Charles University

Faculty of Social Sciences
Institute of Economic Studies

MASTER'S THESIS

Comparison of continuous and frequent batch auctions

Author: **Bc. Oskar Gottlieb**

Supervisor: **RNDr. Martin Šmíd Ph.D.**

Academic Year: **2017/2018**

Declaration of Authorship

The author hereby declares that he compiled this thesis independently, using only the listed resources and literature, and the thesis has not been used to obtain a different or the same degree.

The author grants to Charles University permission to reproduce and to distribute copies of this thesis document in whole or in part.

Prague, May 10, 2018

Signature

Acknowledgments

I would like to thank my master thesis supervisor Martin Šmíd for the guidance, continuous feedback and suggestions which greatly improved the quality of this thesis.

Abstract

We simulate a fragmented market and study three types of agents and their interactions in continuous trading and frequent-batch auctions. We model the markets using the agent-based modeling approach. There are two exchanges on which one asset is being traded by zero-intelligence (ZI) traders, market makers and a latency arbitrageur. The former two agents are marked as slow traders, the arbitrageur is a fast trader - fast trader has perfect information about the market, slow traders are dependent on the (possibly lagged) NBBO information provided by the regulator. Our main metric is the surplus of ZI traders, we also measure other market's characteristics. We then simulate the market for different delays of the NBBO delay and we find that under certain conditions and until certain length, the batch auctions are beneficial to ZI traders, as they reduce the advantage and therefore the profit of the fast trader.

JEL Classification C63, D47, G18, H39

Keywords High frequency trading, latency arbitrage, Zero-intelligence traders, market makers, Agent-based model

Author's e-mail `77460949@fsv.cuni.cz`

Supervisor's e-mail `smid@utia.cas.cz`

Abstrakt

V této práci simulujeme trh se třemi typy agentů a pozorujeme jejich chování v režimu spojitého obchodování a krátkých a častých aukcích. Používáme metodu multiagentního modelování. Základ našeho trhu tvoří dvě burzy, na kterých se obchoduje jeden statek. Naše agenty lze rozdělit na pomalé - Zero intelligence obchodníci a tvůrci trhu a rychlé - arbitrážník. Pomalí obchodníci dostávají informace o stavu na trhu od regulátora, tato informace může mít zpoždění oproti situaci na trhu. Arbitrážník má přístup k oběma burzám s nulovým zpožděním, předpokládáme u něj nekonečnou rychlost. Hlavním ukazatelem podle kterého posuzujeme kvalitu trhu je nadbytek ZI obchodníka, ale sledujeme také další charakteristiky trhu. V trhu simulujeme zpoždění informace od regulátora pomalým obchodníkům a zkoušíme různé délky aukcí. Zjistili jsme, že pokud jsou splněny určité podmínky, má ZI obchodník vyšší nadbytek v režimu krátkých aukcí a to i pokud má zpožděnou informaci od regulátora.

Klasifikace JEL C63, D47, G18, H39

Klíčová slova Vysokofrekvenční obchodování, arbitráž, Multiagentní modelování, Zero-intelligence obchodník

E-mail autora 77460949@fsv.cuni.cz

E-mail vedoucího práce smid@utia.cas.cz

Contents

List of Tables

List of Figures

Acronyms

ABM Agent-Based Modelling

HFT High-frequency trading

LA Latency Arbitrageur

MM Market Maker

NBBO National Best Bid and Offer

ZI Zero-Intelligence

Master's Thesis Proposal

Author	Bc. Oskar Gottlieb
Supervisor	RNDr. Martin Šmíd Ph.D.
Proposed topic	Comparison of continuous and frequent batch auctions

Motivation The financial markets have undergone an enormous change in the last few decades, mainly thanks to advances in computer sciences and information technologies in general. This change is generally regarded as a positive development - for instance the transaction costs have gone down significantly, making the markets accessible not only to large institutions but to retail investors as well. One of the most influential concepts is that of high frequency trading (HFT) - thanks to high-performance computers, we are now able to let computers trade, making tens of thousands of transactions per second.

HFT also became quite a controversial topic of discussion and of interest to many researchers. The proponents of HFT argue that it further helped drive down the transaction costs, as high-frequency traders usually act as market makers, replacing the inefficient human market makers and specialists, narrowing further the bid-ask spread. On the other hand HFT is usually criticized as it is associated with many black-hat methods and with some adverse events such as flash crashes. Just to name a few, high-frequency traders have been many times fined due to techniques such as quote-stuffing, front-running or spoofing.

Leaving the general critique aside, HFT nowadays is almost at its limits when it comes to the physical restrictions. Light, hence information, can only travel so fast. That is why we saw immense investment into the trading infrastructure, the most famous example is that of $300m cable which decreased the time it takes the information to go from New York to Chicago by three milliseconds (that is $3 * 10^{-3}$ seconds). Such investment is incentivized mainly because trading today is conducted on the first-come first serve basis in a continuous fashion. Therefore only a slight time-advantage can be the difference between making and losing money as a HFT firm.

