeness is increasing over time. Odd lots alone contribute 39% to price discovery in the 2010 to 2011 period. The data also show that almost 86% of price discovery is accounted for by trades of 100 shares or less. Furthermore, trades greater than 500 shares now contribute only about 2% of price discovery. Thus, price discovery is shifting to smaller trade sizes, with odd-lot trades playing a very important role in this process.

V. Why Does It Matter? The Impact of Missing Trades on Empirical Research

For researchers, the fact that a large, and growing, fraction of trades are missing from the databases generally used for academic studies is cause for concern. In this section, we discuss how these missing trades can affect the design and interpretation of research. First, we show that several widely used empirical measures have significant bias because of odd-lot truncation, implying that researchers should be cautious in using these measures in the future. Second, we show that odd-lot truncation can also affect the interpretation of results in previous literature.

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 investor sentiment. The measure has been used to explain stock returns (Chordia, Roll, and Subrahmanyam (2002), Chordia and Subrahmanyam (2004)), momentum (Hvidkjaer (2006)), herding (Christoffersen and Tang (2009), Jame and Tong (2009)), the 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 prices.

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 and Chordia, Roll, and Subrahmanyam (2002) and Chordia and Subrahmanyam (2004) use the dollar volume in addition to the first two definitions.

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

Table XI 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).

We consider a trading day for each stock as one observation. The HF data identify buys and sells, so we can calculate the true order imbalance of all trades as true buy imbalance, true balance, and true sell imbalance. TAQ data only record trades of 100 shares or more, so using those trades we define observed buy imbalance, observed balance, and observed sell imbalance. The TAQ data do not indicate buys and sells, so for our purposes here we use the true buy/sell information from the HF data. In general, however, users of TAQ data need to use a signing algorithm such as Lee-Ready, which leads to greater errors in calculating imbalances.

OIBNUMB 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

Table XI

Correctly Signed Order Imbalance

This table reports the percentage of correctly signed buy and sell imbalances and the PIN estimated over all trades and trades greatar than or equal to 100 shares. The table provides a conservative estimation because it is based on the assumption that Lee and Ready (1991) make 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 the daily imbalances we would observe in the TAQ data, if all the buy and sells were correctly signed. OIBNUM is defined as the number of buy trades minus the number of sell trades. OIBSH is buy volume minus sell volume. OIBDOL is buy dollar volume minus sell dollar volume.

	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

imbalances or no imbalance. We also find that 5.52% of imbalances are classified as sells when they are buy imbalances or no imbalance. Finally, some days are classified as no imbalance days when they are actually buy or sell imbalance days (approximately 0.23%). Chordia, Roll, and Subrahmanyam (2002) recommended using the number of trades–based imbalance measure for empirical work, but this is clearly not advisable: the OBINUM measure is seriously biased by missing trades.

Table XI 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 measures are used. We suggest that researchers use such volume- or dollar-volume-based measures for order imbalance measurement.

Missing odd lots have a much larger impact on order imbalances of small trades, which is often used as a proxy for the sentiment of individual traders. TAQ data do not reveal a trader's identity, so Lee and Radhakrishna (2000) propose a \$5,000 cutoff value to identify individual (or retail) trades. This method is used extensively in the literature to study individual traders' behavior (see, for example, Shanthikumar (2004), Lamont and Frazzini (2007), Barber, Odean, and Zhu (2009), Christoffersen and Tang (2009), Jame and Tong (2009)).

The absence of odd-lot trades means that the \$5,000 cutoff generates a second, potentially more severe bias in the data. Because TAQ data do not have trades of less than 100 shares, a stock with a price above \$50 will have zero imputed retail trading. These stocks are then either defined as having no individual trade imbalance, or truncated unintentionally from the sample when order imbalance is taken to be the ratio of buy orders to the sum of buy and sell orders (Barber, Odean, and Zhu (2009)). As a result, any paper that uses the \$5,000 cutoff is actually based on stocks with a price below \$50.

This bias can be substantial. We calculate the incidence of zero imputed retail trades for stock using TAQ data and the \$5,000 trade cutoff for the period 1983 to 2011. Figure 8 shows that, depending upon the time period, up to 15% of stocks have zero imputed retail trades. Those stocks, however, tend to be both larger and more actively traded, so that looking at the percentage of zero individual trading weighted by market capitalization results in almost 70% of the value-weighted index having zero imputed retail trades.

