

Does Algorithmic Trading Improve Liquidity?

TERRENCE HENDERSHOTT, CHARLES M. JONES, and ALBERT J. MENKVELD*

ABSTRACT

Algorithmic trading (AT) has increased sharply over the past decade. Does it improve market quality, and should it be encouraged? We provide the first analysis of this question. The New York Stock Exchange automated quote dissemination in 2003, and we use this change in market structure that increases AT as an exogenous instrument to measure the causal effect of AT on liquidity. For large stocks in particular, AT narrows spreads, reduces adverse selection, and reduces trade-related price discovery. The findings indicate that AT improves liquidity and enhances the informativeness of quotes.

TECHNOLOGICAL CHANGE HAS REVOLUTIONIZED the way financial assets are traded. Every step of the trading process, from order entry to trading venue to back office, is now highly automated, dramatically reducing the costs incurred by intermediaries. By reducing the frictions and costs of trading, technology has the potential to enable more efficient risk sharing, facilitate hedging, improve liquidity, and make prices more efficient. This could ultimately reduce firms' cost of capital.

Algorithmic trading (AT) is a dramatic example of this far-reaching technological change. Many market participants now employ AT, commonly defined as the use of computer algorithms to automatically make certain trading decisions, submit orders, and manage those orders after submission. From a starting point near zero in the mid-1990s, AT is thought to be responsible for as much as 73 percent of trading volume in the United States in 2009.¹

There are many different algorithms, used by many different types of market participants. Some hedge funds and broker-dealers supply liquidity

*Hendershott is at Haas School of Business, University of California Berkeley. Jones is at Columbia Business School. Menkveld is at VU University Amsterdam. We thank Mark van Achter, Hank Bessembinder, Bruno Biais, Alex Boulatov, Thierry Foucault, Maureen O'Hara, Sébastien Pouget, Patrik Sandas, Kumar Venkataraman, the NASDAQ Economic Advisory Board, and seminar participants at the University of Amsterdam, Babson College, Bank of Canada, CFTC, HEC Paris, IDEI Toulouse, Southern Methodist University, University of Miami, the 2007 MTS Conference, NYSE, the 2008 NYU-Courant algorithmic trading conference, University of Utah, the 2008 Western Finance Association meetings, and Yale University. We thank the NYSE for providing system order data. Hendershott gratefully acknowledges support from the National Science Foundation, the Net Institute, the Ewing Marion Kauffman Foundation, and the Lester Center for Entrepreneurship and Innovation at the Haas School at UC Berkeley. Menkveld gratefully acknowledges the College van Bestuur of VU University Amsterdam for a VU talent grant.

¹ See "SEC runs eye over high-speed trading," *Financial Times*, July 29, 2009. The 73% is an estimate for high-frequency trading, which, as discussed later, is a subset of AT.

using algorithms, competing with designated market-makers and other liquidity suppliers (e.g., Jovanovic and Menkveld (2010)). For assets that trade on multiple venues, liquidity demanders often use smart order routers to determine where to send an order (e.g., Foucault and Menkveld (2008)). Statistical arbitrage funds use computers to quickly process large amounts of information contained in the order flow and price moves in various securities, trading at high frequency based on patterns in the data. Last but not least, algorithms are used by institutional investors to trade large quantities of stock gradually over time.

Before AT took hold, a pension fund manager who wanted to buy 30,000 shares of IBM might hire a broker-dealer to search for a counterparty to execute the entire quantity at once in a block trade. Alternatively, that institutional investor might have hired a New York Stock Exchange (NYSE) floor broker to go stand at the IBM post and quietly “work” the order, using his judgment and discretion to buy a little bit here and there over the course of the trading day to keep from driving the IBM share price up too far. As trading became more electronic, it became easier and cheaper to replicate that floor trader with a computer program doing AT (see Hendershott and Moulton (2009) for evidence on the decline in NYSE floor broker activity).

Now virtually every large broker-dealer offers a suite of algorithms to its institutional customers to help them execute orders in a single stock, in pairs of stocks, or in baskets of stocks. Algorithms typically determine the timing, price, quantity, and routing of orders, dynamically monitoring market conditions across different securities and trading venues, reducing market impact by optimally and sometimes randomly breaking large orders into smaller pieces, and closely tracking benchmarks such as the volume-weighted average price (VWAP) over the execution interval. As they pursue a desired position, these algorithms often use a mix of active and passive strategies, employing both limit orders and marketable orders. Thus, at times they function as liquidity demanders, and at times they supply liquidity.

Some observers use the term AT to refer only to the gradual accumulation or disposition of shares by institutions (e.g., Domowitz and Yegerman (2005)). In this paper, we take a broader view of AT, including in our definition all participants who use algorithms to submit and cancel orders. We note that algorithms are also used by quantitative fund managers and others to determine portfolio holdings and formulate trading strategies, but we focus on the execution aspect of algorithms because our data reflect counts of actual orders submitted and cancelled.

The rise of AT has obvious direct impacts. For example, the intense activity generated by algorithms threatens to overwhelm exchanges and market data providers,² forcing significant upgrades to their infrastructures. But researchers, regulators, and policymakers should be keenly interested in the broader implications of this sea change in trading. Overall, does AT have

² See “Dodgy tickers-stock exchanges,” *Economist*, March 10, 2007.

salutary effects on market quality, and should it be encouraged? We provide the first empirical analysis of this question.

As AT has grown rapidly since the mid-1990s, liquidity in world equity markets has also dramatically improved. Based on these two coincident trends, it is tempting to conclude that AT is at least partially responsible for the increase in liquidity. But it is not at all obvious a priori that AT and liquidity should be positively related. If algorithms are cheaper and/or better at supplying liquidity, then AT may result in more competition in liquidity provision, thereby lowering the cost of immediacy. However, the effects could go the other way if algorithms are used mainly to demand liquidity. Limit order submitters grant a trading option to others, and if algorithms make liquidity demanders better able to identify and pick off an in-the-money trading option, then the cost of providing the trading option increases, in which case spreads must widen to compensate. In fact, AT could actually lead to an unproductive arms race, where liquidity suppliers and liquidity demanders both invest in better algorithms to try to take advantage of the other side, with measured liquidity the unintended victim.

In this paper, we investigate the empirical relationship between AT and liquidity. We use a normalized measure of NYSE electronic message traffic as a proxy for AT. This message traffic includes electronic order submissions, cancellations, and trade reports. Because we normalize by trading volume, variation in our AT measure is driven mainly by variation in limit order submissions and cancellations. This means that, for the most part, our measure is picking up variation in algorithmic liquidity supply. This liquidity supply likely comes both from proprietary traders that are making markets algorithmically and from buy-side institutions that are submitting limit orders as part of “slice and dice” algorithms.

We first examine the growth of AT and the improvements in liquidity over a 5-year period. As AT grows, liquidity improves. However, while AT and liquidity move in the same direction, it is certainly possible that the relationship is not causal. To establish causality we study an important exogenous event that increases the amount of AT in some stocks but not others. In particular, we use the start of autoquoting on the NYSE as an instrument for AT. Previously, specialists were responsible for manually disseminating the inside quote. This was replaced in early 2003 by a new automated quote whenever there was a change to the NYSE limit order book. This market structure provides quicker feedback to algorithms and results in more electronic message traffic. Because the change was phased in for different stocks at different times, we can take advantage of this nonsynchronicity to cleanly identify causal effects.

We find that AT does in fact improve liquidity for large-cap stocks. Quoted and effective spreads narrow under autoquote. The narrower spreads are a result of a sharp decline in adverse selection, or equivalently a decrease in the amount of price discovery associated with trades. AT increases the amount of price discovery that occurs without trading, implying that quotes become more informative. There are no significant effects for smaller-cap stocks, but our instrument is weaker there, so the problem may be a lack of statistical power.

Surprisingly, we find that AT increases realized spreads and other measures of liquidity supplier revenues. This is surprising because we initially expected that if AT improved liquidity, the mechanism would be competition between liquidity providers. However, the evidence clearly indicates that liquidity suppliers are capturing some of the surplus for themselves. The most natural explanation is that, at least during the introduction of autoquote, algorithms had market power. Over a longer time period liquidity supplier revenues decline, suggesting that any market power was temporary, perhaps because new algorithms require considerable investment and time to build.

The paper proceeds as follows. Section I discusses related literature. Section II describes our measures of liquidity and AT and discusses the need for an instrumental variables approach. Section III provides a summary of the NYSE's staggered introduction of autoquote in 2003. Section IV examines the impact of AT on liquidity. Section V explores the sources of the liquidity improvement. Section VI studies AT's relation to price discovery via trading and quote updating. Section VII discusses and interprets the results, and Section VIII concludes.

I. Related Literature

Only a few papers address AT directly. For example, Engle et al. (2007) use execution data from Morgan Stanley algorithms to study the effects on trading costs of changing algorithm aggressiveness. Domowitz and Yegerman (2005) study execution costs for a set of buy-side institutions, comparing results from different algorithm providers. Chaboud et al. (2009) study AT in the foreign exchange market and focus on its relation to volatility, while Hendershott and Riordan (2009) measure the contributions of AT to price discovery on the Deutsche Boerse.

Several strands of literature touch related topics. Most models take the traditional view that one set of traders provides liquidity via quotes or limit orders and another set of traders initiates a trade to take that liquidity—for either informational or liquidity/hedging reasons. Many assume that liquidity suppliers are perfectly competitive, for example, Glosten (1994). Glosten (1989) models a monopolistic liquidity supplier, while Biais et al. (2000) model competing liquidity suppliers and find that their rents decline as the number increases. Our initial expectation is that AT facilitates the entry of additional liquidity suppliers, increasing competition.

The development and adoption of AT also involves strategic considerations. While algorithms have low marginal costs, there may be substantial development costs, and it may be costly to optimize the algorithms' parameters for each security. The need to recover these costs should lead to the adoption of AT at times and in securities where the returns to adoption are highest (see Reinganum (1989) for a review of innovation and technology adoption).

As we discuss briefly in the introduction, liquidity supply involves posting firm commitments to trade. These standing orders provide free trading options to other traders. Using standard option pricing techniques, Copeland

and Galai (1983) value the cost of the option granted by liquidity suppliers. Foucault et al. (2003) study the equilibrium level of effort that liquidity suppliers should expend in monitoring the market to reduce this option's cost. Black (1995) proposes a new limit order type that is indexed to the overall market to minimize picking-off risk. Algorithms can efficiently implement this kind of monitoring and adjustment of limit orders.³ If AT reduces the cost of the free trading option implicit in limit orders, then measures of adverse selection depend on AT. If some users of AT are better at avoiding being picked off, they can impose adverse selection costs on other liquidity suppliers as in Rock (1990) and even drive out other liquidity suppliers.

AT may also be used by traders who are trying to passively accumulate or liquidate a large position.⁴ There are optimal dynamic execution strategies for such traders. For example, Bertsimas and Lo (1998) find that, in the presence of temporary price impacts and a trade completion deadline, orders are optimally broken into pieces so as to minimize cost.⁵ Many brokers incorporate such considerations into the AT products that they sell to their clients. In addition, algorithms monitor the state of the limit order book to dynamically adjust their trading strategies, for example, when to take and offer liquidity (Foucault et al. (2010)).