Budish et al. (2015) have proposed a different mechanism - frequent batch auctions. With such model, the marginal speed difference is not as important and traders should focus more on price rather than on the speed. It is similar mechanism to how current price scale works, instead of having a continuous one different markets have different tick sizes - discrete points at which trading can take place. The logic here is that lower price differential than a tick size is economically insignificant and therefore should be ignored. Similarly marginal time differences should be ignored in the same way. In my thesis I would like to look into metrics based on which we could assess suitability of the two models for various trading environment.

Hypotheses

> Hypothesis #1: Batch auctions diminish the importance of the speed of execution.
>
> Hypothesis #2: Batch auctions are more socially optimal way of organizing the markets.
>
> Hypothesis #3: High-frequency traders will be most influenced by the change in market structure.

Methodology Due to the nature of our problem, agent-based model is an appropriate approach - it bypasses the need of having the data and gives us enough flexibility with the actual modeling scenarios. We will be using the zero-intelligence agents Gode et al. (1993), as these traders under given constraint approximate behavior of traders in the real markets. This way we can model the behavior of various actors e.g. liquidity traders (LT) will represent those interested in the underlying asset, high frequency traders will mainly act as market makers, making use of their low-latency between the exchanges and their servers. The model will be studied from point of view of characteristics of the market environment, be it the efficiency of price discovery, short-term liquidity or transaction costs.

Expected Contribution Although some studies on the topic of the ideal market organization have already been published, the problem can still be tackled from many different points of view. Hence the thesis will most likely focus on a subset of conditions/market situations and study the behavior of agents under these constraints. The outcome of this thesis should then should shed some light on the particular subset and its implication towards policy making and exchange regulations.

Outline

1. Motivation - summary of proponents and critics of HTF. Literature review.

2. Methodology - construction of exchanges and of the agent model.
3. Criteria - explanation of metrics based on which the models will be compared.
4. Running the simulation, summary of results
5. Concluding remarks; implications for exchange policies and regulations.

Core bibliography

Budish, E., Cramton, P., & Shim, J. (2015). The high-frequency trading arms race: Frequent batch auctions as a market design response. The Quarterly Journal of Economics, 130(4), 1547-1621.

Brons, Burlakov, Y., Kamal, M., & Salvadore, M. (2012). Optimal limit order execution in a simple model for market microstructure dynamics.

Conrad, J., Wahal, S., & Xiang, J. (2015). High-frequency quoting, trading, and the efficiency of prices. Journal of Financial Economics, 116(2), 271-291.

Gode, D. K., & Sunder, S. (1993). Allocative efficiency of markets with zero-intelligence traders: Market as a partial substitute for individual rationality. Journal of political economy, 101(1), 119-137.

LeBaron, B. (2001). A builder's guide to agent-based financial markets. Quantitative Finance, 1(2), 254-261.

Menkveld, A. J., & Zoican, M. A. (2017). Need for speed? Exchange latency and liquidity. The Review of Financial Studies, 30(4), 1188-1228.

O'Hara, M. (2015). High frequency market microstructure. Journal of Financial Economics, 116(2), 257-270.

Rojcek, J., & Ziegler, A. (2016). High-frequency trading in limit order markets: Equilibrium impact and regulation.

Author

Supervisor

Chapter 1

Introduction

In recent years the financial markets have undergone a rapid change, fueled mainly by the global technological progress. The markets, more specifically the access to them became much more widespread, thanks first to phones, cell phones and later to the internet. This change meant on average lower transaction costs and higher volume and liquidity to the investors and therefore is generally regarded as beneficial. The transformation has changed the playing field and it brought in new market participants who rely heavily on technology and computational power. We refer to these traders for lack of a better word as high-frequency traders, mainly due to the nature of their trading, as they are making a large number of transactions in very short time periods. Due to the fact that computers are able to process information from financial markets at a much higher rate than humans, high-frequency trading (HFT) has been often labeled as manipulative and generally harmful towards other market participants due to its unfair advantage over regular traders and investors. As HFT refers simply to the time frame at which orders are executed, we do not do any generalizing conclusions to all high-frequency traders, rather we focus on a smaller subset that can be studied and evaluated separately from the rest of the market participants.

This thesis focuses on one aspect of the market design, which is the frequency at which market orders are cleared. Today the majority of securities markets work on a serial basis, orders arrive at the exchange and are put in the queue based on the time of their arrival. A trade is made when the buying and selling order have the same price, or if buying (selling) order which arrives to the market as second is above (below) the counterparty. As the pairing of orders is determined by the time of their arrival, even a small speed advantage could, therefore, mean the difference between getting the order executed or having it left in the order book untouched. We refer to this market organization as a continuous double auction (CDA).

The proponents of alternative market designs argue that the CDA by design is not socially optimal. Their proposed alternatives usually revolve around diminishing the importance of speed which they perceive to be the biggest drawback of the CDAs. We simulate a market with the CDAs and batch-auctions clearing. We compare the performance of the two clearing mechanisms by analyzing the surplus/profit of our agents, the transaction costs, the liquidity and the price volatility.

