How do Flows affect Prices? Evidence from an Experiment on the U.S. Stock Market*

Trevor S. Gallen†

Yana Gallen‡

February 2024

Preliminary and Incomplete: Do not cite

Abstract

This paper reports the results of a randomized controlled trial in which we randomly buy and sell nearly $3.5 million of U.S. non-financial common stocks listed on the New York, NASDAQ, and American stock exchanges from 2021-2023. We find significant causal effects of a one-share retail trade on subsequent transaction, bid, and ask prices, as well as cumulative abnormal volume. We report results decomposing our causally-identified impulse-response function into information, inventory, and order-processing costs. Finally, we empirically explore heterogeneity in market impact and its drivers using a causal forest model finding that only spread predicts treatment effects.

I Introduction

In the standard perfect-information asset pricing equation, trades have little scope to move prices without changing discount rates or the net present value of future dividends. However, a large literature attempts to both document and explain the effects of signed trades on prices. Documenting these effects on prices has been hampered by issues of both measurement and causal identification. In this

---

*This research does not qualify as Human Subjects Research under federal human rights research regulations, and does not require IRB approval (Purdue IRB-2021-1353). This research was preregistered on the American Economic Association's Randomized Controlled Trial Registry under ID numer AEARCTR-0008550.

†Correspondence: Purdue University Krannert School of Business, West Lafayette, IN 47906. Tel.: (765) 496-2458. Email: tgallen@purdue.edu. Web: www.tgallen.com

‡Correspondence: University of Chicago, Harris School of Public Policy, Aarhus University, and IZA. Email: yana@uchicago.edu. Web: www.yanagallen.com

paper, we contribute to solving both problems by running a randomized controlled trial, in which we randomly buy and sell nearly \$3.5 million of U.S. non-financial common stocks listed on the New York, NASDAQ, and American stock exchanges from 2021-2023. The experiment yields three new facts: 1. on average, individual transactions have a large (one basis point) impact on subsequent traded prices, on average, 2. the effect is transitory and attenuated to zero by four minutes after the initial trade, and 3. the effect is driven by stocks which have a large spread—no other stock characteristics predict treatment effects beyond spread.

In this setting, and in contrast to related work by Andrea, Israel, and Moskowitz (2018) and Gomes and Waelbroeck (2015), small retail trades are initiated by a retail trader and filled by market makers off-exchange within the framework of payment for order flow (PFOF). These trades are reported only with a delay to the public. That these 'stale' trades affect prices at all is surprising in a perfect information framework, but may be explained by beliefs that an individual trader has private information about future prices. This may also be explained by ... inattention? what? Next, say what you are able to tell us about distinguishing those mechanisms.

In addition to documenting effects of order flow on subsequent transaction prices, this study investigates the impact on bid prices, ask prices, and cumulative abnormal traded volume. The findings of this study reveal statistically significant average effects resulting from even a single-share trade executed by a retail trader, which persist for approximately three minutes in subsequently-traded prices. The peak effect, observed immediately after the random trade, is approximately one basis point for an \$80 trade. It is of course easy to increase the price one pays as a buyer by simply purchasing at the quoted ask price. Consequently, while the price of our own trade is measured, our analysis and results remove it from the price and volume data completely, focusing only on the effect of our trade on subsequent trades made entirely by third parties on both sides. We find that there is not only the effect of subsequently-traded prices, but also a significant impact on cumulative abnormal traded volume, peaking at approximately 30 seconds with a magnitude of nearly 0.01% of daily volume before diminishing in significance. The bid and ask prices are symmetrically affected, exhibiting an increase of nearly 0.4 basis points, which loses significance around the 5-minute mark and returns to the mean level by approximately 10 minutes.

Moreover, the analysis reveals heterogeneity in these effects across stocks, with spread playing a predominant role in predicting this heterogeneity within a causal random forest model, followed by market capitalization. To comprehensively explore both the average effect and the mechanisms driving heterogeneous treatment effects, the experiment employs three distinct treatment arms. The first arm involves random trading of a single share in approximately 15% of U.S. non-financial common stocks listed on the New York, NASDAQ, and American stock exchanges, spanning a total of 85 trading days, while retaining 85% of the market as control. The second arm focuses on the same group of stocks but restricts trades to those priced at $7 or below, with higher quantities purchased and sold: typically around 100 shares. The third arm investigates the potential impact on the ten- and twenty-largest stocks by market capitalization, involving single-share trades with a value approaching $28,000, for a duration of 43 trading days.

Our findings make a contribution towards resolving a longstanding issue in interpreting market impact regressions. This issue is succinctly summarized by Hasbrouck (2007), a seminal textbook in empirical market microstructure, which suggests that orders do not directly 'impact' prices; rather, they serve as indicators that forecast future prices. While the distinction between functional causality and forecast enhancement is crucial, empirical resolution has proven challenging. Similarly, Cochrane (2004), in a review of the literature on market impact, asserts that it remains uncertain whether flows cause price changes or are merely correlated with them. This uncertainty arises from the need to comprehend the origin of order flow before drawing conclusions about its impact on prices. Consequently, trades may not directly impact markets but rather exhibit correlation with information that would have otherwise influenced market conditions. Alternatively, trades may impact prices solely through an informational channel to the market maker. Finally, trades may directly influence prices by, for instance, altering the inelastic demand curve of a stock. Importantly, our paper eliminates the first possibility by providing evidence that trades indeed have a direct impact on subsequent prices.

After establishing the average treatment effect of purchases on subsequent prices, we turn to examine the large heterogeneity underlying market impact. To do so, we first present univariate sorts, finding many variables significantly predict market impact, including spread, market cap, market volume, and previous volatility, among others. However, these variables are correlated, and likely redundant factors in predicting price impact. To nonparametrically summarize the treatment effect heterogeneity, we leverage the causal random forest model of Athey, Tibshirani, and Wagner (2019). We find that spread is highly predictive of market impact, and the significant differences of all univariate sorts after holding other variables at the mean value vanish, suggesting that redundancy is a significant problem. We find evidence of significant treatment effect heterogeneity more generally, and use the method of Erik, Ayush, and Wager (2020) to find the shallow decision tree that optimally selects treatment to maximize treatment effect, finding it selects on spread and market capitalization.

After establishing the causal effect of individual trades and underlying heterogeneity, we look to summarize them economically. To give an economic summary of the causal effect of price impact, we adapt the standard traditional three-fold decomposition of Huang and Stoll (1997) to take advantage of a randomized controlled study. In this model, the dynamic effect of price is split into the immediate effect due to the bid-ask "bounce," a persistent transitory "inventory" effect that goes to zero over time, and a permanent "information" effect. While the typical estimation requires a researcher to estimate the signs of multiple trades, and is identified by changes over time, we instead use the fact that the signs of previous trades before our random trade are equally likely to be buys and sells when comparing treatment and control, and instead identify our model using the causal impulse-response function generated by differencing between treatment and control, rather than over time. We find all of the price impact on subsequent trades is due to inventory effects, with little scope for informational effects. We also conduct this on the “high treatment effect" and “low treatment effect" branches of our optimal policy tree.

Our paper contributes to the extant literature on price impact in several ways. To the best of our knowledge, we are the first to analyze price impact on actual financial instruments using a randomized controlled experiment, with the exception of Camerer (1998), which conducts a similar experiment in pari-mutuel betting on horse races. While others papers documenting the effects of flows on prices make causal arguments, we are the first to use the so-called “gold standard" of causal inference, the randomized controlled trial. This approach enables us to know with certainty, rather than estimate, whether trades were initiated by the buy- or sell-side and also allows us to rule out any association with larger, serially-correlated “meta-orders.” Moreover, it allows for a causal interpretation of our estimated impulse-response functions, given that no private information relevant to stock prices influenced the experimental transactions. We are the second paper, following Ernst, Sokobin, and Spatt (2021), in documenting effects of stale retail trades on prices. We contribute to the buy/sell-measurement literature by producing a dataset of more than 40,000 correctly-labelled buy and sell transactions and comparing measurement accuracy. Finally, we extend the Huang and Stoll (1997) decomposition framework to accommodate panel experimental data, allowing for an experimentalist to distinguish between “herding" and “meta-order" sources of signed trade serial correlation.

The rest of this paper proceeds as follows: Section II provides a review of relevant literature. Section III describes the theoretical model, economic and statistical identification of our adaptation of the Huang and Stoll (1997) three-fold decomposition to an experimental framework. Section IV describes the institutional background. Section V describes our experiment. Section VI describes the data. Section VII describes our results. Section VIII concludes.

II Literature Review

The literature on the price impact of trades is vast, with over 56,000 results on Google Scholar. It originated with Kraus and Stoll (1972), who examined price changes at market close resulting from large institutional 'block' trades categorized as ''buy" or ''sell" based on their relation to the preceding price. However, accurately measuring and interpreting 'buy' and 'sell' orders has proven challenging, as both trades and price changes may be influenced by new information. Building on the work of Kraus and Stoll, researchers have made significant efforts to improve the measurement of buys and sells, provide theoretical explanations for market impact, and estimate market impact using various modeling approaches and new data sources.

The classification of buys and sells has been approached using four main methodologies. The 'tick rule' compares transaction prices with the nearest previous price change, categorizing trades as 'sells' if below the previous price and 'buys' if above. Lee and Ready (1991) proposed an algorithm that classifies a buy (sell) as a price above (below) the bid-ask midpoint, with the previous price used as a tiebreaker in case of a tie. Other studies, such as the PIN model (Easley, Kiefer, and O'Hara, 1997; Easley, Hvidkjaer, and O'Hara, 2002) or the "bulk volume" model Easley, López de Prado, and O'Hara (2012), focus on extracting the probability of information-based trading from the joint distribution of buys and sells. Chakrabarty, Pascual, and Shkilko (2015) find that these three rules correctly identify approximately 90-95% of the volume, depending on parameterization. In this study, we contribute by providing a dataset with known intended buy/sell orders.

Several studies have utilized institutional data to identify signed trades. For example, Gomes and Waelbroeck (2015) analyze a dataset of institutional trades flagged as driven by cash flow requirements rather than fundamentals. They find that non-informed 'cash-flow' trades have no market impact within two to five days, while potentially informed orders do exhibit market impact. Similarly, Almgren et al. (2005) investigate a large dataset from the Citigroup US equity trading desk and identify a concave price impact. Their findings indicate that shares traded as a fraction of daily volume are statistically significant predictors of immediate and long-run price impact, while market capitalization and bid/ask spread show no discernible influence. Andrea, Israel, and Moskowitz (2018) analyze a staggering \$1.7 trillion worth of trades from AQR capital. They examine trades executed by an algorithm designed to minimize trade costs within a specific time frame, providing knowledge of the sign of trade and addressing concerns of mismeasurement. To address potential violations of the exclusion restriction, the authors focus on trades driven by longer-term strategies unlikely to be influenced by short-term price forecasts. In comparison, our study benefits from a control sample, allowing for a comprehensive analysis of deep heterogeneity, including factors such as day fixed effects, time-of-day effects, and a broader scope of traded securities.

Theoretical literature, initiated by Kyle (1985), has extensively explored economic models of price impact and bid/ask spreads. These models present a framework where a competitive market maker, acting as a price-setter, determines the bid and ask spread in the presence of both informed and uninformed traders, often referred to as 'noise' traders. These models form a significant part of modern microstructure textbooks, such as chapters 5-7, 12, and 13 in Hasbrouck (2007), as well as various chapters in O'Hara (1998) and Foucault, Pagano, and Roell (2013).

Microstructure regression methods have been applied in several studies to estimate Kyle's lambda, a parameter used to measure the price impact of informed trades. These studies have found that the value of lambda varies across different markets and time periods. For instance, Hasbrouck and Seppi (2001) estimated lambda for a sample of NASDAQ stocks and observed variations ranging from 0.015 to 0.024, depending on the specific stock and time period analyzed. Furthermore, liquidity measures such as the bid-ask spread and order flow imbalance have been identified as significant predictors of the price impact of informed trades in these regression models. Additional research has explored the relationship between Kyle's lambda and other variables, including volatility, trading volume, and market structure. For example, Ni, Pan, and Poteshman (2008) investigated the impact of informed trading on price impact, finding that the days leading up to earnings announcements saw significantly large price impact. In a separate study, Brennan and Subrahmanyam (1996) utilized Lee-Ready classifications to estimate Kyle's lambda for all stocks in the NYSE from 1984 to 1987. They organized the estimates into a 5x5 table based on the size of the stocks and lambda values. Although their focus was primarily on returns, their findings regarding the variability of lambda within different groups hold significance even in the context of this study.

To the best of our knowledge, this paper represents the first estimation of price impact using a field experiment. Its closest competitor in this regard is Camerer (1998), who publicly placed and removed bets at pari-mutuel horse races, where the bets influenced the observed prices. Camerer found that a $500 bet on a horse, equivalent to roughly five percent of the total "market cap" of a race, resulted in a maximum additional bet of $50 on the horse, which was statistically insignificant and reversed leading up to the close of betting. Contrasting Camerer's work, our paper focuses on a different financial instrument (stocks rather than horse bets), operates at a larger magnitude (trades totaling $3 million compared to $58,000, or $110,000 in 2023 dollars), and avoids deception or even concerns about deception, as our executed trades are irreversible without incurring risk and transaction costs.

Two other papers conducted field experiments by purchasing stocks to understand the effects of payment for order flow (PFOF). Neither examines the price impact on subsequent trades, which is the focus of this paper. Levy (2022) and Schwarz and Odean (2022) both purchase stocks, routing trades across executors to find the distribution of price improvement. Both papers document significant dispersion in both price improvement and in execution price. Levy (2022) conducts a randomized controlled trial, randomly choosing stock and broker, and purchasing or selling approximately $1,000 or \$4,000 in 750 trades over 20 days. Schwarz and Odean (2022) report 85,000 trades over 128 stocks, for \$15.4 million. Unlike this paper or Levy (2022), the authors simultaneously submit identical market orders, set to be approximately \$100, at different brokers over the course of 113 days. To our knowledge, this paper reports the results of the largest randomized controlled trial on the U.S. stock market.