II. Data

We start by characterizing the time-series evolution of AT and liquidity for a sample of NYSE stocks over the 5 years from February 2001 through December 2005. We limit attention to the post-decimalization regime because the change to a one-penny minimum tick was a structural break that substantially altered the entire trading landscape, including liquidity metrics and order submission strategies. We end in 2005 because substantial NYSE market structure changes occur shortly thereafter.

We start with a sample of all NYSE common stocks that can be matched in both the Trades and Quotes (TAQ) and Center for Research in Security Prices CRSP databases. To maintain a balanced panel, we retain those stocks that are present throughout the whole sample period. Stocks with an average share

³ Rosu (2009) develops a model that implicitly recognizes these technological advances and simply assumes that limit orders can be constantly adjusted. Consistent with AT, Hasbrouck and Saar (2009) find that by 2004 a large number of limit orders are cancelled within two seconds on the INET trading platform.

⁴ Keim and Madhavan (1995) and Chan and Lakonishok (1995) study institutional orders that are broken up.

⁵ Almgren and Chriss (2000) extend this optimization problem by considering the risk that arises from breaking up orders and slowly executing them. Obizhaeva and Wang (2005) optimize assuming that liquidity does not replenish immediately after it is taken but only gradually over time. For each component of a larger transaction, a trader or algorithm must choose the type and aggressiveness of the order. Cohen et al. (1981) and Harris (1998) focus on the simplest static choice: market order versus limit order. However, a limit price must be chosen, and the problem is dynamic; Goettler et al. (2009) model both aspects.

price of less than \$5 are removed from the sample, as are stocks with an average share price of more than \$1,000. The resulting sample consists of monthly observations for 943 common stocks. The balanced panel eliminates compositional changes in the sample over time, which could induce some survivorship effects if disappearing stocks are less liquid. This could overstate time-series improvements in liquidity, although the same liquidity patterns are present without a survivorship requirement (see Comerton-Forde et al. (2010)).

Stocks are sorted into quintiles based on market capitalization. Quintile 1 refers to large-cap stocks and quintile 5 corresponds to small-cap stocks. All variables used in the analysis are 99.9 % winsorized: values smaller than the 0.05% quantile are set equal to that quantile, and values larger than the 99.95% quantile are set equal to that quantile.

A. Proxies for AT

We cannot directly observe whether a particular order is generated by a computer algorithm. For cost and speed reasons, most algorithms do not rely on human intermediaries but instead generate orders that are sent electronically to a trading venue. Thus, we use the rate of electronic message traffic as a proxy for the amount of AT taking place.⁶ This proxy is commonly used by market participants, including consultants Aite Group and Tabb Group, as well as exchanges and other market venues.⁷

For example, in discussing market venue capacity limits following an episode of heavy trading volume in February 2007, a *Securities Industry News* report quotes NASDAQ senior vice president of transaction services, Brian Hyndman, who noted that exchanges have dealt with massive increases in message traffic over the past 5 to 6 years, coinciding with algorithmic growth:

“It used to be one-to-one,” Hyndman said. “Then you’d see a customer send ten orders that would result in only one execution. That’s because the black box would cancel a lot of the orders. We’ve seen that rise from 20- to 30- to 50-to-one. The amount of orders in the marketplace increased exponentially.”⁸

In the case of the NYSE, electronic message traffic includes order submissions, cancellations, and trade reports that are handled by the NYSE’s SuperDOT system and captured in the NYSE’s System Order Data (SOD) database. The electronic message traffic measure for the NYSE excludes all specialist quoting, as well as all orders that are sent manually to the floor and are handled by a floor broker.

⁶ See Biais and Weill (2009) for theoretical evidence on how AT relates to message traffic.

⁷ See, e.g., Jonathan Keehner, “Massive surge in quotes, electronic messages may paralyze US market,” <http://www.livemint.com/2007/06/14005055/Massive-surge-in-quotes-elect.html>, June 14, 2007.

⁸ See Shane Kite, “Reacting to market break, NYSE and NASDAQ act on capacity,” *Securities Industry News*, March 12, 2007.

As suggested by the quote above, an important issue is whether and how to normalize the message traffic numbers. The top half of Figure 1 shows the evolution of message traffic over time. We focus on the largest-cap quintile of stocks, as they constitute the vast bulk of stock market capitalization and trading activity. Immediately after decimalization at the start of 2001, the average large-cap stock sees about 35 messages per minute during the trading day. There are a few bumps along the way, but by the end of 2005 there are an average of about 250 messages per minute (more than 4 messages per second) for these same large-cap stocks. We could, of course, simply use the raw message traffic numbers, but there has been an increase in trading volume over the same interval, and without normalization a raw message traffic measure may just capture the increase in trading rather than the change in the nature of trading. Therefore, for each stock each month we calculate our AT proxy, $algo_trad_{it}$, as the number of electronic messages per \$100 of trading volume.⁹ The normalized measure still rises rapidly over the 5-year sample period, while measures of market liquidity such as proportional spreads have declined sharply but appear to asymptote near the end of the sample period (see, e.g., the average quoted spreads in the top half of Figure 2 below), which occurs as more and more stocks are quoted with the minimum spread of \$0.01.

The time-series evolution of $algo_trad_{it}$ is displayed in the bottom half of Figure 1. For the largest-cap quintile, there is about \$7,000 of trading volume per electronic message at the beginning of the sample in 2001, decreasing dramatically to about \$1,100 of trading volume per electronic message by the end of 2005. Over time, smaller-cap stocks display similar time-series patterns.

It is worth noting that our AT proxies may also capture changes in trading strategies. For example, messages and $algo_trad_{it}$ will increase if the same market participants use algorithms but modify their trading or execution strategies so that those algorithms submit and cancel orders more often. Similarly, the measure will increase if existing algorithms are modified to “slice and dice” large orders into smaller pieces. This is useful, as we want to capture increases in the intensity of order submissions and cancellations by existing algorithmic traders, as well as the increase in the fraction of market participants employing algorithms in trading.

B. Liquidity Measures

We measure liquidity using quoted half-spreads, effective half-spreads, 5-minute and 30-minute realized spreads, and 5-minute and 30-minute price impacts, all of which are measured as share-weighted averages and expressed in basis points as a proportion of the prevailing midpoint. The effective spread is the difference between the midpoint of the bid and ask quotes and the actual

⁹ Our results are virtually the same when we normalize by the number of trades or use raw message traffic numbers (see Table IA.4 in the Internet Appendix, available online in the “Supplements and Datasets” section at <http://www.afajof.org/supplements.asp>). The results are also the same when we use the number of cancellations rather than the number of messages to construct the AT measure.

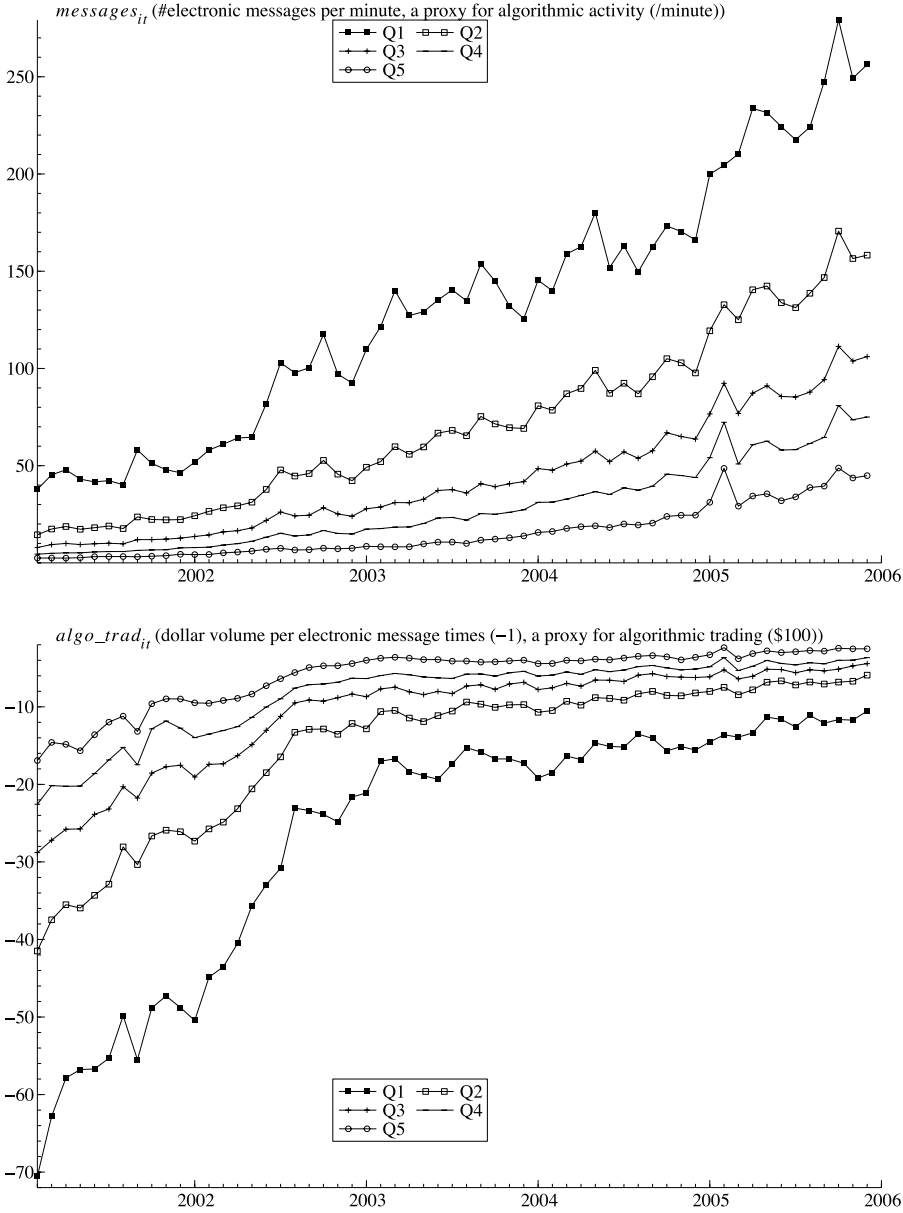


Figure 1. Algorithmic trading measures. For each market-cap quintile, where Q1 is the largest-cap quintile, these graphs depict (i) the number of (electronic) messages per minute and (ii) our proxy for algorithmic trading, which is defined as the negative of trading volume (in hundreds of dollars) divided by the number of messages.

