New Insights on Computerized Trading

Implications of Frequently Revised Trading Decisions

Petter Dahlström
New Insights on Computerized Trading
Implications of Frequently Revised Trading Decisions

Petter Dahlström

Abstract
Computerized trading may be viewed as an aspect of modernization of financial markets. This dissertation contains four articles that in different ways examine to what extent the modernization influences the economics of the markets.

Article 1 investigates transaction costs for large orders which are split up by execution algorithms to be executed in smaller pieces. I find that the costs associated with not being able to execute all pieces are substantial. These costs can be lowered by speeding up the trading pace but at the expense of higher costs for the successfully executed pieces.

Article 2 investigates the strategies trading firms pursue in particular cases, known as toxic arbitrage opportunities. We find that trading firms, that otherwise behave as market makers, morph into liquidity takers as toxic arbitrage opportunities emerge. In contrast to common belief, market makers are net beneficiaries of toxic arbitrage, and this finding puts into question whether the amount of toxic arbitrage leads to wider bid-ask spreads.

Article 3 investigates the information content of limit orders in an alternative way by studying the price impact implied by the depth in the limit order book. I find that the price impact estimates are slightly lower relative to those from a structural vector auto regressive model, but slightly higher compared to those from a price impact regression. Thus, the limit order book implied price impact estimates match those from benchmark models, and this finding contradicts earlier research.

Article 4 investigates the economic rationale behind limit order cancellations. We put forth a model that explains the frequent limit order cancellations seen in today’s markets, and we test its predictions using a unique data set from Nasdaq. Our results points towards that frequent order cancellations is a benign feature of modern market making, as opposed to different types of manipulative behavior.
NEW INSIGHTS ON COMPUTERIZED TRADING
Petter Dahlström
New Insights on Computerized Trading
Implications of Frequently Revised Trading Decisions
Petter Dahlström
Acknowledgements

A road has come to an end, and it is time for new adventures. For the most part, it has been an enjoyable and highly rewarding journey. I spent numerous hours writing this thesis, and I am grateful to everyone that contributed.

First, I would like to express my sincere gratitude to my supervisors Björn Hagströmer and Lars Nordén for always being available, for carefully reading my texts, and for the support over the last five years. I also highly appreciate the encouragement from Frank Hatheway and Björn Hertzberg at Nasdaq.

I will miss my Ph.D. student colleagues and friends. I am grateful to Fatemeh Abrahamian, Ester Felez Vinas, Dong Zhang, and Chengcheng Qu with whom I share research interest, and to Ian Khrashchevskyi and Anton Hasselgren for the discussions on related topics. Thank you all, together we learned a lot.

All the administrative things can be overwhelming and drive one crazy. But thanks to the backing from Linnea Shore, Helene Olofsson, Oskar Sjölander, and in particular Doris Rehnström, things went smoothly. The support I got is beyond my expectations. Furthermore, it has been a pleasure to teach at Stockholm Business School; and Li Malmström, Jarkko Peltomäki and Ai Jun Hou have always been very accommodating.

This thesis is shaped by the comments I got at my milestone seminars, and I want to thank Caihong Xu, Marius Zoican, Iñaki Rodriguez Longarela and Erik Hjalmarsson. It was also valuable for me to visit HEC in Paris, and I am thankful to Thierry Foucault for hosting me.
Lastly, I would like to express my gratitude to all my other beloved friends that encouraged me throughout. Mina sista ord av tacksamhet går till de jag håller av allra mest: Johan, Charlotta, Mamma och Pappa.

