Does Algorithmic Trading Reduce Information Acquisition?

Brian M. Weller
Duke University

I demonstrate an important tension between acquiring information and incorporating it into asset prices. As a salient case, I analyze algorithmic trading (AT), which is typically associated with improved price efficiency. Using a new measure of the information content of prices and a comprehensive panel of 54,879 stock-quarters of Securities and Exchange Commission (SEC) market data, I establish instead that the amount of information in prices decreases by 9% to 13% per standard deviation of AT activity and up to a month before scheduled disclosures. AT thus may reduce price informativeness despite its importance for translating available information into prices. (*JEL* G10, G12, G14)

Received May 21, 2016; editorial decision October 25, 2017 by Editor Itay Goldstein. Authors have furnished an Internet Appendix, which is available on the Oxford University Press Web site next to the link to the final published paper online.

Financial markets aggregate information to determine asset values and allocate resources and risk. Traders contribute to price discovery by (1) acquiring new information and (2) incorporating existing information into prices. In this paper, I investigate a striking conflict between these components of price discovery. Mechanisms or traders that improve price efficiency with respect to *existing* information themselves deter information acquisition and diminish price efficiency with respect to *acquirable* information. This trade-off is important because innovations improving market efficiency classically defined (by, e.g., Fama (1970)) may nonetheless distort asset prices and worsen allocations.

I investigate this tension within price discovery by considering the epochal technological changes associated with algorithmic trading. Algorithmic trading, or the use of computer systems to execute trading strategies, has

I thank the Editor, Itay Goldstein; two anomymous referees; Snehal Banerjee, Eric Budish, Anna Cieslak, Nicolas Crouzet, Ian Dew-Becker, Austin Gerig, Andrei Kirilenko, Robert Korajczyk, Jia Li, Matt Lyle, Joshua Mollner, James Naughton, Ryan Riordan, Matt Spiegel, Jiang Wang, and Mao Ye; and seminar participants at Cornell University, Northwestern Kellogg, Duke University, Temple Fox, UBC Sauder, UIUC College of Business, the NBER Microstructure Meeting, the Texas Finance Festival, the Financial Intermediation Research Society Conference, the Frontiers in Finance Conference, the Western Finance Association Conference, and the 7th Conference on Current Topics in Financial Regulation for valuable comments and discussion. Xu Lu provided excellent research assistance. Supplementary data can be found on *The Review of Financial Studies* Web site.
Send correspondence to Brian M. Weller, Duke University, 213 Social Sciences Building, Box 90097, Durham, NC 27708; telephone: (919) 660-1720. E-mail: brian.weller@duke.edu.

doi:10.1093/rfs/hhx137 Advance Access publication December 7, 2017

come to dominate stock, futures, and Treasury markets, among others, in the United States and globally. Algorithmic traders (AT) rapidly incorporate public information into prices (e.g., Zhang 2017; Chakrabarty, Moulton, and Wang 2017), and AT orders are more likely to translate information into prices through heightened quoting efficiency (Hendershott, Jones, and Menkveld 2011) and greater permanent price impacts (Brogaard, Hendershott, and Riordan 2014). However, these studies exclusively speak to one part of the price discovery equation: AT enhances market efficiency with respect to public information *conditional on that information being revealed by other sources*. It is left unasked whether AT-derived improvements in the extent to which prices reflect acquired information ("informational efficiency") come at the expense of discouraging the acquisition of new information. I focus on algorithmic trading precisely because prior work in this literature overlooks this important component of price discovery in evaluating AT's effects on the information content of prices ("price informativeness") and because AT represents a transformative innovation in information processing in markets.[1]

Several potential mechanisms connect algorithmic trading to information acquisition in addition to information incorporation. Algorithmic liquidity provision and smart execution algorithms both reduce costs for typical trades, thereby increasing the potential trading profits due to an informational edge. However, as noted by Stiglitz (2014) and Han, Khapko, and Kyle (2014), these lower average costs may result from a greater ability of algorithmic liquidity providers to screen order flow and avoid adverse selection.[2] Improved screening of informed order flow in turn deters information acquisition by transferring prospective rents away from information acquirers. More directly, "back-running" algorithms may erode information rents and deter use of costly information sources like in Yang and Zhu (2017), or more colorfully, in Michael Lewis' *Flash Boys*.[3]

My core analysis evaluates the net effect of these mechanisms around 54,879 quarterly earnings announcements from January 2012 through September 2016. Evaluating the determinants of the information content of prices poses an empirical challenge because information sets are difficult to observe. I address this challenge by developing a new measure of relative information acquisition,

[1] Brunnermeier (2005) distinguishes between "informational efficiency" and "informativeness," and I adopt his terminology to cleanly separate the components of price discovery. Prices may reflect acquired information well (high informational efficiency), but nonetheless summarize a low absolute level of information (low informativeness).

[2] Stiglitz (2014) and Han et al. (2014) consider HFT rather than AT, but their screening and stale-quote avoidance mechanisms are not specific to AT. For example, Foucault, Röell, and Sandås (2003) explain how market makers use software to monitor different types of news to avoid being "picked off" by fast traders. This paper predates HFT, and the liquidity providers described are human-algorithm hybrids rather than HFT.

[3] Front- and back-running strategies fall under the umbrella of "order anticipation" strategies and are described in the Securities and Exchange Commission Concept Release on Equity Market Structure U.S. Securities and Exchange Commission (2010). Order anticipators attempt to trade ahead of large and/or informed traders to benefit from near-term price movements in the direction of trade.

the price jump ratio. The price jump ratio divides the return at the time of information's public disclosure to the total return over the pre-announcement period. Large impacts at announcement indicate that information is not discovered until publicly revealed, and this feature is reflected in a high price jump ratio. Unlike most alternative measures, such as the absolute cumulative abnormal return, the price jump ratio captures how much information enters prices early *relative to how much is potentially acquirable*.

I combine the price jump ratio measure with new data from the United States. Securities and Exchange Commission's Market Information Data Analytics System (MIDAS) to analyze the effects of algorithmic trading on information acquisition. Developed in response to analytics deficiencies revealed by the 2010 Flash Crash, MIDAS captures all public messages across all major stock exchanges for most U.S. common stocks and exchange-traded products since January 2012.[4]

The MIDAS data significantly improve identification of AT activity relative to alternative data sources. MIDAS facilitates construction of algorithmic trading proxies including odd lot volume ratios (or the fraction of volume associated with abnormally small trades), trade-to-order volume ratios, cancel-to-trade count ratios, and average trade sizes. With the (recent) exception of odd lot ratios and average trade sizes, these measures cannot be constructed using the standard NYSE Trades and Quotes database, and most alternative sources provide data for a short time series, a small sample of securities, or a single, potentially unrepresentative exchange.[5]

The price jump ratio measure and MIDAS data enable tests of the relationship between algorithmic trading and information acquisition for the first time. The baseline panel regressions of the price jump ratio on algorithmic trading proxies establish a strong negative association between algorithmic trading and information acquisition around earnings announcements. However, these results are potentially biased by causality running in the opposing direction. Algorithmic traders might sort into stocks with less information acquisition, or they may detect informed traders in the market and scale back their activities.

I use lagged log stock prices as a simple instrument to work around this concern. Lagged stock prices shift the incentives to use trading algorithms. The "sub-penny" rule (SEC Rule 612) imposes a minimum tick size of one cent for securities covered by Reg NMS, and this minimum price increment translates into variation in the fineness of the price grid as a function of price. High price stocks have a relatively fine price mesh, which favors algorithms

4 Messages include transactions; quote additions, modifications, and cancellations; order imbalances; and order-book level updates. MIDAS maps these messages—numbering in the hundreds of billions—into publicly available aggregates by stock, date, and exchange.

5 The New York Stock Exchange Trade and Quotes (TAQ) data only deliver order book changes at the best bid and offer rather than for the entire order book; NASDAQ TotalView-ITCH, NYSE OpenBook, and other feeds cover single exchanges in isolation; and proprietary resources, such as HFT-identified NASDAQ data, are typically quite limited in scope or in time.

over humans for continually updating limit orders for liquidity provision or monitoring for stale limit orders in liquidity taking. When the price mesh is relatively sparse, for example, 20 basis points for a \$5 stock as compared to 2 basis points for a \$50 stock, the underlying value of a security can vary more without requiring traders to update their order strategies, all else equal. High prices hence favor highly attentive algorithms over human traders. At the same time, lagged stock prices focus the empirical analysis on the ex ante effects of algorithmic trading on information acquisition rather than the endogenous response of algorithmic activity to informed traders. Conditional on other controls, such as market capitalization and institutional ownership, differences between high-priced stocks with few shares outstanding versus low-priced stocks with many shares outstanding are unlikely to affect the incentives to acquire information outside of the AT channel.[6]

Armed with this instrument, I find that algorithmic trading powerfully undermines pre-announcement information acquisition: a one standard deviation increase in algorithmic trading decreases information acquisition before earnings announcements by 9% to 13%.[7] These estimates hold across market capitalization and time strata, and for both individual securities and portfolios. Consistent with strategic trading models with informed traders, the "information gap" between high- and low-AT stocks grows monotonically as earnings announcements approach. Importantly, AT measurably discourages information acquisition up to *a month* in advance, by contrast with post-announcement price efficiency gains on horizons of (milli-)seconds to minutes typically considered in the literature. To the extent that they identify similar quantities, other measures of information in prices considered in robustness analyses arrive at the same conclusions.

Algorithmic traders impede information acquisition with potentially significant welfare consequences. First, reduced information acquisition facilitates trade at imprecise prices, in the sense that average absolute pricing errors are larger than they would be under a counterfactual low-AT regime. The welfare reduction associated with negative AT effects on price discovery may be quite severe if price uncertainty aggregates to systematic risks (like in O'Hara 2003). Among other channels, less informative prices may undermine efficient capital allocation through noisy costs of capital or budget constraints (Merton 1974; Baker, Stein, and Wurgler 2003) or reduce managerial learning from prices (Baumol 1965; Dow and Gorton 1997; Chen, Goldstein, and Jiang 2007).

Although such welfare effects may be muted around earnings announcements, the technologies and strategies relating AT to reduced information

6 Section 5.2 discusses and rules out potential contamination by other determinants of nominal stock price levels.

7 Incidentally, slightly greater magnitudes of the OLS coefficients offers fresh evidence for the endogenous withdrawal of algorithmic traders when confronted by informed traders, as theorized by Han et al. (2014) and Baldauf and Mollner (2017).

acquisition should be active in other settings for which the information content of prices cannot be readily estimated. Indeed, an analysis of an alternative measure of information in prices around earnings announcements, product releases, executive appointments and resignations, and merger and acquisition (M&A) events suggests that my findings extend to a wide range of economically important firm events.

This paper also contributes to the debate on the role of technology in financial markets. The same technological advances that improve price efficiency reduce the informativeness of prices in the medium run. This finding complements recent work by Bai, Philippon, and Savov (2016) relating the growth of the financial sector to stagnant price informativeness at horizons of a year or less.

1. Related Literature

1.1 Algorithmic trading and price discovery

Empirical work on algorithmic trading focuses primarily on whether algorithmic trading affects trading costs, market resilience, or price discovery.[8] A lively literature suggests that algorithmic trading contributes to faster incorporation of information into prices conditional on information acquisition having taken place. Zhang (2017) finds that high-frequency traders (HFT) are particularly skilled in incorporating hard information, and Chakrabarty, Moulton, and Wang (2017) echo this result in the context of earnings announcements temporarily overlooked by inattentive human traders. Carrion (2013), Hirschey (2013), and Brogaard, Hendershott, and Riordan (2014) demonstrate that HFT trades forecast price changes several seconds ahead and are more likely to have permanent price effects. Although these papers focus on the HFT subset of algorithmic trading, the consensus that HFT improve incorporation of acquired signals carries through to algorithmic trading generally.

By contrast, there is comparatively little work considering whether algorithmic trading encourages information acquisition by human and computerized agents. Grossman and Stiglitz (1980) suggest that changes to the profitability of acquiring information should change the amount of information acquired, and Stiglitz (2014) questions whether high-frequency traders dissuade information acquirers by reducing information rents. More positively, Foucault, Hombert, and Roşu (2016) theoretically exposit a mechanism by which HFT directly contribute to information gathering by trading on Brownian news surprises.

Baldauf and Mollner (2017) develop a model of the market quality trade-off between tight bid-ask spreads and effective price discovery. Their paper

[8] Several papers focus on high-frequency trading as a specific category of algorithmic trading. High-frequency trading is distinguished by ultra-low latencies and extensive automation of decision-making and execution.

is close in spirit to this work. Rather than balancing the desirable features of low transactions costs and informative prices, I focus instead on the conflict between information acquisition and information incorporation components of price discovery. Brunnermeier (2005) describes a similar potential conflict between informational efficiency and price informativeness in the context of theoretical model of information leakage, with an emphasis on the long-run negative effects of information leakage on both quantities rather than on a trade-off between them.

1.2 Measuring information acquisition

A rich literature on the role of traders and trade in incorporating information into prices precedes my work. To isolate my contribution to this literature, I briefly discuss selected related empirical measures. Because (absolute) cumulative abnormal returns (ACAR) are central to the construction of my price jump ratio measure, I revisit ACAR separately in Section 3.

The price jump ratio measure can also be understood as a stock-level variant of the intraperiod timeliness (IPT) measure of McNichols (1984), Freeman (1987), and Alford et al. (1993). IPT compares the performance of two zero-cost “perfect-foresight” portfolios comprised of stocks with different characteristics, such as the level of participation by algorithmic traders. Higher perfect-foresight returns earlier in the pre-event period indicate more pre-event information acquisition for stocks with that characteristic. Because the IPT measure compares two portfolios rather than many assets, it is not suitable for examining several characteristics simultaneously or for controlling for covariates when examining a single characteristic. Nevertheless, I adopt the IPT approach without controls in Section 6.1 to confirm the robustness of my results to the use of this portfolio-level measure with established properties.

The jump ratios of Morse (1981) and Meulbroek (1992) are the most closely related to my price jump ratio measure, and they are applied in a similar context of studying price changes prior to earnings announcements.[9] Both measures quantify the extent to which announcement returns are especially large in absolute value relative to a benchmark return, with the distinction that Meulbroek’s variant enables analysis of an additional determinant of abnormal returns (such as illegal insider trading in her study). Like the price jump ratio, these measures differ from the other informational measures in part because they gauge information acquisition without hinging on a particular model of how information is incorporated into prices. This distinction is important because tests of potential determinants of information acquisition using other measures are in fact joint tests with an auxiliary hypothesis that informed traders behave in a particular way.

[9] Heflin, Subramanyam, and Zhang (2003) use similar measures to evaluate the effects of SEC Regulation Fair Disclosure (“Reg FD”).

Recent work by Collin-Dufresne and Fos (2015) emphases the importance of such "model-free" measures in finding that measures of informed trading do not reflect informed trades. For example, if informed agents signal their intentions suboptimally via their order behavior, liquidity providers should respond by adjusting their quotes rather than suffering losses in trading. Trade-based measures such as trade informativeness (Hasbrouck 1991a, 1991b) or the weighted price contribution (Barclay and Warner 1993) thus may be quite low even as prices increasingly reflect the knowledge of an informed trader. As I detail in Section 3, the price jump ratio and Morse's and Meulbroek's measures sidestep these issues by assuming only that some set of market participants moves prices in response to information acquired before its public disclosure.