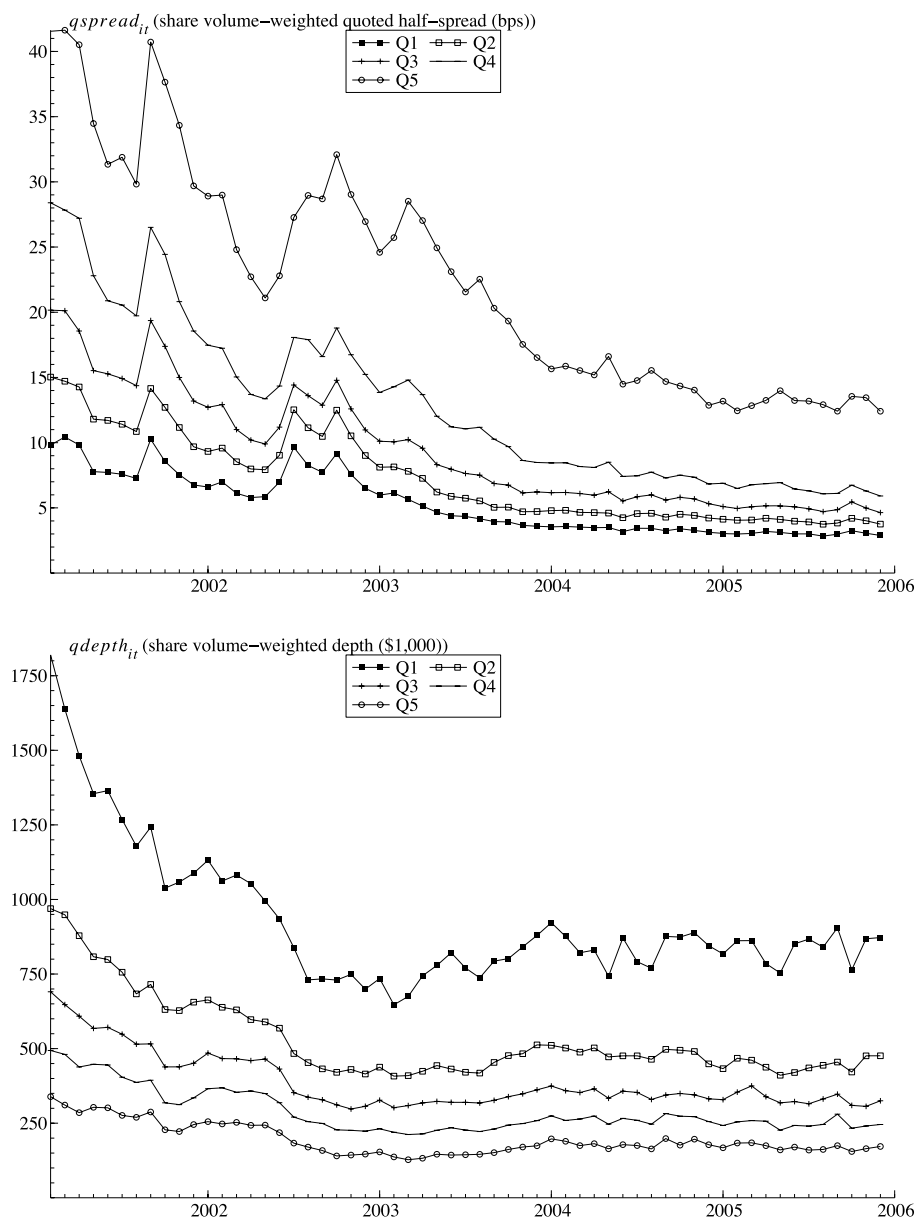


Figure 2. Liquidity measures. These graphs depict (i) quoted half-spread, (ii) quoted depth, and (iii) effective spread. All spread measures are share volume-weighted averages within-firm that are then averaged across firms within each market-cap quintile, where Q1 is the largest-cap quintile.

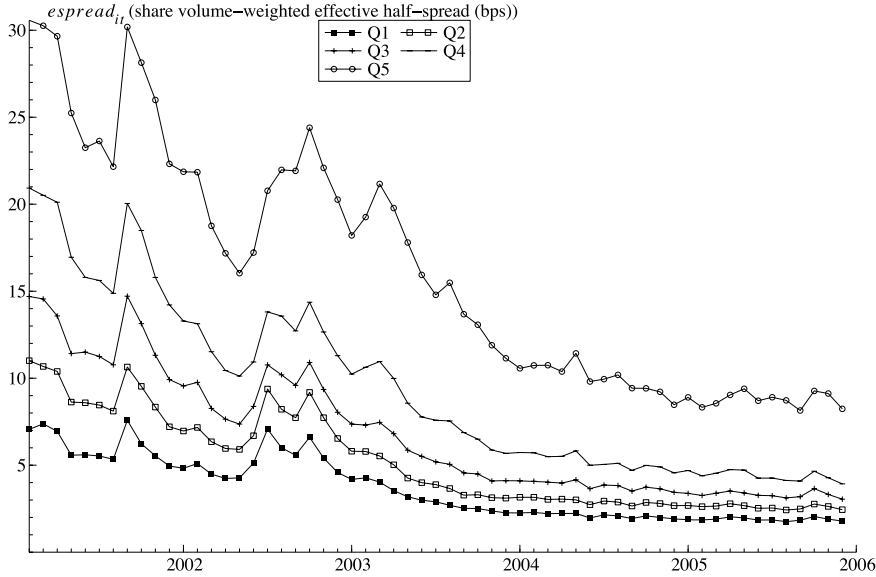


Figure 2. Continued

transaction price. The wider the effective spread, the less liquid is the stock. For the NYSE, effective spreads are more meaningful than quoted spreads because specialists and floor brokers are sometimes willing to trade at prices within the quoted bid and ask prices. For the t^{th} trade in stock j , the proportional effective half-spread, $espread_{jt}$, is defined as

$$espread_{jt} = q_{jt}(p_{jt} - m_{jt})/m_{jt}, \quad (1)$$

where q_{jt} is an indicator variable that equals +1 for buyer-initiated trades and -1 for seller-initiated trades, p_{jt} is the trade price, and m_{jt} is the quote midpoint prevailing at the time of the trade. We follow the standard trade-signing approach of Lee and Ready (1991) and use contemporaneous quotes to sign trades and calculate effective spreads (see Bessembinder (2003), for example). For each stock each day, we use all NYSE trades and quotes to calculate quoted and effective spreads for each reported transaction and calculate a share-weighted average across all trades that day. For each month we calculate the simple average across days. We also measure share-weighted quoted depth at the time of each transaction in thousands of dollars.

Figure 2 shows quite clearly that our measures of liquidity are generally improving over the sample period. Figure 1 shows that AT increases almost monotonically. The spread measures are not nearly as monotonic, with illiquidity spikes in both 2001 and 2002 that correspond to sharp stock market declines and increased volatility over the same sample period (see Figure IA.5 in the Internet Appendix). Nevertheless, one is tempted to conclude that these

two trends are related. The analysis below investigates exactly this relationship using formal econometric tools.

If spreads narrow when AT increases, it is natural to decompose the spread along the lines of Glosten (1987) to determine whether the narrower spread means less revenue for liquidity providers, smaller gross losses due to informed liquidity demanders, or both. We estimate revenue to liquidity providers using the 5-minute realized spread, which assumes the liquidity provider is able to close her position at the quote midpoint 5 minutes after the trade. The proportional realized spread for the t^{th} transaction in stock j is defined as

$$rspread_{jt} = q_{jt}(p_{jt} - m_{j,t+5min})/m_{jt}, \quad (2)$$

where p_{jt} is the trade price, q_{jt} is the buy–sell indicator (+1 for buys, –1 for sells), m_{jt} is the midpoint prevailing at the time of the t^{th} trade, and $m_{j,t+5min}$ is the quote midpoint 5 minutes after the t^{th} trade. The 30-minute realized spread is calculated analogously using the quote midpoint 30 minutes after the trade.

We measure gross losses to liquidity demanders due to adverse selection using the 5-minute price impact of a trade, $adv_selection_{jt}$, defined using the same variables as

$$adv_selection_{jt} = q_{jt}(m_{j,t+5min} - m_{jt})/m_{jt}. \quad (3)$$

The 30-minute price impact is calculated analogously. Note that there is an arithmetic identity relating the realized spread, the adverse selection (price impact), and the effective spread $espread_{jt}$

$$espread_{jt} = rspread_{jt} + adv_selection_{jt}. \quad (4)$$

Figure 3 graphs the decomposition of the two spread components. Both realized spreads, $rspread_{it}$, and price impacts, $adv_selection_{it}$, decline from 2001 to 2005. Most of the narrowed spread is due to a decline in adverse selection losses to liquidity demanders. Depending on the size quintile considered, 75% to 90% of the narrowed spread is due to a smaller price impact.

So far, the graphical evidence shows time-series associations between AT and liquidity. The natural way to formally test this association is by regressing the various liquidity measures (L_{it}) on AT (A_{it}) and variables controlling for market conditions (X_{it}):

$$L_{it} = \alpha_i + \beta A_{it} + \delta' X_{it} + \varepsilon_{it}. \quad (5)$$

The problem is that AT is an endogenous choice made by traders. A trader's decision to adopt AT could depend on many factors, including liquidity. For example, the evidence in Goldstein and Kavajecz (2004) indicates that humans are used more often when markets are illiquid and volatile. Econometrically, this means that the slope coefficient β from estimating equation (5) via OLS is not an unbiased estimate of the causal effect of a change in AT on liquidity. Unless we have a structural model, the only way to identify the causal effect is to find an instrumental variable (IV) that affects AT but is uncorrelated with

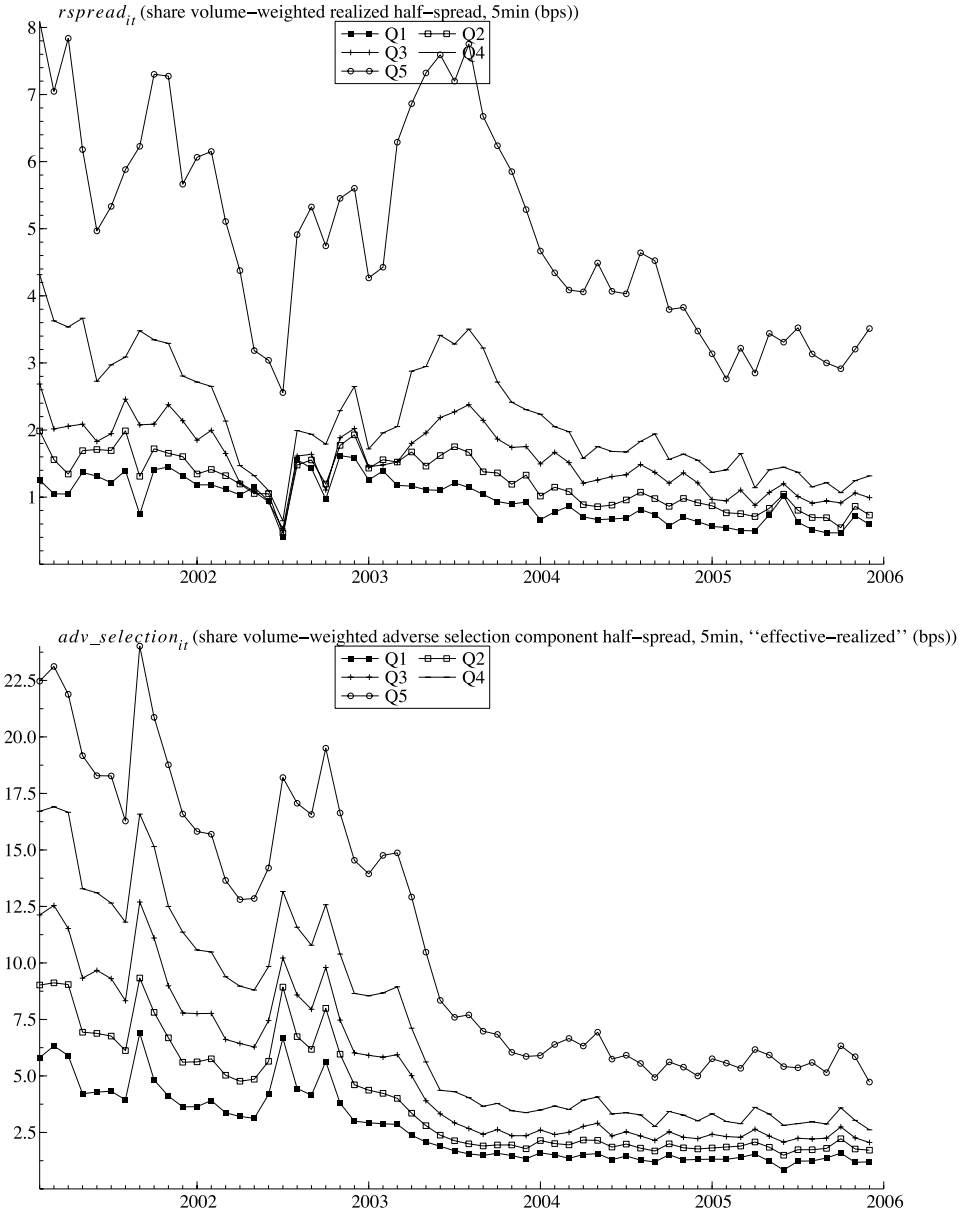


Figure 3. Spread decomposition into liquidity supplier revenues and adverse selection. These graphs depict the two components of the effective spread: (i) realized spread and (ii) the adverse selection component, also known as the (permanent) price impact. Both are based on the quote midpoint 5 minutes after the trade. Results are graphed by market-cap quintile, where Q1 is the largest-cap quintile.