Petter Dahlström

Table of Contents

Introduction 1

Article 1: Transaction Costs of Large Orders, Trading Pace, and the Cost of Non-Execution 19

Article 2: Dr. Jekyll and Mr. Hyde: Market Makers and Toxic Arbitrageurs 61


Article 4: Determinants of Limit Order Cancellations 143
Introduction

Computerized trading can be viewed as an aspect of the modernization of financial markets. In the literature, the definition of algorithmic trading (AT) includes a wide set of different computer algorithms designed to make trading decisions, submit orders, and manage those orders after submission. Smart order routers and order-splitting algorithms are common AT tools that brokers use to implement investment decisions on behalf of their clients. For example, brokers use smart order routers to, automatically and systematically, search for the most favorable prices across trading venues. A subset of AT strategies rely on speed, and therefore specialized traders acquire the latest technology to be faster and better informed than their peers. These traders are called high-frequency traders (HFTs), and their trading strategies rely entirely on computer algorithms for making all trading decisions. Directional strategies, arbitrage, and market making are examples of low latency–dependent strategies that exploit the possibility of either buying or selling whenever short-lived profit opportunities arise. This thesis zooms in on several features of computerized trading and analyzes the extent to which modernization influences the economics of the markets. For example, does this modernization benefit some market participants, in terms of increased profitability, at the expense of others?

All the chapters in this thesis are in the field of market microstructure. This research field studies “the process and outcomes of exchanging assets under explicit trading rules” (O’Hara 1995, 1). Of main concern are liquidity and price discovery, and together these concepts define the quality of financial markets. Liquidity can be defined as the degree to which an order can be executed in a short time frame at a price close to the consensus value (Foucault, Pagano, and Röell 2013). Buy orders tend to push prices
up, while sell orders tend to move prices down. The more liquid a market is, the lower
the impact of buy and sell orders, and the lower the transaction costs paid by investors.
Price discovery is the speed and accuracy at which prices incorporate information
available to market participants (Foucault et al. 2013). A mismatch between the price and
the value of a financial security threatens to weaken investors’ trust in the market and to
disincentive participation.

The consensus view is that transaction costs for small orders have decreased as a
consequence of an increase in AT (Brogaard, Hendershott, and Riordan 2013;
Hendershott, Jones and Menkveld 2011; Menkveld 2013). Only a few studies have
accessed time series data on large orders, but Frazzini, Israel, and Moskowitz (2014)
report that transaction costs have decreased for such orders too.

The first article of this thesis investigates the transaction costs of large orders
when investors use execution algorithms to split the large order (parent order) into
smaller orders (child orders). Some have expressed concern that the transaction costs for
such orders have increased and attribute the increased costs to HFTs (see the discussion
of Menkveld and van Kervel (2018) and the references therein). Menkveld and van Kervel
(2018) report that HFTs trade in the same direction as large informed orders once they
learn about them, thus increasing transaction costs for their counterparts and delaying
price discovery. This finding is consistent with the back-running theory proposed by Yang
orders are executed predictably and finds that transaction costs increase with
predictability. Investors executing large orders must consider such properties.

When executing large orders by execution algorithms, investors specify the
trading pace (the rate at which child orders are sent to market). They want to trade slowly
to reduce the price impact of their own order; however, they also want to trade quickly to reduce the risk of a price impact caused by other investors trading on similar information (e.g., from back-running strategies). This trading pace trade-off is a concern for investors executing large orders and has recently been emphasized in the theoretical model of Kyle, Obizhaeva, and Wang (2018). The literature on the transaction costs of large orders is scarce, and, to my knowledge, there is no empirical evidence of any trading pace trade-off.

My data set contains full details on parent orders, as well as child orders, both executed and unexecuted. Transaction costs associated with not being able to execute all child orders comprise 40% of the total. By ignoring the cost of non-execution, the literature underestimates the total transaction cost of large orders. I provide empirical evidence of a trading pace trade-off. The cost of non-execution can be lowered by speeding up the trading pace, but at the expense of higher costs for successfully executed child orders.

The second article, coauthored with Björn Hagströmer and Lars Nordén, studies two different strategies commonly used by HFTs: arbitrage and market making. Arbitrageurs closely monitor two or more highly correlated financial instruments traded on different markets. When the prices are out of sync, the arbitrageur buys the cheap asset at the ask price and simultaneously sells the expensive asset at the bid price, for a profit. This strategy requires the use of market orders for immediate execution (liquidity taking).