The thesis has the following structure: Chapter 2 presents the motivation for this thesis and Chapter 3 summarizes the already existing literature. Chapter 4 describes the agents and models that are used. Chapter 5 presents the environment and market mechanisms under which the agents operate. Chapter 6 presents the results of our simulation. The conclusion summarizes our findings.

Chapter 2

Motivation

Let us first look at the primary motivation for even considering a different market design than what the CDA offers. In the recent 10 years, we have seen a lot of criticism directed towards high-frequency traders and their behavior on various exchanges. They have been associated with black-hat (malicious) practices such as quote stuffing, front-running of orders, order spoofing, marking the close and have been also accused of being the leading cause of many shflash crashes. Hence establishing a slightly different market environment in which speed is not as important of a factor could diminish the adverse effects of HFT. We shortly look at the already mentioned black hat techniques to illustrate in what ways exactly can the HFT make use of their speed advantage.

Rossi *et al.* (2015) summarize recent investigations which took place as a reaction to various HFT's misconducts. They analyze multiple cases where a trading company was charged with malicious practices. E.g. Athena capital was the main player in some stocks of the NASDAQ exchange. At times, they were responsible for more than 70% of the volume in these stocks. This enabled them to set the closing price, which usually serves as a reference price for other products or contracts which are traded off the exchange. This practice is called marking the close and it is illegal.

Another case mentioned by Rossi *et al.* (2015) is that of Panther Energy Trading LLC. They were charged with spoofing, a technique which consists of using non-bona fide orders - orders whose primary goal is not getting executed but rather create confusion and false impressions. This technique is regarded as harmful towards the quality of the market.

Sar (2017) states that order spoofing misrepresents the order book and argues that section 747 (Antidisruptive practices authority) of the Dodd-Frank

Act is a step in the right direction towards higher market quality. This section addresses both spoofing and marking the close as malicious techniques and lowers the barriers of an investigation against the malpractice. The fact that these practices are included in the Dodd-Frank act shows the regulator's stance on the subject.

The events of the Flash Crash of May 6th, 2010 were studied in detail by Kirilenko *et al.* (2011). During this day, the Dow Jones Industrial Average (DJIA) lost roughly 9% of its value in little over 30 minutes, but it recovered the loss in a similarly short time period. Kirilenko *et al.* (2011) analyzed the actions of all market participants in the E-mini S&P 500 index contract, coming to the conclusion that high-frequency (HF) traders were not the primary cause of the initial downward move. They were however aggressively selling in the latter phase of the downward move.

An article by Hunsader (2013) points at another way in which the markets are being manipulated, this time he points at the physical impossibility of HFT's reaction to a piece of public news. The Federal Reserve (FED) announcement was released from Washington and shortly after the HFT were trading off that news in Chicago. The problem is that their reaction times were extremely quick. According to the author, the only reasonable explanation is that they had to have the information upfront before the announcement. Otherwise, the author argues that the information would need to travel faster than the speed of light. This problem is already systematically tackled and is considered by the SEC as a market manipulation (insider trading).

Front-running is another malicious practice mainly known from days when trading in physical pits was still prevalent. It describes a situation in which a market participant knows of the incoming order flow and can act upon this information before the order actually arrives. This equals to an almost sure profit, knowing that a large buy market order (market order are executed immediately against the resting limit orders) will come onto the exchange means that any trader with this information can buy the asset, only to sell it a short while later for a profit.

Of course, this practice has been established illegal for brokers, where the profit of the broker comes directly as a loss to the client, who in this case gets a worse execution price. Front-running and general faster access to information could have been achieved by buying a seat on a given exchange. Today such an advantage is achievable via co-location, an option offered by most exchanges by placing servers as close to the exchange as possible. This assures that colocated traders are the first to receive information about the order flow and can act upon it. The principle nowadays remains the same, however, the time scale at which front-running takes place has changed.

Tong (2015) examines the impact of HFT activity on the execution costs that arise to large institutional investors. More specifically, she studies the execution shortfall, the difference between order's weighted execution price and the market price at the time of the order's arrival. By controlling for factors such as stock liquidity or corporate events Tong finds HFT to be causing an average increase in cost to the average institutional investor, as the execution shortfall increases with HFT's activity.

The importance of the speed has been thoroughly documented by Baron *et al.* (2016). The authors use the data from the Swedish stock exchange OMX Stockholm and the Transaction Reporting System (TRS) along with a list of firms which they have identified as HFT firms. Baron *et al.* find a significant relation between firm's speed and profitability. They found that firms with lower relative latency have on average higher Sharpe ratio (risk-adjusted return). The important distinction is that even the lower latency (or higher speed), is not measured absolutely but rather relatively - against other market participants.

Similarly to front-running Budish *et al.* (2015) looked into the arbitrage potential which exists between closely correlated markets. At a daily, hourly or even minute timeframe, the arbitrage potential has diminished over the years. However, it stays constant on the very short timeframe (microseconds). Budish *et al.* (2015) argue that the correlation breaks down in a very short timeframe and that this is a negative externality of the continuous double auction, as over time the resulting arbitrage has not been competed away.