ε_{it} . Standard econometrics texts, for example, Greene (2007, Ch. 12), show that under these conditions, the resulting IV estimator consistently estimates the causal effect, in this case the effect of an exogenous change in AT on liquidity. We discuss such an instrument in the next section.

III. Autoquote

In this section we provide an overview of our instrument, which is a change in NYSE market structure that causes an exogenous increase in AT.

As a result of the reduction of the minimum tick to a penny in early 2001 as part of decimalization, the depth at the inside quote shrank dramatically. In response, the NYSE proposed that a “liquidity quote” for each stock be displayed along with the best bid and offer. The NYSE liquidity quote was designed to provide a firm bid and offer for substantial size, typically at least 15,000 shares, accessible immediately.¹⁰

At the time of the liquidity quote proposal, specialists were responsible for manually disseminating the inside quote.¹¹ Clerks at the specialist posts on the floor of the exchange were typing rapidly and continuously from open to close and still were barely keeping up with order matching, trade reporting, and quote updating. In order to ease this capacity constraint and free up specialists and clerks to manage a liquidity quote, the exchange proposed that it “autoquote” the inside quote, disseminating a new quote automatically whenever there was a relevant change to the limit order book. This would happen when a better-priced order arrived, when an order at the inside was canceled, when the inside quote was traded with in whole or in part, or when the size of the inside quote changed.

Note that the specialist’s structural advantages were otherwise unaffected by autoquote. A specialist could still disseminate a manual quote at any time in order to reflect his own trading interest or that of floor traders. Specialists continued to execute most trades manually, and they could still participate in those trades subject to the unchanged NYSE rules. NYSE market share remains unchanged at about 80% around the adoption of autoquote.

Autoquote was an important innovation for algorithmic traders because an automated quote update could provide more immediate feedback about the potential terms of trade. This speedup of a few seconds would provide critical new information to algorithms, but would be unlikely to directly affect the trading behavior of slower-reacting humans. Autoquote allowed algorithmic liquidity suppliers to, say, quickly notice an abnormally wide inside quote and provide liquidity accordingly via a limit order. Algorithmic liquidity demanders could quickly access this quote via a conventional market or marketable limit order or by using the NYSE’s automated execution facility for limit orders of

¹⁰ For more details, the NYSE proposal is contained in Securities Exchange Act Release No. 47091 (December 23, 2002), 68 FR 133.

¹¹ One exception: NYSE software would automatically disseminate an updated quote after 30 seconds if the specialist had not already done so.

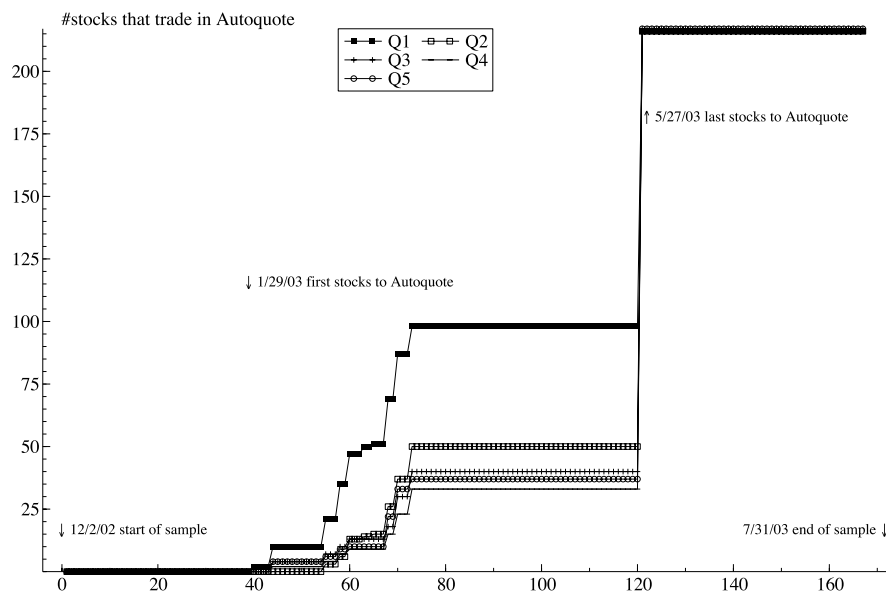


Figure 4. Autoquote introduction. This graph depicts the staggered introduction of autoquote on the NYSE. It graphs the number of stocks in each market-cap quintile that are autoquoted at a given time. Quintile 1 contains largest-cap stocks.

1,099 shares or less. In the next section, we show that autoquote is positively correlated with our AT measure, which is one of the requirements for autoquote to be a valid instrument.

The NYSE began to phase in the autoquote software on January 29, 2003, starting with six active, large-cap stocks. During the next 2 months, over 200 additional stocks were phased in at various dates, and all remaining NYSE stocks were phased in on May 27, 2003.¹² Figure 4 provides additional details on the phase-in process. The rollout order was determined in late 2002. Early stocks tended to be active large-cap stocks because the NYSE felt that these stocks would benefit most from the liquidity quote. Beyond that criterion, conversations with those involved at the NYSE indicate that early phase-in stocks were chosen mainly because the specialist assigned to that stock was receptive to new technology.

The phase-in is particularly important to our empirical design. It allows us to take out all market-wide changes in liquidity, and identify the causal effect of AT by comparing autoquoted stocks to non-autoquoted stocks using a difference-in-differences methodology. The IV methodology discussed below incorporates data before and after each NYSE stock's autoquote adoption so the estimated effect of AT on liquidity incorporates every stock's autoquote

¹² Liquidity quotes were delayed due to a property rights dispute with data vendors, so they did not become operational until June 2003, after autoquote was fully phased in. Liquidity quotes were almost never used and were formally abandoned in July 2005.

transition, whenever it occurs. Thus, even if the phase-in order is determined by other unknown criteria, our empirical methodology remains valid in most cases. For example, there is no bias if the phase-in is determined by the specialist's receptiveness to new technology, and this is correlated with the amount of AT in his stocks. There are a small number of problematic phase-in scenarios, however. We discuss these next.

For the staggered introduction of autoquote to serve as a valid instrument, it must satisfy the exclusion restriction. Specifically, a stock's move to autoquote must not be correlated with the error term in that firm's liquidity equation (equation (5)). This does not mean that the autoquote rollout must be assigned randomly. The liquidity equation includes a firm fixed effect, calendar dummies, and a set of control variables. The instrument remains valid even if the rollout schedule is related to these particular explanatory variables. For instance, if the stocks chosen for early phase-in tend to have high mean liquidity, this would be picked up by the firm fixed effect and the exclusion restriction would still hold. In fact, due to the explanatory variables, the exclusion restriction is violated only if the autoquote phase-in schedule is somehow related to contemporaneous changes in firm-specific, idiosyncratic liquidity that are not due to changes in AT.

Thus, it is quite helpful that the rollout schedule for autoquote was fixed months in advance, as it seems highly unlikely that the phase-in schedule could be correlated with idiosyncratic liquidity months into the future. The only way this might happen is if there are sufficiently persistent but temporary shocks to idiosyncratic liquidity. For example, if temporarily illiquid stocks are chosen for early phase-in, these stocks might still be illiquid when autoquote begins, and their liquidity would improve post-autoquote as they revert to mean liquidity, thereby overstating the causal effect.¹³ To investigate this possibility, we study the dynamics of liquidity using an AR(1) model of effective spreads for each firm in the sample. Table IA.3 of the Internet Appendix shows that the average AR(1) coefficient is 0.18, corresponding to a half-life of less than a day.¹⁴ We also do not find statistical support for the conjecture that stocks that migrate experience unusual liquidity just ahead of the migration. More precisely, the predicted effective spread based on all information up until the day before the introduction, including liquidity covariates, is not significantly different from its unconditional mean. All of this supports the exogeneity of our instrument.

Lastly, the exclusion restriction requires autoquote to affect liquidity only via AT. We have argued that autoquote's time scale is only relevant for algorithms and that autoquote does not directly affect liquidity via nonalgorithmic trading.¹⁵ However, we cannot test this conjecture using the available data.

¹³ Late phase-in stocks will not offset this effect. Even if late phase-in stocks are temporarily liquid when chosen, this temporary effect has more time to die out by the time autoquote is implemented for them.

¹⁴ Also because the average daily AR(1) coefficient is quite small, there is little scope for the bias that can arise in dynamic panel data models with strong persistence. See, for example, Arellano (2003, Ch. 6.2).

¹⁵ For example, autoquote could simply make the observed quotes less stale. We investigate

Thus, it is important to emphasize that our conclusions on causality rely on the intuitively appealing but ultimately untestable assumption that autoquote affects liquidity only via its effect on AT.

IV. AT's Impact on Liquidity

To study the effects of autoquote, we build a daily panel of NYSE common stocks. The sample begins on December 2, 2002, which is approximately 2 months before the autoquote phase-in begins, and it extends through July 31, 2003, about 2 months after the last batch of NYSE stocks moves to the autoquote regime. We use standard price filters: stocks with an average share price of less than \$5 or more than \$1,000 are removed. To make our various autoquote analyses comparable, we use the same sample of stocks throughout this section. The Hasbrouck (1991a,b) decomposition (discussed below in Section VI) has the most severe data requirements, so we retain all stocks that have at least 21 trades per day for each day in the 8-month sample period. This leaves 1,082 stocks in the sample. The shorter time period for the autoquote sample allows for a larger balanced panel compared to the 5-year balanced panel used to create Figures 1–3.

Next, we sort stocks into quintiles based on market capitalization. Quintile 1 refers to large-cap stocks and quintile 5 corresponds to small-cap stocks. Table I contains means by quintile and standard deviations for all of the variables used in the analysis. All variables used in the analysis are 99.9% winsorized: values smaller than the 0.05% quantile are set equal to that quantile, and values larger than the 99.95% quantile are set equal to that quantile.