Market makers post limit orders and are willing to buy at a lower price (the bid price) and simultaneously sell at a higher price (the ask price). They profit from the price difference between the bid and ask prices, the bid–ask spread. The fastest market makers
make higher profits, since they avoid adverse price movements (Menkveld 2013) and relax their inventory constraints (Brogaard, Hagströmer, Nordén and Riordan, 2015).

The extent to which HFTs split their activity between these strategies has implications for liquidity. Most theoretical models (e.g., Ait-Sahalia and Saglam 2014; Foucault, Kozhan, and Tham 2017; Hoffmann 2014; Menkveld and Zoican 2017) show that liquidity improves and bid-ask spreads decrease when HFTs engage in market-making strategies rather than liquidity-taking strategies.¹ Thus, the finding that HFT is beneficial for market quality and improves liquidity in these models depends on what kinds of strategies HFTs pursue. Budish, Cramton, and Shim (2015) argue that the effect of HFTs could be negative, regardless of whether HFTs are market makers or liquidity takers. They show theoretically that all trading firms would be better off if they collectively committed to not invest in speed. The authors argue that, since both market makers and liquidity takers invest heavily in speed to be faster than each other, the end result is a socially wasteful race for speed.

Empirical evidence show that HFTs split their order flow almost equally between limit orders and market orders (Brogaard et al. 2015; Hagströmer and Nordén 2013) and generate revenue both through market-making and liquidity-taking strategies (Baron, Brogaard, Hagströmer and Kirilenko (2019). However, the US Securities and Exchange Commission (2014) states that the literature so far examines only a relatively small amount of HFT activity, and they express a desire for further research on multimarket strategies in particular.

¹There are, however, models that achieve lower bid–ask spreads, even when HFTs act as liquidity takers (e.g., Foucault, Hombert, and Rosu 2016).
We study multimarket strategies by looking at toxic arbitrage opportunities. Toxic arbitrage occurs when two markets that trade the same security respond to news with different degrees of latency. We find that trading firms that otherwise behave as market makers morph into liquidity takers as toxic arbitrage opportunities emerge. In contrast to common belief, market makers are net beneficiaries of toxic arbitrage. We point out that it is market makers are unlikely to charge higher spreads during normal times if they make profits from toxic arbitrage opportunities.

The third article in this thesis investigates the information content of limit orders in an alternative way by studying the price impact implied by the depth in the limit order book (LOB). The idea is that LOB depth reflects market makers’ beliefs about the price impact of market orders. This link between the LOB depth and expected price impact relies on the models of Sandås (2001) and Glosten (1994).

To determine whether this methodology is justified, I compare the implied price impact to two different benchmark models. I find that the implied price impact estimates are lower than those from a structural vector autoregressive (SVAR) model, but higher than those from a price impact regression. Thus, the LOB-implied price impact estimates lie within the range from benchmark models. My finding contradicts earlier research, since the same methodology does not work well using data from the 1990s; however, my results suggest that the way markets work nowadays better suits the model’s assumptions.

My results are line with the literature on the information content of limit orders. Hendershott, Jones, and Menkveld (2011) document that an increase in AT participation results in greater price discovery through quotes, as opposed to trades (i.e., limit orders are more informative). Chaboud, Hjalmarsson, and Zikes (2018) show that the
information share of limit orders has increased over time. In addition, they show that price discovery becomes faster, thus suggesting improvements in market efficiency. Brogaard, Hendershott, and Riordan (2019) explicitly link informative limit orders to HFTs.