Table XII presents evidence on the magnitude of these two types of biases for our sample stocks. Based on order numbers, 9.61% of imbalances are misclassified, with 4.77% of buy imbalances classified as sell imbalances and 4.58% of sell imbalances classified as buy imbalances, 0.11% of stock days classified as buy imbalance days although there is a balance of trades, and 0.15% of stock days misclassified as sell imbalance days although 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.

The problem is much more severe when we observe zero individual trades. Across all 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 returns, this can cause one of two problems. If order imbalance is treated as missing because there are no observed trades, this will lead to a 17% truncation of the regression sample. If order imbalance is treated as zero because zero buy and zero sell implies zero order imbalance, this will result in 17% of the 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 order imbalances as sell order imbalances has a 50% chance of being correct. These results

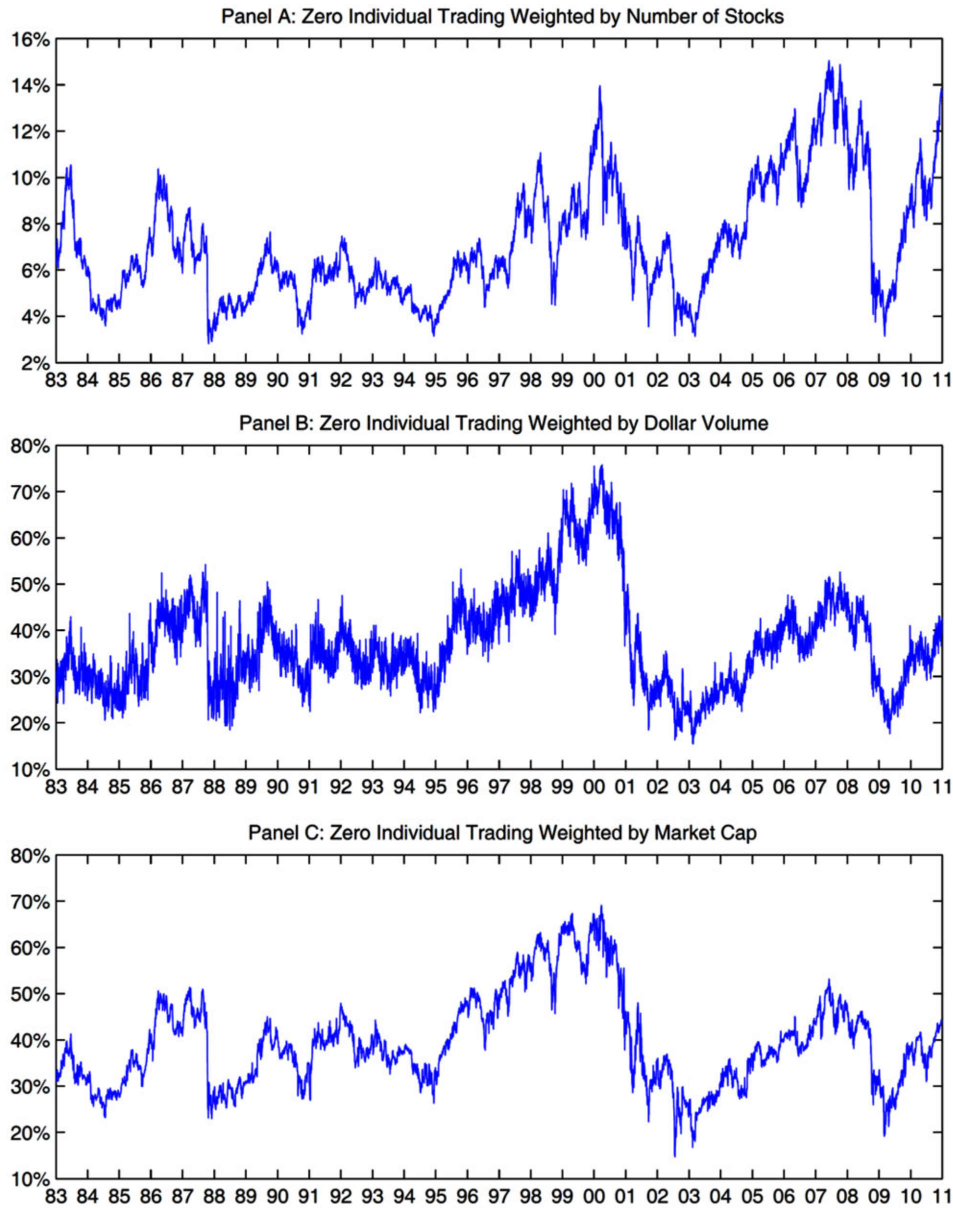

Figure 8. Stocks showing zero individual trading as a result of $5,000 cutoff value. This figure displays the percentage of stocks showing zero individual trades by applying Lee and Radhakrishna's $5,000 cutoff to the TAQ data. Because TAQ does not report trades of less than 100 shares, we observe zero trading for individual trades for stocks with a price higher than $50. The graph is computed through CRSP. Panel A depicts the percentage of stocks with zero individual trades. Panel B weights each stock by its dollar volume, and Panel C provides the value-weighted average.

Table XII

The Percentage of Correctly Signed Order Imbalance for Individual Trades

This table reports the percentage of correctly signed buy and sell imbalances based on Lee and Radhakrishna's $5,000 cutoff 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 the daily imbalances we would observe in the TAQ data if all the buy and sells are correctly signed. OIBNUM is the number of buy trades minus the number of sell trades. OIBSH is buy volume minus sell volume. OIBDOL is buy dollar volume minus sell dollar volume. The sample period is 2008 to 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

strongly suggest that researchers avoid using investor sentiment proxies based on order imbalance or trade-size cutoffs in future work.

One interesting fact about this truncation is that it is independent of the actual magnitude of odd lots—we do not even use the level of odd-lot activity in the replication. The truncation is actually based on price level. But it is because there are no odd-lot trades in TAQ/ISSM that using cutoffs for retail trades leads to the removal of high-priced stocks that constitute a significant part of the value-weighted portfolio. The truncation then generates significant return patterns by truncating high-priced stocks.