Autoquote clearly leads to greater use of algorithms. Figure 1 shows that message traffic increases by about 50% in the most active quintile of stocks as autoquote is phased in. It is certainly hard to imagine that autoquote would change the behavior of humans by anything close to this magnitude. However, nowhere in the paper do we rely on this time-series increase in AT. Instead, we include stock fixed effects and time fixed effects (day dummies), so that we identify the effect of the market structure change via its staggered introduction. The presence of these two-way fixed effects means that we are comparing the changes experienced by autoquoted stocks to the changes in not-yet-autoquoted control stocks.

We begin by estimating the following first-stage regression:

$$M_{it} = \alpha_i + \gamma_t + \beta Q_{it} + \varepsilon_{it}, \quad (6)$$

where M_{it} is the relevant dependent variable, for example, the number of electronic messages per minute, Q_{it} , is the autoquote dummy set to zero before the

this possibility in Section I of the Internet Appendix and find that this mechanical explanation is unlikely to account for our results.

Table II
Autoquote Impact on Messages, Algorithmic Trading Proxy, and Covariates

This table shows the impact of autoquote on other variables, and the second column can be interpreted as the first-stage instrumental variables regression when $algo_trad_{it}$ is the dependent variable. The analysis is based on daily observations from December 2002 through July 2003, which covers the phase-in of autoquote. We regress each of the variables used in the IV analysis on the autoquote dummy ($auto_quote_{it}$) using the following specification:

$$M_{it} = \alpha_i + \gamma_t + \beta Q_{it} + \varepsilon_{it},$$

where M_{it} is the relevant dependent variable, for example, the number of electronic messages per minute, Q_{it} is the autoquote dummy set to zero before the autoquote introduction and one afterward, α_i is a stock fixed effect, and γ_t is a day dummy. There are also separate regressions for each size quintile, and statistical significance is based on standard errors that are robust to general cross-section and time-series heteroskedasticity and within-group autocorrelation (see Arellano and Bond (1991)). Table I provides other variable definitions. */** denote significance at the 95%/99% level.

	$messages_{it}$	$algo_trad_{it}$	$share_turnover_{it}$	$volatility_{it}$	$1/price_{it}$	$ln_market_cap_{it}$
Slope coefficient from regression of column variable on $auto_quote_{it}$						
All	2.135**	0.291**	-0.016**	0.001	0.000**	-0.003**
Q1 (largest cap)	6.286**	0.414**	0.016*	-0.003	0.000**	-0.005**
Q2	0.880**	0.396**	-0.029*	0.007	-0.000**	0.003**
Q3	0.944**	0.292**	0.002	-0.001	0.000	-0.004**
Q4	0.223**	0.029	-0.006	-0.003	-0.000	0.002
Q5 (smallest cap)	-0.031	0.219**	-0.080**	0.003	0.002**	-0.013**

autoquote introduction and one afterward; α_i is a stock fixed effect; and γ_t is a day dummy. There are also separate regressions for each size quintile.

Table II reports the slope coefficients for this specification. When the dependent variable M_{it} is the number of electronic messages per minute for stock i on day t , we find a significant positive relationship. The coefficient of 2.135 on autoquote implies that autoquote increases message traffic by an average of two messages per minute. In December 2002, the month before autoquote begins its rollout, our sample stocks average 36 messages per minute, so autoquote causes a 6% increase in message traffic on average. Associations are stronger for large-cap stocks, consistent with the conventional wisdom that AT was more prevalent at the time for active, liquid stocks.

Table II also shows that there is a significant positive relationship between the autoquote dummy and our preferred measure of AT, $algo_trad_{it}$, which is the negative of dollar volume in hundreds per electronic message. Thus, it is clear that autoquote leads to more AT in all but the smallest quintiles.¹⁶ There

¹⁶ In the IV regressions in Tables III to V we report F statistics that reject the null that the instruments do not enter the first-stage regression. Bound et al. (1995, p. 446) mention that “ F statistics close to 1 should be cause for concern.” Our F statistics range from 5.88 to 7.32, and we are thus not afflicted with a weak instruments problem.

is no consistent relationship between autoquote and any other variable, such as turnover, volatility, and share price.

Our principal goal is to understand the effects of algorithmic liquidity supply on market quality, and so we use the autoquote dummy as an instrument for AT in a panel regression framework. Our main instrumental variables specification is a daily panel of 1,082 NYSE stocks over the 8-month sample period spanning the staggered implementation of autoquote. The dependent variable is one of five liquidity measures: quoted half-spreads, effective half-spreads, realized spreads, or price impacts, all of which are share volume-weighted and measured in basis points, or the quoted depth in thousands of dollars. We include fixed effects for each stock as well as time dummies, and we include share turnover, volatility based on the daily price range (high minus low, see Parkinson (1980)), the inverse of share price, and the log of market cap as control variables. Results are virtually identical if we exclude these control variables. Based on anecdotal information that AT was relatively more important for active large-cap stocks during this time period, we estimate this specification separately for each market-cap quintile.

The estimated equation is

$$L_{it} = \alpha_i + \gamma_t + \beta A_{it} + \delta' X_{it} + \varepsilon_{it}, \quad (7)$$

where L_{it} is a spread measure for stock i on day t , A_{it} is the AT measure $algo_trad_{it}$, and X_{it} is a vector of control variables, including share turnover, volatility, the inverse of share price, and log market cap. We always include fixed effects and time dummies. The set of instruments consists of all explanatory variables, except that we replace $algo_trad_{it}$ with $auto_quote_{it}$. Inference is based on standard errors that are robust to general cross-section and time-series heteroskedasticity and within-group autocorrelation (see Arellano and Bond (1991)). Section II of the Internet Appendix shows that the IV regression is unaffected by the use of a proxy for AT, as long as the noise in the proxy is uncorrelated with the autoquote instrument.

The results are reported in Panel A of Table III and the most reliable effects are in larger stocks. For large-cap stocks (quintiles 1 and 2), the autoquote instrument shows that an increase in algorithmic liquidity supply narrows both the quoted and effective spread. To interpret the estimated coefficient on the AT variable, recall that the AT measure $algo_trad_{it}$ is the negative of dollar volume per electronic message, measured in hundreds of dollars, while the spread is measured in basis points. Thus, the IV estimate of -0.53 on the AT variable for quintile 1 means that a unit increase in AT, for example, from the sample mean of \$1,844 to \$1,744 of volume per message, implies that quoted spreads narrow by 0.53 basis points.¹⁷ The average within-stock standard deviation for $algo_trad_{it}$ is 4.54, or \$454, so a one-standard deviation

¹⁷ Table IA.4 in the Internet Appendix contains additional analysis showing that the message traffic component of $algo_trad_{it}$ drives the decline in spreads.

Table III
Effect of Algorithmic Trading on Spread

This table regresses various measures of the (half) spread on our algorithmic trading proxy. It is based on daily observations from December 2002 through July 2003, which covers the phase-in of autoquote. The nonsynchronous autoquote introduction instruments for the endogenous $algo_trade_{it}$ to identify causality from algorithmic trading to liquidity. The specification is

$$L_{it} = \alpha_i + \gamma_l + \beta A_{it} + \delta X_{it} + \varepsilon_{it},$$

where L_{it} is a spread measure for stock i on day t , A_{it} is the algorithmic trading measure $algo_trade_{it}$, and X_{it} is a vector of control variables, including share turnover, volatility, 1/price, and log market cap. Coefficients for the control variables and time dummies are quintile-specific. Market cap-weighted coefficients are reported for the control variables. Fixed effects and time dummies are included. The set of instruments consists of all explanatory variables, except that $algo_trade_{it}$ is replaced with $auto_quote_{it}$. There are separate regressions for each size quintile, and t -values in parentheses are based on standard errors that are robust to general cross-section and time-series heteroskedasticity and within-group autocorrelation (see Arellano and Bond (1991)). */** denote significance at the 95%/99% level. In Panel B the suffix indicates the effective spread for a particular trade size category: "1" if 100 shares \leq trade size \leq 499 shares; "2" if 500 shares \leq trade size \leq 9,999 shares; "3" if 2,000 shares \leq trade size \leq 4,999 shares; "4" if 5,000 shares \leq trade size \leq 9,999 shares; "5" if 9,999 shares $<$ trade size.

	Coefficient on $algo_trade_{it}$					Coefficients on Control Variables			
	Q1	Q2	Q3	Q4	Q5	$share_turnover_{it}$	$volatility_{it}$	$1/price_{it}$	$ln_mkt_cap_{it}$
Panel A: Quoted Spread, Quoted Depth, and Effective Spread									
$qspread_{it}$	-0.53** (-3.23)	-0.42** (-2.21)	-0.43 (-1.44)	-0.21 (-0.06)	9.92 (1.22)	-2.81** (-2.98)	0.90** (9.71)	108.40** (7.42)	-3.61** (-2.28)
$qdepth_{it}$	-3.49** (-2.51)	-1.43 (-1.16)	-1.99 (-1.07)	15.60 (0.39)	0.61 (0.19)	-5.22 (-0.64)	-1.64* (-1.86)	-3.44 (-0.02)	12.01 (0.82)
$espread_{it}$	-0.18** (-2.67)	-0.32** (-2.23)	-0.35 (-1.56)	-1.67 (-0.42)	4.65 (1.16)	-1.01** (-2.30)	0.69** (9.39)	72.77** (10.80)	-1.30 (-1.46)

(continued)

Table III—Continued

	Coefficient on $algo_trade_{it}$					Coefficients on Control Variables			
	Q1	Q2	Q3	Q4	Q5	$share_turnover_{it}$	$volatility_{it}$	$1/price_{it}$	$ln_mkt_cap_{it}$
Panel B: Effective Spread by Trade Size Category									
$espread1_{it}$	-0.12** (-3.06)	-0.14** (-2.02)	-0.17 (-1.09)	-1.83 (-0.44)	4.99 (1.20)	-0.83** (-2.70)	0.28** (3.91)	50.45** (12.43)	-1.10 (-1.61)
$espread2_{it}$	-0.22** (-3.25)	-0.30** (-2.61)	-0.41 (-1.62)	-4.21 (-0.45)	4.21 (1.16)	-1.62** (-2.68)	0.43** (2.80)	53.84** (8.15)	-2.24 (-1.58)
$espread3_{it}$	-0.25** (-2.76)	-0.26 (-1.42)	-0.66 (-1.57)	-3.27 (-0.36)	6.89 (1.13)	-1.85** (-2.67)	0.64** (4.03)	61.48** (7.55)	-1.96 (-1.29)
$espread4_{it}$	-0.11 (-1.31)	-0.29 (-0.90)	-0.24 (-0.62)	643.14 (0.00)	-3.27 (-0.38)	28.99 (0.00)	-10.35 (-0.00)	224.78 (0.01)	93.04 (0.00)
$espread5_{it}$	-0.04 (-0.45)	-0.32 (-1.04)	-0.36 (-0.86)	1.45 (0.26)	10.48 (0.25)	-0.21 (-0.25)	1.05** (4.01)	73.01** (8.06)	0.12 (0.05)
Panel C: Spread Decompositions Based on 5-Min and 30-Min Price Impact									
$rsread_{it}$	0.35** (3.52)	0.76** (3.97)	1.03** (2.06)	14.25 (0.46)	15.88 (1.36)	3.13* (1.92)	-1.06** (-2.15)	45.81** (4.14)	5.05 (1.18)
$adv_selection_{it}$	-0.53** (-3.56)	-1.07** (-4.08)	-1.39** (-2.06)	-15.51 (-0.47)	-11.21 (-1.33)	-4.12** (-2.23)	1.76** (3.28)	26.65* (1.84)	-6.30 (-1.34)
$rsread_30m_{it}$	0.33** (2.82)	0.47* (1.94)	0.91 (1.61)	11.11 (0.47)	12.63 (1.31)	2.69** (1.98)	-2.33** (-5.99)	52.24** (4.21)	2.83 (0.83)
$adv_selection_30m_{it}$	-0.51** (-3.43)	-0.81** (-2.76)	-1.27* (-1.80)	-12.60 (-0.47)	-8.28 (-1.25)	-3.66** (-2.33)	3.02** (6.91)	20.21 (1.35)	-4.10 (-1.05)

#observations: 1,082 * 167 (stock * day)

 F test statistic of hypothesis that instruments do not enter first-stage regression: 7.32 ($F(5, 179,587)$), p -value: 0.0000

change in our AT measure is associated with a $4.54 * 0.53 = 2.41$ basis point change in proportional spreads. This represents nearly a 50% decline from the mean quoted spread of 5.19 basis points for quintile 1.