The fourth article, coauthored with Björn Hagströmer and Lars Nordén, investigates the economic rationale behind limit order cancellations. One implication of the modernization and computerization of markets is that limit orders are updated more frequently (Angel, Harris, and Spatt 2015). Hasbrouck (2018) reports increasing excess short-term volatility in bid and ask quotes. The author also notes market participants’ fears of quote volatility arising from manipulative behavior such as quote stuffing (canceling and submitting orders to delay or confuse others) or spoofing (submitting and canceling orders to create a false impression of trading interest). To curb excessive cancellations, exchanges and regulators have imposed cancellation fees and a minimum quote life (once an order has been submitted, it cannot be cancelled until a time threshold is met). However, the evidence so far suggests that quote volatility arises from legitimate trading strategies. Hasbrouck and Saar (2009) suggest that traders systematically update orders in response to price changes, and van Kervel (2015) finds that traders cancel orders in response to information transmitted from other markets. Hasbrouck (2018) shows that sellers incrementally undercut each other’s orders on price (so-called Edgeworth cycles).

We put forth a model that explains the frequent limit order cancellations seen in today’s markets and test its predictions using a unique data set from Nasdaq. The model stipulates that liquidity suppliers cancel orders whenever the expected revenue no longer exceeds the expected cost. The predictions find strong support in the data for both
HFTs and other trading firms. This result suggests that HFTs are not different from other trading firms with respect to their economic rationale. Our results show that frequent order cancellations are a benign feature of modern market making. The evidence suggests that policies aimed at curbing order cancellations (and excessive volatility of the limit order flow) instead could limit liquidity supply.

The following presents detailed summaries of these articles.

**Article 1: Transaction Costs of Large Orders, Trading Pace, and the Cost of Non-Execution**

This article studies the transaction costs of large orders handled by execution algorithms. The execution algorithms split the total trading demand (parent order) into several smaller orders (child orders) and send these to the market over time. Investors choose the desired parent order size and trading pace. The trading pace is the rate at which child orders are sent to the market, measured as a participation rate per unit of time. For example, investors can specify their pace as a target participation rate of 10% of the market volume in each hour. The higher the participation rate, the earlier the parent order is filled.

I hypothesize that investors executing large orders face a trade-off when choosing the trading pace. On the one hand, they want to trade slowly to reduce the price impact of their own order; on the other hand, trading quickly can reduce the risk of the price impact caused by other investors trading on similar information. This trading pace trade-off is known to market participants and emphasized in the theoretical models of Kyle et al. (2018) and Garleanu and Pedersen (2013), but has not, to my knowledge, been verified or quantified empirically.
Perold (1988) suggests a measure of total transaction costs called implementation shortfall, defined as the price difference between actual trades and a hypothetical trade (with no transaction costs) at the time of the trading decision. The implementation shortfall measure consists of execution costs (the costs of the executed child orders) and opportunity costs (the costs of unexecuted child orders). Since the implementation shortfall metric includes the costs of unexecuted child orders, it captures the trading pace trade-off. If investors trade slowly, they experience a lower price impact on their executed child orders and therefore incur lower execution costs. However, the slow pace increases the risk of adverse price movements, such that the initial profit opportunity vanishes, and child orders remain unexecuted. The opportunity cost increases with the level of such adverse price movements.

I access a unique data set of large orders that were taken to the market by execution algorithms, provided by an anonymous broker. My data sample consists of 3,293 parent orders posted between December 1, 2013, and August 31, 2016. The data set covers 115 stocks listed on the stock exchanges in Denmark, Finland, and Sweden. It was important that I observe all child orders, both executed and unexecuted. This was essential for the estimation of opportunity costs and the trading pace and allowed me to test the trading pace hypothesis.

Due to the difficulties of observing unexecuted orders, little is known about opportunity costs. I show that such costs are economically significant, constituting 40% of the total implementation shortfall. I contribute to the literature by providing empirical evidence of a trading pace trade-off. Investors can lower the opportunity costs of non-execution by speeding up the trading pace. The drawback of speeding up the pace is that the execution costs for successfully filled child orders increase. Both execution and
opportunity costs are economically significant. My estimations show that a one standard deviation increase in the trading pace increases execution costs by 5.4 basis points (1/100th of a percent) and decreases opportunity costs by 2.1 basis points. The net effect is an increase of 3.3 basis points in the total implementation shortfall.