My price jump ratio departs from Morse's and Meulbroek's measures in two important respects. First, I extend application of jump ratios to a panel context by constructing a variant that is estimated separately for each stock-quarter event. This simple extension greatly expands jump ratios' empirical applicability. Second, my measure normalizes by a different benchmark, the total abnormal return associated with each announcement event. This benchmark facilitates analysis of the determinants of the information content of prices *relative to acquirable information*, the primary aim of my study.

2. Hypothesis Development and Empirical Setting

Several potential mechanisms link algorithmic trading to information acquisition, and the net effect of many different algorithms acting in concert is uncertain a priori. First, algorithmic traders may be better than human traders at extracting price-relevant information from public signals. For example, Brogaard et al. (2014) find that liquidity-taking algorithmic traders are more responsive to information in order books and public announcements. Foucault et al. (2016) generate a similar theoretical prediction that fast liquidity takers may improve pre-announcement price discovery via superior signal processing. More generally, algorithmic "analysts" may augment or supplant their human counterparts, potentially skewing the type or quantity of information acquired (Zhang 2017).

Algorithmic traders also may encourage or discourage human traders from acquiring information. From a Grossman and Stiglitz (1980) perspective in which information rents motivate information acquisition, lower trading costs associated with algorithmic trading—either from "smart" algorithmic execution or through lower bid-ask spreads—may encourage the collection and incorporation of information. However, this mechanism may be reversed if the decline in spreads results from algorithmic traders' improved ability to avoid being adversely selected (Han et al. 2014; Baldauf and Mollner 2017). Likewise, if aggressive algorithmic traders anticipate or "piggyback" on informed orders, as Stiglitz (2014) and Yang and Zhu (2017) suggest, profits accruing to information acquisition decline, and AT decreases equilibrium information acquisition.[10]

These competing theories motivate my main empirical tests of how and when algorithmic trading contributes to information acquisition, and I investigate these relations in the context of quarterly earnings announcements. Earnings announcements provide a near-ideal setting for this investigation for three reasons. First, earnings announcements are among the recurring firm events with the greatest average effects on stock prices. The importance of earnings announcement news relative to day-to-day price movements boosts the signal-to-noise ratio of my measure of relative information acquisition. Second, earnings announcements are scheduled, public events. Kyle (1985) and other models of informed trading offer crisp predictions for how prices should evolve in this setting. Third, quarterly earnings announcements are ubiquitous, which contributes to rich cross sections and minimal selection effects over which information events are observed.

Notwithstanding these advantages, the earnings announcements setting has two notable drawbacks related to its external validity. For one, market participants may be able to mitigate the effects of changes in pre-announcement information acquisition. For example, firms may avoid conditioning their decisions on market prices shortly before announcements in favor of waiting until after public disclosures. Consequently, the economic cost of delayed information incorporation around earnings announcements is likely to be muted relative to other events I cannot analyze. For this reason, I only take limited steps toward assessing the potential welfare impact of AT on information acquisition.

Second, I can only indirectly assess whether the effects of algorithmic trading around earnings announcements apply across a wider set of events. However, given the prominence of earnings as a focus of analyst and market attention, earnings announcements (1) are important in their own right and (2) represent a best case for information acquisition by market participants. Moreover, the mechanisms linking algorithmic trading and information acquisition apply equally well for the informational events that I cannot observe as a researcher.

3. The Price Jump Ratio as a Measure of Information Acquisition

3.1 Defining the price jump ratio

The absolute cumulative abnormal return (ACAR) of Fama et al. (1969) and Ball and Brown (1968) is a standard measure of the incorporation of information into prices. This measure constructs pre-event price drift net of a predicted

[10] Harris (2013) emphasizes that this channel concerns algorithmic trading generally rather than high-frequency trading specifically: “the successful implementation of [order anticipation] strategy depends less on low-latency communications than on high-quality pattern-recognition algorithms.”

returns from a factor model, for example,

$$CAR_{it}^{(k_1,k_2)} = \sum_{t=k_1}^{k_2} \left(r_{it} - \alpha_i - \sum_{m=1}^{M} \beta_{im} r_{mt} \right) = \sum_{t=k_1}^{k_2} \epsilon_{it}, \tag{1}$$

where r_{it} is the log return of stock i on date t and α_i and β_i are estimated from a M-factor model of realized returns. The cumulative abnormal return (CAR) cumulates the abnormal return from dates k_1 to k_2 around the announcement date T, and the ACAR is the absolute value of this number.

The ACAR answers the question of how much pre-announcement information enters prices prior to an earnings announcement *without consideration for how much information might be available to acquire*. Rather than using the ACAR, I proceed along the lines of Meulbroek (1992) and normalize the CAR by a measure of total announcement-related variation. Specifically, I use the ratio of post-announcement price variation to total variation before and including the earnings announcement:

$$jump_{it}^{(a,b)} = \frac{CAR_{it}^{(T-1,T+b)}}{CAR_{it}^{(T-a,T+b)}}, \tag{2}$$

with $a > 1$ to capture pre-announcement variation and $b \geq 0$ to capture post-announcement drift.

I use a normalized price jump measure for two reasons. First, the normalized measure naturally accounts for differences in the expected magnitude of abnormal returns among stocks and over time. For example, panel analysis using the ACAR measure implicitly assign large weights to stocks in the "dusty corners" of the market because market capitalization and variance of earnings announcement impact are strongly negatively correlated.[11] Second, the price jump ratio has a different economic interpretation from the ACAR. The price jump ratio quantifies the *share* of information acquired and incorporated into prices pre-announcement. If acquirable return information is proportional to the total realized return—which holds, for example, when market participants can learn all price-relevant information before the announcement—this ratio directly measures the fraction of acquirable information incorporated into prices before its public disclosure.

To build intuition for the price jump ratio measure, consider the price path of a hypothetical stock with and without information acquisition before an earnings announcement or other scheduled public news event. Without an informed trader, the price may be efficient with respect to all information acquired by non-insiders, but no market participant is aware of the information content of the impending disclosure event. By contrast, in the presence of

[11] Chakrabarty, Moulton, and Wang (2017) note precisely this feature of ACAR in a concurrent study of high-frequency trading and investor attention after public announcements.

an informed trader, the price drifts toward the post-announcement asset value on account of her order submission and trading strategy: strategic informed trading models have a common implication of smooth convergence of prices to post-announcement values *before* information is publicly revealed (e.g., Kyle 1985; Back 1992). The difference in price paths with and without informed traders represents an "information gap," or the extent to which prices better approximate the post-announcement asset value when informed traders are active. Importantly, several measures of price efficiency like pricing error variance (Hasbrouck 1993), serial autocorrelation in returns, and variance ratios speak only to price efficiency with respect to acquired information rather than to price informativeness and the information gap.

The price jump ratio by contrast is the empirical counterpart to the (unobservable) information gap. A high price jump ratio corresponds to a large announcement-date jump relative to pre-announcement drift, whereas a low price jump ratio corresponds to a small announcement-date jump. Aggressive informed trading drives the price jump ratio toward 0, and the absence of informed trading precipitates a price jump ratio close to 1. Higher values of the price jump ratio thus represent less information in prices relative to the post-announcement information set.

Strictly speaking, the price jump ratio measures the pre-announcement information content of prices rather than information acquisition. Market participants could acquire information but not push stock prices appreciably toward their post-announcement values. However, such trading behavior around scheduled announcements is dominated by an alternative trading strategy in which informed traders enforce near or complete convergence of prices to the post-announcement values. Otherwise, the residual discrepancy of prices an instant before the announcement date represents foregone profits from trading against uninformed resting orders. For this reason, the optimal interim price path may vary with liquidity conditions, but the anticipated post-announcement value should remain an informed trader's target. Moreover, from a welfare perspective, reduced information acquisition and more "smart trading" have similar distorting effects for the allocation of risk and resources because prices remain uninformative with respect to acquirable information.

3.2 Implementing the price jump ratio measure

I estimate abnormal returns relative to a Fama and French (1992) three-factor model using daily returns over a 365-calendar day window ending 90 days before the earnings announcement. Observations with estimation windows with fewer than 63 valid preceding trading days (one calendar quarter) are dropped. I select a 21-day pre-announcement window ($a = 21$ in Equation (2)) to balance resolution on earnings announcement price effects against the possibility of earnings-related information entering earlier than the period considered, although results are robust to this choice.[12] I define the cumulative variation associated with the earnings announcement to include an additional two trading days after the announcement ($b=2$) because prices may exhibit post-earnings announcement drift.

In practice noise in informed traders' signals on the ex post value of an asset moves the realized price jump ratio away from the 0 or 1 outcomes delivered by a standard Kyle model. This noise introduces potentially sizable deviations in the price jump ratio's numerator and denominator. For this reason an individual price jump ratio realization reveals little about information acquisition around any particular announcement. However, relating conditional (truncated) means or medians to covariates reveals the determinants of information acquisition, and this is the path my analysis takes.

Unlike the unnormalized (A)CAR, the price jump ratio has an undefined conditional mean because its denominator may be close to zero. Throughout the empirical analysis, I resolve this problem by dropping observations with small values of the price jump ratio denominator, namely nonevents from the market's perspective. I define the cutoff for retaining an announcement event in terms of return volatility over the preceding month, estimated simply as the square root of the daily return variance. Each event retained satisfies

$$\left|CAR_{it}^{(T-21,T+2)}\right| > \sqrt{24}\hat{\sigma}_{it}. \tag{3}$$

Events exceeding the value on the right represent large announcement period returns (relative to scaled daily volatility) indicative of material earnings announcement information. This filter is strict, and only 45.5% of observations survive. Appendix A plots the distribution of the price jump ratio and compares stock-quarters that do and do not survive this volatility cutoff. Overall the observable characteristics of included and excluded observations are very much alike. Likewise, Appendix B verifies that results are not driven by differences in market capitalization or relative importance of earnings news $\left|CAR_{it}^{(T-21,T+2)}\right|/\hat{\sigma}_{it}$ among stocks with different levels of AT.

I choose this approach for economic and statistical reasons. Economically, near-zero price impact events have limited potential informational distortions because prices with or without informed traders are similar throughout the pre-announcement period. From a statistical perspective, these small-denominator events also have very low signal-to-noise ratios. All coefficient estimates for this restricted sample nonetheless are virtually identical to those estimated by median regression or weighted least squares on the full sample of stock-quarter events. The Online Appendix tabulates results and provides additional discussion.

[12] I also conduct a separate analysis in which I control for price changes in the more distant past. Although the main results are effectively unchanged, AT appear to counteract behavioral or strategic disclosure responses to negative news. I present these results with additional discussion in the Online Appendix.

4. Data

This paper combines several data sources for the main analysis: the Center for Research in Securities Prices (CRSP); NYSE Trades and Quotes (TAQ); the Thomson Reuters Institutional Brokers' Estimate System (I/B/E/S) and SEC Form 13F Institutional Holdings Database; and the SEC Market Information Data Analytics System (MIDAS). CRSP provides abnormal and cumulative abnormal returns[13] for each earnings event as well as daily prices, market capitalization, and return volatility. Monthly TAQ (2012–2014) and Daily TAQ (2015–2016) contribute quoted spreads via derived national best bid and offer (NBBO) tables. I/B/E/S provides quarterly earnings announcement dates and times, as well as the number of analyst estimates for each earnings announcement, and the 13F database delivers institutional holdings information. SEC MIDAS enables construction of algorithmic trading proxies and order volume shares by exchange, stock, and date.

In the months following the May 6, 2010, Flash Crash, the Securities and Exchange Commission sought to enhance its ability to monitor markets in real time and rapidly reconstruct market past events. The Market Information Data Analytics System (MIDAS), launched in January 2013, provides the SEC with order book information from all major U.S. stock exchanges with microsecond timestamps, as well as the capacity to process these data quickly. To the best of my knowledge, this paper is the first to use SEC MIDAS data for academic research.

The SEC provides limited summary statistics by security and exchange for public use. The summary metrics include lit trade and order volume,[14] hidden volume, odd lot volume, and counts of trades and cancellations (full or partial), as well as several quantities derived from these metrics. MIDAS also provides distributions of quote survival lifetimes by market capitalization terciles for stocks and exchange traded products.

MIDAS collects order data across all major U.S. stock exchanges to present a comprehensive view of lit market activity. By contrast with the now-standard market data found in TAQ, the MIDAS data incorporates quotes and cancellations from all levels of the order book. It also reports odd lot trading, which only recently entered the TAQ data. Most relevant for this paper, public availability of order book summary data for the construction of algorithmic trading proxies at the stock-day frequency is an important step forward for avoiding potential sample biases endemic to the specialized or proprietary data typically used by researchers.[15]

---

[13] Factor return information is provided by Kenneth French through his website at http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html.

[14] "Lit" orders are visible to other market participants before a trade, and lit trades are executions resulting from lit orders. "Hidden" orders are not lit, and hidden volume references shares exchanged using hidden orders.

[15] U.S. Securities and Exchange Commission (2014) surveys data sources used in the analysis of algorithmic and high-frequency trading and their respective limitations.

The main analysis includes all CRSP common stocks (CRSP share code 10 or 11) matched in the I/B/E/S and MIDAS databases.[16] Stock events include all quarterly earnings announcements from January 1, 2012, through September 30, 2016. Announcements recorded by I/B/E/S as occurring after market close are assigned to the next trading day, as returns are computed from market closing prices. I exclude observations for which earnings announcements are less than 45 days or more than 150 days from the time of the previous earnings announcement. The resulting sample has 3,839 unique securities and 54,879 security-quarter observations.

For each stock-earnings announcement and date pair it, I construct rolling 21-day averages of price, market capitalization, and quoted spreads over the interval $[T-42, T-22]$ (measured in days relative to the earnings announcement date T). I then take the log of price and market capitalization values. Log return volatility is computed as the log of the standard deviation of daily returns over the same 21-day horizon.

Daily quoted spread values from TAQ are "time-in-force" or duration-weighted averages from 9:35 a.m. to 4:00 p.m., where the first five minutes of trading are dropped to avoid incorporating abnormal values associated with market open. Bid-ask spreads use daily TAQ when available (2015–2016) and monthly TAQ for 2012–2014. Daily TAQ NBBOs are coarsened to seconds rather than milliseconds to equate observation frequencies across datasets. Quoted spreads less than zero or greater than 25% of the concurrent price are excluded from rolling averages over the same $[T-42, T-22]$ interval.

I also include two controls for institutional attention and activity. Analyst coverage for each observation is the log of the largest number of reporting analysts in the Thomson Reuters I/B/E/S database associated with each stock-quarter announcement. The institutional ownership ratio is the total number of shares held by 13F filing institutions at the end of the preceding calendar quarter divided by the total number of shares outstanding at that time.