In the spirit of an event study, we also estimate an analogous non-IV panel regression with the autoquote dummy directly on the right-hand side. We do not report these results, but quoted and effective spreads are reliably narrower for the three largest quintiles. For quintile 1, quoted spreads are 0.50 basis points smaller ($t = -9.18$) after the autoquote introduction, and effective spreads are 0.17 basis points smaller ($t = -4.33$). Effective spreads narrow even more for quintiles 2 and 3 by 0.21 and 0.23 basis points, respectively.

The IV estimate on $algo.trad_{it}$ is statistically indistinguishable from zero for quintiles 3 through 5. This could be a statistical power issue. Figure 4 above shows that most small-cap stocks were phased in at the very end, reducing the nonsynchronicity needed for econometric identification. Perhaps as a result, the autoquote instrument is only weakly correlated with AT in these quintiles. Alternatively, it could be the case that algorithms are less commonly used in these smaller stocks, in which case the introduction of autoquote might have little or no effect on these stocks' market quality.

Quoted depth also declines with autoquote. One might worry that the narrower quoted spread simply reflects the smaller quoted quantity, casting doubt on whether liquidity actually improves after autoquote is introduced. Here, a calibration exercise is useful. The results for quintile 1 indicate that a one-unit increase in the AT variable—a \$100 decrease in trading volume per message—reduces the quoted spread by 10%, as the average quoted spread from Table I is 5.19 basis points. The same change reduces the quoted depth by about 5%, based on an average quoted depth of \$71,220. A small liquidity demander is unaffected by the depth reduction and is unambiguously better off with the narrower spread. A liquidity demander who trades the average quoted depth of \$71,220 is probably better off as well. She pays 10% less on 95% of her order, and as long as she pays less than 290% of the original spread on the remaining 5% of her order, she is better off overall. Based on the \$40.01 average share price for this quintile, the average 5.19 basis point quoted spread translates to 2.1 cents. For these stocks, it seems extremely unlikely that the last 5% of her trade executes at a spread of more than 6.1 cents. Most likely, this last 5% would execute only one cent wider. This makes it quite clear that the depth reduction is small relative to the narrowing of the spread, at least for these trade sizes.

To further explore the decline in depth, in Panel B of Table III we check to see whether autoquote is associated with narrower effective spreads after controlling for trade size. Specifically, we sort trades into bins based on trade size, calculate effective spreads for these trades alone, and estimate the IV specification for each trade size bin. For the larger quintiles, AT significantly narrows the effective spread for all trade sizes below 5,000 shares. Point estimates go in the same direction for the largest trade sizes but are not reliably different from zero. In any case, we only have transaction-level data, not order-level data, so it is not possible to conduct the analysis on large orders, given that large or-

ders are typically broken up into many smaller pieces and executed over time. If more AT means large orders are broken up more, then even if the effective spread narrows for individual trades, the aggregate price impact of the entire order could increase.

V. Sources of Liquidity Improvement

As discussed earlier in the paper, narrower effective spreads imply either less revenue per trade for liquidity providers, smaller gross losses due to informed liquidity demanders, or both. In Table III Panel C, we decompose effective spreads into a realized spread component and an adverse selection or price impact component in order to understand the source or sources of the improvement in liquidity under autoquote. These spread components are calculated using a 5-minute and a 30-minute horizon, and the IV regressions are repeated using each component of the spread.

The results are somewhat surprising. For large- and medium-cap stocks, quintiles 1 through 3, the realized spread actually increases significantly after autoquote, indicating that liquidity providers are earning greater net revenues. These greater revenues are offset by a larger decline in price impacts, implying that liquidity providers are losing far less to liquidity demanders after the introduction of autoquote. As before, nothing is significant for the two smallest-cap quintiles. Panel C also shows that the results are the same at both 5-minute and 30-minute horizons.

We describe these results as surprising because they do not match our priors going into the analysis. We thought that if autoquote improved liquidity, it would be because algorithmic liquidity suppliers were low-cost providers who suddenly became better able to compete with the specialist and the floor under autoquote, thereby improving overall liquidity by reducing aggregate liquidity provider revenues. Instead, it appears that liquidity providers in aggregate were able to capture some of the surplus created by autoquote.

However, this liquidity supplier market power reflected in larger realized spreads appears to have been fleeting. Figure 3 shows that while liquidity provider revenues increased during the autoquote introduction in the first half of 2003, realized spreads decline during the second half of 2003. Existing algorithms could have found themselves with a distinct competitive advantage in response to the increased information flow, given that new algorithms take considerable time and expense to build and test. Thus, one can interpret the data as a temporary increase in algorithms' market power under autoquote that disappears as new entrants developed competing algorithms.

Which liquidity providers benefit? We do not have any trade-by-trade data on the identity of our liquidity providers, but we do know specialist participation rates for each stock each day, so we can see whether autoquote changed the specialist's liquidity provision market share. We conduct an IV regression with the specialist participation rate on the left-hand side, and the results in Table IV confirm that, at least for the large-cap quintile of stocks, specialists

Table IV
Effect of Algorithmic Trading on Nonspread Variables

This table regresses nonspread variables on our algorithmic trading proxy. It is based on daily observations from December 2002 through July 2003, which covers the phase-in of autoquote. The nonsynchronous autoquote introduction instruments for the endogenous *algo.trad_{it}* to identify causality from algorithmic trading to these nonspread variables. The specification is

$$M_{it} = \alpha_i + \gamma_t + \beta A_{it} + \varepsilon_{it},$$

where *M_{it}* is a nonspread variable for stock *i* on day *t*, and *A_{it}* is the algorithmic trading measure. Fixed effects and time dummies are included. There are separate regressions for each size quintile, and *t*-values in parentheses are based on standard errors that are robust to general cross-section and time-series heteroskedasticity and within-group autocorrelation (see Arellano and Bond (1991)). */** denote significance at the 95%/99% level.

	Coefficient on <i>algo.trad_{it}</i>				
	Q1	Q2	Q3	Q4	Q5
<i>share_turnover_{it}</i>	0.04 (1.02)	-0.07* (-1.77)	0.01 (0.07)	-0.20 (-0.26)	-0.36** (-2.89)
<i>trades_{it}</i>	0.58** (2.60)	-0.01 (-0.23)	-0.01 (-0.15)	-0.51 (-0.33)	-0.15** (-2.60)
<i>trade_size_{it}</i>	-2.04** (-4.64)	-0.80** (-3.23)	-0.33 (-0.69)	2.27 (0.20)	-0.22 (-0.60)
<i>specialist_particip_{it}</i>	-0.59** (-2.22)	-0.23 (-1.24)	-0.92 (-1.43)	-13.24 (-0.29)	-1.89** (-2.02)
#observations: 1,082 * 167 (stock * day)					
<i>F</i> test statistic of hypothesis that instruments do not enter first-stage regression: 5.88(<i>F</i> (5,179,607)), <i>p</i> -value: 0.0000					

appear to participate less under autoquote, suggesting that it is other liquidity providers who capture the surplus created by autoquote.

Table IV also puts a number of other nonspread variables on the left-hand side of the IV specification. The most interesting is trade size. At least for the two largest quintiles, the autoquote instrument confirms most observers' strong suspicions that the increase in AT is one of the causes of smaller average trade sizes in recent years.¹⁸

The decomposition of the effective spread used above has the advantage of being simple, but it also has distinct disadvantages. In particular, it chooses an arbitrary time point in the future, 5 minutes and 30 minutes in this case, and implicitly ignores other trades that might have happened in that 5- or 30-minute time period. Lin et al. (1995) develop a spread decomposition model that is estimated trade by trade and accounts for order flow persistence (the empirical fact, first noted by Hasbrouck and Ho (1987), that buyer-initiated

¹⁸ Turnover has negative and significant coefficients in Q2 and Q5 in Table IV. To ensure that possible endogeneity of turnover is not affecting the *algo.trad* coefficients in Table III, Table IA.5 in the Internet Appendix repeats Table III omitting turnover as a control variable. Including turnover does not change the results.

trades tend to follow buyer-initiated trades).¹⁹ Results in Section III of the Internet Appendix show that our spread decomposition results continue to hold using their approach.

VI. AT and Price Discovery: Trades versus Quotes

In this section, we investigate whether AT changes the nature of price discovery. We use the framework of Hasbrouck (1991a,b), who introduces a model based on a vector autoregression (VAR) that makes almost no structural assumptions about the nature of information or order flow, but instead infers the nature of information and trading from the observed sequence of quotes and trades. In this framework, all stock price moves end up assigned to one of two categories: they are either associated or unassociated with a recent trade. Though the model does not make any structural assumptions about the nature of information, we usually refer to price moves as private information-based if they are associated with a recent trade. Price moves that are orthogonal to recent trade arrivals are sometimes considered based on “public information,” as in Jones et al. (1994) and Barclay and Hendershott (2003).