Due to the size of the potential profit (Budish estimates it to be lower nine-digits, Lewis & Baker assess it in the vicinity of $20B USD), the investment that goes into building an infrastructure capable of these operations is enormous and it ends up being a socially not optimal race towards the fastest operations. Budish *et al.* (2015) mark this as a consequence of a market design where traders with marginally lower latency get the majority of the profit. They, therefore, propose a different approach - a frequent batch auction which should ease the pressure which is currently put on the speed of the market participants, making the markets more socially optimal in the process.

Most of the critique of the HF traders was related to the market regulation and imposing policies which ease the ex-post analysis of the market participants' activity. This is also one the main goals of the Dodd-Frank Act when it comes to malicious trading activities. The work by Budish *et al.* (2015) is different, as it suggests a new market design which in theory should eliminate some of the drawbacks of the current design. The paper raises interesting questions such as what criteria should be used to determine which type of auction is better or what the length of the auction should be.

Chapter 3

Literature review

In our model, we compare the standard limit continuous trading in a fragmented market with the batch auction approach which clears the market periodically in intervals of length δ. In this section, we first shortly summarize the literature covering HFT's relation with market quality. Next, we go over the list of alternative market clearing mechanisms. Finally, we cover the concept of zero-intelligence traders - agents which form the core of the model.

HFT's impact on market quality

High-frequency trading only describes one attribute of the trader's strategy - its speed, yet it encompasses a variety of strategies and approaches - as Biais *et al.* (2014) describe it, HFT strategies are heterogeneous in nature and their impact on the market depends on the strategy used.

Brogaard *et al.* (2014) use the standard NASDAQ dataset to study the effect HFT has on the price discovery process. They find that high-frequency traders often supply liquidity (submit limit orders) on the thinner side of the order book and demand liquidity (market orders, or limit orders with high chances of immediate execution) on the other side. The authors studied the market during and following the financial crisis of 2008, where the volatility has been abnormally high. They analyze the behavior around news announcement and order book imbalances - and conclude that HFT firms improve price discovery and that they were profitable, but there was no direct effect on price's instability.

Boehmer *et al.* (2015) study the effect of algorithmic trading more generally - as their dataset does not include direct identification of HFT firms. The proxy they use for the average rate of algorithmic trading in the market is the rate of order submission and cancellation. This proxy variable should represent the speed with which the participants react to new market information and adjust their orders. Given their limited dataset the interpretation of HFT's impact is not straightforward (they do not have a way to distinguish between market with HFT firm and one with a larger proportion of traders using the general algorithmic trading). The effect of algorithmic trading is generally positive with regards to liquidity and efficiency, with exception of days with excessive returns and small-cap stocks.

Brogaard *et al.* (2017) focus on extreme events and HFT's activity during these events which are characterized by high volatility. They use the same dataset as in Brogaard *et al.* (2014). The events they study are characterized by large moves in the National Best Bid and Offer (NBBO). During these events, the HFT firms on average provide liquidity and therefore ease the process of price reversal back to its original value. The exceptions are the situations where there are multiple events happening at once in multiple stocks. Then HFT firms take more liquidity from the market than they provide. This last statement is in line with the events of the Flash Crash, where the entire index S&P 500 dropped also partially due to HFT firms' activity.

A comprehensive analysis of HFT related papers has been conducted by Brogaard *et al.* (2014) and includes a thorough of HFT's influence on the market.

Alternative market clearing mechanisms

There were multiple alternatives proposed to the prevailing CDA structure, their main motivation is usually addressing the drawbacks of the current clearing mechanism.

Hoffmann (2014) claims that the current continuous structure of the market leads to significant gains for the faster traders, who are willing to invest substantial amounts into the infrastructure in order to further increase their speed advantage. The problem, Hoffmann (2014) argues, is in the nature of the market opportunities, where only the fastest market participant is eligible to them. Therefore from an individual point of view, it does make sense to invest heavily in a literal speed arms race. In his, model Hoffmann says that even a cancellation fee for orders in the limit order book might be a partial solution to his problem, which however would not fully deal with the adverse selection of the resting limit orders.

Similar conclusions are made by Biais *et al.* (2015), their proposed mechanisms for dealing with the negative effects of fast traders are either a creation of exchanges for slow traders only, the other being the introduction of the Pigovian tax, as an attempt to deal with the negative externalities produced by fast traders. None of these options is truly optimal, as according to the authors, the former would force the investment into technology to be below its optimal level, the latter can be optimized for by the companies in a way that would not make the final equilibrium optimal.

Kyle & Lee (2017) also argue that the current standard limit order book model which is employed by most exchanges can be exploited by faster traders. They propose a change to this model by making it completely continuous in all dimensions other than time only. In the standard model, the HF trader has three main speed-related advantages which the new model mitigates. He can pick off resting orders faster than slow traders, cancel his resting orders faster than a slow trader can execute against them and he has a better position in the queue, as the orders are sorted in the order in which they arrived onto the exchange.