4.1 Construction of algorithmic trading proxies

I construct four algorithmic trading proxies using the MIDAS data: the odd lot volume ratio, or the total volume executed in quantities smaller than 100 shares divided by total volume traded;[17] the trade-to-order volume ratio, or the total volume traded divided by the total volume across all orders placed; the cancel-to-trade ratio, or the number of full or partial cancellations divided by the number of trades; and the average trade size, or the number of shares traded divided by the number of trades. Existing literature suggests that higher odd lot and cancel-to-trade ratios indicate more algorithmic trading, and higher

[16] To facilitate matching, I use the WRDS IBES to CRSP linking table and retain observations with matching 6-digit CUSIPs and date ranges (99.3% of the data).

[17] Three tickers have round lot sizes different from 100: BH has a round lot size of 10, and BRK.A and SEB have a round lot size of 1.

trade-to-order volume ratios and average trade sizes indicate less algorithmic trading.

I follow MIDAS's prescription for constructing the odd lot, trade-to-order, and cancel-to-trade ratios. Of the 12 stock exchange feeds captured by MIDAS,[18] NYSE and NYSE MKT are excluded from the odd lot ratio, the cancel-to-trade ratio, and average trade size measures because the level-book reporting method of these two exchanges does not allow for comparable accounting of trades with the order-based feeds used by the other 10 exchanges.[19] For each stock, I average each algorithmic trading proxy across dates $[T-21, T-1]$ and take the log of the result to eliminate significant right skew in the averaged values. I drop observations with ratios less than the 1st percentile or greater than the 99th percentile of AT proxies to guard against potential reporting errors in the MIDAS data.[20]

Table 1 presents summary statistics for AT proxies and controls. All AT proxies exhibit significant dispersion across stocks. The top portion of Panel A quantifies this distributional information for the pooled sample, and the bottom portion of panel A reports similar statistics for the set of controls. Panel B provides pairwise correlations among key independent variables (top) and controls (bottom). Of note on top are the strong correlations among algorithmic trading measures. As previous studies suggest, odd lot and cancel-to-trade ratios covary positively with each other and covary negatively with trade-to-order volume ratios and average trade sizes. These correlations are nonetheless far from one, and the measures appear to embed separate information content.

5. Main Results

5.1 Algorithmic trading and the information content of prices

My main empirical analysis investigates whether algorithmic trading is associated with differences in pre-announcement information acquisition around quarterly earnings announcements. I start with a simple panel regression relating price jump ratios to algorithmic trading proxies,

$$jump_{it}^{(21)} = \alpha + \beta x_{it} + \gamma \times controls_{it} + \epsilon_{it}, \tag{4}$$

where $jump^{(21)}$ is shorthand for $jump^{(21,2)}$ and x_{it} stands in for the odd lot ratio, the trade-to-order volume ratio, the cancel-to-trade ratio, or the average trade

[18] I exclude IEX because it starts trading all stocks as an exchange in September 2016, the last month of my sample.

[19] NYSE and NYSE MKT report the trade size of the initiating order, whereas the other exchanges separate trades by initiating and contra orders. For this reason the number of trades and trade size distributions for NYSE stocks are not comparable to their counterparts on other exchanges. Notwithstanding these differences, the pairwise correlations between log odd lot ratios, cancel-to-trade ratios, and average trade sizes for NYSE/NYSE MKT and all other exchanges are quite high at 88.0%, 55.6%, and 76.7%, respectively, and using only NYSE and NYSE MKT variables on the left-hand side delivers similar results.

[20] Additional MIDAS details and discussion of exchange exclusions are provided on the MIDAS website at http://www.sec.gov/marketstructure/mar_methodology.html.

Table 1
Summary statistics for algorithmic trading proxies and controls

A. Summaries by stock-event observations

AT proxies	Odd lots	$\frac{\text{Trades}}{\text{Orders}}$	$\frac{\text{Cancels}}{\text{Trades}}$	Trade size
Mean	−2.43	−3.76	3.31	4.71
SD	0.80	0.63	0.58	0.45
10%	−3.59	−4.61	2.62	4.31
Median	−2.26	−3.72	3.25	4.59
90%	−1.58	−3.00	4.09	5.33
N	53,796	53,844	53,844	54,879

Controls	Market cap.	Price	Ret. vol.	Quoted spr.	#Analysts	*IOR*
Mean	20.71	2.90	−3.96	0.50	1.71	0.57
SD	1.89	1.21	0.58	0.84	0.96	0.28
10%	18.28	1.22	−4.64	0.05	0.00	0.00
Median	20.65	3.06	−4.00	0.23	1.79	0.64
90%	23.20	4.27	−3.22	1.15	2.94	0.88
N	54,879	54,879	54,879	54,595	54,879	54,439

B. Pairwise correlations by stock-event observations

AT proxies	Odd lots	$\frac{\text{Trades}}{\text{Orders}}$	$\frac{\text{Cancels}}{\text{Trades}}$
Trade-to-order ratio	−0.56		
Cancel-to-trade ratio	0.42	−0.83	
Trade size	−0.94	0.52	−0.34

Controls	Market cap.	Price	Ret. vol.	Quoted spr.	#Analysts
Price	0.77				
Ret. vol.	−0.50	−0.56			
Quoted spr.	−0.60	−0.48	0.33		
#Analysts	0.74	0.51	−0.25	−0.53	
IOR	0.42	0.44	−0.24	−0.40	0.43

This table presents descriptive statistics (top panel) and correlations (bottom panel) among key explanatory and control variables. The first table in each panel summarizes algorithmic trading proxies (all in logs). The second table in each panel presents corresponding summaries for dollar market capitalization, share price, and return volatility (all in logs) from the CRSP data. Additional summaries include the median end-of-minute quoted bid-ask spreads in percent from the TAQ NBBO files (one-second version); the number of analysts as the log of the maximum number of reporting analysts for each stock-event pair in Thomson Reuters I/B/E/S; and the institutional ownership ratio (*IOR*) as the fraction of shares held by 13F filing institutions at the end of the preceding calendar quarter. All proxy variables are rolling 21-day mean values from $[T-21, T-1]$, and control variables are lagged an additional 21 days prior to the announcement date ($[T-42, T-22]$).

size for stock i at date t. As Section 4 describes, these proxies alternate between strongly increasing in algorithmic trading activity (odd lot and cancel-to-trade ratios) and strongly decreasing in algorithmic trading activity (trade-to-order volume ratio and average trade size). Throughout, standard errors are two-way clustered by stock and month to account for correlated errors arising from differences in firm characteristics, such as disclosure policies, or timing within the sample period.

Table 2 presents results from estimation of Equation (4). The first specification for each algorithmic trading measure includes the AT proxy and lagged log market capitalization. I include lagged log market capitalization as a control in all specifications because market capitalization is among the strongest covariates with AT, and it also has the potential to influence the amount of information that enters prices pre-announcement through channels other than AT.

Table 2
Determinants of announcement price impact

	x = Odd lot ratio			x = Trade-to-order ratio			x = Cancel-to-trade ratio			x = Avg. trade size		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
x	0.0741***	0.0898***	0.144***	−0.0725***	−0.0889***	−0.142***	0.0718***	0.113***	0.144***	−0.152***	−0.150***	−0.263***
	(0.00683)	(0.00937)	(0.0111)	(0.00863)	(0.0103)	(0.0107)	(0.00858)	(0.00988)	(0.0113)	(0.0122)	(0.0176)	(0.0248)
Market cap.	0.00244	−0.0183***	−0.0478***	0.00783***	−0.0280***	−0.0131	0.0135***	−0.0210***	−0.00424	0.000119	−0.0220***	−0.0423***
	(0.00318)	(0.00542)	(0.0121)	(0.00290)	(0.00572)	(0.0118)	(0.00314)	(0.00556)	(0.0119)	(0.00318)	(0.00529)	(0.0124)
Price		−0.0290***	X		−0.00510	X		−0.0105	X		−0.0151*	X
		(0.00867)			(0.00876)			(0.00835)			(0.00857)	
Ret. vol.		−0.0163	−0.000812		−0.0106	0.00528		−0.00193	0.00933		−0.0194	−0.00387
		(0.0124)	(0.0123)		(0.0117)	(0.0122)		(0.0115)	(0.0120)		(0.0120)	(0.0120)
Quoted spr.		−0.0324***	−0.0169		−0.0572***	−0.0363***		−0.0713***	−0.0372***		−0.0222***	−0.00209
		(0.00745)	(0.0123)		(0.00779)	(0.0120)		(0.00871)	(0.0129)		(0.00725)	(0.0115)
#Analysts		0.0521***	0.0326**		0.0546***	0.0361***		0.0564***	0.0334**		0.0505***	0.0305**
		(0.00754)	(0.0136)		(0.00742)	(0.0133)		(0.00746)	(0.0133)		(0.00728)	(0.0132)
IOR		0.123***	0.0239		0.123***	0.0421		0.139***	0.0550		0.106***	0.00781
		(0.0231)	(0.0358)		(0.0222)	(0.0349)		(0.0222)	(0.0347)		(0.0228)	(0.0338)
Constant	0.599***	X	X	0.0403	X	X	−0.0426	X	X	1.179***	X	X
	(0.0729)			(0.0692)			(0.0736)			(0.107)		
Month FEs	No	Yes	Yes	No	Yes	Yes	No	No	Yes	No	No	Yes
Stock FEs	No	No	Yes	No	No	Yes	No	No	Yes	No	No	Yes
R^2	0.0132	0.0337	0.0174	0.00904	0.0330	0.0198	0.00762	0.0352	0.0190	0.0157	0.0331	0.0158
N	24,107	23,814	23,517	24,068	23,800	23,503	24,120	23,842	23,539	24,512	24,201	23,910

$^{*}p<.10$, $^{**}p<.05$, $^{***}p<.01$

This table presents results from a regression of price jump ratios on a set of algorithmic trading proxies:

$$jump_{it}^{(21)}=\alpha+\beta x_{it}+\gamma\times controls_{it}+\epsilon_{it}.$$

For each stock i and quarterly earnings announcement t from January 2012 through September 2016, the price jump ratio ($jump_{it}^{(21)}$) is measured as the ratio of the announcement response divided by the total variation in the pre- and post-announcement period: $CAR_{it}^{(T-1,T+2)}/CAR_{it}^{(T-21,T+2)}$. Cumulative abnormal returns (in logs) are net of Fama and French (1992) three-factor implied returns over the same interval. Observations with cumulative net price impact not exceeding a minimum multiple of prior return volatility (described in the main text) are dropped. Market capitalization, share price, and return volatility are the log of daily averages over $[T-42, T-22]$, and the quoted spread is average of the time-weighted bid-ask spread over the same interval (reported in percent). The number of analysts is the log of the largest number of reporting analysts in the Thomson Reuters I/B/E/S database associated with each stock-quarter announcement. The institutional ownership ratio (*IOR*) is the fraction of shares held by 13F filing institutions at the end of the preceding calendar quarter. Construction of algorithmic trading proxies (x_{it}) derived from SEC MIDAS data are described in the main text. All standard errors are clustered by security and month and are reported in parentheses. The log price is dropped from stock fixed effects regression on account of near-collinearity with market capitalization. Reported R^2 are between-group values in stock fixed effects specifications. The number of observations for the within-stock specifications is slightly smaller on account of dropping securities with only one observation.

The key entry in each column is the coefficient on x_{it}. Interpreting the coefficient in the upper left of the table, a one unit increase in the log odd lot ratio increases the price jump ratio by 0.074 (relative to a median price jump ratio of 0.455). Equivalently, a one standard deviation increase in the log odd lot ratio (0.80) is associated with a 13% decrease in the fraction of earnings announcement price impact that occurs pre-announcement.

The second and third specifications for each algorithmic trading proxy include additional controls. The second specification augments lagged log market capitalization with log share price, log return volatility, average quoted bid-ask spreads, institutional ownership share (all lagged), as well as the number of reporting analysts and month fixed effects. All nondummy variables except for analyst coverage clearly vary in response to information acquisition in the pre-announcement period and must be lagged to ensure predetermination. Notwithstanding these additions, the coefficient on the odd lot ratio typically strengthens slightly across specifications (1) and (2).

The final column adds stock fixed effects. This control absorbs potentially relevant fixed differences in disclosure policy and investor base not captured by the other controls. I drop log price from this specification because of the near-collinearity between log price and log market capitalization within stocks (99.6% correlation net of fixed effects).[21] The resulting within-stock coefficient is larger in absolute value, indicating that more algorithmic trading activity is associated with diminished information acquisition both across securities and within a given security.

The estimated effects of algorithmic trading on information acquisition have comparable economic and statistical magnitudes across all four algorithmic trading proxies. To illustrate, I multiply the point estimates for the first regression specification by each measure's respective standard deviation. By comparison with the 13% decrease in the pre-announcement share of earnings announcement price impacts estimated using odd lot ratios, I estimate a 10% decrease using trade-to-order volume ratios, a 9% decrease using cancel-to-trade ratios, and a 15% decrease using average trade sizes. Similar comparability is achieved across all regression specifications.

5.2 Identifying exogenous variation in algorithmic trading

Interpreting these estimates causally requires exogenous variation in AT participation because the preceding regressions potentially suffer from reverse causality and omitted variable problems. Algorithmic traders may respond to information entering prices, for example, by detecting the presence of informed traders and withdrawing from the market to avoid adverse selection. This

[21] The within-stock coefficient is identified using deviations from stock means, which raises the question of what causes the residual variation in algorithmic trading. Residual drivers of algorithmic trading may include time-varying substitutability with other securities (e.g., convertible debt and highly correlated stocks), index inclusion status, and institutional hedging demands, to name a few.

explanation accords with Hendershott, Jones, and Menkveld's (2011) finding of lower adverse selection with more AT, and such a response would contribute to the observed negative relationship between algorithmic trading and the price jump ratio. Likewise, trading algorithms may be deployed more or less frequently for stocks with less information to acquire. Such selection is another source of bias for the estimated AT-information acquisition relation.

Although I control for lagged market capitalization and bid-ask spreads, other liquidity variables not in the regression also may contribute to an omitted variable bias. A decrease in uninformed trading shortly before an announcement—arising, for example, from fear of adverse selection—increases the price impact of informed trades and decreases algorithmic market making revenues. Mechanisms such as these can generate a spurious relation between AT and the price jump ratio measure unrelated to AT's effects on information acquisition.

To address these concerns, I use the log of the average stock price from 42 to 22 days before each earnings announcement as an instrument for algorithmic trading activity. Controlling for covariates such as market capitalization and institutional ownership, variation in lagged stock prices should relate little to the incentives of market participants to acquire information or to the amount of information available to acquire.[22] I make this argument below. The use of lagged information also focuses the empirical tests squarely on the ex ante effects of algorithmic traders on information acquisition rather than on the responses of algorithmic traders to information already acquired.

The exclusion restriction requires that the component of lagged share prices orthogonal to market capitalization and other covariates must not affect the incentives of market participants to acquire information. For example, market capitalization increases the potential profits associated with acquiring information, but the component of prices orthogonal to market capitalization should have no clear direct effect on the desirability of learning about a stock. The literature on stock splits offers several theories for why firms may manage their share prices, and their potential impact on the exclusion restriction merits discussion.