To separate price moves into trade-related and trade-unrelated components, we construct a VAR with two equations: the first describes the trade-by-trade evolution of the quote midpoint, while the second describes the persistence of order flow. Continuing our earlier notation, define q_{jt} to be the buy–sell indicator for trade t in stock j (+1 for buys, −1 for sells), and define r_{jt} to be the log return based on the quote midpoint of stock j from trade $t - 1$ to trade t . The VAR picks up order flow dependence out to 10 lags:

$$r_t = \sum_{i=1}^{10} \alpha_i r_{t-i} + \sum_{i=0}^{10} \beta_i q_{t-i} + \varepsilon_{rt}, \quad (8)$$

$$q_t = \sum_{i=1}^{10} \gamma_i r_{t-i} + \sum_{i=1}^{10} \phi_i q_{t-i} + \varepsilon_{qt}, \quad (9)$$

where the stock subscripts j are suppressed from now on. The VAR is inverted to get the vector moving average (VMA) representation:

$$y_t = \begin{bmatrix} r_t \\ q_t \end{bmatrix} = \theta(L)\varepsilon_t = \begin{bmatrix} a(L) & b(L) \\ c(L) & d(L) \end{bmatrix} \begin{bmatrix} \varepsilon_{rt} \\ \varepsilon_{qt} \end{bmatrix}, \quad (10)$$

where $a(L)$, $b(L)$, $c(L)$, and $d(L)$ are lag polynomial operators. The permanent effect on price of an innovation is given by $a(L)\varepsilon_{rt} + b(L)\varepsilon_{qt}$, and because we

¹⁹ See Barclay and Hendershott (2004) for a discussion of Lin et al. (1995) versus other spread decomposition models.

include contemporaneous q_t in the return equation, $\text{cov}(\varepsilon_{rt}, \varepsilon_{qt}) = 0$ and the variance of this random-walk component can be written as

$$\sigma_w^2 = \left(\sum_{i=0}^{\infty} a_i \right)^2 \sigma_r^2 + \left(\sum_{i=0}^{\infty} b_i \right)^2 \sigma_q^2. \quad (11)$$

The second term captures the component of price discovery that is related to recent trades, and the first term captures price changes that are orthogonal to trading. The VAR is estimated in transaction time; we follow Hasbrouck (1991b) and multiply the resulting variances by the number of trades each day to get a decomposition of the daily variance of log changes in the efficient price.

The information content of individual trades can also be calculated using the VAR. As shown in Hasbrouck (1991a), the cumulative response of the quoted price to a one-time unit shock in the order flow equation is a measure of adverse selection that accounts for the persistence in order flow as well as possible positive or negative feedback trading. The cumulative impulse response is easiest to see in the VMA representation and is equal to $(\sum_{i=0}^{\infty} b_i)$. As discussed in Hasbrouck (1991a,b), the VAR approach is robust to price discreteness, lagged adjustment to information, and lagged adjustment to trades.

The VAR, the cumulative impulse response, and the trade-related and non-trade-related standard deviations are estimated for each stock each day. As in the prior graphs, we calculate monthly averages for each quintile and graph these in Figure 5. The most striking feature of the graphs is the decline in the cumulative impulse response and the trade-related standard deviation during the first half of 2003, while the non-trade-related standard deviations do not change much as autoquote is introduced. This suggests that under autoquote much more information is being incorporated into prices without trade, consistent with Boulatov and George (2007), who show that mean-squared error in the quote midpoint is smaller when informed traders place limit orders, compared to a world where liquidity is only provided by perfectly competitive, uninformed, risk-neutral market makers.

While these time-series effects appear large, again we prefer to identify the effect using the staggered autoquote instrument. The IV panel regression is estimated first with the daily trade-related standard deviation as the dependent variable. We then repeat the regression using the non-trade-related standard deviation on the left-hand side. The panel regressions continue to include stock fixed effects, calendar dummies, and the same set of control variables.

The results can be found in Table V, and at least for the two largest quintiles they confirm the results from the time-series graphs. When a large-cap stock adopts autoquote and experiences an exogenous increase in AT, the impulse response measure of adverse selection decreases. In addition, there is much less trade-correlated price discovery, and much more price discovery that is uncorrelated with trading. Consistent with other methodologies, we do not find consistently reliable effects for the smaller-cap quintiles.

For the largest quintiles, AT has an economically important effect on the nature of price discovery. During the autoquote sample period, the within

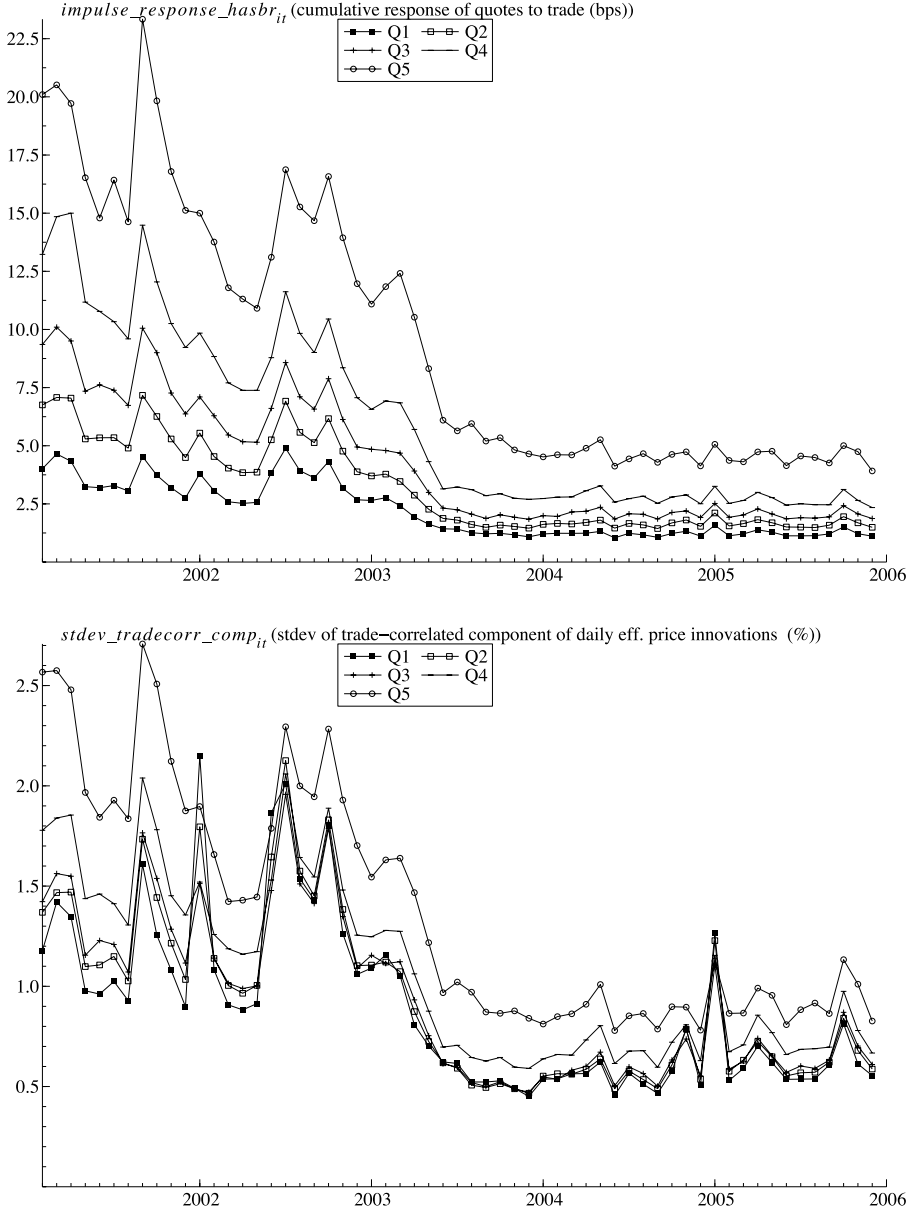


Figure 5. Trade-correlated and trade-uncorrelated information. These graphs illustrate the estimation results of the Hasbrouck (1991a,b) VAR model for midquote returns and signed trades. The top graph illustrates the time-series pattern of the long-term response of the midquote price to a unit impulse in the signed trade variable. The bottom two graphs illustrate the decomposition of the daily percentage variance of changes in the efficient price into a trade-related ($stdev_trade_{corr_comp_{it}}$) and trade-unrelated ($stdev_nontrade_{corr_comp_{it}}$) component (see Hasbrouck (1991a,b) for details on the methodology). The graph includes the autoquote sample period, which runs from December 2002 through July 2003. Results are graphed by market-cap quintile, where Q1 is the largest-cap quintile.

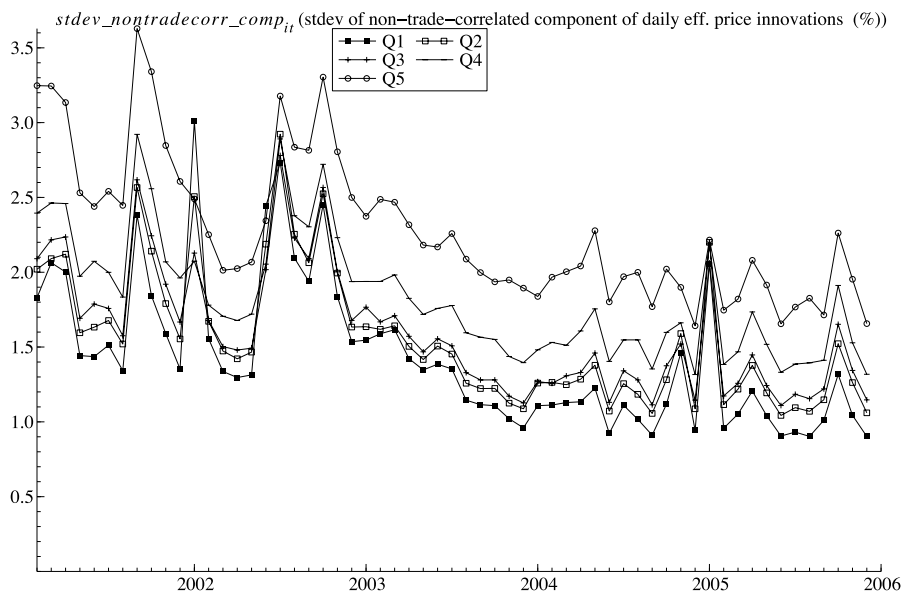


Figure 5. Continued

standard deviation in our AT variable is 4.54, so a one standard deviation increase in AT during this sample period leads to an estimated change in trade-correlated price discovery equal to a $4.54 \times 0.22 = 1.00$ percentage point reduction in the daily standard deviation of trade-correlated returns for the largest-cap stocks. Figure 5 shows that this is the same order of magnitude as the actual level of trade-correlated standard deviations, so this is indeed a substantial change in how prices are updated to reflect new information over time.

VII. Discussion and Interpretation

Why is the nature of price discovery changing? It seems likely that algorithms respond quickly to order flow and price information, updating their limit orders to prevent them from becoming stale and being picked off. We suspect that two kinds of information are of first-order importance, though it is hard to be sure, given the opacity of algorithm creators and providers. First, we think algorithms can easily take into account common factor price information and adjust trading and quoting accordingly. For example, if there is an upward shock to the S&P futures price, an algorithmic liquidity supplier in IBM that currently represents the inside offer may decide to cancel its existing sell order before it is picked off by an index arbitrageur or another trade, replacing the sell order with a higher-priced ask. Shocks to other stocks in the same industry could cause similar reactions from algorithms. Second, some algorithms are designed to sniff out other algorithms or otherwise identify order flow and other information patterns in the data. For example, if an algorithm identifies

a sequence of buys in the data and concludes that more buys are coming, an algorithmic liquidity supplier might adjust its ask price upward. Information in new swires can even be parsed electronically in order to adjust trading algorithms.²⁰

To help understand the counterintuitive realized spread result, it is important to consider how cost structures differ for human and algorithmic trading. If monitoring costs are sufficiently high that humans do not always monitor the market, limit orders submitted by humans will not always reflect all public information and may become stale. While algorithms have large fixed development costs, one of their main advantages is that there is virtually no marginal cost in monitoring public information and adjusting their orders or quotes. Thus, an increase in algorithmic activity causes more changes in the efficient price to be revealed through a quote update rather than via trade. To develop this intuition further, Section IV of the Internet Appendix develops a very simple generalized Roll model that is a slight variation on the one developed in Hasbrouck (2007).