III Theoretical Model and Identification

III.1 Theoretical Identification

In this section, we build upon the basic model proposed by Huang and Stoll (1997) and introduce three extensions. First, we incorporate the assumption that dealer inventories follow an autoregressive process of order one. This inclusion aligns with the insights of Hasbrouck (1991) that long-run effects in market dynamics are driven by informational factors. Secondly, we expand the model to account for serial correlation in buy and sell orders, which can arise from both meta-orders and herding behavior. Importantly, the randomized controlled experiment employed in this study allows for the differentiation between these two sources of serial correlation. Lastly, we modify the identification methodology by employing a portfolio-differencing approach, as opposed to the conventional time-differencing method. This modification alleviates several issues associated with accurately measuring trade directionality. Despite these extensions, the derivation that follows is similar in structure to that of Huang and Stoll (1997).

The model considers a perfectly competitive market maker who forms beliefs about the fundamental value of a stock, denoted as V_t, prior to posting bid and ask quotes at time t. The current fundamental value V_t is a function of the market maker's previous belief V_{t-1}, the impact of the buy or sell order Q_{t-1} from the previous period (taking a value of 1 for a "buy" order and -1 for a "sell" order), and a serially uncorrelated private information shock ϵ_t:

$$V_t = V_{t-1} + \lambda(Q_{t-1} - E_{t-2}(Q_{t-1})) + \epsilon_t \tag{1}$$

Here, the Kalman gain λ represents the change in price resulting from the market maker's informational update, which is often associated with adverse selection in the spread. The direction of trades is influenced by two factors: "meta-orders" and "herding." Meta-orders refer to large trades that are broken down into smaller orders to minimize market impact, while herding captures the behavior of market participants reacting to trades. The conditional expected value of the previous buy order $E_{t-2}(Q_{t-1})$ is determined by the probability of another herding order (π_H) and another piece of a meta-order (π_M), with the remaining probability $(1 - \pi_H - \pi_M)$ indicating an equal probability of buys and sells:

$$E_{t-2}(Q_{t-1})) = (\pi_H + \pi_M)Q_{t-2} \tag{2}$$

Importantly, Equation 2 is valid from the market maker's information set, but not from the perspective of the experimentalist's or a noise traders. In expectation, the market maker will be surprised by the lack of follow-up meta-orders from an experimental trade and adjusts the price accordingly. This provides an avenue for identifying the probabilities π_H and π_M based solely on price data, without the need to directly measure Q_t. However, if trades can be accurately signed, Equation 2 also allows for direct estimation of π_H and π_M. By incorporating Equation 2 into Equation 1, we obtain the market maker's update to the fundamental value as expressed in Equation 3:

$$V_t = V_{t-1} + \lambda(Q_{t-1} - (\pi_H + \pi_M)Q_{t-2}) + \epsilon_t \tag{3}$$

Letting k_t be the market maker's inventory of the stock, the market maker induces reduction in positive holdings by raising the midpoint price in proportion to the inventory. This adjustment is consistent with a model of linear disutility of the market maker over the standard deviation of future wealth. Instead of assuming that the change in inventories is equal to Q_{t-1} in equilibrium, we introduce an assumption that inventories follow an AR(1) process with a shock term η_{t-1}, as described in Equation **??**:

$$M_t = V_t + \beta k_t$$

While Huang and Stoll (1997) and other decompositions take the difference in the midpoint, assuming that the change in inventories is equal to Q_{t-1} in equilibrium, we instead assume that inventories are subject to an AR(1) process, with shock equal to η_{t-1}, as in Equation 4:

$$k_t = \rho k_{t-1} + \eta_{t-1} \tag{4}$$

Consequently, the midpoint M_t is given by Equation III.1:

$$M_t = V_{t-1} + \lambda(Q_{t-1} - (\pi_H + \pi_M)Q_{t-2}) + \beta(\rho k_{t-1} + \eta_{t-1}) + \epsilon_t$$

Finally, the transaction price at time t accounts for order-processing costs denoted by γ:

$$P_t = \begin{cases} M_t + \gamma & \text{if buy} \\ M_t - \gamma & \text{if sell} \end{cases}$$

Apart from the modifications in Equations 2 and 3, our derivation of the midpoint follows a similar approach to Huang and Stoll (1997). While M_t is observable, Q_t, k_{t-1}, η, and ϵ are often unknown, although Q_t can be estimated using a buy/sell assignment rule, and dealer inventories are sometimes known as in Lyons (1995). While previous studies such as Huang and Stoll (1997) typically estimate the parameters of interest (λ, β, γ) using the change in price over time, our randomized controlled experimental setup allows us to identify these parameters using the portfolio-differencing method described in Equation 5:

$$\overline{\Delta} M_{t+\tau} = M_{t+\tau}^{Treatment} - M_{t+\tau}^{Control} \tag{5}$$

The identification in Equation 5 leverages the assumption that, on average, inventories before our trading event k_t, the last buy-sell direction Q_{t-1} before our trading event, and shocks to all processes (except our experimental shock to η and Q at $t = 0$) are equal in expectation and can be differenced across portfolios.

The intuition behind our portfolio-differencing identification is that before the experiment, prices are expected to be equal. When our experimental trade occurs, Q_0 increases to one for the treatment group, while remaining zero in expectation for the control group. Going forward, the trade direction Q_t remains higher for the treatment group temporarily due to herding effects. However, these effects are likely to be transitory, as even with $\pi_h + \pi_m = 0.6$, the expected difference in trade direction between treatment and control diminishes after a few trades. Similarly, with $0 < \rho < 1$, the inventory differences converge to zero in the long run, exhibiting transitional effects in the short run. Finally, the permanent effect does not converge to λ, but rather $\lambda(1 - \pi_m)$, which represents the final update in beliefs after the initial trade and the revision in the market maker's beliefs when they do not expect follow-up meta-orders, as typically seen in non-experimental data.

Table 1 summarizes the differences between treatment and control implied by our experimental adaptation of Huang and Stoll (1997). We leverage the theoretical moments provided in Table 1 to estimate the parameters in this section."

Table 1: Moment Conditions for Identification

t	Q_t^T	Q_{t-1}^T	$E_{t-2}(Q_{t-1}^T)$	$Q_{t-1}^T - E_{t-2}(Q_{t-1}^T)$	k_t	V_t	M_t	P_t
-2	0	0	0	0	0	0	0	0
-1	0	0	0	0	0	0	0	0
0	1	0	0	0	0	0	0	0
1	π_h	1	0	1	1	λ	$\lambda+\beta$	$\lambda+\beta+\gamma$
2	$\pi_h(\pi_h+\pi_m)$	π_h	$\pi_h+\pi_m$	$-\pi_m$	$\rho+\pi_h$	$\lambda(1-\pi_m)$	$\lambda(1-\pi_m)+\beta(\rho+\pi_h)$	$\lambda(1-\pi_m)+\beta(\rho+\pi_h)+\pi_h\gamma$
3	$\pi_h(\pi_h+\pi_m)^2$	$\pi_h(\pi_h+\pi_m)$	$\pi_h(\pi_h+\pi_m)$	0	$\rho(\rho+\pi_h)+\pi_h(\pi_h+\pi_m)$	$\lambda(1-\pi_m)$	$\lambda(1-\pi_m)+\beta(\rho(\rho+\pi_h)+\pi_h(\pi_h+\pi_m))$	$\lambda(1-\pi_m)+\beta(\rho(\rho+\pi_h)+\pi_h(\pi_h+\pi_m))+\gamma(\pi_h(\pi_h+\pi_m)^2)$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
t large	0	0	0	0	0	$\lambda(1-\pi_m)$	$\lambda(1-\pi_m)$	$\lambda(1-\pi_m)$

Table 1: This table depicts the expected difference between treatment and control in trade direction, expected trade direction, the market maker's "surprise," market maker's holding, value, price midpoint, and realized price over time. The price moments P_t, are used to estimate all parameters except transaction cost parameter γ, which is estimated separately using level bid-ask spread at $t=-5$.

Figure 1 provides a simplified graphical representation of the mean differenced (treatment-control) effect on the price level over time. Order-processing costs (γ) occur immediately and vanish in the next trade. Inventory costs (β) also have an immediate effect but decline gradually over time. Information effects (λ) are permanent. Figure 1 simplifies the analysis by omitting the effects of metaorders, herding, and market-maker surprise due to the lack of follow-up metaorders, which can complicate the impulse response in the initial periods. These factors may lead to a rise in inventory costs (due to herding), a decrease in the “permanent effect" (due to the market maker’s surprise at the lack of follow-up orders), and a prolonged duration of order-processing costs (due to herding). However, the overall shape of the identification process aligns with the pattern depicted in Figure 1.

III.2 Statistical Identification

Statistically, detecting an effect in a random walk process may seem challenging due to the small size of our transactions compared to market capitalization or volume. However, the power to detect an effect increases exponentially as the timespan studied decreases because the variance in prices decreases linearly with time. The minimum detectable effect, denoted as δ is a function of the t-statistic for statistical significance ($t_{\alpha/2}$), the t-statistic for power (t_β), the time since the event (t), the probability a stock is randomly assigned to treatment (p), the number of treated and untreated stocks (N), and the continuous-time variance of the random walk (σ^2), and is given by:

$$\delta_\tau = (t_{\alpha/2} + t_\beta)\sqrt{\frac{\tau\sigma^2}{Np(1-p)}} \tag{6}$$

Crucially, halving time period of interest τ has the same effect on the minimum detectable effect size as doubling sample size. Denoting the cumulative distribution of the standard normal with Φ, the formula for power is:

$$\text{Power} = \Phi^{-1}\left(\delta\sqrt{\frac{N(p^2-p)}{\tau\sigma^2}} - t_{\alpha/2}\right) \tag{7}$$

Equation 7 makes clear that the t-statistic follows an inverse power law as a function of τ: as τ goes to zero, power “quickly" goes to one, even for small samples, so long as $N > 2$ and $0 < p < 1$. While realized stock transaction prices are not continuous, the intuition of Equation 7 is useful for understanding why seemingly economically small experimental treatments might be statistically measurable.

To give a quantitative feel for power as a function of time-from-trade, assume a treatment fraction of 0.15, a sample size of 200,000 stocks (with 30,000 treated stocks), and a significance level of 95%, which broadly approximates this study, it is only left to find σ^2. Taking all stocks in the TAQ on average, the standard deviation of five-minute price changes in TAQ on a random day was 45

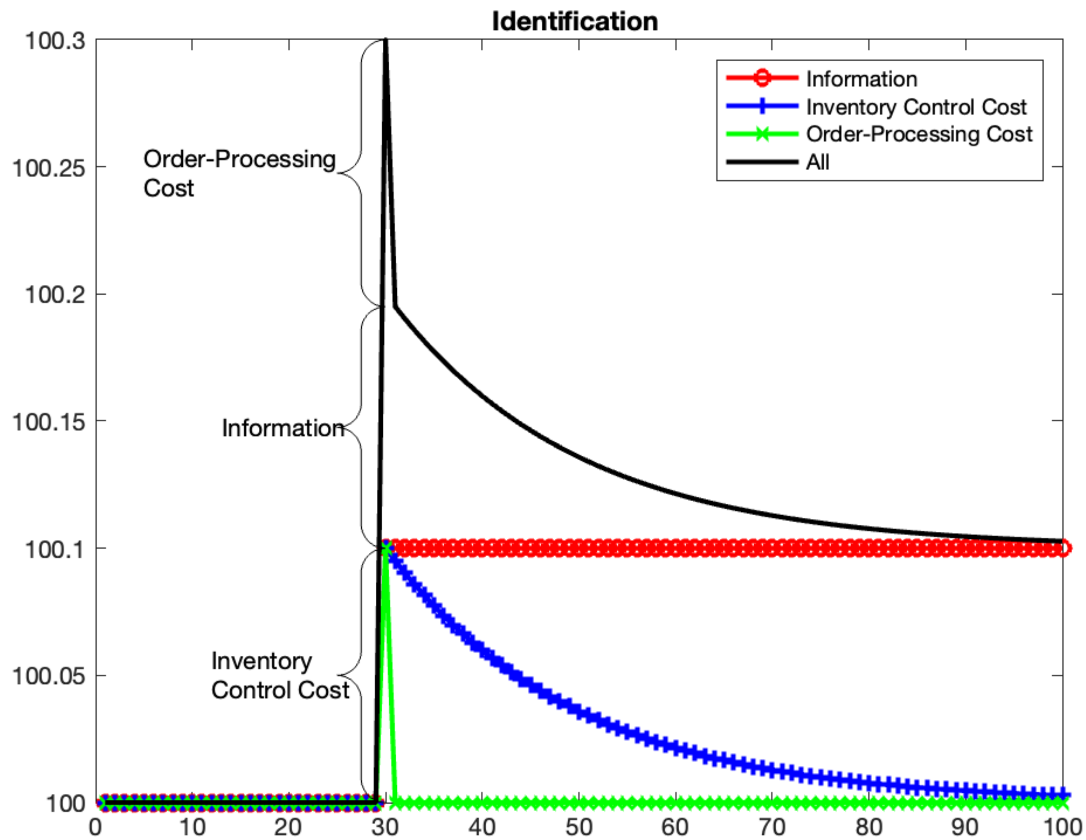

Figure 1: This figure depicts a stylized identification scheme from the impulse response function of a buy order in the spirit of Huang and Stoll (1997) and Hasbrouck (1991). The red line, “information," depicts the IRF of price if only information was present, a permanent shock. The blue line “inventory," depicts the IRF if only inventory effects are present, an AR(1) process. And the green line “Order-processing" depicts the IRF of price if only order-processing costs were present, and a the buy order followed and was immediately followed by a sell order, so that the effect reverts immediately. Because of herding effects, metaorders, and market-marker surprise at the lack of follow-up orders, all three lines in the true IRF deviate slightly from this idealized form.

basis points. Assuming variance increases linearly with time as in a random walk, this gives a σ^2 of 6.75e-08, and gives a numerical relationship between statistical power and minimum detectable effect. With these parameters, Figure 2 displays two results for Equation 7. The first, black line sets the minimum detectable effect size to be one basis point, and examines how much power the study has as a function of time. The following three re, blue and green lines describe power as a function of time for significance levels of 95%, 99%, and 99.9%. What the first line makes clear that we might confidently detect a one-basis point effect approximately 480 seconds after an event. The comparison between statistical significance levels as a function of time makes clear the usefulness of the inverse power law with respect to time: all converge to one as τ falls to zero.