These results demonstrate why it is important for researchers to be aware of the fact that TAQ/ISSM data do not have trades for less than 100 shares. This omission will bias any study using arbitrary trade-size cutoffs to proxy for particular trader groups. We also suggest caution in interpreting existing results due to the sample selection biases that may have been present. In the

Internet Appendix to this paper, we show that odd-lot truncation can reconcile the differences in results between two papers on retail trading (Hvikjaer (2008), Barber, Odean, and Zhu (2009)). Given the increasing incidence of odd-lot trades, these truncation problems may represent an even greater problem going forward.

VI. Conclusion

In this paper, we investigate the changing role and incidence of odd-lot trades in equity markets. We demonstrate that odd-lot trades are a large and growing fraction of trades, reflecting the new dynamics of high-frequency markets. While traditionally used by retail traders, odd-lot trades are now much more likely to come from high-frequency traders, and their incidence is increased by practices such as pinging and order shredding. Moreover, we show that these odd-lot trades are highly informative, contributing 39% to price discovery. With round-lot trades contributing 50% of price discovery, the vast majority of price discovery is now taking place in very small trades.

Due to traditional trade reporting rules, however, none of these odd-lot trades are visible to the market due to their exclusion from the consolidated tape. Because TAQ data are derived from the tape, these missing trades are also 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, this problem is acute, with 50% or more of trades missing from the data. Equally important, order imbalance or imputed trader identity or sentiment measures can be severely biased, and analyses of issues related to return or market efficiency are also subject to error. As we show, these biases can result in spurious inferences being drawn from the data.

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 in incidence. For researchers using TAQ and other market 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 flawed. A firm-varying cutoff based on firm price, such as is used in Hvidkjaer (2008), may mitigate the problem by ensuring that small trades exist for all stocks. In addition, the development of new, more complete databases 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) raises 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?

Our paper provides answers to these important questions and is, to our knowledge, the first to do so. As we demonstrate, market data are biased because of the reporting rules. 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. Particularly unsettling is that, 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 contain odd lots, and other market venues also 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).[27] The market thus looks very different to those relying on the consolidated tape than it does to those buying proprietary data feeds. Even the SEC faces challenges knowing the true state of the market because 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.