I explicitly control for bid-ask spreads and institutional ownership to address potential liquidity or catering rationales for controlling nominal share prices (see, e.g., Conroy, Harris, and Benet 1990 and Angel 1997 for liquidity rationales; Baker, Greenwood, and Wurgler 2009 for a catering explanation). Under norm-based theories of choosing a target price range such as Rozeff (1998) and Weld, Michaely, Thaler, and Benartzi (2009), changing the numeraire for trading does not affect economic fundamentals or the incentives to acquire information.

[22] The sample universe excludes pink sheets and other issues for which low stock prices may translate into different disclosure regimes. All coefficient estimates are comparable with minimum stock price thresholds of \$1, \$20, and \$50.

Signaling theories like in Brennan and Copeland (1988) and Asquith, Healy, and Palepu (1989) conjecture that numeraire changes are a costly signal of firm value. Empirical evidence suggests that firm earnings and profitability significantly increase before stock splits, but not thereafter (Lakonishok and Lev 1987; Asquith et al. 1989). Because the change in numeraire does not translate into future shifts in fundamentals, it is not likely that split-induced variation in prices alters the incentives to acquire information. Moreover, Easley, O'Hara, and Saar (2001) find no evidence of reductions in information asymmetries around stock splits, as would be consistent with firms signaling private information. Because information asymmetries do not decline, either (1) the incentives to acquire information remain the same (and the exclusion restriction is satisfied) or (2) signaling by the firm is precisely counterbalanced by a change in the set of information available to acquire. In the Online Appendix I confirm that possible signaling through changes in price levels does not affect my inference by discarding all observations within three months of a stock split declaration or effective date.

The "sub-penny" rule (SEC Rule 612) mandates a minimum price increment of one cent for displayed orders in stocks covered by Reg NMS, and this minimum increment plays a critical role in creating dispersion in algorithmic trading activity across stocks. As stock prices increase, this minimum increment tightens the grid of tradable prices as a fraction of the midpoint price. For example, the minimum price increment on a \$10 stock equates to 10 basis points, whereas the minimum price increment on a \$100 stock equates to 1 basis point. For a given percent change in the underlying value of the asset, the \$100 stock is more likely to require quote revisions than the \$10 stock. A 5-basis-point change in the \$10 stock's underlying value may require no action by traders—after all, the movement is half of the smallest price increment—whereas the \$100 stock's value moves five price ticks. Because algorithmic liquidity providers are better equipped to continually update quotes than human traders are, and algorithmic liquidity takers are similarly better able to take advantage of momentarily "stale" quotes (in the sense of Foucault, Röell, and Sandås 2003), we would expect algorithmic traders to comprise a greater share of trading in stocks with higher prices all else equal.

Yao and Ye (2017) argue for the opposite relation between relative tick size and high-frequency trading, a subset of AT. Because tick sizes are larger than the equilibrium price of liquidity for low-priced stocks, inframarginal units of quoted depth are profitable. HFT have an advantage in obtaining time priority and enjoying these inframarginal rents, whereas slower traders tend toward the back of the queue at the same price or at a less competitive price. Using HFT-identified NASDAQ data, they find evidence for this opposing relation.

In my data, the quote-updating channel dominates the time-priority channel, and the positive relationship between algorithmic trading proxies and the log price is quite strong. Table 3 reports correlations of algorithmic trading proxies and the lagged log stock price instrument. All correlations are similar

Table 3
Correlations with lagged log price instrument

Sample	Odd lots	Trades/orders	Cancels/trades	Trade size	AT PCF_1
Full sample	0.730***	−0.432***	0.251***	−0.766***	0.640***
Regression sample	0.721***	−0.456***	0.293***	−0.752***	0.640***
Net of mkt. cap.	0.787***	−0.587***	0.579***	−0.763***	0.779***
Net of all controls	0.769***	−0.550***	0.560***	−0.750***	0.752***

$*p<.10$, $**p<.05$, $***p<.01$
This table reports raw correlations of the lagged log stock price and algorithmic trading proxies for the full sample and for the regression sample, as well as correlations of these measures net of variation spanned by market capitalization and other control variables for each stock i and quarterly earnings announcement t from January 2012 through September 2016. Construction of algorithmic trading proxies (x_{it}) derived from SEC MIDAS data is described in the main text. PCF_1 denotes the first principal component factor of the AT proxies.

between full and regression samples, and the correlations only strengthen when orthogonalizing the log price with respect to the obvious confound of market capitalization. Importantly, these relations are not driven by extreme prices in the sample: averages of all AT proxies vary monotonically with deciles of the lagged log price and the lagged log price orthogonalized with respect to the controls (Figure 1).

Having established the necessary conditions for the lagged log stock price instrument, I now augment Equation (4) with a first stage for x_{it},

$$x_{it}=\zeta+\eta lprice_{it}+\theta\times controls_{it}+\delta_{it},$$

$$jump_{it}^{(21)}=\alpha+\beta\hat{x}_{it}+\gamma\times controls_{it}+\epsilon_{it}. \tag{5}$$

Specifications parallel the noninstrumented regressions, with the exception that lagged log price is necessarily excluded as an independent variable in the second-stage regression. I omit stock fixed effect specifications because stock-level averages absorb much of the variation in my instrument (stock prices are persistent across quarters).

Table 4 presents results from this analysis.[23] These IV estimations reinforce the OLS results: an increase in algorithmic trading causes a sharp reduction in the information content of prices around earnings announcements. Moreover, the estimated magnitudes are in line with the OLS estimates of Table 2 in the baseline specification, and coefficients are effectively unchanged with the addition of other controls in the IV.

The specifications with multiple controls differ in the use of lagged log stock price as an instrument rather than a control variable, and this distinction is revealing about how algorithmic traders may endogenously respond to information entering stock prices. Comparing coefficients in Tables 2 and 4, the part of AT captured by log prices (net of the other controls) has a weaker

[23] Throughout I use the (cluster-robust) Kleibergen and Paap (2006) *rk* Lagrange multiplier (LM) and Wald F statistics to evaluate instrument strength for each two-stage regression. All LM tests reject underidentification at the 1% significance level, and all specifications feature first-stage F statistics exceeding the critical value corresponding to a 10% maximal IV size (16.38).

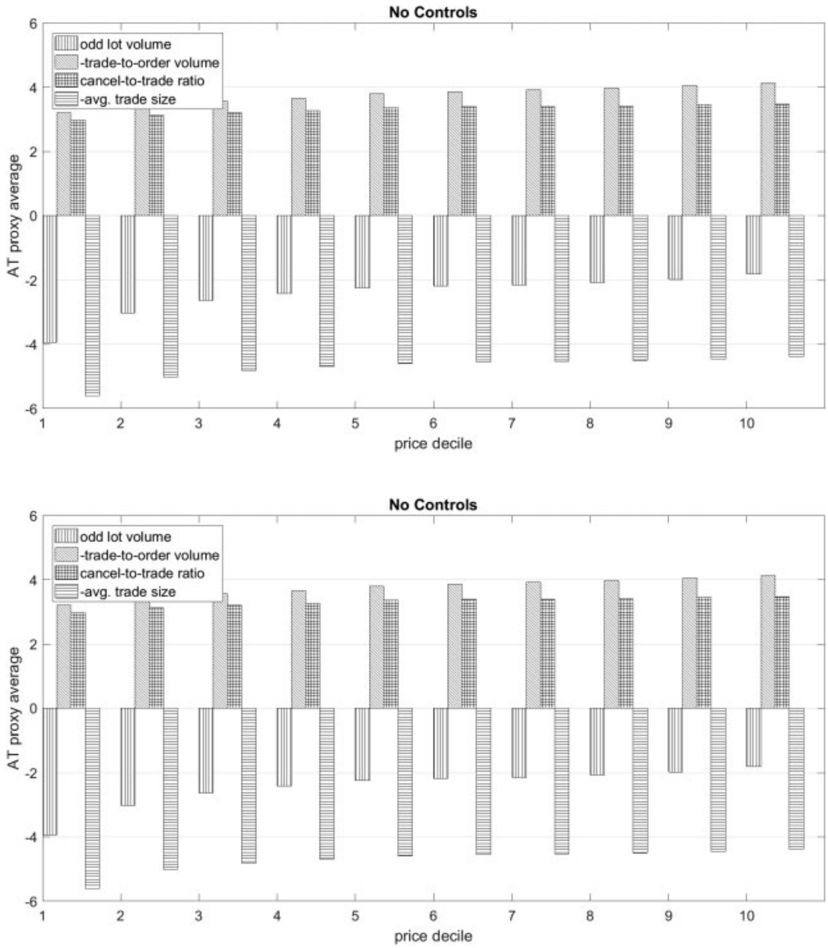

Figure 1
Algorithmic trading proxies by log price decile
This figure presents averages of the four algorithmic trading proxies for each decile of the log price and the component of log price orthogonal to all controls. “No controls” specification (top figure) consists of the algorithmic trading proxy averaged within log price deciles, where deciles are assigned by calendar month. “All controls” specification (bottom figure) averages AT proxies by residual log price decile, where the residual is constructed by regressing the log price on return volatility, quoted spread, number of analysts, institutional ownership ratio, and month fixed effects. AT proxies are normalized to be increasing in AT. Market capitalization, share price, and return volatility are the log of daily averages over $[T-42, T-22]$, and the quoted spread is average of the time-weighted bid-ask spread over the same interval (reported in percent). The number of analysts is the log of the largest number of reporting analysts in the Thomson Reuters I/B/E/S database associated with each stock-quarter announcement. The institutional ownership ratio (IOR) is the fraction of shares held by 13F filing institutions at the end of the preceding calendar quarter. Construction of algorithmic trading proxies derived from SEC MIDAS data is described in the main text.

Table 4
Determinants of announcement price impact with lagged log price instruments

	x = Odd lot ratio		x = Trade-to-order ratio		x = Cancel-to-trade ratio		x = Avg. trade size	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
x	0.0605***	0.0507***	−0.0920***	−0.0778***	0.104***	0.0874***	−0.123***	−0.108***
	(0.00775)	(0.00841)	(0.0127)	(0.0138)	(0.0149)	(0.0158)	(0.0154)	(0.0175)
Market cap.	0.00422	−0.0281***	0.00719**	−0.0296***	0.0146***	−0.0254***	0.00269	−0.0272***
	(0.00320)	(0.00443)	(0.00302)	(0.00439)	(0.00311)	(0.00435)	(0.00321)	(0.00430)
Ret. vol.		−0.0202		−0.0120		−0.00644		−0.0209*
		(0.0127)		(0.0124)		(0.0126)		(0.0123)
Quoted spr.		−0.0395***		−0.0560***		−0.0660***		−0.0292***
		(0.00787)		(0.00792)		(0.00881)		(0.00815)
#Analysts		0.0485***		0.0534***		0.0537***		0.0488***
		(0.00764)		(0.00789)		(0.00775)		(0.00747)
IOR		0.128***		0.123***		0.136***		0.113***
		(0.0227)		(0.0222)		(0.0218)		(0.0227)
Constant	0.529***	X	−0.0193	X	−0.170**	X	0.991***	X
	(0.0747)		(0.0678)		(0.0814)		(0.120)	
Month FEs	No	Yes	No	Yes	No	Yes	No	Yes
Stock FEs	No	No	No	No	No	No	No	No
N	24,107	23,814	24,068	23,800	24,120	23,842	24,512	24,201
K-P rk LM	42.67***	41.92***	42.03***	41.14***	41.44***	41.00***	42.12***	41.44***
K-P rk Wald F	4,226.2	4,554.4	1,493.9	1,189.7	1,001.3	899.2	2,765.0	2,808.9

$^{*}p<.10$, $^{**}p<.05$, $^{***}p<.01$

This table reports results from an instrumental variables regression of price jump ratios on a set of algorithmic trading proxies:

$$x_{it}=\zeta+\eta lprice_{it}+\theta\times controls_{it}+\delta_{it},$$

$$jump_{it}^{(21)}=\alpha+\beta\hat{x}_{it}+\gamma\times controls_{it}+\epsilon_{it}.$$

For each stock i and quarterly earnings announcement t from January 2012 through September 2016, the first-stage regression instruments algorithmic trading measures using the log of the average end-of-day stock price from $T-42$ to $T-22$. The second-stage regression for which results are reported relates predicted algorithmic trading proxies to the price jump ratio ($jump_{it}^{(21)}$). The price jump ratio is measured as the ratio of the announcement response divided by the total variation in the pre- and post-announcement period: $CAR_{it}^{(T-1,T+2)}/CAR_{it}^{(T-21,T+2)}$. Cumulative abnormal returns (in logs) are net of Fama and French (1992) three-factor implied returns over the same interval. Observations with cumulative net price impact not exceeding a minimum multiple of prior return volatility (described in the main text) are dropped from both stages. Market capitalization, share price, and return volatility are the log of daily averages over $[T-42,T-22]$, and the quoted spread is average of the time-weighted bid-ask spread over the same interval (reported in percent). The number of analysts is the log of the largest number of reporting analysts in the Thomson Reuters I/B/E/S database associated with each stock-quarter announcement. The institutional ownership ratio (*IOR*) is the fraction of shares held by 13F filing institutions at the end of the preceding calendar quarter. Construction of algorithmic trading proxies (x_{it}) derived from SEC MIDAS data is described in the main text. All standard errors are clustered by security and month and are reported in parentheses. K-P refers to the Kleibergen and Paap (2006) rk LM and Wald F statistics. In these specifications, a 10% maximal IV size corresponds to a critical value of 16.38.

association with reduced information acquisition than does the residual part of AT unspanned by log prices. One interpretation of this difference is that some algorithmic traders endogenously reduce their participation in stocks with greater information acquisition (like in Han et al. 2014; Baldauf and Mollner 2017), or conversely, that firm-quarters with larger price jump ratios tend to attract algorithmic traders.

5.3 When does AT deter information acquisition?

Pinpointing the timing of deterred information acquisition is critical for assessing whether informational distortions in financial markets might translate into suboptimal production or investment decisions. To assess the timing of information distortions, I decompose the price jump ratio for the entire pre-announcement period into daily *price response ratios* for earnings information. The price response ratio divides the cumulative abnormal return through date $T-k$ by the total abnormal return variation in the pre- and post-announcement period,

$$responseratio_{it}^{(k,21)} = \frac{CAR_{it}^{(T-21,T-k)}}{CAR_{it}^{(T-21,T+2)}}. \tag{6}$$

Lower price response ratios are associated with less pre-announcement information acquisition.

For each pre-announcement date k, I run the regressions of Equation (5) using the price response ratio for date k as the dependent variable in the second stage:

$$x_{it}^{(k)} = \zeta + \eta lprice_{it}^{(k)} + \theta \times controls_{it}^{(k)} + \delta_{it},$$

$$responseratio_{it}^{(k,21)} = \alpha + \beta^{(k)} \hat{x}_{it}^{(k)} + \gamma \times controls_{it}^{(k)} + \epsilon_{it}, \; \forall k = 1,\dots,21. \tag{7}$$

$x_{it}^{(k)}$ is the k-day lagged algorithmic trading proxy for stock i at date t. I instrument each lagged AT proxy using the lagged log stock price with corresponding additional lags, that is, the date k lagged log price is the log of the average price over $[T-42-k, T-22-k]$ measured in days relative to the earnings announcement date T. I simultaneously estimate the system (7) for all $k=1,\dots,21$ using two-step GMM to test hypotheses involving βs across multiple dates. In particular, I test hypotheses using Wald tests on $\beta^{(k)}$s for several pre-announcement dates k. Coefficients jointly different from zero indicate that algorithmic trading changes information acquisition as of a given point in the pre-announcement period.