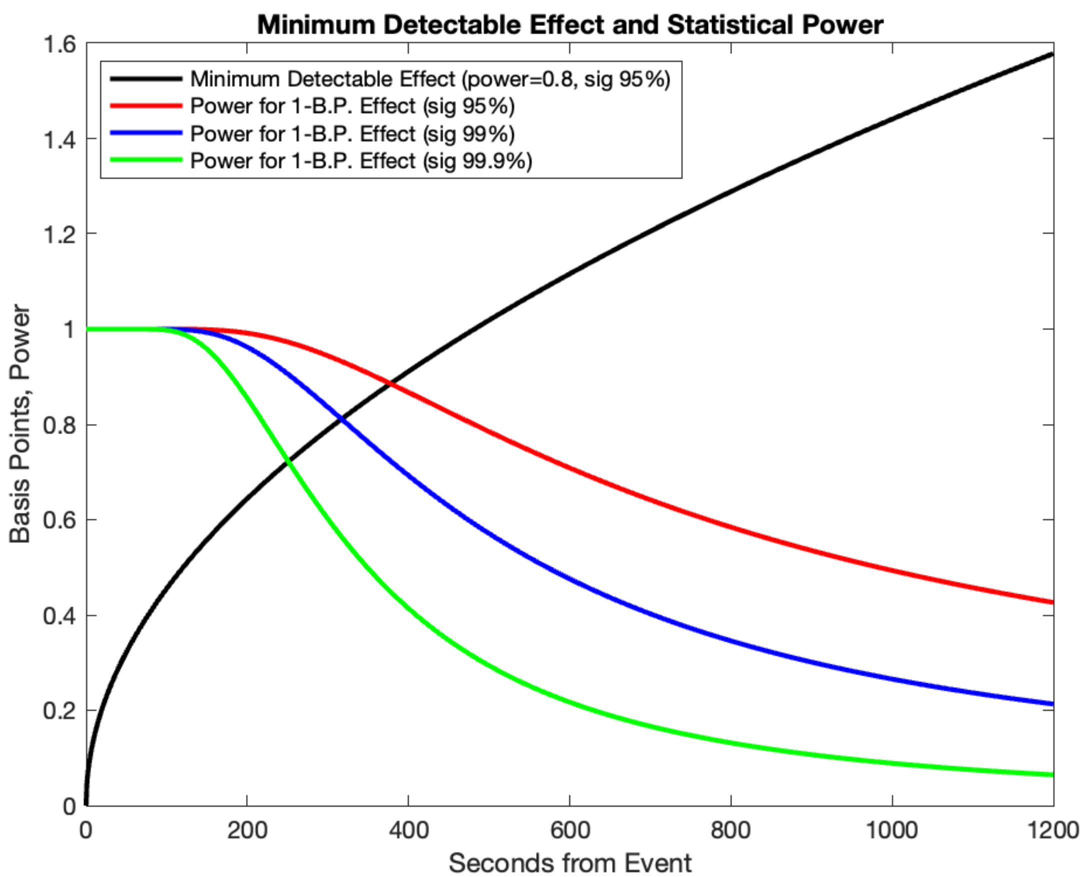

Figure 2: This figure displays the the power and minimum detectable effects as a function of time, derived from Equation 7 and assumptions given in the text. The red, blue, and green lines give power at 95%, 99%, and 99.9% significance levels. The black line gives the minimum detectable effect in basis points when power is 80% and significance is 95%.

IV Institutional Background

All trades were executed as a retail customer using TD Ameritrade's Application Programming Interface (API) in Python, utilizing funds allocated to a custodial Health Savings Account (HSA). As a retail customer, our trades did not have direct market access, and as a custodial HSA account, we could not route orders to a specific market. Instead, TD Ameritrade routed the orders to affiliate TD Ameritrade Clearing, which then sold the order flow to various market makers, internalizers, and wholesalers. According to Ameritrade (2021), almost all (99.96%) of TD Ameritrade Clearing, Inc's orders in July 2021 were non-directed orders, which include our orders. Among these orders, around 21% were market orders, and 6.5% were buy limit orders at or above the ask (or sell below the bid), which matches the type of orders we placed

These orders were sold to five different venues, with Citadel Securities (29%), Virtu Americas (44%), and G1 Execution Services (21%) receiving the lion's share of marketable limit orders. When executing these orders, market makers are required to offer a price better than the National Best Bid Offer (NBBO), which represents the lowest ask and highest bid among at least 100 shares on all markets. However, due to the prevalence of "odd-lot" trades consisting of less than 100 shares, market makers can arbitrage odd-lot bid and ask orders within the NBBO boundaries, generating profit and providing price improvement relative to the NBBO to traders. It is worth noting that our study does not directly examine the effects of Payment for Order Flow (PFOF) on price improvement, which has been analyzed experimentally in Levy (2022) and Schwarz and Odean (2022).

Two important developments made this experiment feasible. First, the removal of commissions in 2019, approximately one year before the start of the experiment, significantly reduced trading costs. Without this change, the experiment would have been prohibitively expensive, exceeding the account's value by more than tenfold. Second, the majority of trades in this experiment consisted of odd-lot one-share trades. These trades, which were not reported to the consolidated tape until 2013, as highlighted by O'Hara, Yao, and Ye (2014), now contribute significantly to trading volume. As a result, our reported trades have the potential to impact the market, as we document.

Due to the nature of off-exchange trades, there is a reporting delay associated with their visibility. Trades are reported to a Trade Reporting Facility (TRF), which then transmits the trade information to the Securities Information Processor (SIP) for public dissemination. In Ernst, Sokobin, and Spatt (2021), it was found that "dark" trades, similar to mine, are reported with an average latency of 2,500 microseconds, compared to 0.3 microseconds for on-exchange trades. Despite the delayed and seemingly uninformed nature of these quotes, the authors discovered significant increases in trading and quote revisions, indicating the transmission of information through the reporting of stale uninformed quotes, similar to the ones generated in our experiment. The authors utilized differences in geographic latencies for their identification strategy.

While all trades must be reported "as soon as practicable but no later than 10 seconds after execution" (FINRA 6622), retail participants can only access information about trades of 100 shares or more. For instance, platforms like TD Ameritrade, Charles Schwab, or Google Finance would not see any change in price or record of a transaction from a 99-share trade. We exploit this difference in information in our quantity treatments.

V Experimental Description

In this section, we describe three separate, but related experimental treatment arms.

In the first experiment, referred to as the "baseline" experiment, we utilized a sample of U.S.-based, non-financial common stocks available in both the most recent CRSP dataset and the TAQ dataset. Using TD Ameritrade's API, we attempted to locate the stock tickers for these stocks, resulting in a dataset of approximately 2800 stocks. From this sample, 450 stocks (480 after July 2022), accounting for around 16%-17% of the total, were randomly assigned to the treatment group, while the remaining stocks were assigned to the control group in a randomized order. The stocks in the treatment group were rapidly purchased, with approximately one stock bought per second between 11:30 A.M. and 12:30 P.M. (after July 22nd, 2022, purchases were spread out between 9:30 A.M. and 4:00 P.M.). This experiment was repeated 43 times, conducted between December 15th, 2021 and December 29th, 2022. A detailed calendar of the treatment days can be found in Appendix D.[1]

During the experiment, there were cases where certain stocks could not be treated, occurring typically two to four times per day for the treatment group. This can be attributed to two primary reasons. Firstly, some stocks were subject to trading halts due to pending material news.[2] Secondly, stocks delisted on trading day typically remained in the CRSP and TAQ datasets from the previous day and were therefore in our dataset. To address this issue, we queried TD Ameritrade for every control stock as if it were a treatment stock. Finally, TD Ameritrade sometimes linked old symbols to new (for instance, attempting to purchase FB, Facebook's old ticker, would automatically change to META, its new ticker. However, if TD Ameritrade changed this link in between the randomization phase on the previous day and the trading day, a trade would sometimes fail. Stocks that could not be treated or had the potential to be treated were removed from both the treatment and control groups. Consequently, the final dataset consisted of slightly fewer than 450 or 480 treated stocks per treatment day, while ensuring that the un-treatable stocks were removed from both groups. Additionally, there were a few rare instances where trades could not be executed, such as internet outages at Purdue, registration

[1]Because trading, while automated, required heavy supervision, and would sometimes crash, trades were monitored by the experimenter. Consequently, we did not pre-specify any of the trading days, as they depended on researcher availability. We believe it is unlikely that researcher availability varied significantly with price impact of treatment relative to control.

[2]A complete list of trading halts for the NYSE can be found at https://www.nyse.com/trade-halt-historical, and for the NASDAQ at https://www.nasdaqtrader.com/trader.aspx?id=TradingHaltHistory.

failures for certain stocks, or account liquidity issues. Fortunately, these incidents are unlikely to be related to the treatment effect.

After completing the trading phase, we obtained a new dataset comprising successfully treated and controlled stocks by excluding stocks that could not be treated and could not have been treated. Subsequently, we re-randomized the trading order and waited for at least one day before proceeding to sell the stocks, maintaining the same treatment and control assignments as in the purchasing phase.

The one-share treatment described above serves two purposes. Firstly, it maximizes the effect size per dollar traded if the effect is concave in the number of shares purchased. Secondly, it is less expensive per treated stock, allowing for randomization over the entire universe of U.S. common stocks on the NYSE, NASDAQ, and AMEX. However, it is important to note that retail market participants do not see these trades, and there is no variation in the number of shares traded.

To address the limitations of the one-share treatment and to explore additional microstructure phenomena, the second treatment arm involves varying the number of stocks purchased. With a budget of approximately $30,000 on each treatment day, it is not possible to purchase 100 shares of stocks valued at more than $300. Additionally, if 20 stocks need to be treated with 100 shares each, only stocks with a price less than $15 can be purchased. Therefore, the same stock selection procedure as in the one-share treatment is followed, but stocks with a price less than $7 are selected. The number of shares purchased can be 1, 98, 99, 100, 101, 200, or two consecutive trades of 50 shares. Initially, 14 stocks are chosen in each size bin, although some may fail to trade, as discussed later.

Table 2 provides a summary of the experimental treatment arms, including the number of days, the number of stock-days controlled and treated, the mean price of treated stocks, and the total traded funds. These treatments allow for the exploration of several microstructure phenomena, including the reporting status of trades and the impact of trade size on market behavior.

In order to address skepticism regarding the effects on large-cap stocks, two additional rounds of treatments were added after the initial experimental preanalysis plan to explicitly test the hypothesis. In the first round, on twenty different treatment days, the lesser of 100 shares or the maximum affordable with the available funds was purchased/sold for a randomly-selected stock from List A (see Appendix B). These treatments took place between June 6th 2022 and December 2nd 2022. Preliminary evidence suggested these effects were significant near the ten percent level. Consequently, a second round of the same experiment was conducted on the twenty-one largest stocks, as listed in List B (see Appendix B). For this round, all available funds, typically around $28,000, were used to purchase a single stock, while the other twenty stocks remained in the control group. This experiment was repeated twenty-one times, treating each of the twenty largest-cap stocks. While adding additional observations beyond the pre-analysis plan may be concerning prima facie evidence of “data mining" for significant results, we note that the additional observations moved our surprising effect on mega-cap stocks from mildly

Table 2: Experimental Treatment Arms

Concept	One-Share	Many-Share	Mega-Cap
# Days	85	30	43
# Stock-Days Controlled	188,623	23,982	614
# Stock-Days Treated	36,846	3,698	42
Mean Price of Treated Stock	$53.46	$4.97	$203.44
Total Traded Funds	$1,985,865	$669,287	$987,295
U.S. Stocks	x	x	x
Non-financial stocks	x	x	x
Price Limit	None	≤ $7	None
Number of shares traded	1	{1,50+50,98,99,100,101,200}	$\{100, max\ (28000/P_t)\}$

Table 2: This table describes our three treatments arms in brief. Ten of our mega-cap purchases were for 100 shares, (or the maximum we could purchase if it was above our funds, typically around $27,000), and twenty were for the maximum we could purchase. The many-share treatment had seven potential quantities, one share, two fifty-share trades in quick succession, a 98, 99, 100, 101, as well as 200 share trades.

significant to insignificant. Nonetheless, we acknowledge this experimental arm is potentially less informative than our pre-specified ones.

Our randomization successfully balances relevant covariates for each treatment. Appendix C describes balance in price, bid, ask, and cumulative daily volume levels at time of treatment, in their growth 30 seconds before treatment to treatment, as well as other variables of interest such as market capitalization and volume. Of the thirty-six treatment-control differences we test, we find three significant at the five percent level all in the many-share treatment. The cumulative volume as a fraction of total daily volume before treatment is 45% in control, but 44% in treatment, significant at the 3% level, the change in cumulative volume in the thirty seconds leading up to treatment is 0.099% in control compared to 0.068% in treatment, significant at the 5% level, and time at treatment, which is 4.03 hours after open for control, but 3.94 hours after open for treatment, significant at the 2% level. All three are linked: if treatment takes place slightly before control on average, then cumulative volume is also likely to be commensurately lower. We do not anticipate this five-minute average difference in treatment time to effect our results, and our multivariate analysis confirms that non-parametrically controlling for time, treatment effect is largely unchanged.