**Article 2: Dr. Jekyll and Mr. Hyde: Market Makers and Toxic Arbitrageurs**

Together with Björn Hagströmer and Lars Nordén, I study the toxic arbitrage opportunities that arise when two markets that trade the same security respond to news with different degrees of latency. For a moment, the quotes of the slower market become stale, and arbitrageurs rush in to trade on the price difference. Relatively slow market makers who have not yet revised their outstanding quotes can then incur losses. The aim of this paper is to investigate the extent to which market makers and toxic arbitrageurs operate within the same trading firms (which we call the activity mix). The activity mix is important, we argue, because it potentially influences liquidity. We also argue that if the market makers are picking off each other’s quotes, they can recoup the losses made as the passive counterparty in toxic arbitrage trades, by profits made on the active side.

We investigate arbitrage activity between an equity index futures contract and an exchange-traded fund, both tracking the Swedish blue-chip index OMXS 30. We access public data on quotes and trades and combine these with proprietary transaction records provided to us by the Swedish Financial Supervisory Authority (*Finansinspektionen*). The proprietary database is key to our analysis, since it includes trading firm identifiers for each trade.

We find that trading firms that otherwise behave as market makers morph into liquidity takers as toxic arbitrage opportunities emerge. Trading firms identified as
futures market makers earn 44.0% of the toxic arbitrage profits. The same traders incur 31.9% of the toxic arbitrage losses. Trading firms identified as ETF market makers earn 45.5% of the profits and lose 29.7% of the money at stake. It is evident that the profits of both market maker categories exceed their respective losses incurred in toxic arbitrage.

The model of Foucault et al. (2017) distinguishes market makers and arbitrageurs as distinct trader types. When toxic arbitrage occurs, the market maker incurs losses if the arbitrageur is faster and picks off her stale quote. In equilibrium, the market maker sets the bid–ask spread such that she breaks even, thus the bid–ask spread increases with her toxic arbitrage losses.

We find that, in contrast to common belief, market makers are net beneficiaries of toxic arbitrage. In light of our results, toxic arbitrage is unlikely to lead market makers to charge higher spreads. We contribute to the arbitrage literature by highlighting that arbitrageurs and market makers should not be viewed as polar cases, but, rather, as pursuing different strategies within the same trading firms. This finding is important to keep in mind when interpreting theoretical models on fast trading (e.g., Foucault et al. 2017), where these roles are treated as distinct.

Our results have practical implications for policymakers at exchanges and regulatory agencies. Initiatives such as delays to incoming market orders, aimed to protect investors from being picked off, could be detrimental to market makers.

**Article 3: What does the order book depth tell us about price impact?**

The information content of trades is central in the market microstructure literature. Traditionally, the empirical strategy is to measure price impacts ex post by the extent to which prices react to arriving marker orders or trades. The traditional
methodology has limitations, since traders nowadays use order-splitting strategies to spread out trading over time. Given such behavior, the full trading interest is not fully revealed at a given point in time, and the information content of individual trades could be low (O'Hara (2015)). Moreover, traditional methods lack the ability to account for the information content of limit orders, which has increased in recent years (Chaboud et al. 2018; Hendershott et al. 2011).

In this paper, I estimate the price impact by using the depth offered in the LOB. The idea is that, if the LOB is in equilibrium, the depth reflects market makers' beliefs about the price impact of market orders. This link between the LOB depth and the expected price impact relies on the models of Sandås (2001) and Glosten (1994). Sandås also presents empirical evidence of a price impact implied by the LOB that is, on average, almost 10 times higher than its benchmark, based on Glosten and Harris’ (1988) price impact regression. The parameter for order processing costs is estimated to be negative. Order processing costs must cover the expenses market makers incur in trading and clearing fees, and these expenses are generally expected to be positive.