VII. Postscript

As we suggested in this paper, the SEC began requiring odd-lot reporting to the tape as of December 9, 2013. On that day, 17.5% of trades and 2.9% of volumes were odd lots. Odd-lot trading was most pronounced on NASDAQ, where 25% of trades and 5.5% of volumes were odd lots, and on BATS, where 21.2% of trades and 6.4% of volumes were odd lots. Odd-lot trading on the NYSE amounted to 15.5% of trades and 2.9% of volumes. Individual stocks showed wide variance in odd-lot trades. For Google, odd lots amounted to 67.5% of trades and 23.1% of volumes, for Amazon odd lots amounted to 53.9% of trades and 12.5% of volumes, and for Apple 45.7% and 13%, respectively.

Going forward, odd-lot trades will be printed on the tape, but no odd-lot data will be provided retrospectively so researchers should be careful in interpreting

[27] These data feeds are not inexpensive. NASDAQ ITCH data, for example, cost from $500 per port/per month for the basic data to $2,500 per port/per month for the multicast ITCH/FPGA feed.

and combining data from prior periods. Odd-lot limit orders will not be included in the consolidated quote montage for pricing purposes.

Initial submission: January 23, 2012; Final version received: January 10, 2014
Editor: Campbell Harvey

REFERENCES

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

Barber, Brad M., Terrance Odean, and Ning Zhu, 2009, Do noise traders move markets? *Review of Financial Studies* 22, 151–186.

Barclay, Michael, and Terrance Hendershott, 2003, Price discovery and trading after hours, *Review of Financial Studies* 16, 1041–1073.

Barclay, Micheal, Terrance Hendershott, and Timothy McCormick, 2003, Competition among trading venues: Information and trading on electronic communications networks, *Journal of Finance* 58, 2637–2666.

Barclay, Michael J., and Jerold B. Warner, 1993, Stealth trading and volatility: Which trades move prices? *Journal of Financial Economics* 34, 281–305.

Baron, Matthew, Jonathan Brogaard, and Andrei Kirilenko, 2012, The trading profits of high frequency traders, Working paper, University of Washington.

Bloomfield, Robert, Maureen O'Hara, and Gideon Saar, 2013, Hidden liquidity: Some new light on dark trading, *Journal of Finance,* Forthcoming.

Brogaard, Jonathan, Terrence Hendershott, and Ryan Riordan, 2013, High frequency trading and price discovery, *Review of Financial Studies,* Forthcoming

Busse, Jeffrey A., and T. Clifton Green, 2002, Market efficiency in real time, *Journal of Financial Economics* 65, 415–437.

Buti, Sabrina, Barbara Rindi, and Ingrid Werner, 2011a, Dark pool trading strategies, Working paper, Fisher College of Business, Ohio State University.

Buti, Sabrina, Barbara Rindi, and Ingrid Werner, 2011b, Diving into dark pools, Working paper, Fisher College of Business, Ohio State University.

Chaboud, Alain, Benjamin Chiquoine, Erik Hjalmarsson, and Clara Vega, 2014, Rise of the machines: Algorithmic trading in the foreign exchange market, *Journal of Finance* 69, 1645–1684.

Chakravarty, Sugato, 2001, Stealth-trading: Which traders' trades move stock prices? *Journal of Financial Economics* 61, 289–307.

Chan, Kalok, and Wai-Ming Fong, 2000, Trade size, order imbalance, and the volatility-volume relation, *Journal of Financial Economics* 57, 247–273.

Choe, Hyuk, and Oliver Hansch, 2005, Which trades move stock prices in the Internet age? Working paper, Pennsylvania State University and Seoul National University.

Chordia, Tarun, Amit Goyal, and Narasimhan Jegadeesh, 2011, Buyers versus sellers: Who initiates trades and when? Working paper, Emory University and University of Lausanne.

Chordia, Tarun, Richard Roll, and Avanidhar Subrahmanyam, 2002, Order imbalance, liquidity, and market returns, *Journal of Financial Economics* 65, 111–130.

Chordia, Tarun, and Avanidhar Subrahmanyam, 2004, Order imbalance and individual stock returns: Theory and evidence, *Journal of Financial Economics* 72, 485–518.

Christie, William G., and Paul H. Schulz, 1994, Why do NASDAQ market makers avoid odd-eighth quotes? *Journal of Finance* 49, 1813–1840.

Christoffersen, Susan Kerr, and Ya Tang, 2009, Institutional herding and information cascades: Evidence from daily trades, Working paper, McGill University.