VI Data

This section provides a brief overview of the data construction and datasets used in the study. For a more detailed description, please refer to the Preanalysis Plan.

The first step of the data construction process involves linking the most recent CRSP header and price datasets, TAQ quote and trade datasets, and trade data collected through TD Ameritrade's API. Each individual trade is located in the TAQ dataset and matched based on price, size, and second. In cases where identical trades occur in terms of size, price, and second, the first trade is removed. However, this choice does not impact the analysis, and any other trade could have been chosen for removal instead. For each individual stock, the second trade in the TAQ dataset is identified, or the theoretical trade if it is a control observation. This second trade typically occurs two seconds after we submit a trade to TD Ameritrade. Using this information, an event study dataset is created.

To construct the event study dataset, a complete dataset at the one-second level is generated from 9:30 AM to 4:00 PM on the trading day. The "last transaction price" is determined by taking the last transaction price at each second. If a second does not have a transaction price (excluding our own trade), the price is filled forward in time. If a trade occurs near 9:30 AM, the first observed price is filled backward, and if it occurs near 4:00 PM, the last observed price is filled forward. This ensures that at least 5 minutes of data is available before a trade, and 10 minutes of data is available after the trade. In cases where there are multiple trades per second, the value-weighted mean price is taken as the new transaction price.

The National Best Bid and Offer (NBBO) data series are constructed using the NBBO, quote, and trade datasets from TAQ, following the methodology outlined in Holden and Jacobsen (2014). When multiple best bid or ask quotes occur in the same second, the mean within that second is taken as the resolved quote.

Cumulative abnormal trading volume is calculated by taking the cumulative sum of volume traded in the TAQ dataset by second (excluding our own trade) and dividing it by the total volume for that day. This differs from the initially registered intention of using a sixty-second moving average of volume. The final dataset consists of stacked event studies, with each event study representing one stock.

Implied volatility data, used as a control variable, is sourced from the OptionMetrics 2022 update available on WRDS, while data on short interest, institutional ownership, and earnings report dates were gathered from CompuStat data.

VII Results

The primary pre-registered objective of this study is to present the causally-identified impulse response function for various outcomes of interest: (1) last transaction price (excluding our trade), (2) cumulative excess volume as a percentage of total daily volume, (3) bid price, and (4) ask price, separately for the one-share, many-share, and mega-cap trades. To effectively illustrate the dynamics, we estimate and report causal impulse response functions using simple differences between the treatment and control groups. As outlined in our pre-registration plan, we employ randomization inference to calculate standard errors, which is similar to bootstrapping but involves re-assigning treatment through sampling from the possible treatment assignment space. It is important to note that these standard errors are conservatively biased upwards in the presence of treatment effect heterogeneity (Athey and Imbens, 2017).

To investigate the underlying heterogeneity behind the treatment effects, we conduct univariate sorts of treatment effect five seconds after treatment against several covariates of interest. The univariate sorts reveal a significant relationship between effect size and the time of day for the one-share treatment. However, the many-share treatment yields nearly identical point estimates that are statistically insignificant, with a mean effect early in the day of two basis points, declining to less than half by 11:00 AM. While no relationship is found with respect to the number of days until earnings, we observe that stocks with higher implied volatility, lower market capitalization, lower trading volume, higher realized volatility, and higher spread tend to exhibit statistically significant treatment effects. There is only modest evidence supporting the presence of date fixed effects.

To address the potential collinearity or interaction among variables and capture treatment heterogeneity, we employ a generalized random forest (GRF) model as outlined in Athey, Tibshirani, and Wagner (2019). GRF extends the random forest algorithm to enable the estimation of causal effects and treatment heterogeneity. By combining the strength of random forests with the framework of causal inference, GRF provides valuable insights into causal questions and estimates treatment effects.

To provide an economic interpretation of the impulse response functions, we further analyze and report the results of the three-fold decomposition outlined in Section III. The findings indicate that while the significance of the results diminishes over time, the more highly powered one-share and many-share trades exhibit statistically significant "informational" or "permanent" effects, whereas the mega-cap treatment remains unidentified as it lacks an initial effect.

In our analysis, we find that certain variables, such as spread, market capitalization, and time of day, retain their explanatory power and survive the variable selection process implemented by the GRF model. On the other hand, some variables that were significant in the univariate sorts lose their explanatory power when accounting for other covariates. This demonstrates the ability of the GRF model to identify the key variables associated with treatment effects while mitigating the potential confounding effects of collinearity or interaction among variables.

VII.1 Causal Impulse-Response Functions of Flow

In this subsection, we present the impulse-response functions for various outcomes, including stock price excluding our trade, cumulative excess volume as a percentage of total daily volume, bid price, and ask price. To estimate the causal impulse response function for outcome Y at time t, we calculate the mean difference between the treatment group (λ_t) and the control group (α_t), as shown in Equation

8:

$$Y_{it} = \alpha_t + \lambda_t D_i^{\text{Treat}} + \epsilon_t \tag{8}$$

When analyzing price outcomes, it is common practice to use signed volume, multiplying it by negative one for sell orders and one for buy orders. However, for convenience, we multiply Y by negative one when it represents a price (e.g., last price excluding our trade, bid price, or ask price), but not when it denotes volume. When treated separately (not shown), buy and sell effects on all four variables are statistically indistinguishable up to sign.

Previous studies have used different approaches in formulating Equation 8, such as focusing on the occurrence of a signed trade or using signed volume or the square root of signed volume in place of D_t^{Treat}. For our one-share trade, the treatment in Equation 8 is equivalent to all these approaches. In the case of the many-share trade, our main analysis reports results for treatments above 98 shares, excluding one-share trades, but they are combined as a single treatment, and coefficients could be scaled by 100 or 10 to retrieve the relevant estimate. We analyze size effects separately in Appendix F. As outlined in our pre-analysis plan, we report standard errors estimated through randomization inference. This involves randomly reassigning treatment 200 times while maintaining treatment-control balance across dates and calculating the 95% confidence interval based on these 200 repetitions. We compare the randomization inference standard errors with homoskedastic, heteroskedastic-robust, and bootstrapped standard errors in Appendix Section E.

While most studies focus on price or midpoint bid/ask as the primary outcomes of interest, we also include bid price, ask price, and subsequent cumulative traded volume as a fraction of daily volume as additional outcome variables of interest. While some authors highlight the importance of considering only the signed volume in excess of expectations and defining price changes relative to some baseline expected price, a randomized controlled trial effectively controls for these counterfactual baselines through the control term α_t. The advantage of Equation 8 becomes apparent when estimating the distribution of outcomes, even if the mean counterfactual y is close to zero for short time horizons.

Figure 3 presents the causal impulse response estimated using Equation 8 for the one-share treatment, while Figure 4 displays the same analysis for the many-share treatment. The equivalent results for the mega-cap treatment are reported in Appendix Figure 10, with no significant findings and large standard errors. All standard errors are calculated through randomization inference, with clustering at the date level.

Both Figure 3 and Figure 4 reveal a similar pattern: the last price excluding the treatment price, cumulative volume, bid price, and ask price are significantly influenced by a signed flow. The immediate average effect of a single-share trade on subsequent prices is close to one basis point, while trades between 98 and 200 shares exhibit an effect of nearly six basis points. It is important to note

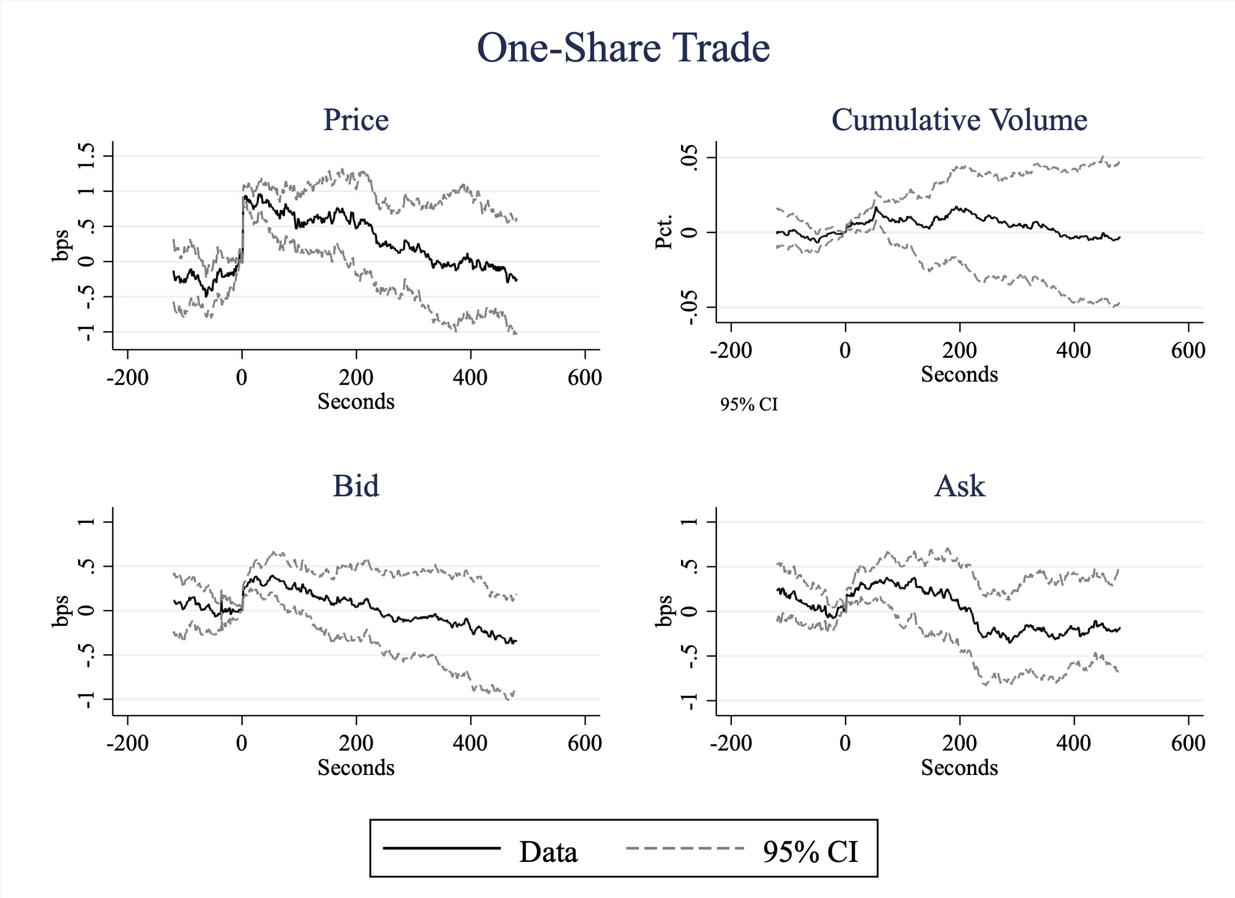

Figure 3: This figure depicts the mean effect of a one-share purchase of treated stocks on observed transaction price (excluding the experimental trade), cumulative volume as a percent of total daily volume, bid price, and ask price. All price effects are measured in basis points, while cumulative volume is measured in percent. All standard errors are calculated via randomization inference.

that the displayed price outcomes do not represent the price the experimentalist pays, as it is removed from the time series. These results clearly demonstrate a direct causal impact of flows on prices, indicating that the act of trading itself affects prices, even when the trade is random. This finding does not distinguish whether flows affect market maker information as a mediator or directly contribute to "price formation" rather than solely "price discovery" Bouchaud et al. (2018).

While the majority of empirical literature has focused on prices, both treatments in our study show significant effects on trading volume. For the one-share trade, the highest point estimate suggests a 0.017% increase in volume traded as a fraction of the daily traded volume 52 seconds after the treatment. The largest point estimate for the many-share treatment before continuous insignificance is a 0.135% higher volume at 61 seconds after the treatment. The one-share trade remains consistently insignificant shortly after one minute, while the many-share trade becomes persistently insignificant