Kyle & Lee (2017) propose a model which is not based on individual messages as we know it today. Instead, a trader defines the price and quantity at which he is willing to buy, where both parameters are discrete in nature. In contrast to the continuous model, a trader defines the minimum and maximum price at which he is willing to trade along with a quantity and maximum rate of execution. This model is in nature quite similar to the supply and demand functions which are in microeconomics usually drawn as linear functions of price. It also interestingly gets rid of discrete prices, this action in itself would however not be advantageous to the investor as it would mean that HF traders could now make use of their speed in two dimensions - time and price.

The paper by Harris (2013) offers yet another way of leveling the playing field. Once an order is received by the exchange, the author says that the exchange should add a random delay drawn from a uniform distribution before the order is processed. He claims that this change should add a level of randomness to the execution such that the time advantage of faster traders is not as noticeable. However, there is no model nor a theoretical background that we could build on.

Brown & Yang (2016) analyzed the role of a speed bump in a market similar to the financial exchanges. They looked at the mechanism of Betfair - an online betting exchange, where an individual can buy or sell his bet, bypassing the traditional bookmakers. Betfair has been using this mechanism for more than 10 years now. The underlying mechanism is similar to the one used on stock exchanges, there is a limit order book for a particular bet. One can submit orders to the limit order book or can trade immediately with a market order. Brown & Yang (2016) focus on the adverse selection that is in the presence of a significant predictive power in the market orders for the future direction of the market moves as well as on the market quality (volatility, liquidity) and cancellation of limit orders. They argue that although there can be seen a protection of the slower traders, the influence on other metrics related to market quality are less clear.

The paper by Budish *et al.* (2015) is also inclined towards a different market structure, saying that the current organization of the markets is flawed as it is in direct conflict even with the weak version of the Efficient Market Hypothesis (EMH), as there is the rent implicitly in the market's structure for whoever is it the fastest participant. Budish *et al.* propose a model with n liquidity snipers. All of these participants have the same goal, to make money of the stalling quotes of other traders. With a continuous model, any trader who is slower than the liquidity sniper would lose money due to his lack of speed (the sniper is faster at hitting a stalling quote than the trader is at canceling it).

By replacing the serial processing of orders with frequent auctions, the authors concluded that this has greatly increased the probability of canceling the quote before a liquidity sniper could have made the trade. The relative importance of speed decreased substantially in a way that there most likely would not be as large of an incentive to invest heavily in speed. Policies other than a batch-auction regime were met with strong arguments.

Budish *et al.* (2015) assert that a Tobin-tax (tax proposed to reduce excessive trading by imposing a small tax on every single trade) would address the snipping quote problem but at a cost that would have to be born by the individual investor. Other policies which are directly aimed at the perceived negative externalities of HFT according to the authors are also not the solution because they focus on the effect of the market structure rather than on the cause. Some of them such as the minimum resting time would go against the problem at hand - under these rules, the order would have to stay in the order book for a minimum duration and if there were a jump during that time, the resting order would be filled with a probability of 1, as there would be nothing the trader could do against the liquidity snipers.

Bishop (2017) has shown in a fairly transparent way the way in which the

IEX has coped with order snipping in the past. IEX today works as a dark pool - it is an exchange where the limit order book is not visible. The only information that the exchange makes visible to the public is the last price traded for a given security. She shows the way in which the exchange has protected its clients against something they call "Quote crumbling". The US securities market is very fragmented and a quote is so-called crumbling once we see a change in bid and ask prices quoted over multiple exchanges - the quote is most likely about to change from one set of price to another one.

Given that in the US an order can not trade at a price worse than what is offered by the NBBO, the exchange must route the order if it's possible to get a better execution at a different venue. Stalling quote in this regard would be one that is in the order book at one of the exchanges and can be picked off by a different trader who is able to identify a crumbling order sooner. According to the authors, such execution goes against the purpose of the NBBO. At the time of the execution, the actual NBBO information that the exchange sees is not the one that exists in the market. A faster market participant can make use of his speed advantage and can pick off the resting orders before the true NBBO information gets to them. The exchange here offers protection via a speed bump. this time is used for an update of the NBBO, after which the incoming snipping order is no longer relevant. Bishop shows a rather simple model that the exchange has used in the past and it has decreased the percentage of adversely selected orders substantially.

The performance of the NBBO has been more thoroughly analyzed by Ding *et al.* (2014). By comparing the data of the NBBO from the Securities Information Processor (SIP) and of an aggregate of the NBBO (authors have been gathering the data and creating their own datafeed of the NBBO) they found discrepancies between the two streams of information. This discrepancy is the result of the way the NBBO is currently calculated - a market participant (after having done a larger initial investment) can aggregate the data ahead of the NBBO and can make use of that information ahead of the general public. Ding *et al.* found that there was a positive relationship between the price and daily volume of the stock and the number of such discrepancies. Yet the authors have not done a quantitative analysis as to how one can make use of this speed advantage, they only conclude that the NBBO is actually quite slow given the number of discrepancies observed. They do raise the question as to the necessity of having an NBBO even in a fragmented market, given the frequency at which the NBBO does not correspond to reality.