Dyl, Edward A., and Edwin 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, David, Marcos M. Lopez de Prado, and Maureen 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, David, Marcos M. Lopez de Prado, and Maureen O'Hara, 2012a, Flow toxicity and liquidity in a high frequency world, *Review of Financial Studies* 25, 1457–1493.

Easley, David, Marcos M. Lopez de Prado, and Maureen O'Hara, 2012b, The volume clock: Insights into the high frequency paradigm, *Journal of Portfolio Management* 39, 19–29.

Easley, David, Maureen O'Hara, and Liyan Yang, Differential access to price information in financial markets, *Journal of Financial and Quantitatiue Analysis,* Forthcoming

Engle, Robert F., Robert Ferstenberg, and Jeffery Russell, 2006, Measuring and modeling execution cost and risk, *The Journal of Portfolio Management* 38, 14–28.

Gai, Jiading, Chen Yao, and Mao Ye, 2012, The externalities of high frequency trading, Working paper, The University of Illinois.

Harris, Larry, 1991, Stock price clustering and discreteness, *Review of Financial Studies* 4, 389–415.

Hasbrouck, Joel, 1991a, Measuring the information content of stock trades, *Journal of Finance* 46, 179–207.

Hasbrouck, Joel, 1991b, The summary informativeness of stock trades: An econometric analysis, *Review of Financial Studies* 4, 571–595.

Hasbrouck, Joel. 1996. Order characteristics and stock price evolution: An application to program trading, *Journal of Financial Economics* 41, 129–149.

Hasbrouck, Joel, and Gideon Saar, 2011, Low-latency trading, Working paper, Cornell University.

Hendershott, Terrence, Charles M. Jones, and Albert J. Menkveld, 2011, Does algorithmic trading improve liquidity? *Journal of Finance* 66, 1–33.

Hvidkjaer, Soeren, 2006, A trade-based analysis of momentum, *Review of Financial Studies* 19, 457–491.

Hvidkjaer, Soeren, 2008, Small trades and the cross-section of stock returns, *Review of Financial Studies* 21, 1123–1151.

Jame, Russell, and Qing Tong, 2009, Retail investor industry herding, Working paper, Emory University.

Lakonishok, Josef, and Edwin Maberly, 1990, The weekend effect: Trading patterns of individual and institutional investors, *Journal of Finance* 49, 231–243.

Lamont, Owen A., and Andrea Frazzini, 2007, The earnings announcement premium and trading volume, Working paper, Yale University and Harvard University.

Lee, Charles M.C., and Balkrishna Radhakrishna, 2000, Inferring investor behavior: Evidence from TOR Q data, *Journal of Financial Markets* 3, 83–111.

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

Malkiel, Burton, 1981, *A Random Walk Down Wall Street* (Norton, New York).

Moulton, Pamela 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.

NYSE, 2007, Odd Lot Order Requirements, Information Memo 07-60.

O'Hara, Maureen, and Mao Ye, 2011, Is market fragmentation harming market quality? *Journal of Financial Economics* 3, 459–474.

Ritter, Jay R., 1988, The buying and selling behavior of individual investors at the turn of the year, *Journal of Finance* 43, 701–717.

Rozeff, Michael S., 1985, The tax-loss selling hypothesis: New evidence from share shifts, Working paper, University of Iowa.

The Securities and Exchange Commission (SEC), 2010, Concept release on equity market structure, Securities and Exchange Commission Release No. 34-61358. Available at http://www.sec.gov./rules/concept/2010/34-61358.pdf.

Shanthikumar, Devin, 2004, Small and large trader behavior: Reactions to information in financial markets, Dissertation, Stanford University.

Sias, Richard W., 1997, Price pressure and the role of institutional investors in closed-end funds, *Journal of Financial Research* 20, 211–229.

Wu, Hsiu-Xwang, 1972, Odd lot trading in the stock market and its market impact, *Journal of Financial and Quantitative Analysis* 7, 1321–1344.

Ye, Mao, 2011, A glimpse into the dark: Price formation, transaction cost and market share of the crossing network, Working paper, University of Illinois at Urbana-Champaign.

Supporting Information

Additional Supporting Information may be found in the online version of this article at the publisher's website:

Appendix S1: Internet Appendix.