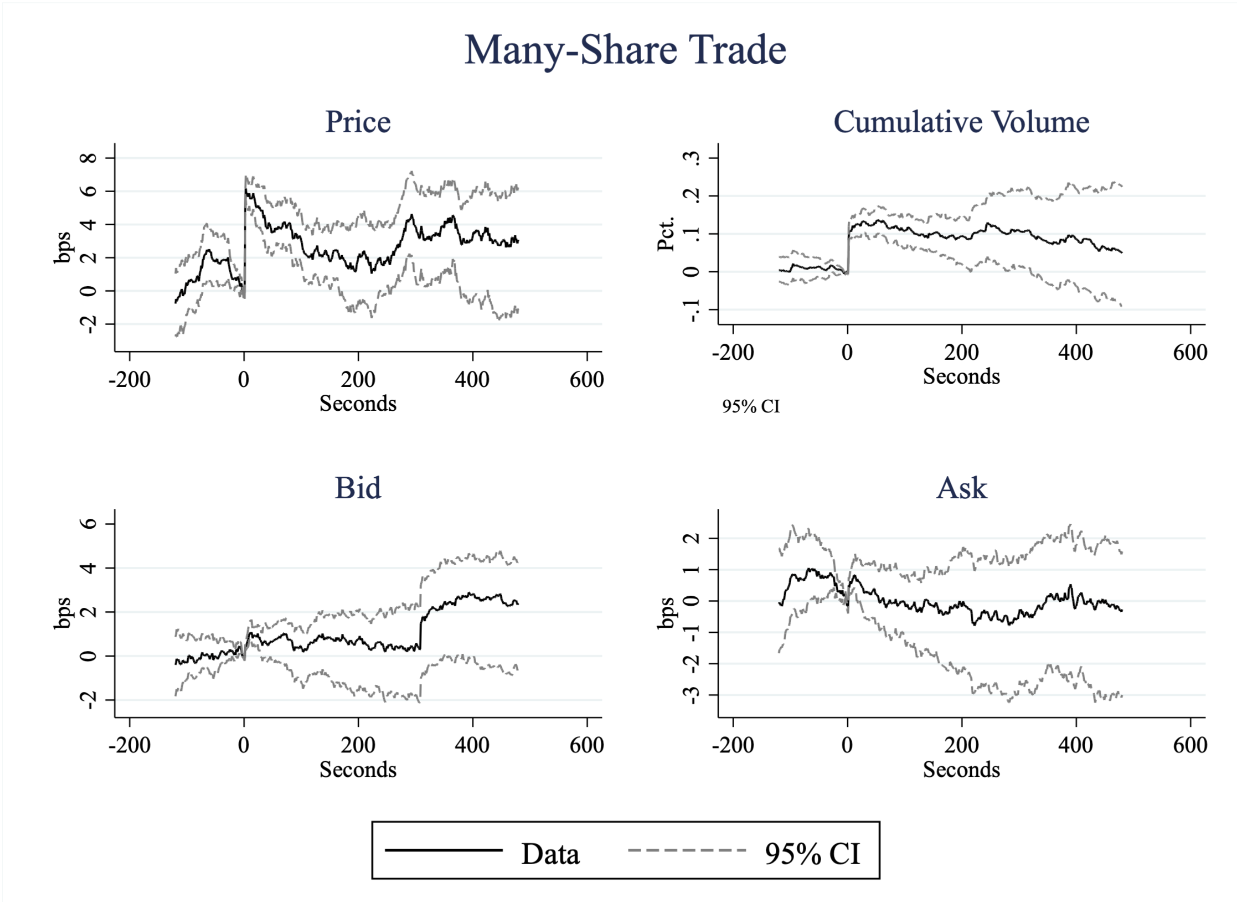

Figure 4: This figure depicts the mean effect of a many-share purchase of treated stocks on observed transaction price (excluding the experimental trade), cumulative volume as a percent of total daily volume, bid price, and ask price. Trades of size 98 shares or greater are combined into one treatment, with trades of size zero used as control. Trades of size one are excluded from this graph. All price effects are measured in basis points, while cumulative volume is measured in percent. All standard errors are calculated via randomization inference.

after five minutes. Both treatments gradually converge toward zero, with the many-share trade's point estimates intersecting with zero after 13 minutes. These results indicate that trading behavior tends to be temporarily displaced towards a signed trade. It does not appear that securities experiencing larger volume increases exhibit higher or lower residual price increases, as indicated by a correlation of 0.02 between residual price and residual size effects at five seconds, with similar and insignificant results across the entire time horizon.

In both the one-share and many-share treatments, the bid and ask prices increase, although to a lesser extent than the last traded price. For the one-share trade, the bid (ask) price increases by approximately 0.40 (0.34) basis points at 53 seconds after the trade. While this increase is overshadowed by the 1.07 basis point increase in the last traded price, it is worth noting that the average spread is nearly 100 basis points. As a result, the midpoint and last traded price remain relatively similar in relation to the bid-ask spread. Furthermore, due to the nearly symmetric movement of the bid and ask prices, the spread (not shown) is never statistically or economically significantly affected by a trade.

VII.2 Univariate Sorts

To explore the underlying heterogeneity in price impact, we conduct univariate sorts of the price impact by potential explanatory variables. In Figure 5, we present the mean treatment effect and confidence intervals by quintile for implied volatility, market cap, volatility, and in Figure 6 we display the same information for spread, days until earnings reported, percent of the stock held short, and the fraction institutionally held. It is important to note that the quintiles are created within each treatment, so they may not correspond to the same levels of the variable. Additionally, for visual convenience, we divide the many-share price results by six.

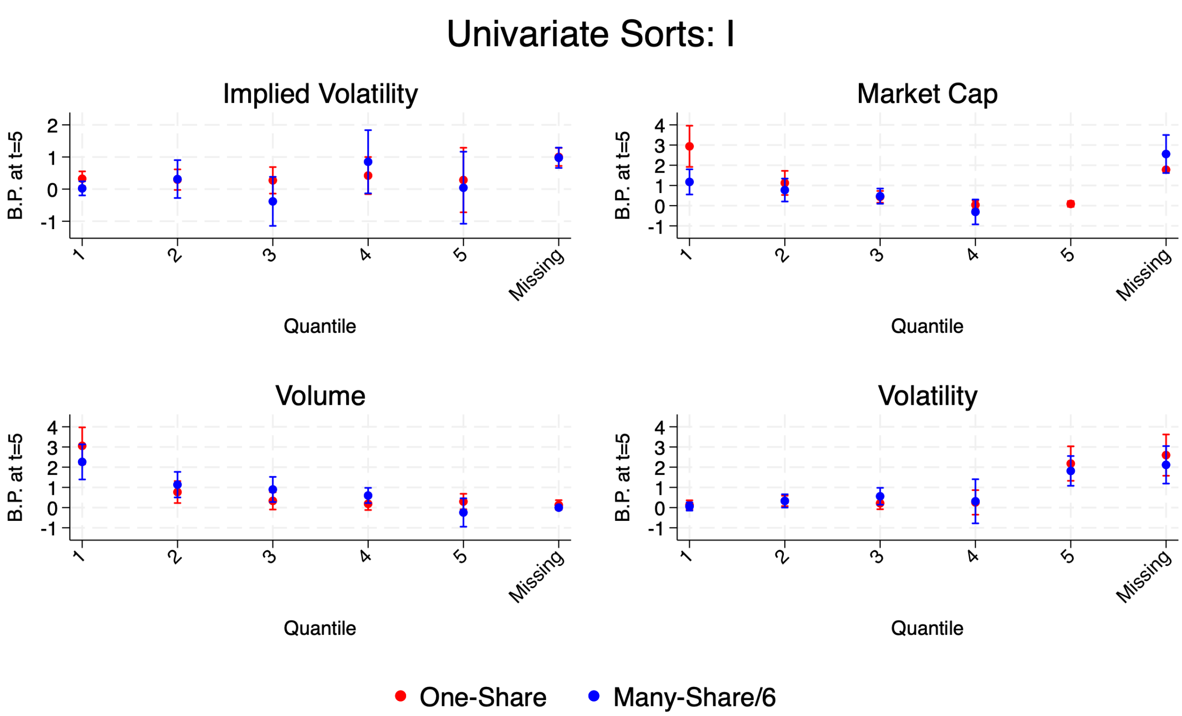

Figure 5: This figure depicts the estimated point coefficients and confidence intervals of the treatment effect on five univariate sorts: by market cap, by volume, by historical volatility, by option-implied volatility, and by spread. Coefficients and confidence intervals for the many-share treatment are scaled down by a factor of four, for convenience. Only treatment observations with quantities greater than or equal to 98 shares were included in the many-share treatment.

Figure 5 reveals several patterns. Market capitalization and volume are negatively linked to price impact, while volatility is positively linked. Implied volatility is not significantly linked to price impact,

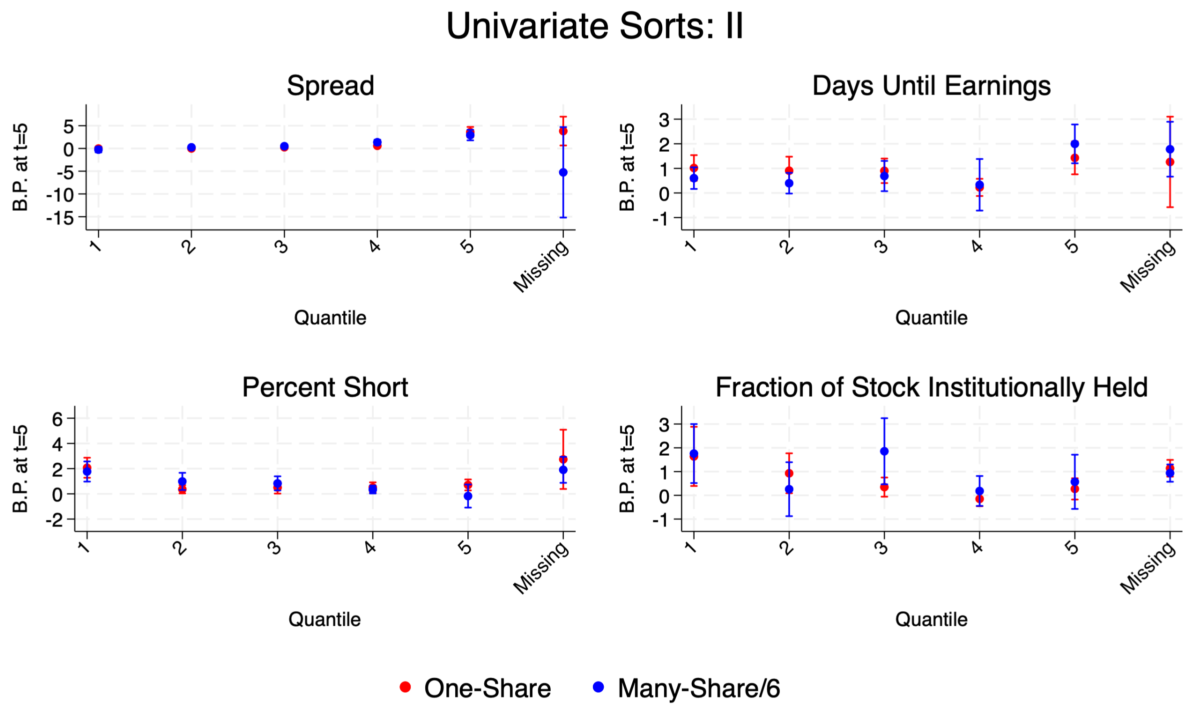

Figure 6: This figure depicts the estimated point coefficients and confidence intervals of the treatment effect on five univariate sorts: by market cap, by volume, by historical volatility, by option-implied volatility, and by spread. Coefficients and confidence intervals for the many-share treatment are scaled down by a factor of four, for convenience. Only treatment observations with quantities greater than or equal to 98 shares were included in the many-share treatment.

though this may be linked to a far larger fraction of stocks missing implied volatility data than any other variable. Similarly for Figure 6: market impact is positively associated with spread, negatively associated with the percent of the stock short, and is not clearly related to days until earnings or the fraction of the stock institutionally held. Surprisingly, after rescaling, the many-share trade shows similar rescaled price impacts for both groups.

Of the eight sorts we display in Figures 5 and 6, five align with our expectations. More valuable companies, with larger market capitalization, are expected to have smaller price impacts when holding the price constant, and the results confirm this expectation. Similarly, more liquid companies, as measured by higher trading volume or smaller spreads, are less affected by price impacts, which is in line with our intuition. Additionally, more volatile companies, which may exhibit higher uncertainty or lower liquidity, tend to have larger impacts, as anticipated. Finally, iour finding that price impact occurs primarily on stocks with little held short is consistent with the idea that the variable reflects the presence of informed parties willing to correct prices.

We also provide a nonparametric estimate of time-of-day effects in Figure 7, which we estimate as

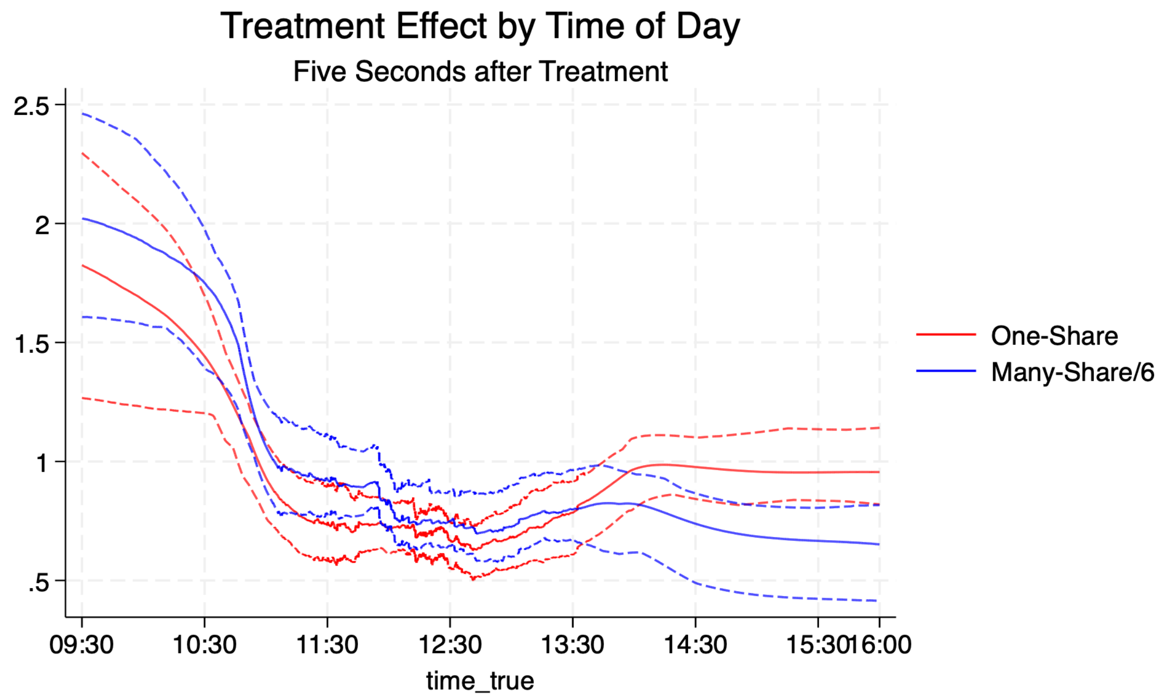

Figure 7: This figure depicts the thirty-minute moving average of treatment effect by order time of day. Only treatment observations with quantities greater than or equal to 98 shares were included in the many-share treatment.

a 30-minute moving average. We find enormous variation in the price impact by time of day: in the first hour of the day, our average price impact for a one-share trade was more than twice as large as during the middle of the trading day for both one-share and many-share stocks.