Wah & Wellman (2017) study the effects of latency arbitrage under various market regimes. They base their model on the US securities market where the fragmentation resulted in the creation of multiple venues on which the same asset is traded. Latency arbitrageur is an agent who is infinitely fast (he knows of any price change exactly at the time such price change happens) and can make use of the price discrepancy between the venues in case there is a profit for arbitrage - best bid (ask) price of one exchange overlaps best ask (bid) of the other exchange. For simplicity sake, the authors employed only two exchanges, the role of the individual investor has been passed onto ZI traders. The trader decides to trade according to a Poisson process with intensity λ.

Two market system Wah and agent-based modelling

The idea of the batch-auctions as proposed by Budish *et al.* (2015) has been modeled in the work of Wah & Wellman (2017). They found that the presence of a latency arbitrageur in a standard CDA decreases the efficiency of the market, as his existence in the market is of no real benefit to his trading counterparts. By replacing the two exchanges with a central call market, the authors saw an increase in the overall market efficiency. An increase was also found if the agent of latency arbitrageur was removed completely. Unfortunately, even though both of these options would reduce inefficiencies, it is unlikely that these would be applied in real life as the latency arbitrageur can freely enter the markets at any time and US markets are fragmented by design.

Wah *et al.* (2016) further develop the model, simulating an environment with fast and slow traders and both batch auction exchange and standard CDA exchange. Both types of traders do have the choice of picking either market, according to their preference. The authors found the frequent batch auction to serve as a better structural ground for slower traders, as their welfare is higher in these markets. Faster traders in a way prey onto the existing slower traders and they are willing to trade on both markets, even though their welfare is higher when the trading is done continuously.

Zero-Intelligence traders

The Zero-Intelligence traders are the core of our model. Gode & Sunder (1993) compared Zero-Intelligence (ZI) traders with human traders, both were trading the same asset and both received an initial endowment. Forcing a restriction upon the ZI traders in form of a budget constraint has shifted zero intelligence from diverging to converging towards an equilibrium price. Human traders were still converging faster and their allocation was more efficient, but the main point of the author's paper was that learning, profit, motivation or intelligence were not necessary conditions of convergence towards an equilibrium price.

Gode & Sunder claim that the structure of the market and market constraints are sufficient for effective allocation of an asset. The concept of ZI traders has since been used in various agent-based models which work with financial data e.g. Duffy & Ünver (2006) use ZI as the basis of their model, simulating price bubbles and crashes quite similar to those caused by human traders. Farmer *et al.* (2005) apply ZI principles in their study of the London Stock Exchange (LSE) equity markets, where they treat the incoming orders as revealed preferences of ZI traders.

However, the ZI approach has also been subject to some critique. Walia *et al.* (2003) modeled the short-term market for the electricity, where flexibility costs (costs due to changes to earlier commitments) are expensive. They argue that due to changes in trader's valuation, the CDA as initially proposed does not deliver as efficient of an outcome. Cliff *et al.* (1997) claim that the market structure alone is not sufficient condition for the efficient allocation, as the transaction price is only close to its theoretical equilibrium if the supply and demand functions are symmetric. Instead, they propose slightly enhanced agents - Zero-intelligence Plus (ZIP) traders which adjust their trading quotes to current market situation and to their personal inventory. Cliff *et al.* (1997) went on to show that ZIP's trader performance resembles the human trader's behavior more than ZI did.

Chapter 4

Methodology

Our model closely follows Wah & Wellman (2017). It is modeled using the two exchanges on which a single security is traded by three types of traders. In section 4.1 we introduce the general concepts of the two trading venues model. Section 4.2 describes the agents and defines their behavior, finally in section 4.3 we define the metrics and characteristics which we then analyze. This approach has been chosen because of two reasons. Firstly, batch auctions from the theoretical point of view seem to be the best approach to counter most of the drawbacks of the current market design. Secondly, the previous research of batch auctions does not work with any existing data but rather simulates the data on the spot. This is very practical, because any real high-frequency data provided by an exchange are hard to get and extremely costly. Also, any historical data could not be used reliably as the batch auction regime as of today is not fully adopted on any exchange.

4.1 Two exchanges model

4.1.1 Order book

The standard limit order book in our model represents the exchange. An order book of a given security is a collection of buying and selling limit orders. The word limit means that every order has a specified maximum (minimum) price at which the trader is willing to buy (sell) the asset. An example of the limit order book is shown in Figure 4.1. The buy (sell) side of the market is called the bid (ask) side and the orders on both sides of the market are ordered with a price-time priority. We refer to the price of the bid and ask orders which are first in their respective queues as best bid and best ask respectively. In our example, the best bid is 100 and best ask is 101. The difference between the best bid and best ask is referred to as the bid-ask spread, in Figure 4.1 this spread equals 1.