Figure 2 plots regression coefficients by date for each of the four algorithmic trading proxies. To facilitate comparison across measures of algorithmic trading, I scale all coefficients by the standard deviation of the predicted AT proxy $\hat{x}^{(k)}$ from the first-stage regression of $x^{(k)}$ on shifted log price and controls. The top plot's regressions include a single control (lagged

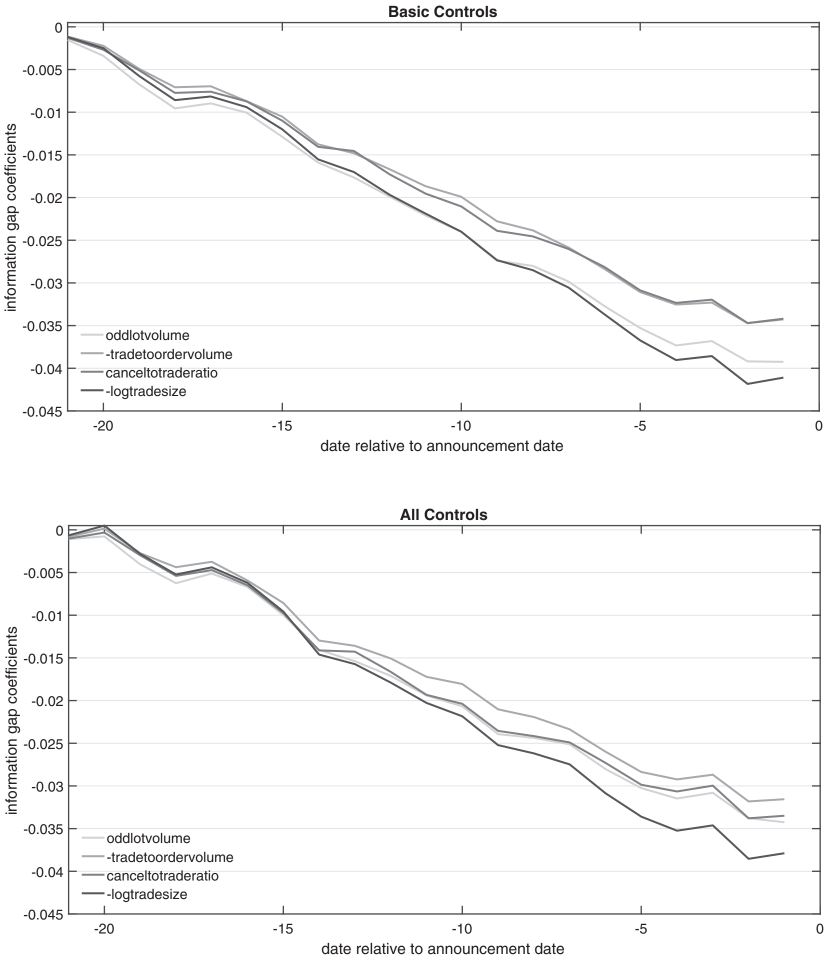

Figure 2
Price response ratios for algorithmic trading proxies
This figure presents coefficients from regressions of price response ratios on the four algorithmic trading proxies using the methodology of Table 5. To facilitate comparison across AT measures, I standardize all coefficients by normalizing the instrumented AT proxy $\hat{x}^{(k)}$ to have a unit standard deviation and to be increasing in AT. For each stock i and quarterly earnings announcement t from January 2012 through September 2016, the price response ratio is measured as the ratio of the cumulative price response through date k prior to the announcement date divided by the total variation over the information incorporation window: $CAR_{it}^{(T-21,T-k)}/CAR_{it}^{(T-21,T+2)}$. Cumulative abnormal returns (in logs) are net of Fama and French (1992) three-factor implied returns over the same interval. Observations with cumulative net price impact not exceeding a minimum multiple of prior return volatility (described in the main text) are dropped. "Basic" specification (top figure) consists of the algorithmic trading proxy and market capitalization. "All" specification (bottom figure) adds return volatility, quoted spread, number of analysts, institutional ownership, and month fixed effects. Market capitalization, share price, and return volatility are the log of daily averages over $[T-42-k, T-22-k]$, and the quoted spread is average of the time-weighted bid-ask spread over the same interval (reported in percent). The number of analysts is the log of the largest number of reporting analysts in the Thomson Reuters I/B/E/S database associated with each stock-quarter announcement. The institutional ownership ratio (*IOR*) is the fraction of shares held by 13F filing institutions at the end of the preceding calendar quarter. Construction of algorithmic trading measures derived from SEC MIDAS data is described in the main text.

market capitalization), and the bottom plot's regressions include all controls of specification (2) from Table 2.

The figure demonstrates that the relation between algorithmic trading and information acquisition becomes increasingly negative throughout the pre-announcement period. Coefficients for all algorithmic trading proxies follow very similar paths, both within each plot and between plots. The stable downward slope from 21 days prior to announcement is consistent with a near-constant difference between high- and low-AT stocks in the share of information being acquired and impounded into prices each day. Smooth incorporation of information is a cornerstone of strategic informed trading models (e.g., Kyle 1985; Back 1992), so it should be expected that deterring information acquirers manifests as ever-widening pre-announcement information gaps.[24]

Table 5 supplements these graphical results with formal tests of the hypothesis that information gaps are present at various stages of the pre-announcement period. The top panel uses price response ratios as the dependent variable in the second stage and tests whether total information gaps within a period are distinguishable from zero. Tests reject the hypothesis of no information gap for both two week subperiods two to four weeks and zero to two weeks pre-announcement, as well as for the whole pre-announcement period (not tabulated). Rejection of no information gap is comparably strong for each algorithmic trading proxy independently as well as for the first principal component factor of the four AT proxies.

The bottom panel examines the incremental information gap by day for each subperiod. I again test the hypothesis that sets of $\beta^{(k)}s$ are jointly zero in Equation (7), where the second-stage dependent variable is now replaced by the *daily change* in the cumulative price response ratio. In both 2-week subperiods, as for the whole sample, the incremental information gaps are distinguishable from zero on average, indicating that algorithmic trading is associated with continual reductions in relative information acquisition for the entire pre-announcement month.

6. Alternatives to the Price Jump Ratio Measure

6.1 Price nonsynchronicity

This section considers two alternative measures of information acquisition—price nonsynchronicity and intraperiod timeliness—to confirm that findings are robust to how information acquisition is measured. Roll (1988) proposes an R^2-based measure of firm-specific information in prices, and this idea is further developed into a "price nonsynchronicity" measure by Morck, Yeung, and Yu (2000) and Durnev, Morck, Yeung, and Zarowin (2003).

---

[24] Interpretation of the price response ratio coefficients is unchanged by allowing for multiple informed traders like in Holden and Subrahmanyam (1992) or Foster and Viswanathan (1993). The Online Appendix provides additional discussion of this point.

Table 5
Algorithmic trading and the timing of information acquisition

		Δ: 2–4 weeks pre		Δ: 0–2 weeks pre	
		(3)	(4)	(5)	(6)
Controls		Basic	All	Basic	All
Degrees of freedom		10	10	10	10
x = Odd lot ratio	χ^2	53.62***	35.47***	98.61***	54.38***
	p	0.000	0.000	0.000	0.000
x = Trade-to-order ratio	χ^2	45.79***	34.89***	93.80***	56.27***
	p	0.000	0.000	0.000	0.000
x = Cancel-to-trade ratio	χ^2	46.73***	35.56***	91.18***	56.19***
	p	0.000	0.000	0.000	0.000
x = Avg. trade size	χ^2	46.56***	35.11***	97.44***	55.66***
	p	0.000	0.000	0.000	0.000
x = First PCF of AT measures	χ^2	48.78***	34.61***	89.39***	53.79***
	p	0.000	0.000	0.000	0.000

		Δ: 2–4 weeks pre		Δ: 0–2 weeks pre	
		(3)	(4)	(5)	(6)
Controls		Basic	All	Basic	All
Degrees of freedom		10	10	10	10
x = Odd lot ratio	χ^2	51.18***	35.00***	42.01***	22.93**
	p	0.000	0.000	0.000	0.011
x = Trade-to-order ratio	χ^2	44.03***	34.89***	41.34***	24.44***
	p	0.000	0.000	0.000	0.007
x = Cancel-to-trade ratio	χ^2	45.69***	35.96***	41.32***	25.42***
	p	0.000	0.000	0.000	0.005
x = Avg. trade size	χ^2	45.89***	35.47***	50.22***	28.08***
	p	0.000	0.000	0.000	0.002
x = First PCF of AT measures	χ^2	46.30***	34.06***	35.27***	22.20**
	p	0.000	0.000	0.000	0.014

$*p<.10$, **p<.05, $***\ p<.01$

This table presents results from IV regressions of price response ratios on a set of algorithmic trading proxies:

$$x_{it}^{(k)} = \zeta + \eta lprice_{it}^{(k)} + \theta \times controls_{it} + \delta_{it},$$

$$responseratio_{it}^{(k,21)} = \alpha + \beta^{(k)} \hat{x}_{it}^{(k)} + \gamma \times controls_{it} + \epsilon_{it},\ \forall k = 1,\ldots,21.$$

For each stock i and quarterly earnings announcement t from January 2012 through September 2016, the price response ratio is measured as the ratio of the cumulative price response (first subtable) through date k prior to the announcement date divided by the total variation over the information incorporation window: $CAR_{it}^{(T-21,T-k)}/CAR_{it}^{(T-21,T+2)}$. Second subtable replaces the cumulative price response with the concurrent price response $\left(CAR_{it}^{(T-21,T-k)} - CAR_{it}^{(T-21,T-k-1)}\right)/CAR_{it}^{(T-21,T+2)}$, where I set $CAR_{it}^{(T-21,T-22)}$ equal to zero. Cumulative abnormal returns (in log) are net of Fama and French (1992) three-factor implied returns over the same interval. Observations with cumulative net price impact not exceeding a minimum multiple of prior return volatility (described in the main text) are dropped.

Table entries correspond with cross-equation Wald test statistics and p-values for hypotheses on sets of $\beta^{(\cdot)}$ estimated using two-step GMM. "Basic" specification consists of algorithmic trading proxy and market capitalization. "All" specification adds share price, return volatility, quoted spread, number of analysts, and month fixed effects. Market capitalization, share price, and return volatility are the log of daily averages over $[T-42, T-22]$, and the quoted spread is average of the time-weighted bid-ask spread over the same interval (reported in percent). The number of analysts is the log of the largest number of reporting analysts in the Thomson Reuters I/B/E/S database associated with each stock-quarter announcement. The institutional ownership ratio (IOR) is the fraction of shares held by 13F filing institutions at the end of the preceding calendar quarter. Construction of algorithmic trading proxies ($x_{it}^{(k)}$) derived from SEC MIDAS data is described in the main text, and the first principal component factor of these proxies is denoted by "first PCF." All standard errors are clustered by security and month.

Intuitively, the less idiosyncratic information is produced and incorporated into the stock's price, the better a stock's return is approximated by aggregate information. Following Morck et al. (2000) and Durnev et al. (2003), I estimate price nonsynchronicity for each stock-quarter pair as $1-R^2$ from the regression

$$r_{it} = \alpha + \beta_{im} r_{mt} + \gamma r_{ind,t} + \epsilon_{it}, \tag{8}$$

where stock i's industry return is the value-weighted average return for stocks sharing stock i's three-digit SIC code.

One advantage of the price nonsynchronicity measure is the ability to examine a broader set of events for which the price jump ratio is not well-suited. In particular, I no longer require a large and prescheduled information shock because the price nonsynchronicity measure does not require an estimate of event impact, instead normalizing idiosyncratic shocks by the sum of firm plus aggregate shocks. By contrast, the price jump ratio normalizes by total (firm-specific) event impact, which is key for the price jump ratio's interpretation as information acquired relative to information available.

I use RavenPack News Analytics data to construct a wider range of economically important firm-specific events. RavenPack provides real-time collection and analysis of thousands of news sources for tens of thousands of companies in the United States and internationally. Critically for this analysis, RavenPack algorithmically links news items to companies, identifies distinct news events, and classifies items into categories ranging from earnings news to labor issues. Table 6 summarizes the set of events comprising more than 1% of RavenPack equity news coverage, excluding reports on trading or prices (technical analysis signals, stock price movements, order imbalance reports) and announcements of future disclosure dates (investor relations items). Of these items, 17.7% are earnings news, and more than 30% are earnings, dividend, or revenue items (often announced together). The remaining events that exceed the 1% threshold are distributed mostly among product releases, executive appointments and resignations (broadly, "labor issues"), M&A activity, analyst updates, and insider trading reports.

Informed by Table 6, I add three sets of common firm events that provide material information on firms' scope, inputs, or outputs: M&A activity, executive appointments and resignations, and product releases. I construct the set of stock-event pairs as follows. I first build a list consisting of each event with RavenPack "group" equal to one of these categories, relevance score exceeding 90 (highly relevant), and global novelty score equal to 100 (first occurrence of reporting within 24 hours).[25] Then, for each stock, I roll backward through the

[25] From the RavenPack Version 4.0 User Guide, "a relevance value of at least 90 indicates that the entity is referenced in the main title or headline of the news item." I exclude events with relevance scores lower than 90 on a 0–100 scale to ensure that news items have bearing on the algorithmically tagged companies. I also impose a novelty filter using RavenPack's proprietary "global novelty score" (GNS), and I require that the item be the first instance of reporting (GNS=100). The novelty filter prevents double counting of single news events reported in multiple publications.