VII.3 Multivariate Analysis

In addition to the empirical evidence presented, there is also theoretical support for the expectation of heterogeneous price impact. Gabaix and Koijen (2021) provide theoretical justification based on the dependency of price impact on volume traded as a fraction of market capitalization. On the other hand, Andrea, Israel, and Moskowitz (2018) estimate price impact in relation to the purchase of daily traded volume. The univariate sorts we conducted offer initial evidence that price impact heterogeneity may depend on factors beyond volume, such as time of day. However, due to potential collinearity between variables like market capitalization, volume, and percent spread, it is valuable to conduct a joint analysis of price impact heterogeneity.

To nonparametrically estimate the heterogeneous treatment effects, we employ the "generalized random forest" estimator, specifically a "causal forest," as discussed in Athey, Tibshirani, and Wagner

(2019). Causal forests combine the principles of random forests with causal inference methods. While regression forests use an ensemble of decision trees to predict outcome variables by averaging the predictions, causal forests utilize the trees to identify a neighborhood of similar points and apply random forests as an adaptive nearest neighbor method. The advantage of this model is that it provides computationally feasible non-parametric estimates, even with a relatively dense set of predictive covariates, and allows us to remain agnostic about the specific sources of heterogeneity. For instance, if market impact is high for low-cap, low-volume stocks early in the day, the generalized random forest model does not require the specification of a specific set of interactions that may be challenging to identify as researchers.

Specifically, we estimate a causal forest with 7,000 trees on a dataset of covariates and price at time t, with each forest only conditioning on one time period. Each time period is independently estimated. To avoid large price changes unrelated to my trades driving results I require any given node to have at least 100 observations. We focus on the effect five seconds after trading, and run the analysis separately for the one-share and many-share treatments.

To better understand heterogeneity while avoiding overfitting, we split the sample into two halves. The first half is used as a "training" dataset, which I use to order stocks by predicted impact, or search for a parsimonious set of rules that predict the largest impact on prices. Of course, largest-effect stocks are likely due to measurement error, and my estimates for their mean effect will be biased. To obtain unbiased estimates, we use the rule generated by the first dataset and evaluate it using the second dataset. Because we are interested in the potential explanatory power of variables outside of spread, we run the analysis both including and excluding spread from the dataset of variables our random forest has access to.

Using the method discussed in Yadlowsky et al. (2021), I produce a sorting rule for price impact as a function of observed covariates, what they call the "rank-weighted average treatment effect". I then evaluate that rule on the untrained-half of the dataset, and depict the mean effect and its standard errors in Figure 8 for both the one- and many-share treatments, with and without access to spread as a predictive variable.

Both figures show that there is significant heterogeneity explained by observables. In the one-share trade, those stocks predicted to have the top 10% of treatment effect are four basis points, or nearly four times higher than the mean effect of one basis point. However, spread is an important predictor of the largest effect sizes: without access to spread, this same group has a slightly more than two basis point predicted effect.

Similarly, the many share trade also displays significantly higher point estimates when the forest has access to spread. However, unlike the one-share treatment, the highest predicted 10% of treatment appears to be driven primarily by noise: while the treatment effect of the top 10% in the training

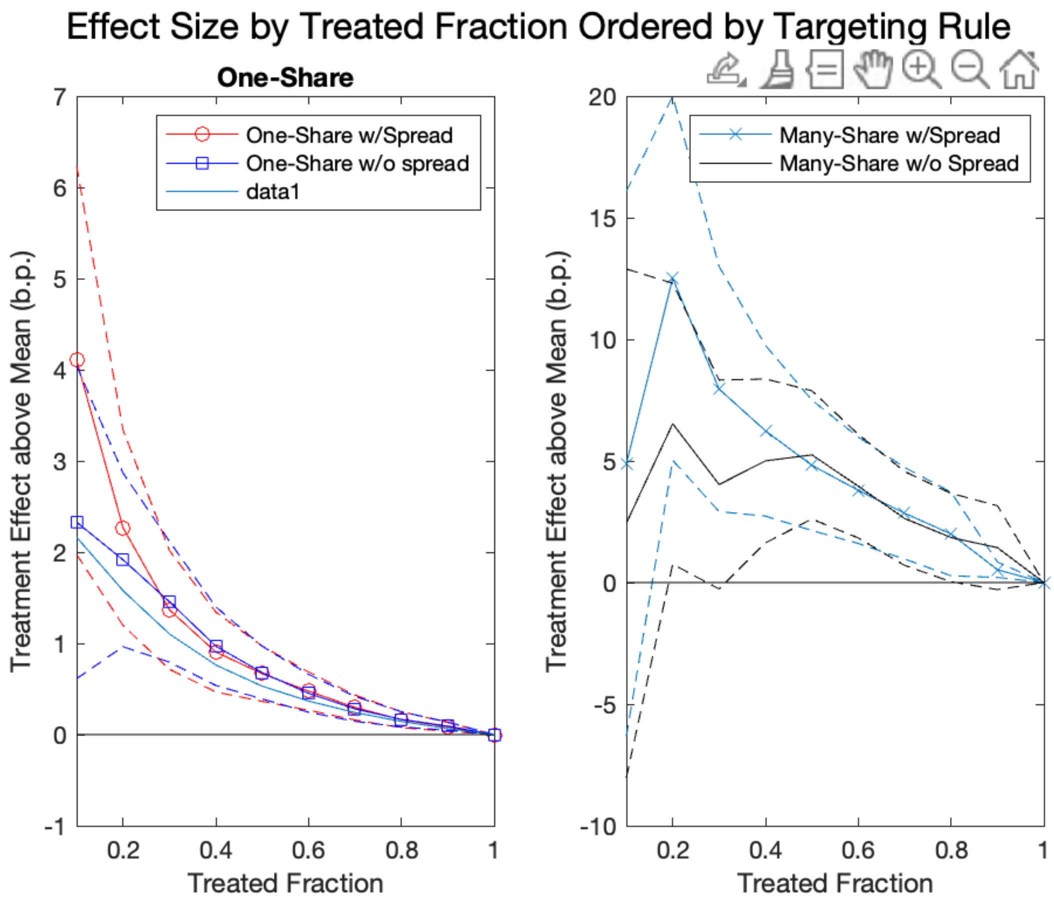

Figure 8: This figure shows the mean treatment effect as a function of the fraction of treated population, ordered by targeting rule $S(X_i)$, where X_i is a function of observables (spread, market capitalization, past volume traded, days until earnings, and so on. Data was split randomly into two halves and trained and ranked separately on each half. The was evaluated on the half of the data it was not trained on, to avoid overfitting bias.

dataset is nearly 23 basis points, it is only five basis points and insignificant in the evaluation dataset. However, we detect significant heterogeneity predictable from covariates as the treated fraction rises.

While Figure 8 has the benefit of describing the treatment effect heterogeneity that can be non-parameterically predicted by observables, it does not help evaluate which variables are most important. To do so, we look for a simple decision tree that maximizes the total effect on those treated relative to those not treated, the “policy tree" discussed in Erik, Ayush, and Wager (2020). As previously, the dataset is split in half, with the first half used to fit the policy tree, and the second used to test and evaluate it.

How deep should our policy tree be? We allow our evaluation dataset to determine the depth: while the effect of allowing more binary decisions will always monotonically increase the mean treatment

Table VII.3: Mean Effect Sizes for One-Share Trees

Depth	Test Data	Low Mean	Low S.E.	High Mean	High SE
Panel A: Training Data					
1	0	-3.83	0.51	8.99	3.74
2	0	-4.78	0.71	10.5	3.43
3	0	-5.27	0.77	19.4	4.74
Panel B: Test Data					
1	1	-4.43	1.49	10.3	3.32
2	1	-3.55	1.57	7.86	3.07
3	1	-3	1.45	11.1	4.14

effect on treated in the training dataset, it will not in the evaluation dataset. Indeed, we find that for both one-share and many-share treatments, allowing access to spread or not, there is no improvement in overall fit going past a tree of depth one. We consequently only report results among trees of depth one.

Our forest is made up of 7000 trees, each with approximately 3200 nodes: while this allows for flexibility that captures signifciant heterogeneity of Figure 8, we now turn to the question of whether there is a simple rule that summarizes a significant proportion of our heterogeneity. To do so, we search for a simple binary rule based on observable variables that chooses treatment, maximizing the total treatment effect minus the cost for a stock's inclusion, which we take to be the average treatment effect. The policy tree therefore earns points for including observations above the average, and loses points for including observations below the average.

We search for rules of depth one, two, and three that splits our sample into "high treatment effect" and "low treatment effect." To do so, we use the exact search algorithm of Erik, Ayush, and Wager (2020), though we round our values so that, for instance, there are 147 (186) unique values for market cap (volume), rather than more than 220,000 for each. Rounding values for each of our ten variables of interest dramatically reduces the exact search space. For the tree of depth three, we use hybrid tree search, which only looks ahead two splits, for computational feasibility. We evaluate the policy by training a policy tree on half the data, and then evaluating the rule using the other half.

Table VII.3 reports the estimated treatment effect size and standard error on the "low treatment effect" and "high treatment effect" groups. The first panel reports the effects on the training dataset, while the second reports the effects on the evaluation dataset. Tree decision rules that correspond to the depths reported in Table VII.3 are summarized in Table VII.3.

In the training sample, the mean "high" effect from a simple rule putting those stocks with a spread above 1.1% in the "high" group and the rest in the "low" group is nearly 9 basis points above the mean, while the "low" effect is nearly -4 basis points below the mean, both highly significantly different from one another and the group mean. This result holds up well in the evaluation sample. However, little is

Table VII.3: Policy Trees

Treatment	Access to Spread?	Split Variable	Low Mean	Low S.E.	High Mean	High SE
One-Share	No	Volume	-0.66	0.79	0.16	0.44
One-Share	Yes	Spread	-4.43	1.49	10.3	3.32
Many-Share	No	Market Cap	-4.09	4	2.12	1.91
Many-Share	Yes	Spread	-4.41	10.35	1.49	3.32

gained from additional policy depth: allowing for three splits (a depth of two) rather than one does not increase the test data's mean, and a depth of three appears clearly driven by measurement error, as the evaluation dataset's effect difference is not increased.

Of course, saying that spread predicts price impact may hardly be surprising. We then conduct a similar analysis but do not allow the policy tree to refer to spread, finding that for the one-share dataset volume is most predictive, while in the quantity treatment market cap is most predictive. Again, nothing is gained for the evaluation dataset by allowing the policy tree access to any depth above one: in each case the point difference falls significantly. Because depth greater than one is not useful for a simple rule, Table VII.3 summarizes the means and standard errors for the one-share and many-share treatments, with and without access to spread as a splitting variable for the evaluation dataset only.

VII.4 Decomposing the IRFs

While we believe the causal impulse response functions and underlying heterogeneity are interesting in their own right, it is also useful to add economic interpretation to our results. To this end, we use the three-fold decomposition of the causal impulse response functions we identify, decomposing price impact into permanent informational or adverse selection effects, AR(1) inventory effects, and short-lived order-processing costs as discussed in Section III. To isolate the long-run informational channel, we estimate using the model outlined in that section using an optimal minimum distance (OMD) estimator. With an eye to the heterogeneity identified in Sections VII.2 and VII.3, we also apply this estimator to the "high treatment effect" and "low treatment effect" groups identified in Figure **??**.

Estimating the basic model assuming homogeneity is a straightforward application of OMD. To identify the parameters ρ, π_m, π_h, λ, β, and γ off the causal impulse response, we match our model parameters to best fit the the mean treatment effect calculated in Figures 3 or 4, weighted by the inverse of the sample variance-covariance matrix, which is the sum of the variance-covariance matrix of price moments calculated for treatment and control separately.[3]. Parameters are reported in Table 3, while data and model fits are reported in Figure 9

[3]This assumes that the covariance of treatment with control between any two time periods is zero, which is consistent with our randomization.

Table 3: Three-Way Decomposition

Parameter	Description	One-Share	Many-Share	One-Share: HTE	One-Share: LTE
ρ	Inventory Decay	0.996 (0.004)	0.984 (0.012)	0.811 (8062.183)	0.961 (731.527)
π_m	Meta-Order	0.712 (0.923)	0.992 (0.010)	0.209 (0.102)	0.897 (0.167)
π_h	Herding	0.021 (0.071)	0.008 (0.009)	0.602 (8062.385)	0.063 (731.496)
λ	Information	0.042 (0.143)	-1.481 (0.305)	2.959 (0.685)	-0.040 (0.031)
β	Inventory	0.758 (0.169)	7.224 (0.850)	0.573 (0.541)	0.247 (0.065)
γ	Order-Processing	84.57 (0.320)	172.77 (1.492)	304.437 (1.105)	31.623 (0.073)

Table 3: This table depicts the results of a three-fold decomposition on the one-share and many-share treatments. "HTE" denotes the "high treatment effect" sample of stocks prescribed by the policy tree in Figure **??**. "LTE" denotes the "low treatment effect" stocks that the policy recommends not treating. At levels of ρ not near one, ρ and π_h become multicollinear, so standard errors for both are large.

Results from Table 3 for the one-share pooled treatment are perhaps not surprising. Our impulse-response function shows a clear trend toward zero, and we are able to reject any permanent informational effect outside the -0.24 to +0.32 basis-point interval. Instead, the transitory increase in price caused by our trade is inventory, though order-processing costs dwarf either of the other two components, even more so than Huang and Stoll (1997), who also found order-processing costs dwarfing the other two components, though only making up 62% of the spread rather than 99% of it. Similarly, while we are able to reject significant herding effects, we are unable to distinguish the level of expected meta-orders, which were identified primarily by a rapid price reversion not present in the data. Figure 9 shows that the model visually conforms closely to the data.