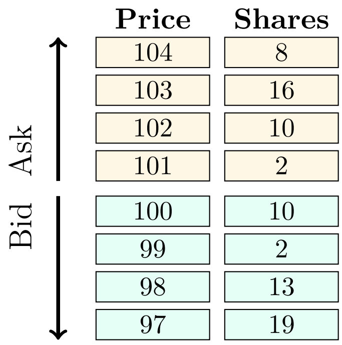


Figure 4.1: An example of a limit order book

Price-time priority

To illustrate the price-time priority, let us now have a market with three traders A, B and C. AS the market opens, traders A and C already have their limit buy and sell orders in at prices 100 and 101 respectively. Trader B enters the market at time t_1 by submitting a buying order at price 100. Given that there already is a limit order at this price, trader B is currently at the second place of the queue of bid orders. Figure 4.1 depicts the situation. At time t_2 trader C decides he wants to sell his asset immediately. He, therefore, submits a new sell order at time t_2 at price 100 and given trader A was the first to submit his order at this particular price level, he is the one whose order got cleared. The last order book in Figure 4.1 shows the state after the clearing took place at time t_2. This price-time priority holds for both market clearing mechanisms, given the times t refer to different clearing times.

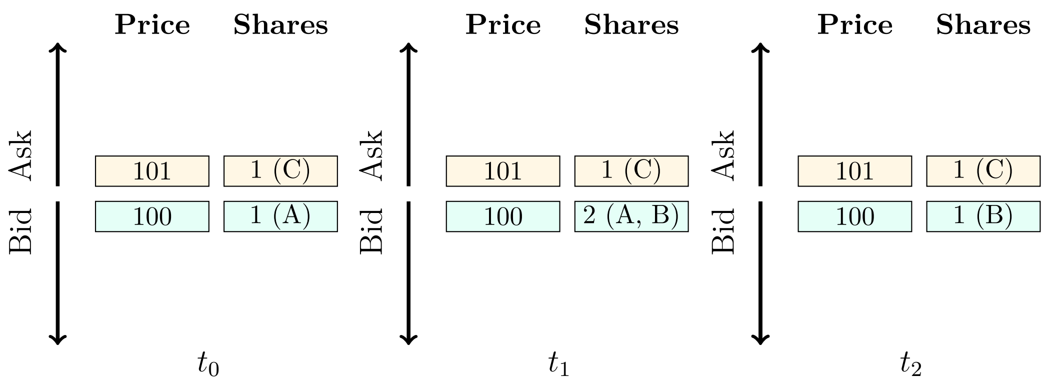


Figure 4.2: An example of the price-time priority

Continuous double auction

Nowadays the CDA is the standard market clearing, implemented at most of the stock exchanges in the world. When the exchange is open, it enables anyone to trade virtually at any time as the time is treated as a continuous variable. As the price-time priority still puts enough emphasis on time, even a marginal time difference ($t_1 - t_0 = \epsilon > 0$) means that due to the serial processing of orders, one order does have a priority in the queue simply because it arrived marginally faster. In practice, one way how some of the exchanges are trying to mitigate the effects of serial processing is weighing the executions against resting limit orders by their respective weights. For instance, orders which are at the end of the queue also can get partially executed if they are large enough.

In our model all limit orders are of quantity 1, therefore no weighting is applied. We also impose a restriction that each order has a unique arrival time - no two new orders can be processed at the same time. Therefore the order in the queue, as well as the order in which the orders get executed, follows the First-In-First-Out (FIFO) principles.

After an order is received by the order book, there are two possible outcomes - it can either be cleared against a resting limit order or it can be added into the order book. Let us recall our example in Figure 4.1 and let's assume that the exchange is clearing using continuous trading. At time t_1 we see the first case as trader B's order is added into the order book because the order's price is below that of the best ask. The second case takes place at time t_2. The new order submitted by trader C crosses the spread (is submitted at price better than or equal to that of the best bid) and it gets cleared at price 100. The clearing quantity equals to one, clearing price is that of the resting limit order. If trader C were to submit his order at price lower than 100, he would still get the execution price of the resting limit order.

Frequent-batch auctions

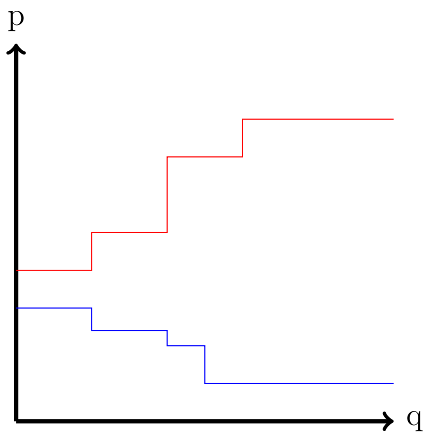
The frequent-batch auction design suggests a different treatment on the micro level and is slightly more complicated due to the various situations which might arise. On the macro level, the order book still consists of two queues of waiting limit orders. But instead of having a continuous clearing of the order book we now have the trading session separated into a sequence of n equally long segments during which the order book accumulates all of the changes and then the market clears at the end of each segment. Such segments should be very short in nature, in other words, their frequency is expected to be in the vicinity of $5 * 10^2 \sim 10^3$ auctions every second.