Table 6
RavenPack news items

Rank	Group	Type	N	%	Cumulative %
1	Earnings	Earnings	3,903,482	17.7	17.7
2	Products-services	Product-release	1,475,898	6.7	24.4
3	Labor-issues	Executive-appointment	1,235,878	5.6	30.1
4	Products-services	Business-contract	1,184,949	5.4	35.4
5	Revenues	Revenue	876,281	4.0	39.4
6	Analyst-ratings	Analyst-ratings-change	875,978	4.0	43.4
7	Acquisitions-mergers	Scquisition	865,047	3.9	47.3
8	Marketing	Conference	766,278	3.5	50.8
9	Credit-ratings	Credit-rating-change	670,356	3.0	53.8
10	Earnings	Earnings-per-share	595,781	2.7	56.6
11	Investor-relations	Conference-call	577,830	2.6	59.2
12	Insider-trading	Insider-sell	571,758	2.6	61.8
13	Dividends	Dividend	541,002	2.5	64.2
14	Partnerships	Partnership	490,267	2.2	66.5
15	Earnings	Earnings-guidance	480,604	2.2	68.6
16	Insider-trading	Sell-registration	472,851	2.1	70.8
17	Acquisitions-mergers	Unit-acquisition	393,994	1.8	72.6
18	Insider-trading	Insider-buy	370,203	1.7	74.3
19	Price-targets	Price-target	354,128	1.6	75.9
20	Labor-issues	Executive-resignation	302,980	1.4	77.2
21	Assets	Facility	268,681	1.2	78.5
22	Revenues	Revenue-guidance	254,041	1.2	79.6
23	Analyst-ratings	Analyst-ratings-set	251,252	1.1	80.8

This table presents all RavenPack company news categories with shares exceeding 1% of the January 2012–March 2016 RavenPack sample. Technical analysis signals, stock price movements, order imbalance reports, and investor relations items (typically announcements of future information revelation dates) are excluded.

data to find any of these events within the last week and take the first event in the string of RavenPack reports. This step is necessary because a 24-hour gap in coverage generates a new, but often redundant event in the RavenPack data. Finally, I move each news announcement date to the first date in which the market is open after the announcement. Each element of this collapsed list represents a potential information shock with a tilt toward idiosyncratic or industry news relative to a typical day in the life of a stock. RavenPack contributes 72,544 events in total.[26] I stack these events with the 54,879 earnings announcement events to obtain a consolidated set of important firm events with which to assess the representativeness of my earnings announcement results.

Equipped with these data, one may be tempted to simply regress price nonsynchronicity around each event on algorithmic trading proxies and controls,

$$nonsynch_{it} = \alpha + \beta x_{it} + \gamma \times controls_{it} + \epsilon_{it}. \tag{9}$$

However, such an approach faces two challenges. First, high R^2 indicates that there is little idiosyncratic information to acquire (case 1) or that idiosyncratic innovations are not discovered and incorporated into stock returns (case 2). AT may prefer stocks with more or less potentially acquirable information so that

26 The RavenPack sample is available only through March 2016, so the date range is slightly shorter than that accommodated by MIDAS, CRSP, and other resources.

an association of AT with price nonsynchronicity reflects case 1 rather than case 2. This concern requires that AT proxies be cleansed of their correlation with the amount of information acquirable. This problem is amenable to an instrumental variables approach like in Section 5.2. For the exclusion restriction to be satisfied, the lagged log stock price must now be unrelated to the prevalence of potentially acquirable *idiosyncratic* information relative to factor information (conditional on other covariates). As before, there is no obvious channel by which the lagged price should relate to market-industry R^2s holding fixed covariates such as lagged market capitalization (e.g., large firms may be more representative of their industries) and institutional ownership ratios.

There is also a second challenge specific to my setting that requires the price nonsynchronicity measure to be treated with additional care. Algorithmic traders may better enforce true factor relationships than their human counterparts (by, e.g., statistical arbitrage), thereby increasing the R^2 of a market-industry model. In this case, it would be wrong to interpret this increase in R^2 as less idiosyncratic information entering prices because it may well reflect more industry or index information entering prices instead. As noted by Kirilenko and Lo (2013), index arbitrage is indeed quite common among algorithmic traders, so we would expect to encounter a negative relation between price nonsynchronicity and AT in estimating Equation (9).[27] For this reason, a significant negative association between AT and price nonsynchronicity is informative only insofar as a strong positive association would be difficult to reconcile with my previous results.

I address this problem by exploiting the fact that the period before major firm news may feature more *firm-specific* acquirable information than the typical span over the preceding year. If factor and industry betas do not change much from typical days to news periods, then subtracting price nonsynchronicity around "typical" periods from price nonsynchronicity around "event" periods nets out the effects of AT statistical arbitrage. A high *price nonsynchronicity difference* reflects larger increases in return responsiveness to increases in firm-specific news. Correspondingly, I define the price nonsynchronicity difference as $1-R^2$ of regression (8) for the $[T-21, T-1]$ period minus $1-R^2$ of regression (8) for the $[T-252, T-64]$ period.

Table 7 reports the results of IV regressions of price nonsynchronicity differences on algorithmic trading proxies,

$$x_{it} = \zeta + \eta lprice_{it} + \theta \times controls_{it} + \delta_{it},$$

$$nonsynch_{it}^{event} - nonsynch_{it}^{ave} = \alpha + \beta \hat{x}_{it} + \gamma \times controls_{it} + \epsilon_{it}. \tag{10}$$

Netting out the effects on the price nonsynchronicity measure arising from index arbitrage algorithms, more algorithmic trading decreases firm-specific

27 Indeed, Gerig (2015) finds evidence of an association between HFT and increased price synchronization across securities.

Table 7
Determinants of announcement price impact: Difference of price nonsychronicities

	x = Odd lot ratio		x = Trade-to-order ratio		x = Cancel-to-trade ratio		x = Avg. trade size	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
x	−0.0170***	−0.00736***	0.0286***	0.0115***	−0.0312***	−0.0121***	0.0332***	0.0141***
	(0.00294)	(0.00227)	(0.00492)	(0.00385)	(0.00567)	(0.00434)	(0.00576)	(0.00456)
Market cap.	0.00630***	0.00688***	0.00515***	0.00698***	0.00318	0.00639***	0.00696***	0.00671***
	(0.00192)	(0.00200)	(0.00189)	(0.00198)	(0.00194)	(0.00197)	(0.00194)	(0.00199)
Ret. vol.		−0.0106**		−0.0117**		−0.0124**		−0.0103**
		(0.00481)		(0.00502)		(0.00505)		(0.00481)
Quoted spr.		−0.00309		−0.000796		0.000119		−0.00464
		(0.00362)		(0.00413)		(0.00440)		(0.00352)
#Analysts		−0.00131		−0.00171		−0.00158		−0.00104
		(0.00278)		(0.00276)		(0.00273)		(0.00271)
IOR		−0.00311		−0.00463		−0.00579		−0.00123
		(0.00728)		(0.00754)		(0.00752)		(0.00721)
Constant	−0.239***	X	−0.0674	X	−0.0313	X	−0.368***	X
	(0.0397)		(0.0411)		(0.0448)		(0.0510)	
Month FEs	No	Yes	No	Yes	No	Yes	No	Yes
Stock FEs	No	No	No	No	No	No	No	No
N	53,679	53,004	53,728	53,113	53,727	53,108	54,757	54043
K-P rk LM	43.05***	42.24***	42.54***	41.68***	41.57***	40.90***	42.39***	41.72***
K-P rk Wald F	4,527.7	5,021.6	1,442.0	1,123.1	1,193.1	974.3	3,027.8	3,077.3

* $p<.10$, ** $p<.05$, *** $p<.01$

This table presents results from a regression of price nonsynchronicity differences on a set of algorithmic trading proxies:

$$x_{it}=\zeta+\eta lprice_{it}+\theta\times controls_{it}+\delta_{it},$$

$$nonsynch_{it}^{event}-nonsynch_{it}^{ave}=\alpha+\beta\hat{x}_{it}+\gamma\times controls_{it}+\epsilon_{it}.$$

For each stock i and event date t from January 2012 through September 2016, the average price nonsynchronicity ($nonsynch_{it}^{ave}$) is measured as the average of one minus the R^2 of a regression of returns of stock i on market and sector returns for the year preceding the date of the news event (trading days $T-252$ through $T-1$). The event-period price nonsynchronicity ($nonsynch_{it}^{event}$) is the same quantity measured over the pre-event period of $T-21$ to $T-1$ days. Market capitalization, share price, and return volatility are the log of daily averages over $[T-42, T-22]$, and the quoted spread is average of the time-weighted bid-ask spread over the same interval (reported in percent). The number of analysts is the log of the largest number of reporting analysts in the Thomson Reuters I/B/E/S database associated with the earnings announcement in the same stock-quarter. The institutional ownership ratio (*IOR*) is the fraction of shares held by 13F filing institutions at the end of the preceding calendar quarter. Construction of algorithmic trading proxies (x_{it}) derived from SEC MIDAS data is described in the main text. All standard errors are clustered by security and month and are reported in parentheses. The log price is dropped from stock fixed effects regression on account of near-collinearity with market capitalization. K-P refers to the Kleibergen and Paap (2006) rk LM and Wald F statistics. In these specifications, a 10% maximal IV size corresponds to a critical value of 16.38.

news in prices before major firm events.[28] This reduction in the discovery of idiosyncratic information around a range of important news events generalizes the evidence obtained before earnings announcements using the price jump ratio. Notably, these estimates are conservative because statistical arbitrage if anything becomes less attractive in the presence of more idiosyncratic news (e.g., information ratios fall as idiosyncratic risk increases). Consequently, algorithmic traders would be expected to pare back such strategies before idiosyncratic events, thereby pushing the price nonsynchronicity difference upward and biasing β up for proxies increasing in AT.

6.2 Intraperiod timeliness

I construct "perfect foresight" returns for the IPT measure by investing $+1/N_{up}$ dollars in all N_{up} stocks that increase in value around the earnings announcement and $-1/N_{down}$ dollars in all N_{down} stocks that decrease in value around the earnings announcement. These zero-cost or "hedge" portfolios are constructed separately for stocks with high and low values of a specific characteristic, namely algorithmic trading activity or lagged log stock prices in this application. I then compare (1) price jump ratio analogues at the portfolio level and (2) "areas under the curve" plotting cumulative abnormal returns relative to total abnormal returns for each pre-announcement date. Large areas under the curve represent earlier incorporation of information.

Figure 3 and Table 8 present results from portfolio price jump ratio and intraperiod timeliness tests, where I split AT measures into quintiles in constructing hedge portfolios. The figure plots scaled return differences between high AT and low AT quintiles using annual hedge portfolios. For every algorithmic trading measure and annual cut, increased algorithmic trading is associated with reduced portfolio-level price response ratios and intraperiod timeliness. Differences between high- and low-lagged price portfolios are comparably large. The table reports tests of the hypothesis of equal responsiveness in the high and low AT portfolios with quarterly observations, and equality is easily rejected for all measures for both CARs and intraperiod timeliness.

The IPT approach circumvents the undefined conditional mean problem of the price jump ratio if all portfolio abnormal returns are "large enough" in absolute value, as they are in this application. However, the comparison of the share of absolute cumulative abnormal returns accrued up to date $T-k$ (or its integral) is not valid if the portfolios are not "all else equal." A major drawback to this analysis is the inability to add multiple control variables to enforce this condition. Unfortunately, adjusting the methodology to overcome this issue is not straightforward because the IPT approach necessarily examines a small

[28] If the effect of additional index arbitrage on price nonsynchronicity is multiplicative rather than additive, the price nonsynchronicity difference should be replaced by $nonsynch_{it}^{event}/nonsynch_{it}^{ave}$ in regression (10). Using this ratio delivers the same qualitative relationships with similar levels of statistical significance.

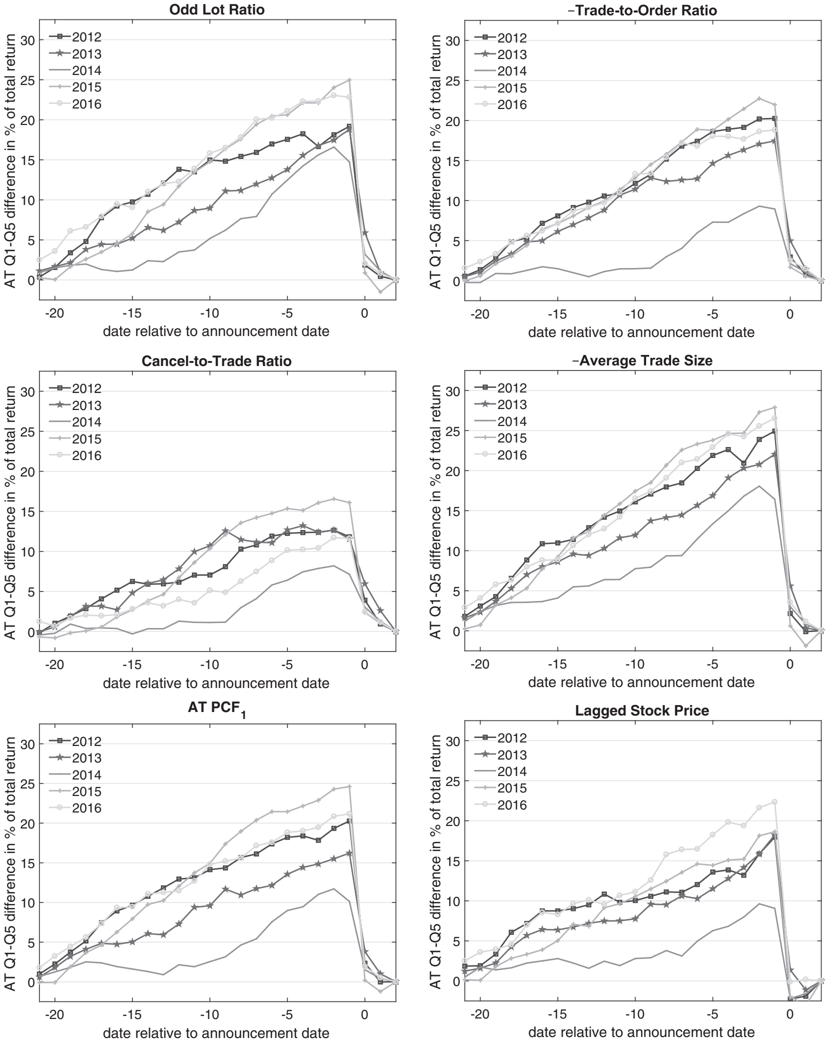

Figure 3
Differences in intraperiod timeliness
This figure presents differences in intraperiod timeliness in the mode of McNichols (1984) and Alford et al. (1993). I sort stocks into quintiles for each algorithmic trading proxy for each calendar year. For securities in quintiles one or five, I construct the "perfect foresight" return on a zero-cost portfolio with scale of one dollar. At date $T-21$ relative to announcement date T, I invest $+1/N_{up}$ dollars in each stock that earns a positive return through $T+2$, and I invest $-1/N_{down}$ dollars in each stock that earns a negative return through $T+2$, where N_{up} and N_{down} are the counts of securities in each group. For each date, I then compute the abnormal portfolio return for each quintile (net of Fama and French 1992 three-factor returns) divided by the total abnormal portfolio return as of $T+2$. Figures plot the difference between the relative returns as of date k preceding the announcement on the low-AT portfolio net of the high-AT portfolio. PCF_1 denotes the first principal component factor of the AT proxies.