While some results in Table 3 for our many-share treatment are similar, others are starkly different. The many-share treatmnet displays similar slow rate of inventory decay, uncertainty about the level of meta-orders, and certainty about the low level of order herding. However, its point estimate differs sharply for permanent informational effects, being significantly negative at the 5% level. A reader might be surprised, as a visual inspection of Figure 9 suggests little room for a negative permanent effect, given that the data never converges to zero! This is caused by the weighting by the inverse of the variance-covariance matrix, which has both diagonals and off-diagonals significantly higher in the latter portion of our data. Consequently, our model primarily fits to the downward slope in the first 200 seconds, yielding a negative point estimate for expected effect on price that is within the data's

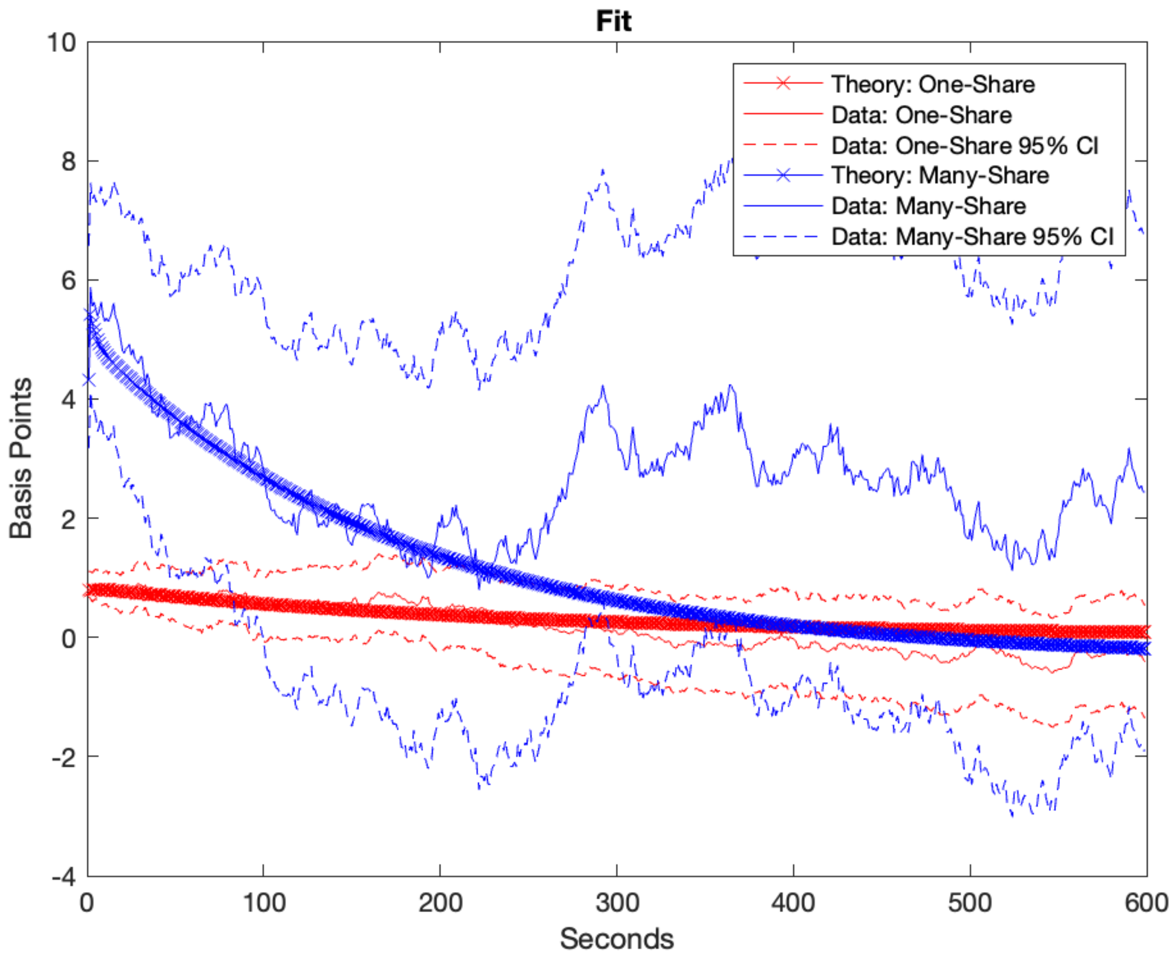

Figure 9: This figure graphically depicts the model fit to the causal impulse-response function.

95% confidence intervals, though below the mean data.

Our heterogeneity analysis finds that for the “high treatment effect" group significant evidence of information effects as large as three basis points, exceeding the inventory effects by an order of more than five. However, even our “low treatment effect" group is significantly impacted in a transitory dimension.

VIII Conclusion

This article makes several contributions to the academic literature and suggests avenues for future research.

The first contribution is methodological, as field experiments in finance are rare, particularly in the study of price dynamics. The recent development of commission-free trading and the surprising finding that a single-share retail trade can have detectable and lasting effects on the market contribute to the novelty of this research. The use of a pre-specified, randomized controlled field experiment provides a strong foundation for estimating causal market impact.

The main contribution of this article is the estimation of causal market impact using the gold standard of causal inference. The analysis demonstrates that not only transaction prices, but also volume, bid prices, and ask prices respond to a one-share trade for a duration measurable in minutes.

Another contribution is the economic interpretation of the causal impulse response functions, adapting previous identification techniques to eliminate measurement error and provide insights into the role of information in market impact.

The examination of heterogeneity in price impact is a fourth contribution of this research. Significant heterogeneity is found in relation to market capitalization, volume, implied volatility, historical volatility, and time-of-day, which are consistent across both the one-share and many-share treatment arms. However, our analysis reveals that the predictive power of many of these variables fails in multivariate nonparametric analysis. Instead, we find that spread alone explains a significant portion of price impact, with simple rules based on market capitalization or volume having no additional explanatory power beyond spread.

There are several directions for future research using the methods presented in this article. One such direction is to investigate heterogeneity in market impact around specific events such as earnings announcements, central bank statements, economic data releases, and news events. Additionally, exploring the impact of retail trades on other financial instruments beyond U.S. equities, such as cryptocurrencies, stock options, and off-the-run treasuries, could provide valuable insights. Lastly, further investigation into the effects of retail trading on volume and herding behavior would contribute to a more comprehensive understanding of market dynamics.

References

Almgren, Robert, Chee Thum, Emmanuel Hauptmann, and Hong Li. 2005. "Direct estimation of equity market impact." *Risk* 18 (7):58–62.

Ameritrade, TD. 2021. "TD Ameritrade Clearing, Inc.'s SEC 606 Order Disclosure." Https://www.tdameritrade.com/content/dam/tda/retail/marketing/en/pdf/cftc/tdac-TDA2054-q3-2021.pdf.

Andrea, Frazzini, Ronen Israel, and Tobias J. Moskowitz. 2018. "Trading Costs." *Working Paper* .

Athey, S. and G.W. Imbens. 2017. "Chapter 3 - The Econometrics of Randomized Experimentsa."

In *Handbook of Field Experiments*, *Handbook of Economic Field Experiments*, vol. 1, edited by Abhijit Vinayak Banerjee and Esther Duflo. North-Holland, 73–140.

Athey, Susan, Julie Tibshirani, and Stefan Wagner. 2019. “Generalized Random Forests.” *The Annals of Statistics* 47 (2):1148–1178.

Bouchaud, Jean-Philippe, Julius Bonart, Jonathan Donier, and Martin Gould. 2018. *Trades, quotes and prices: financial markets under the microscope*. Cambridge University Press.

Brennan, Michael J. and Avanidhar Subrahmanyam. 1996. “Market microstructure and asset pricing: On the compensation for illiquidity in stock returns.” *Journal of Financial Economics* 41 (3):441–464.

Camerer, Colin F. 1998. “Can Asset Markets Be Manipulated? A Field Experiment with Racetrack Betting.” *Journal of Political Economy* 106 (3):457–482.

Chakrabarty, Bidisha, Roberto Pascual, and Andriy Shkilko. 2015. “Evaluating trade classification algorithms: Bulk volume classification versus the tick rule and the Lee-Ready algorithm.” *Journal of Financial Markets* 25:52–79.

Cochrane, John H. 2004. “Asset pricing: Liquidity, trading, and asset prices.” *NBER Reporter Online* (Winter 2004/05):1–12.

Easley, David, Soeren Hvidkjaer, and Maureen O’Hara. 2002. “Is Information Risk a Determinant of Asset Returns?” *The Journal of Finance* 57 (5):2185–2221.

Easley, David, Nicholas M. Kiefer, and Maureen O’Hara. 1997. “The information content of the trading process.” *Journal of Empirical Finance* 4 (2):159–186. High Frequency Data in Finance, Part 1.

Easley, David, Marcos M. López de Prado, and Maureen O’Hara. 2012. “Flow Toxicity and Liquidity in a High-frequency World.” *The Review of Financial Studies* 25 (5):1457–1493.

Erik, Sverdrup, Kanodia Ayush, and Zhengyuan ZhouSusan AtheyStefan Wager. 2020. “policytree: Policy Learning via Doubly Robust Empirical Welfare Maximization over Trees.” *The Journal of Open Source Software* 5 (50):1–6.

Ernst, Thomas, Jonathan Sokobin, and Chester Spatt. 2021. “The Value of Off-Exchange Data.”

Foucault, Thierry, Marco Pagano, and Ailsa Roell. 2013. *Market Liquidity: Theory, Evidence, and Policy*. Oxford University Press.

Gabaix, Xavier and Ralph S. J Koijen. 2021. “In Search of the Origins of Financial Fluctuations: The Inelastic Markets Hypothesis.” Working Paper 28967, National Bureau of Economic Research.

Gomes, C. and H. Waelbroeck. 2015. “Is market impact a measure of the information value of trades? Market response to liquidity vs. informed metaorders.” *Quantitative Finance* 15 (5):773–793.

Hasbrouck, Joel. 1991. “Measuring the Information Content of Stock Trades.” *The Journal of Finance* 46 (1):179–207.

Hasbrouck, Joel and Duane J. Seppi. 2001. “Common factors in prices, order flows, and liquidity.” *Journal of Financial Economics* 59 (3):383–411.

Hasbrouck, Joel R. 2007. *Empirical Market Microstructure: The Institutions, Economics, and Econometrics of Securities Trading*. Oxford University Press.

Holden, Craig W. and Stacey Jacobsen. 2014. “Liquidity Measurement Problems in Fast, Competitive Markets: Expensive and Cheap Solutions.” *The Journal of Finance* 69 (4):1747–1785.

Huang, Roger D. and Hans R. Stoll. 1997. “The Components of the Bid-Ask Spread: A General Approach.” *The Review of Financial Studies* 10 (4):995–1034.

Kraus, Alan and Hans R. Stoll. 1972. “Price Impacts of Block Trading on the New York Stock Exchange.” *The Journal of Finance* 27 (3):569–588.

Kyle, Albert S. 1985. “Continuous Auctions and Insider Trading.” *Econometrica* 53 (6):1315–1335.

Lee, Charles M. C. and Mark J. Ready. 1991. “Inferring Trade Direction from Intraday Data.” *The Journal of Finance* 46 (2):733–746.

Levy, Bradford. 2022. “Price Improvement and Payment for Order Flow: Evidence from A Randomized Controlled Trial.”

Lyons, Richard K. 1995. “Tests of microstructural hypotheses in the foreign exchange market.” *Journal of Financial Economics* 39 (2):321–351.

Ni, Sophie X., Jun Pan, and Allen M. Poteshman. 2008. “Volatility Information Trading in the Option Market.” *The Journal of Finance* 63 (3):1059–1091.

O’Hara, Maureen. 1998. *Market Microstructure Theory: 1st Edition*. Wiley.

O’Hara, Maureen, Chen Yao, and Mao Ye. 2014. “What’s Not There: Odd Lots and Market Data.” *The Journal of Finance* 69 (5):2199–2236.

Schwarz, Brad Barber Xing Huang Philippe Jorion, Christopher and Terrance Odean. 2022. “The “Actual Retail Price" of Equity Trades.”

Yadlowsky, Steve, Scott Fleming, Nigam Shah, Emma Brunskill, and Stefan Wager. 2021. "Evaluating Treatment Prioritization Rules via Rank-Weighted Average Treatment Effects."

A Randomization Procedure

1. Most recent TAQ and CRSP datasets are merged, with only successful merges kept. (9370 securities)
2. Only stocks without a suffix, e.g. excluding convertible stock, warrants Class A shares, etc. (9295 stocks)
3. Require membership in NYSE, NASDAQ, or AMEX (6815 stocks)
4. Require U.S. stock incorporation (4164 stocks)
5. Drop NAICS 52 (financial firms, blank-check companies) (3282 stocks)
6. Require successful location of TD-Ameritrade symbol, excluding companies that changed ticker from last CRSP (3280 stocks)
7. Randomly treat 450 (later 480) stocks, leaving approximately (2800 control stocks)

The procedure for creating the sample was as follows. We first acquire the most recent TAQ dataset (typically around 12,000 securities), and merge it with CRSP (typically around 9500 securities), keeping only successful merges. This will exclude new stocks that were listed after the last CRSP dataset was released, with the previous month's dataset released typically around the middle of the month. Consequently, CRSP data ranged from 12 to 33 days "stale." It will also exclude stocks that did not have TAQ data because they were not traded. This reduces the dataset to approximately 9370 stocks. We then drop stocks in the joint CRSP-TAQ dataset with "suffixes," denoting convertible stock, warrants, preferred shares, and other unusual securities, reducing the sample to 9295 stocks. We require the stock to be listed on the NYSE, NASDAQ, or AMEX, reducing the sample to 6815 stocks. We require the stock to be incorporated in the United States, reducing the sample to 4164 stocks. We exclude the NAICS code (52), reducing the stock to 3282 stocks. Finally, we require that all stocks be locatable with TD Ameritrade, which excludes stocks that have recently changed name or ticker (≈2 stocks).