Sandås (2001) estimates the price impact implied by the LOB depth by using two different sets of moment conditions. In the first set of moment conditions, the price impact is captured by the slope coefficient for the linear relations between the bid-ask spread and the cumulative depth for the two best levels in the LOB (called break-even conditions). In addition, a set of updating conditions specifies how the LOB depth should respond to market orders. Similar to Sandås, I estimate the break-even and updating conditions jointly and separately by the generalized method of moments estimator. The data sample consists of 10 of the most liquid stocks listed on the London Stock Exchange, for which I access tick-by-tick data on trades and quotes from Refinitive.
The implied price impact that I find is of a reasonable magnitude. For an average market order, the implied price impact ranges between 12% and 22% of the spread for the different stocks. The main benchmark model is an SVAR model with returns, market orders, and a limit order flow. In addition, I estimate a Glosten–Harris (1988) price impact regression, since it provides me with the opportunity for comparison against Sandås (2001). I find that the implied price impact estimates are 18% lower than the estimates from the SVAR model and 80% higher than those from the Glosten–Harris price impact regression. The estimates thus line up within the benchmark estimates.

I find that the methodology, which did not work well using data from the 1990s, works better today. The methodology offers an alternative way of analyzing limit order information. I shed new light on limit order informativeness, thus contributing to the contemporary literature on the increased information content of limit orders.

**Article 4: Determinants of Limit Order Cancellations**

Together with Björn Hagströmer and Lars Nordén, I investigate the economic rationale behind limit order cancellations from the perspective of liquidity suppliers. We predict that investors cancel limit orders when they become unprofitable, that is, when the expected revenue of the order no longer exceeds the expected cost. Our empirical approach is based on the fact that variation in the expected profitability of a limit order can be gauged from changes in the state of the order book. A novel prediction driven by adverse selection costs is that changes in the order queue position are an important determinant of cancellations.

We use a proprietary data set that contains all the LOB updates on Nasdaq Stockholm during May 2014. The data allow us to track each limit order from submission
to either cancellation or execution and to trace the order time priority, which is key to testing the model predictions. We estimate a stratified proportional duration hazard model, which allows us to analyze the impact of changes in the expected profits during the limit order lifetime while controlling for executions, order-specific market conditions, and stock-level effects. Our findings strongly support the prediction that a limit order is more likely to be cancelled following a negative shock to its expected profit.

We identify HFTs at the trading firm level and analyze their limit order activities relative to other trading firms. Orders submitted by HFTs have shorter lifetimes, but we find that the determinants of cancellations are the same as for other trading firms. The main difference relative to other trading firms is that HFTs respond more strongly to changes in the expected profits of their limit orders.

A key aspect of our contribution to the understanding of limit order cancellations is that we adopt the perspective of a liquidity supplier. The literature suggests alternative explanations for limit order cancellations. Both theoretical and empirical studies suggest that frequent cancellations are consistent with liquidity demand strategies, such as order price revisions. An order price revision is a case in which a new order is submitted at a different price immediately after a cancellation. We show that at most 15.6% of all cancellations in our sample are part of order price revisions.

Our empirical evidence points to frequent order cancellations as a feature of modern market making. Liquidity suppliers closely monitor even marginal fluctuations in fundamental values and cancel orders that are not expected to be profitable. HFTs, known to invest in low-latency technology, adhere particularly strongly to the economics of limit order cancellations.
These results have implications for initiatives aimed at curbing the frequency of cancellations. The fact that traders, both fast and slow, adhere to a supply-related economic rationale in their order management implies that curbs on cancellations are likely to restrict the ability of market makers to fine-tune their outstanding orders. Hence, policies aimed at curbing order cancellations instead risk curbing liquidity supply.
References