Table 8
Intraperiod timeliness tests

A. Test of equality of cumulative abnormal portfolio return as of T

$x=$	Odd lots (1)	$\frac{\text{Trades}}{\text{Orders}}$ (2)	$\frac{\text{Cancels}}{\text{Trades}}$ (3)	Trade size (4)	AT PCF_1 (5)	log price (6)
Q5–Q1	−20.36***	−16.63***	−11.71***	−23.40***	−18.71***	−16.47***
	(14.19)	(11.38)	(8.85)	(15.06)	(11.62)	(10.84)

B. Test of equality of IPT as of T

$x=$	Odd lots (1)	$\frac{\text{Trades}}{\text{Orders}}$ (2)	$\frac{\text{Cancels}}{\text{Trades}}$ (3)	Trade size (4)	AT PCF_1 (5)	log price (6)
Q5–Q1	−230.21***	−195.23***	−136.23***	−268.61***	−221.78***	−175.57***
	(11.51)	(10.22)	(7.33)	(13.13)	(10.05)	(8.39)

$^*p<.10$, $^{**}p<.05$, $^{***}p<.01$
This table presents results from intraperiod timeliness tests in the style of McNichols (1984) and Alford et al. (1993). I sort stocks into quintiles for each algorithmic trading proxy for each calendar quarter. For securities in quintiles one or five, I construct the “perfect foresight” return on a zero-cost portfolio with scale of one dollar. At date $T-21$ relative to announcement date T, I invest $+1/N_{up}$ dollars in each stock that earns a positive return through $T+2$, and I invest $-1/N_{down}$ dollars in each stock that earns a negative return through $T+2$, where N_{up} and N_{down} are the counts of securities in each group. For each date, I then compute the abnormal portfolio return for each quintile (net of Fama and French (1992) three-factor returns) divided by the total abnormal portfolio return as of $T+2$. Panel A tests the hypothesis that the average relative returns (in %) are equal in portfolios 1 and 5 as of date T across calendar quarters. Panel B supplements this t test with an IPT comparison by cumulating the area under each curve and comparing average areas as of date T between portfolios 1 and 5. PCF_1 denotes the first principal component factor of the AT proxies. Standard errors are reported in parentheses.

number of portfolios in place of the price jump ratio’s richer cross section of stock-quarter observations.[29]

7. Conclusion

Friction between the building blocks of price discovery—acquiring new information and impounding existing information into prices—is more than a theoretical curiosity. I demonstrate a conflict between these forces in the context of major technological innovations in trading. Algorithmic trading simultaneously increases price efficiency with respect to acquired information while reducing the available information to which prices respond. These competing effects of algorithmic trading complicate the evaluation of technological advances in financial markets.

On balance, my findings reflect a large negative effect of algorithmic trading on information acquisition. Two channels are consistent with this result. First, AT order anticipation or learning from order flow may erode rents to information acquirers (Yang and Zhu 2017; Stiglitz 2014). Second, more sophisticated signal processing by algorithmic liquidity providers may reduce adverse selection costs, but at the same time, it may also reduce the profits

[29] In principle adding sorting dimensions or reweighting observations could accommodate additional controls. However both approaches face a curse of dimensionality with the number of sorting or matching dimensions, and adding more than one or two controls is not feasible given my sample size.

accruing to an informational advantage (Baldauf and Mollner 2017). I offer one attempt to quantify the relative importance of these channels in the Online Appendix, but more work is needed to assess the precise mechanisms by which improved trading technology reduces the information content of prices.

In general, weighing the trade-off between information acquisition and price efficiency requires additional structure from models of trader behavior. However, in the specific case of algorithmic trading, the net effect on social welfare through informational channels is almost surely a loss. With homogenous households, production or investment decisions must be affected for information acquisition to affect social welfare, and these decisions do not depend on split-second price changes. Unlike improvements in price efficiency that occur over horizons of seconds or milliseconds, deterrence of information acquisition persists over (at least) a 1-month horizon prior to the earnings announcements. In light of the allocative externalities associated with information-rich prices, formalizing the welfare trade-offs associated with information acquisition and market quality remains an important challenge for future research and policy guidance.

Appendix A. Sample Selection

Figure A1 plots the distribution of price jump ratios for the set of sufficiently economically important announcements for several choices of the minimum event size cutoff, $\left|CAR_{it}^{(T-21,T+2)}\right| > \sqrt{24}\hat{\sigma}_{it} \times$ cutoff. The plot reveals the trade-off between retaining a larger portion of the sample and reducing the fraction of price jump ratios in the $[0,1]$ interval. For the most permissive cutoff considered, more than 63% of retained observations are in the unit interval in which the "true" price jump ratio should lie absent idiosyncratic noise. I select a cutoff of one to balance retaining data and identifying coefficients off of important earnings announcements, but results do not vary much across a range of cutoffs exceeding 1/2 (or less).

A significant fraction of the sample is thus omitted from the main analysis. I now assess whether the observations included in the main analysis systematically differ from the observations excluded by the volatility cutoff rule. Table A1 presents summary statistics for the included and the excluded groups for every control variable in my analysis. Distributions of included and excluded observations are very similar along the dimensions of market capitalization, analyst coverage (excepting the left tail), and institutional ownership. Excluded observations have slightly higher volatility, as would be expected given the volatility-based cutoff. Log price distributions are shifted leftward by 10% to 20% for excluded observations, and correspondingly, relative bid-ask spreads are 20% to 30% larger for that set. If anything, the analysis seems to focus on a set of securities with marginally more predicted AT than the general set of stocks. Notwithstanding this difference, the overlap between price and bid-ask spread distributions is still quite substantial, for example, the 10th (25th) [50th] percentile of price among the included observations is between the 10th (25th) [50th] and 25th (50th) [75th] percentiles of price among the excluded observations.

Figure A2 presents corresponding breakdowns by industry. I classify stocks by SIC code into Fama and French's 49 industries (industry definitions are provided at Ken French's website). Industry membership is broadly similar, with a few notable exceptions. Included firm-quarters are somewhat more likely than excluded firm-quarters to be in the retail or software industries, and they are somewhat less likely than excluded firm-quarters to be in the banking, pharmaceutical, or financial industries. Dropping these five industries from the analysis strengthens my results slightly.

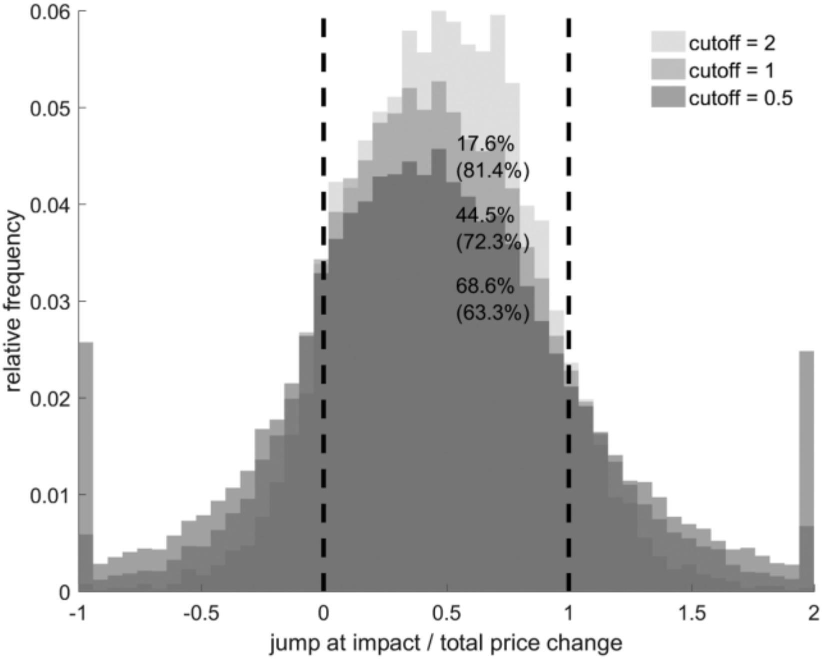

Figure A1
Range of announcement impact price jump ratio
This figure depicts the empirical distribution of earnings announcement price jump ratios for several minimum thresholds for the total price movement over the interval. The price jump ratio is given by $CAR_{it}^{(T-1,T+2)}/CAR_{it}^{(T-21,T+2)}$ for Fama and French (1992) three-factor residualized changes in log price. Stock-quarter observations enter the sample for $\left|CAR_{it}^{(T-21,T+2)}\right| > \sqrt{24}\hat{\sigma}_{it} \times$ cutoff, where $\hat{\sigma}_{it}$ is the daily return volatility from $T-42$ to $T-22$ days before an earnings announcement. For each choice of cutoff in $\{1/2, 1, 2\}$, the top number in the histogram is the proportion of the sample that exceeds the relevance cutoff. The bottom number is the fraction of the conditional ratio of post-announcement date variation to total variation that is bounded in the $[0,1]$ interval. Values exceeding -1 on the left and 2 on the right are winsorized at these respective values for visual clarity.

Table A1
Summary statistics of key variables for included and excluded data

Included	Market cap.	Price	Ret. vol.	Quoted spr.	#Analysts	*IOR*
Mean	20.89	36.55	−4.06	0.40	1.81	0.60
SD	1.81	56.73	0.52	0.67	0.92	0.27
10%	18.61	4.27	−4.68	0.05	0.69	0.16
25%	19.65	10.48	−4.41	0.09	1.10	0.44
Median	20.84	23.92	−4.09	0.21	1.95	0.66
75%	22.04	46.47	−3.73	0.43	2.48	0.79
90%	23.27	75.39	−3.39	0.87	2.94	0.88
Excluded	**Market cap.**	**Price**	**Ret. vol.**	**Quoted spr.**	**#Analysts**	***IOR***
Mean	20.56	31.52	−3.88	0.58	1.63	0.55
SD	1.94	50.20	0.60	0.96	0.98	0.28
10%	18.11	2.81	−4.59	0.05	0.00	0.10
25%	19.15	7.88	−4.30	0.10	1.10	0.35
Median	20.46	19.32	−3.93	0.26	1.79	0.62
75%	21.85	40.29	−3.51	0.61	2.40	0.77
90%	23.13	68.55	−3.10	1.41	2.89	0.87

Table presents descriptive statistics for all control variables in my study. The top panel summarizes dollar market capitalization, share price, and return volatility (all in logs) from CRSP data for the observations satisfying the $\left|CAR_{it}^{(T-21,T+2)}\right| > \sqrt{24}\hat{\sigma}_{it}$ cutoff ("included"). Additional summaries include the median end-of-minute quoted bid-ask spreads in percent from the TAQ NBBO files (one-second version); the number of analysts as the log of the maximum number of reporting analysts for each stock-event pair in Thomson Reuters I/B/E/S; and the institutional ownership ratio (*IOR*) as the fraction of shares held by 13F filing institutions at the end of the preceding calendar quarter. All control variables are lagged an additional 21 days prior to the announcement date relative to the AT proxies ($[T-42, T-22]$). The bottom table summarizes the corresponding quantities for observations not satisfying the CAR cutoff ("excluded").

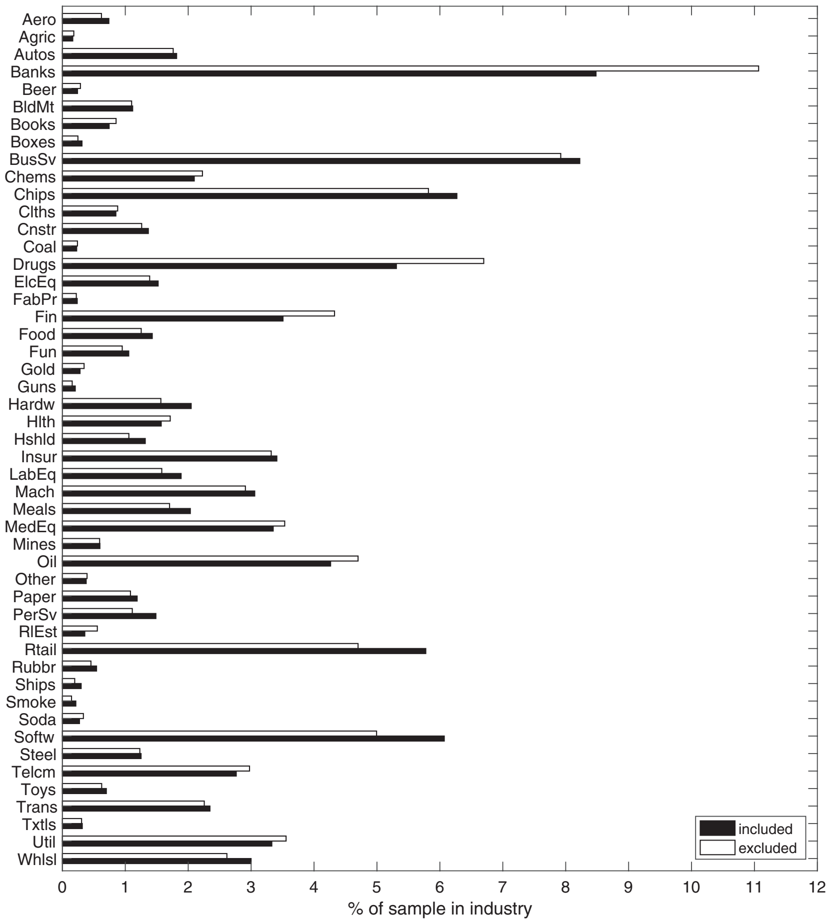

Figure A2
Industry membership for included and excluded data
The figure presents industry classifications for the included and the excluded observations. Black bars represent the share of observations in each Fama-French 49 industry category for "included" observations satisfying the $\left|CAR_{it}^{(T-21,T+2)}\right| > \sqrt{24}\hat{\sigma}_{it}$ cutoff. White bars represent the corresponding shares of "excluded" observations not satisfying this cutoff.

Appendix B. Nonrandom Variation in the Importance of Earnings Announcements

Idiosyncratic returns unrelated to earnings announcements add noise to the price jump ratio measure. This noise biases the estimated relation between AT and information acquisition if the relative importance of earnings news is correlated with algorithmic trading activity, as may be the case, for example, if algorithmic trading affects the steady-state information content of prices.

I address this potential bias using a matched pairs approach to control for the relative scale of announcement information innovations across stocks. The key grouping variables are (1) the first principal component factor of the four AT measures and (2) and the lagged log price AT instrument. For each calendar quarter, I generate pairs from these groups by minimizing the Mahalanobis distance between observations matched on lagged market capitalization and cumulative announcement impact relative to lagged stock return volatility, $\left|CAR_{it}^{(T-21,T+2)}/\hat{\sigma}_{it}\right|$. Matching on these characteristics provides a nonparametric control for the relative variation associated with earnings announcements as well as for market capitalization, which itself subsumes several liquidity-related variables.

I then conduct matched pairs analyses using (1) all stocks in the regression sample and (2) all stocks in the highest and lowest algorithmic trading activity quintiles (subject to the same restriction). In the first analysis, I match without replacement observations in the top half of algorithmic trading activity with observations in the bottom half of algorithmic trading activity. Observations with higher (lower) than median loadings on the first principal component factor of the AT measures are assigned to the "high AT" ("low AT") group. I retain the top 300 matches by distance in each quarter. In the second analysis, I assign matches between the top and bottom quintiles of the first factor loading and retain the top 100 matches by distance for each quarter. I then repeat both analyses using matches between observations within halves and quintiles of the lagged log stock price distribution. By sorting on this exogenous determinant of algorithmic trading, this second analysis provides an "instrumental variables" counterpart to the AT-sorted matched pairs studies.