If the algorithmic traders have more market power than the nonalgorithmic traders, an increase in AT leads to larger imputed revenue to liquidity suppliers (larger realized spread). Is this market power argument plausible? As autoquote was implemented in 2003, the extant algorithms might have found themselves with a distinct competitive advantage in trading in response to the increased information flow, given that new algorithms take considerable time and expense to build and test. While we view the differential market power argument as possible, we do not have any additional evidence to support it. Over the longer run, liquidity supplier revenues decline (see Figure 3), suggesting that any market power was temporary.

Ideally, we would also directly analyze AT behavior to better understand the specific AT order submission and trading strategies that lead to these results. Unfortunately, no U.S. stock exchange has yet been able and willing to identify AT. However, Hendershott and Riordan (2009) can accurately identify AT on the Deutsche Boerse. They find that algorithmic traders generally place more efficient quotes, and algorithms supply more liquidity when spreads are wide. While they describe equilibrium behavior and cannot measure the causal effect of AT on market quality, Hendershott and Riordan (2009) do provide a natural mechanism by which algorithmic traders would improve liquidity and quote efficiency.

VIII. Conclusions

The declining costs of technology have led to its widespread adoption throughout financial industries. The resulting technological change has revolutionized financial markets and the way financial assets are traded. Many institutions now trade via algorithms, and we study whether AT at the NYSE improves liquidity. In the 5 years following decimalization, AT has increased, and markets

²⁰ See "Ahead of the tape—algorithmic trading," *Economist*, June 23, 2007.

have become more liquid. To establish causality we use the staggered introduction of autoquoting as an instrumental variable for AT. We demonstrate that increased AT lowers adverse selection and decreases the amount of price discovery that is correlated with trading. Our results suggest that AT lowers the costs of trading and increases the informativeness of quotes. Surprisingly, the revenues to liquidity suppliers also increase with AT, though this effect appears to be temporary.

While we have not studied it here, it seems likely that AT can also improve linkages between markets, generating positive spillover effects in these other markets. For example, when computer-driven trading is made easier, stock index futures and underlying share prices are likely to track each other more closely. Similarly, liquidity and price efficiency in equity options probably improve as the underlying share price becomes more informative.

A couple of caveats are in order, however. Our overall sample period covers a period of generally rising stock prices, and stock markets are fairly quiescent during the 2003 introduction of autoquote. While we do control for share price levels and volatility in our empirical work, it remains an open question whether AT and algorithmic liquidity supply are equally beneficial in more turbulent or declining markets. Like NASDAQ market makers refusing to answer their phones during the 1987 stock market crash, algorithmic liquidity suppliers may simply turn off their machines when markets spike downward. With access to the right data, the 2007 and 2008 stock markets could prove to be a useful laboratory for such an investigation.

A second caveat relates to trading by large institutions. Some market participants complain that the decline in depth has hampered investors' ability to trade large amounts without substantial costs. While the rise of AT has contributed to the decline in depth, we are optimistic that other technological innovations can offset some of these effects. For instance, some "dark pools" such as LiquidNet and Pipeline represent a modern version of an upstairs market, allowing traders with large orders to electronically search for counterparties without revealing their trading interest (see, e.g., Bessembinder and Venkataraman (2004)).

Finally, our results have important implications for both regulators and designers of trading platforms. For example, the U.S. Securities and Exchange Commission's Regulation NMS (SEC (2005)) is designed to increase competition among liquidity suppliers. Our results highlight the importance of algorithmic liquidity suppliers and the benefits of ensuring vigorous competition among them. Of course, markets need not leave this problem to the regulator. Trading venues can attract these algorithms by lowering development and implementation costs. For example, exchanges and other trading platforms can calculate useful information and metrics to be fed into algorithms, distributing them at low cost. A market can also allow algorithmic traders to co-locate their servers in the market's data center. Finally, offering additional order types, such as pegged orders, can lessen the infrastructure pressures that algorithms impose.

REFERENCES

- Almgren, Robert, and Neil Chriss, 2000, Optimal execution of portfolio transactions, *Journal of Risk* 3, 5–39.
- Arellano, Manuel, 2003, *Panel Data Econometrics* (Oxford University Press, New York).
- Arellano, Manuel, and Stephen R. Bond, 1991, Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations, *Review of Economic Studies* 58, 277–297.
- Barclay, Michael J., and Terrence Hendershott, 2003, Price discovery and after trading hours, *Review of Financial Studies* 16, 1041–1073.
- Barclay, Michael J., and Terrence Hendershott, 2004, Liquidity externalities and adverse selection: Evidence from trading after hours, *Journal of Finance* 59, 681–710.
- Bertsimas, Dimitris, and Andrew W. Lo, 1998, Optimal control of execution costs, *Journal of Financial Markets* 1, 1–50.
- Bessembinder, Hendrik, 2003, Issues in assessing trade execution costs, *Journal of Financial Markets* 6, 233–257.
- Bessembinder, Hendrik, and Kumar Venkataraman, 2004, Does an electronic stock exchange need an upstairs market? *Journal of Financial Economics* 73, 3–36.
- Biais, Bruno, David Martimort, and Jean-Charles Rochet, 2000, Competing mechanisms in a common value environment, *Econometrica* 68, 799–837.
- Biais, Bruno, and Pierre-Olivier Weill, 2009, Liquidity shocks and order book dynamics, Working paper, Toulouse University, IDEI.
- Black, Fischer, 1995, Equilibrium exchanges, *Financial Analysts Journal* 51, 23–29.
- Boulatov, Alex, and Thomas J. George, 2007, Securities trading when liquidity providers are informed, Working paper, University of Houston.
- Bound, John, David A. Jaeger, and Regina M. Baker, 1995, Problems with instrumental variables estimation when the correlation between the instruments and the endogenous explanatory variable is weak, *Journal of the American Statistical Association* 90, 443–450.
- Chaboud, Alain, Benjamin Chiquoine, Erik Hjalmarsson, and Clara Vega, 2009, Rise of the machines: Algorithmic trading in the foreign exchange market, Working paper, Federal Reserve Board.
- Chan, Louis K.C., and Josef Lakonishok, 1995, The behavior of stock prices around institutional trades, *Journal of Finance* 50, 1147–1174.
- Cohen, Kalman, Steven Maier, Robert Schwartz, and David Whitcomb, 1981, Transaction costs, order placement strategy and existence of the bid-ask spread, *Journal of Political Economy* 89, 287–305.
- Comerton-Forde, Carole, Terrence Hendershott, Charles M. Jones, Mark S. Seasholes, and Pamela C. Moulton, 2010, Time variation in liquidity: The role of market maker inventories and revenues, *Journal of Finance* 65, 295–331.
- Copeland, Thomas E., and Dan Galai, 1983, Information effects on the bid-ask spread, *Journal of Finance* 38, 1457–1469.
- Domowitz, Ian, and Henry Yegerman, 2005, The cost of algorithmic trading: A first look at comparative performance, in Brian R. Bruce, ed.: *Algorithmic Trading: Precision, Control, Execution* (Institutional Investor London).
- Engle, Robert F., Jeffrey R. Russell, and Robert Ferstenberg, 2007, Measuring and modeling execution cost and risk, Working paper, New York University.
- Foucault, Thierry, Ohad Kadan, and Eugene Kandel, 2010, Liquidity cycles and make/take fees in electronic markets, Working paper, HEC Paris.
- Foucault, Thierry, and Albert J. Menkveld, 2008, Competition for order flow and smart order routing systems, *Journal of Finance* 63, 119–158.
- Foucault, Thierry, Ailsa Roëll, and Patrik Sandas, 2003, Market making with costly monitoring: An analysis of the SOES controversy, *Review of Financial Studies* 16, 345–384.
- Glosten, Lawrence R., 1987, Components of the bid ask spread and the statistical properties of transaction prices, *Journal of Finance* 42, 1293–1307.

- Glosten, Lawrence R., 1989, Insider trading, liquidity, and the role of the monopolist specialist, *Journal of Business* 62, 211–235.
- Glosten, Lawrence R., 1994, Is the electronic limit order book inevitable? *Journal of Finance* 49, 1127–1161.
- Goettler, Ronald L., Christine A. Parlour, and Uday Rajan, 2009, Informed traders in limit order markets, *Journal of Financial Economics* 93, 67–87.
- Goldstein, Michael A., and Kenneth A. Kavajecz, 2004, Trading strategies during circuit breakers and extreme market movements, *Journal of Financial Markets* 7, 301–333.
- Greene, William H., 2007, *Econometric Analysis* (Prentice Hall, London).
- Harris, Lawrence, 1998, Optimal dynamic order submission strategies in some stylized trading problems, *Financial Markets, Institutions, and Instruments* 7, 1–76.
- Hasbrouck, Joel, 1991a, Measuring the information content of stock trades, *Journal of Finance* 46, 179–207.
- Hasbrouck, Joel, 1991b, The summary informativeness of stock trades: An econometric analysis, *Review of Financial Studies* 4, 571–595.
- Hasbrouck, Joel, 2007, *Empirical Market Microstructure* (Oxford University Press, New York).
- Hasbrouck, Joel, and Thomas Ho, 1987, Order arrival, quote behavior and the return generating process, *Journal of Finance* 42, 1035–1048.
- Hasbrouck, Joel, and Gideon Saar, 2009, Technology and liquidity provision: The blurring of traditional definitions, *Journal of Financial Markets* 12, 143–172.
- Hendershott, Terrence, and Pamela C. Moulton, 2010, Automation, speed, and stock market quality: The NYSE's hybrid, Working paper, University of California, Berkeley.
- Hendershott, Terrence, and Ryan Riordan, 2009, Algorithmic trading and information, Working paper, University of California, Berkeley.
- Jones, Charles M., Gautam Kaul, and Marc L. Lipson, 1994, Information, trading, and volatility, *Journal of Financial Economics* 36, 127–154.
- Jovanovic, Boyan, and Albert J. Menkveld, 2010, Middlemen in limit-order markets, Working paper, New York University, New York.
- Keim, Donald B., and Ananth Madhavan, 1995, Anatomy of the trading process: Empirical evidence on the behavior of institutional traders, *Journal of Financial Economics* 37, 371–398.
- Lee, Charles M.C., and Mark J. Ready, 1991, Inferring trade direction from intraday data, *Journal of Finance* 46, 733–746.
- Lin, Ji-Chai, Gary C. Sanger, and G. Geoffrey Booth, 1995, Trade size and components of the bid-ask spread, *Review of Financial Studies* 8, 1153–1183.
- Obizhaeva, Anna, and Jiang Wang, 2005, Optimal trading strategy and supply/demand dynamics, Working paper, MIT.
- Parkinson, Michael, 1980, The extreme value method for estimating the variance of the rate of return, *Journal of Business* 51, 61–65.
- Reinganum, Jennifer F., 1989, The timing of innovation: Research, development, and diffusion, in Richard Schmalensee and Robert Willig, ed.: *Handbook of Industrial Organization* (Elsevier Publishing, Amsterdam).
- Rock, Kevin, 1990, The specialist's order book and price anomalies, Working paper, Harvard University.
- Rosu, Ioanid, 2009, A dynamic model of the limit order book, *Review of Financial Studies* 22, 4601–4641.
- Securities and Exchange Commission, 2005, Release No. 34-51808.