After selecting our sample, we randomly assign 450 (480 after June 2022) stocks to be in the "treatment" group, while approximately 2800 stocks are in assigned to the control group. However, during purchase, a small number of stocks were unavailable to be traded, typically because of a delisting event. For instance, on August 24th, 2022 we attempted to treat stock ticker "SMED," but received only the error: "This security is currently unavailable. Contact us for details." Further inspection revealed that SMED was delisted on August 23rd, after being purchased by Aurora Capital Partners in July. This occurs on average two to four times. Such stocks are dropped from both the treatment and control datasets, as they would have been excluded before randomization if data were available.

B Mega-Cap Stock List

Table B1: Mega-Cap Stocks

Symbol	Group A: Initial 10 Stocks	Group B: Second 21 Stocks
META	x	
GOOG	x	
JNJ	x	x
AAPL	x	x
WMT	x	x
TSLA	x	x
NVDA	x	x
AMZN	x	x
PG	x	x
MSFT	x	x
ABBV		x
ORCL		x
PFE		x
HD		x
AVGO		x
MRK		x
COST		x
PEP		x
CVX		x
XOM		x
LLY		x
TMO		x
KO		x
Dates	Jun 2022-Dec 2022	Feb 2023-Apr 2023

Table B1: This table lists the tickers for the two waves of mega-cap stock treatments.

C Balance Tables

Table C1 Balance in One-Share Treatment

Variable	Control	Treatment	P-Value
Price (t=-2)	52.81	52.08	0.39
	(0.35)	(0.71)	
Bid (t=-2)	51.96	51.11	0.30
	(0.33)	(0.66)	
Ask (t=-2)	52.65	52.07	0.49
	(0.34)	(0.71)	
Cumulative volume before treatment (thousands) (t=-2)	726.73	720.43	0.76
	(8251.40)	(18088.30)	
Cumulative volume before treatment	0.34	0.34	0.46
as fraction of total daily volume	(0.00)	(0.00)	
Change in price from t=-30 to t-2 (b.p.)	0.04	-0.15	0.24
	(0.06)	(0.16)	
Change in bid from t=-30 to t-2 (b.p.)	-0.04	-0.05	0.89
	(0.03)	(0.08)	
Change in ask from t=-30 to t-2 (b.p.)	-0.03	-0.09	0.52
	(0.04)	(0.08)	
Change in cumulative volume from t=-30 to t-2 (percent)	-0.0767	-0.0773	0.82
	(0.0011)	(0.0025)	
Market Capitalization (billions)	11.50	11.46	0.94
	(0.17)	(0.38)	
Shares outstanding (millions)	151.08	152.11	0.72
	(1.17)	(2.70)	
Time at treatment (hours since market open)	3.97	3.97	0.92
	(0.00)	(0.00)	
Number of observations	196954	38438	

Table C1: This table describes the balance of treatment and control stocks in the one-share treatment for multiple variables of interest, as well as the p-value for their pooled two-way difference.

Table C2 Balance in Quantity Treatment

Variable	Control	Treatment	P-Value
Price (t=-2)	2.83	2.82	0.95
	(0.02)	(0.06)	
Bid (t=-2)	2.80	2.78	0.78
	(0.02)	(0.05)	
Ask (t=-2)	2.84	2.82	0.73
	(0.02)	(0.05)	
Cumulative volume before treatment (thousands) (t=-2)	657.61	519.81	0.19
	(34279.24)	(47386.19)	
Cumulative volume before treatment	0.45	0.44	0.03
as fraction of total daily volume	(0.00)	(0.01)	
Change in price from t=-30 to t-2 (b.p.)	-0.07	1.45	0.10
	(0.29)	(0.91)	
Change in bid from t=-30 to t-2 (b.p.)	-0.07	0.00	0.89
	(0.15)	(0.45)	
Change in ask from t=-30 to t-2 (b.p.)	0.01	0.93	0.09
	(0.17)	(0.55)	
Change in cumulative volume from t=-30 to t-2 (percent)	-0.0989	-0.0684	0.05
	(0.0050)	(0.0099)	
Market Capitalization (billions)	0.22	0.18	0.12
	(0.01)	(0.01)	
Shares outstanding (millions)	73.69	67.37	0.10
	(1.25)	(2.58)	
Time at treatment (hours since market open)	4.03	3.94	0.02
	(0.00)	(0.00)	
Number of observations	19786	2177	

Table C2: This table describes the balance of treatment and control stocks in the many-share treatment for multiple variables of interest, as well as the p-value for their pooled two-way difference.

Table C3 Balance in Mega-Cap Treatment

Variable	Control	Treatment	P-Value
Price (t=-2)	234.56	203.30	0.46
	(10.76)	(25.18)	
Bid (t=-2)	231.93	202.07	0.49
	(11.14)	(25.23)	
Ask (t=-2)	232.04	202.14	0.48
	(11.16)	(25.25)	
Cumulative volume before treatment (thousands) (t=-2)	11516.98	17477.14	0.06
	(7.7e+05)	(4.6e+06)	
Cumulative volume before treatment	0.42	0.46	0.14
as fraction of total daily volume	(0.01)	(0.03)	
Change in price from t=-30 to t-2 (b.p.)	-0.09	0.54	0.53
	(0.25)	(0.98)	
Change in bid from t=-30 to t-2 (b.p.)	-0.04	0.08	0.92
	(0.30)	(0.97)	
Change in ask from t=-30 to t-2 (b.p.)	-0.07	-0.11	0.97
	(0.29)	(0.94)	
Change in cumulative volume from t=-30 to t-2 (percent)	-0.1349	-0.1550	0.34
	(0.0052)	(0.0240)	
Market Capitalization (billions)	600.09	664.08	0.48
	(22.74)	(89.05)	
Shares outstanding (millions)	4002.18	4521.42	0.41
	(158.16)	(581.98)	
Time at treatment (hours since market open)	4.29	4.26	0.93
	(0.00)	(0.00)	
Number of observations	640	43	

Table C3: This table describes the balance of treatment and control stocks in the mega-cap treatment for multiple variables of interest, as well as the p-value for their pooled two-way difference.

D Calendar

Calendar of Treated Dates
2021

December

S	M	T	W	T	F	S
			01	02	03	04
05	06	07	08	09	10	11
12	13	14	15	16	17	18
19	20	21	22	23	24	25
26	27	28	29	30	31	

2022

January

S	M	T	W	T	F	S
						01
02	03	04	05	06	07	08
09	10	11	12	13	14	15
16	17	18	19	20	21	22
23	24	25	26	27	28	29
30	31					

February

S	M	T	W	T	F	S
		01	02	03	04	05
06	07	08	09	10	11	12
13	14	15	16	17	18	19
20	21	22	23	24	25	26
27	28					

March

S	M	T	W	T	F	S
		01	02	03	04	05
06	07	08	09	10	11	12
13	14	15	16	17	18	19
20	21	22	23	24	25	26
27	28	29	30	31		

April

S	M	T	W	T	F	S
					01	02
03	04	05	06	07	08	09
10	11	12	13	14	15	16
17	18	19	20	21	22	23
24	25	26	27	28	29	30

May

S	M	T	W	T	F	S
01	02	03	04	05	06	07
08	09	10	11	12	13	14
15	16	17	18	19	20	21
22	23	24	25	26	27	28
29	30	31				

June

S	M	T	W	T	F	S
			01	02	03	04
05	06	07	08	09	10	11
12	13	14	15	16	17	18
19	20	21	22	23	24	25
26	27	28	29	30		

July

S	M	T	W	T	F	S
					01	02
03	04	05	06	07	08	09
10	11	12	13	14	15	16
17	18	19	20	21	22	23
24	25	26	27	28	29	30
31						

August

S	M	T	W	T	F	S
	01	02	03	04	05	06
07	08	09	10	11	12	13
14	15	16	17	18	19	20
21	22	23	24	25	26	27
28	29	30	31			

September

S	M	T	W	T	F	S
				01	02	03
04	05	06	07	08	09	10
11	12	13	14	15	16	17
18	19	20	21	22	23	24
25	26	27	28	29	30	

October

S	M	T	W	T	F	S
						01
02	03	04	05	06	07	08
09	10	11	12	13	14	15
16	17	18	19	20	21	22
23	24	25	26	27	28	29
30	31					

November

S	M	T	W	T	F	S
		01	02	03	04	05
06	07	08	09	10	11	12
13	14	15	16	17	18	19
20	21	22	23	24	25	26
27	28	29	30			

December

S	M	T	W	T	F	S
				01	02	03
04	05	06	07	08	09	10
11	12	13	14	15	16	17
18	19	20	21	22	23	24
25	26	27	28	29	30	31

2023

January

S	M	T	W	T	F	S
01	02	03	04	05	06	07
08	09	10	11	12	13	14
15	16	17	18	19	20	21
22	23	24	25	26	27	28
29	30	31				

February

S	M	T	W	T	F	S
			01	02	03	04
05	06	07	08	09	10	11
12	13	14	15	16	17	18
19	20	21	22	23	24	25
26	27	28				

March

S	M	T	W	T	F	S
			01	02	03	04
05	06	07	08	09	10	11
12	13	14	15	16	17	18
19	20	21	22	23	24	25
26	27	28	29	30	31	

April

S	M	T	W	T	F	S
						01
02	03	04	05	06	07	08
09	10	11	12	13	14	15
16	17	18	19	20	21	22
23	24	25	26	27	28	29
30						

One-Share	Many-Share
FAANG	SPY/QQQ
Buy	Sell

E Standard Errors

Figure 10: This figure depicts standard errors of the mean price effect for the one-share treatment calculated four different ways. “OLS/homoskedastic" assumes spherical errors. “Huber-White" calculates heteroskedasticity-robust standard errors using Huber-White’s method. “Bootstrap" calculates bootstrapped standard errors, stratified by date and treatment, and using 200 repetitions. “Randomization inference" calculates standard errors via randomization inference, stratified by date and treatment, also using 200 repetitions. A fifth method, equivalent to one-step GMM standard errors using {symbol x treatment} as an instrument, so that there is heteroskedasticity between symbols but homoskedasticity within symbol, is visually indistinguishable from Huber-White standard errors, and therefore not shown.

F Size Effects

As discussed in Section V, our “Quantity" experiment had several different levels of quantities purchased: a single share, 98 shares, 99 shares, 100 shares, 101 shares, 200 shares, and two successive 50 share trades. While the trades larger than one share are pooled

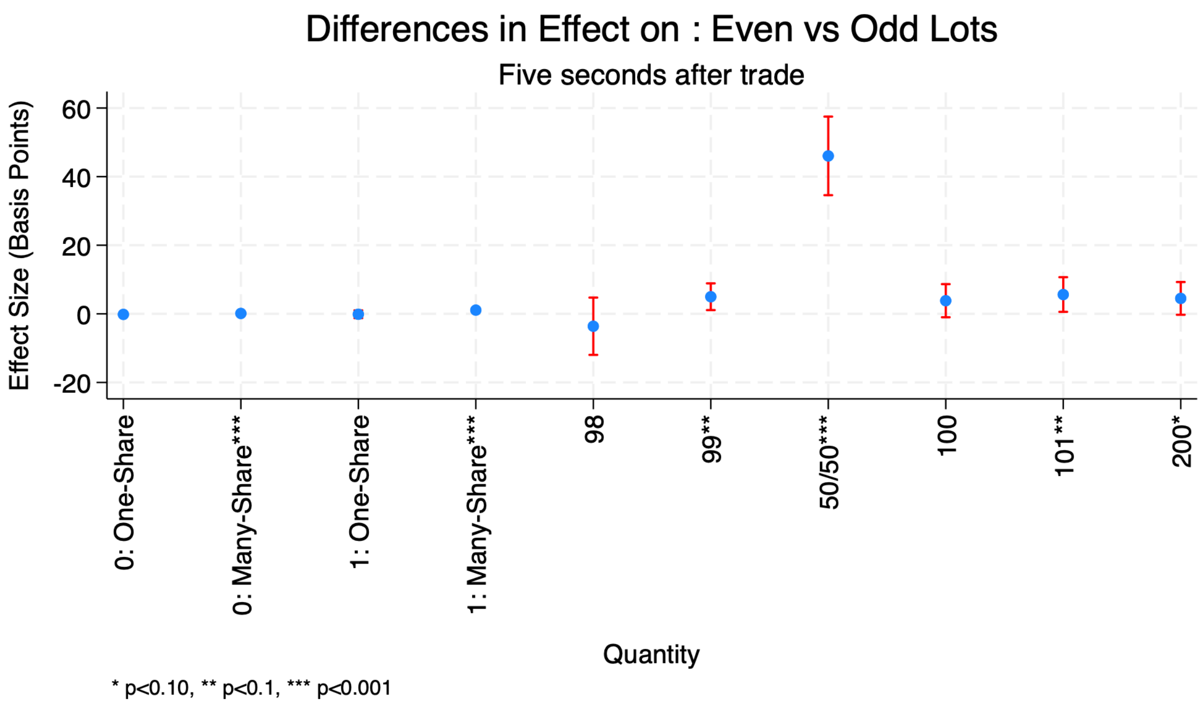

Figure 11: This figure depicts the mean effect of the many-share “Quantity" purchase on treated stocks by size, compared to control. Each reports the coefficients, with 95% standard error bars, of the effect of a purchase of a size compared to the control group. The x-axis label of “50/50" shows the effect of purchasing 50 shares twice in two “odd lot" trades, rather than a single 100 share even lot trade. Coefficients for 98 and purchase of treated stocks on observed transaction bid and ask prices. The lefthand figure depicts the difference between the mean bid price of treated stocks and the mean bid price of control stocks, normalized to zero at $t = -2$ seconds. All bid and ask price data for sells is multiplied by negative one. The righthand figure depicts the same difference, but for ask prices rather than bid prices. All quotes exclude the experimental quote. All confidence errors are calculated using randomization inference (200 samples), clustered at the date level.