Table B1 reports results from these matched pairs analyses. Panels A.1 and A.2 report summary statistics to confirm similarity between groups on the matching variables. Distributions of matched variables appear virtually identical across all specifications, and sorting variables exhibit large dispersion between groups. Importantly, stocks with a higher lagged log stock price in panel A.2 also have higher concentrations of AT activity—this relationship is the matched-pairs counterpart to the requirement of instrument relevance in IV analysis.

Panels B.1 and B.2 present pairwise differences in means and comparisons of distributions of the price jump ratio between high and low algorithmic trading groups and between high and low lagged log stock price groups, respectively. I estimate differences in means using a pairwise t-test and differences in distributions using a Wilcoxon signed rank test. To account for error in the matching step, I follow Abadie and Spiess (2016) and cluster or resample by security, month, and match index for t-tests and Wilcoxon signed-rank tests, respectively.[30] Consistent with Table 2, the high AT group has significantly larger price jump ratios on average, both statistically and economically. These differences are especially large for pairings between extreme AT quintiles, which indicates that extreme AT activity has an even larger negative association with information acquisition. Sorting on the lagged stock price instrument delivers similar results with the interpretation of the

[30] I construct three-way clustered variance-covariance matrices for the Wilcoxon signed rank test following Cameron, Gelbach, and Miller (2011) and in the spirit of Rosner, Glynn, and Lee (2006). Specifically, I first resample 1,000 times with match index clusters, month clusters, and security clusters. I then resample 1,000 times with match index-month clusters, match index-security clusters, and month-security clusters. Finally I resample 1,000 times clustering on all three variables. I calculate the three-way clustered variance matrix of the test statistic as the sum of variances on the single clusters and triple cluster minus the sum of variances on the double clusters.

Table B1
Determinants of announcement price impact: Matched pairs

A.1. Matched sample properties

Matched pairs (halves 2-1)								
	Low AT				High AT			
Group Variable	Impact	Market cap.	AT PCF_1	*lprice*	Impact	Market cap.	AT PCF_1	*lprice*
Mean	1.75	20.95	−0.62	2.64	1.75	20.95	0.78	3.51
SD	0.68	1.42	0.64	0.89	0.68	1.42	0.53	0.77
25%	1.25	19.90	−0.94	2.05	1.25	19.91	0.38	3.00
50%	1.57	20.89	−0.47	2.70	1.57	20.88	0.70	3.54
75%	2.06	21.88	−0.15	3.29	2.06	21.88	1.09	4.04

Matched pairs (quintiles 5-1)								
	Low AT				High AT			
Group Variable	Impact	Market cap.	AT PCF_1	*lprice*	Impact	Market cap.	AT PCF_1	*lprice*
Mean	1.75	20.54	−1.28	2.02	1.75	20.55	1.26	3.57
SD	0.68	1.34	0.55	0.80	0.69	1.33	0.42	0.78
25%	1.26	19.54	−1.58	1.48	1.25	19.54	0.95	3.03
50%	1.57	20.41	−1.16	2.06	1.57	20.42	1.19	3.57
75%	2.04	21.45	−0.87	2.58	2.04	21.45	1.50	4.11

B.1. Matched pairs comparison tests (means and medians)

Matched pairs (halves 2-1)		High AT	Low AT	High-Low AT test
$jump^{(21)}$	Mean	0.50	0.44	0.07*** (6.07)
	Median	0.49	0.42	6.90*** (6.62)
Matched pairs (quintiles 5-1)		**High AT**	**Low AT**	**High-Low AT test**
$jump^{(21)}$	Mean	0.49	0.39	0.10*** (5.06)
	Median	0.48	0.37	6.14*** (5.80)

A.2. Matched sample properties

Matched pairs (halves 2-1)								
	Low lagged log stock price				High lagged log stock price			
Group Variable	Impact	Market cap.	AT PC_1	*lprice*	Impact	Market cap.	AT PC_1	*lprice*
Mean	2.08	20.83	−0.43	2.55	2.07	21.02	0.69	3.72
SD	1.06	1.01	0.85	0.55	1.05	0.95	0.69	0.45
25%	1.33	20.17	−0.93	2.27	1.32	20.47	0.23	3.39
50%	1.77	20.66	−0.40	2.68	1.74	20.97	0.69	3.65
75%	2.47	21.34	0.13	2.94	2.46	21.45	1.16	3.98

Matched pairs (quintiles 5-1)								
	Low lagged log stock price				High lagged log stock price			
Group Variable	Impact	Market cap.	AT PC_1	*lprice*	Impact	Market cap.	AT PC_1	*lprice*
Mean	2.13	20.02	−1.08	1.63	2.03	21.44	0.95	4.29
SD	1.12	0.85	0.80	0.46	1.09	0.67	0.61	0.35
25%	1.34	19.47	−1.59	1.38	1.29	20.99	0.56	4.09
50%	1.81	19.84	−1.04	1.73	1.70	21.53	0.96	4.22
75%	2.56	20.35	−0.56	1.95	2.36	21.91	1.35	4.42

(continued)

Table B1
Continued

B.2. Matched pairs comparison tests (means and medians)

Matched pairs (halves 2-1)		High *lprice*	Low *lprice*	High-Low *lprice* Test
$jump^{(21)}$	Mean	0.51	0.48	0.03*** (2.73)
	Median	0.50	0.46	3.24*** (3.65)
Matched pairs (quintiles 5-1)		**High *lprice***	**Low *lprice***	**High-Low *lprice* Test**
$jump^{(21)}$	Mean	0.52	0.45	0.07*** (3.88)
	Median	0.51	0.43	4.21*** (3.65)

This table presents results from four matched pairs comparisons. For each calendar quarter, I generate matched pairs of observations by minimizing the Mahalanobis distance between lagged market capitalization and cumulative announcement impact, $\left|CAR_{it}^{(T-21,T+2)}/\hat{\sigma}_{it}\right|$. In the first set of matched pairs analyses, I construct matches between "high AT" and "low AT," groups, where high (low) AT observations feature above-median (below-median) factor loadings on the first principal component factor of the AT measures. Of matched pairs, I retain the 300 pairs in each quarter with the smallest distances. The bottom subpanel restricts the matching procedure to all observations in quintiles five ("very high AT") and one ("very low AT") and retains the 100 observations in each group with the minimum matching distances. For both matching groups, I consider only observations with cumulative announcement impacts exceeding the cutoff value established in the main text.

Panel A.1 reports summary statistics for matching variables and the algorithmic trading proxy for each matched pair group. Panel B.1 presents pairwise comparisons of means and medians between price jump ratios for high and low AT proxy groups. "High-Low AT" estimates differences in means by regressing pairwise differences on a constant and differences in distributions using the Wilcoxon signed-rank test (bootstrapped with 1,000 replications). Panels A.2 and B.2 repeat these analysis using matches between high lagged price and low lagged price groups. Standard errors are clustered by match index, security, and month and are reported in parentheses.

lagged stock price now acting as an exogenous treatment. High lagged stock price securities in the top-and-bottom-half matches have bigger price jump ratios holding fixed for market capitalization and cumulative announcement impact, and much higher lagged stock price securities in the quintile matches deliver even larger differences in information acquisition.

References

Abadie, A., and J. Spiess. 2016. Robust post-matching inference. Working Paper.

Alford, A., J. Jones, R. Leftwich, and M. Zmijewski. 1993. The relative informativeness of accounting disclosures in different countries. *Journal of Accounting Research* 31:183–223.

Angel, J. J. 1997. Tick size, share prices, and stock splits. *Journal of Finance* 52:655–81.

Asquith, P., P. Healy, and K. Palepu. 1989. Earnings and stock splits. *Accounting Review* 64:387–403.

Back, K. 1992. Insider trading in continuous time. *Review of Financial Studies* 5:387–409.

Bai, J., T. Philippon, and A. Savov. 2016. Have financial markets become more informative? *Journal of Financial Economics* 122:625–54.

Baker, M., R. Greenwood, and J. Wurgler. 2009. Catering through nominal share prices. *Journal of Finance* 64:2559–90.

Baker, M., J. C. Stein, and J. Wurgler. 2003. When does the market matter? Stock prices and the investment of equity-dependent firms. *Quarterly Journal of Economics* 118:969–1005.

Baldauf, M., and J. Mollner. 2017. High-frequency trading and market performance. Working Paper.

Ball, R., and P. Brown. 1968. An empirical evaluation of accounting income numbers. *Journal of Accounting Research* 6:159–78.

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

Baumol, W. J. 1965. *The stock market and economic efficiency*. New York, NY: Fordham University Press.

Brennan, M. J., and T. E. Copeland. 1988. Stock splits, stock prices, and transaction costs. *Journal of Financial Economics* 22:83–101.

Brogaard, J., T. Hendershott, and R. Riordan. 2014. High-frequency trading and price discovery. *Review of Financial Studies* 27:2267–306.

Brunnermeier, M. K. 2005. Information leakage and market efficiency. *Review of Financial Studies* 18:417–57.

Cameron, A. C., J. B. Gelbach, and D. L. Miller. 2011. Robust inference with multiway clustering. *Journal of Business & Economic Statistics* 29:238–49.

Carrion, A. 2013. Very fast money: High-frequency trading on the NASDAQ. *Journal of Financial Markets* 16:680–711.

Chakrabarty, B., P. C. Moulton, and X. F. Wang. 2017. Attention effects in a high-frequency world. Working Paper.

Chen, Q., I. Goldstein, and W. Jiang. 2007. Price informativeness and investment sensitivity to stock price. *Review of Financial Studies* 20:619–50.

Collin-Dufresne, P., and V. Fos. 2015. Do prices reveal the presence of informed trading? *Journal of Finance* 70:1555–82.

Conroy, R. M., R. S. Harris, and B. A. Benet. 1990. The effects of stock splits on bid-ask spreads. *Journal of Finance* 45:1285–95.

Dow, J., and G. Gorton. 1997. Stock market efficiency and economic efficiency: Is there a connection? *Journal of Finance* 52:1087–129.

Durnev, A., R. Morck, B. Yeung, and P. Zarowin. 2003. Does greater firm-specific return variation mean more or less informed stock pricing? *Journal of Accounting Research* 41:797–836.

Easley, D., M. O'Hara, and G. Saar. 2001. How stock splits affect trading: A microstructure approach. *Journal of Financial and Quantitative Analysis* 36:25–51.

Fama, E. F. 1970. Efficient capital markets: A review of theory and empirical work. *Journal of Finance* 25:383–417.

Fama, E. F., L. Fisher, M. C. Jensen, and R. Roll. 1969. The adjustment of stock prices to new information. *International Economic Review* 10:1–21.

Fama, E. F., and K. R. French. 1992. The cross-section of expected stock returns. *Journal of Finance* 47:427–65.

Foster, F. D., and S. Viswanathan. 1993. The effect of public information and competition on trading volume and price volatility. *Review of Financial Studies* 6:23–56.

Foucault, T., J. Hombert, and I. Roşu. 2016. News trading and speed. *Journal of Finance* 71:335–82.

Foucault, T., A. Röell, and P. Sandås. 2003. Market making with costly monitoring: An aalysis of the SOES Controversy. *Review of Financial Studies* 16:345–84.

Freeman, R. N. 1987. The association between accounting earnings and security returns for large and small firms. *Journal of Accounting and Economics* 9:195–228.

Gerig, A. 2015. High-frequency trading synchronizes prices in financial markets. Working Paper.

Grossman, S. J., and J. E. Stiglitz. 1980. On the impossibility of informationally efficient markets. *American Economic Review* 70:393–408.

Han, J., M. Khapko, and A. S. Kyle. 2014. Liquidity with high-frequency market making. Research Paper, Swedish House of Finance.

Harris, L. 2013. What to do about high-frequency trading. *Financial Analysts Journal* 69:6–9.

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

———. 1991b. The summary informativeness of stock trades: An econometric analysis. *Review of Financial Studies* 4:571–95.

———. 1993. Assessing the quality of a security market: A new approach to transaction-cost measurement. *Review of Financial Studies* 6:191–212.

Heflin, F., K. R. Subramanyam, and Y. Zhang. 2003. Regulation FD and the financial information environment: Early evidence. *Accounting Review* 78:1–37.

Hendershott, T., C. M. Jones, and A. J. Menkveld. 2011. Does algorithmic trading improve liquidity? *Journal of Finance* 66:1–33.

Hirschey, N. 2017. Do high-frequency traders anticipate buying and selling pressure? Working Paper.

Holden, C. W., and A. Subrahmanyam. 1992. Long-lived private information and imperfect competition. *Journal of Finance* 47:247–70.

Kirilenko, A. A., and A. W. Lo. 2013. Moore's law versus Murphy's law: Algorithmic trading and its discontents. *Journal of Economic Perspectives* 27:51–72.

Kleibergen, F., and R. Paap. 2006. Generalized reduced rank tests using the singular value decomposition. *Journal of Econometrics* 133:97–126.

Kyle, A. S. 1985. Continuous auctions and insider trading. *Econometrica* 53:1315–35.

Lakonishok, J., and B. Lev. 1987. Stock splits and stock dividends: Why, who, and when. *Journal of Finance* 42:913–32.

Lewis, M. 2014. *Flash boys: A Wall Street revolt*. New York, NY: W. W. Norton & Company.

McNichols, M. F. 1984. *The anticipation of earnings in securities markets*. PhD Thesis, University of California, Los Angeles.

Merton, R. C. 1974. On the pricing of corporate debt: The risk structure of interest rates. *Journal of Finance* 29:449–70.

Meulbroek, L. K. 1992. An empirical analysis of illegal insider trading. *Journal of Finance* 47:1661–99.

Morck, R., B. Yeung, and W. Yu. 2000. The information content of stock markets: Why do emerging markets have synchronous stock price movements? *Journal of Financial Economics* 58:215–60.

Morse, D. 1981. Price and trading volume reaction surrounding earnings announcements: A closer examination. *Journal of Accounting Research* 19:374–83.

O'Hara, M. 2003. Presidential address: Liquidity and price discovery. *Journal of Finance* 58:1335–54.

Roll, R. 1988. R^2. *Journal of Finance* 43:541–66.

Rosner, B., R. J. Glynn, and M.-L. T. Lee. 2006. The Wilcoxon signed rank test for paired comparisons of clustered data. *Biometrics* 62:185–92.

Rozeff, M. S. 1998. Stock splits: Evidence from mutual funds. *Journal of Finance* 53:335–49.

Stiglitz, J. 2014. Tapping the brakes: Are less active markets safer and better for the economy? Working Paper.

U.S. Securities and Exchange Commission. 2010. Concept release on equity market structure. Technical Report.

U.S. Securities and Exchange Commission. 2014. Equity market structure literature review, part II: High frequency trading. Technical Report, Securities and Exchange Commission.

Weld, W. C., R. Michaely, R. H. Thaler, and S. Benartzi. 2009. The nominal share price puzzle. *Journal of Economic Perspectives* 23:121–42.

Yang, L., and H. Zhu. 2017. Back-running: Seeking and hiding fundamental information in order flows. Working Paper.

Yao, C., and M. Ye. 2017. Why trading speed matters: A tale of queue rationing under price controls. Working Paper.

Zhang, S. 2017. Need for speed: An empirical analysis of hard and soft information in a high frequency world. Working Paper.