Let us define this formally, we have a trading session which is T time units long. There are τ auctions during every time unit. The total number of batch auctions then simply equals the number of time units multiplied by the number of auctions per unit of time - $N = T * \tau$. Orders which arrive during the interval $(t_i - t_{i+1}]$ are treated with the same time priority. Processing new orders at the end of each auction can result in a total of three cases which are represented in Figure 4.3. Blue and red line represent the cumulative buying and selling orders respectively.

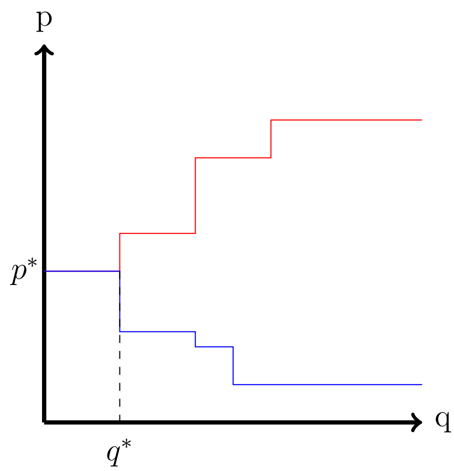
First the bid and ask orders do not cross at all - the bid-ask spread is strictly larger than zero. This situation is illustrated in figure 4.3a. At this point, all the orders are added as limit orders into the order book. If any two orders are added into the orde rbook at the same price, their real order in the queue is picked randomly. The orders are added to the order book on top of already remaining resting orders (at the respective price levels). Therefore on a macro level, the price-time priority still holds from one auction to the next. In the real world we would also have to account for the fact that the size is different across orders. In our model, we do not have to weight the orders as we only work with orders of size 1.

The second option is that the bid and ask columns overlap exactly at one price only. This establishes the clearing price as the one where the bid and ask columns overlap. The number of cleared contracts corresponds to the quantity $min(q_{BID}, q_{ASK})$ where q_{BID}, q_{ASK} are the quantities of the bid and ask orders respectively at the clearing price level.

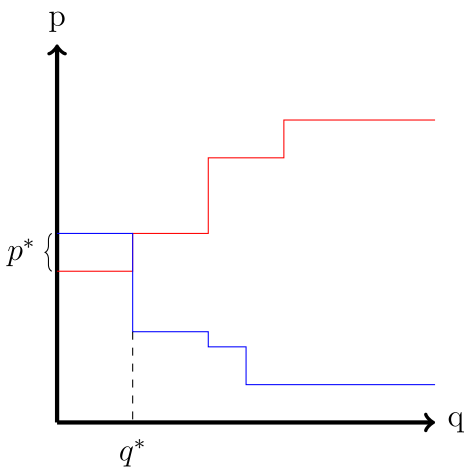
Any remaining orders which have been submitted at this clearing price and cannot be cleared as there are not enough limit orders on the other side of the market are then added to the order book. Also, other orders which have been added at a worse price and were not cleared are added to the order book at top of already existing orders. Finally if at the clearing price there are still any limit orders left which have been added during the previous auction, they simply remain in the order book. Again if there are multiple orders being added at one price level during one auction, their position in which they are processed is picked randomly.



a No market clearing



b Crossing at one price level only



c Clearing across multiple price levels

Figure 4.3: Three states of frequent batch auction

Finally only in batch auctions is it possible to have the best bid above best ask price. That is of course only temporairly before the market clears at the end of the auction. In continuous trading, such state would not be possible, because any bid (ask) order submitted above (below) current ask (bid) price would be immediately paired. In such a situation we still need to define one clearing price and it corresponds to the mean of the highest bid price of an order which would clear with the lowest ask price.

All new orders are added into the order book in the same way we have already defined. Next, we take each best order at both bid and ask side of the market and we compare their prices. If the bid-ask spread is lower than or equal to zero, both orders clear during this auction and the market keeps on clearing orders until the bid-ask spread is larger than zero. There is a single clearing price for the entire auction and it equals to

$$\frac{p_{BID} + p_{ASK}}{2} \tag{4.1}$$

where p_{BID} and p_{ASK} represent the best bid and ask price of orders which are cleared during the auction. Therefore if the price of the best bid order equals that of the best ask order (as we have specified in the previous case), the clearing price is the same. In our model, we again do not apply any weighting when deriving the clearing price. In practice, a more robust way how the clearing price could be calculated is the mean of the weighted prices of the bid and ask orders which get cleared during the auction.

An important final distinction between continuous and batch auction trading is that during the order accumulation phase, the information about the deleted and newly added orders is not available to the public. During the accumulation, the exchange behaves as a dark pool - an exchange with no publicly available information about its bid and ask prices. Only once the market clears do we see the new state of the exchange.

National Best Bid and Offer

Both exchanges fall under one regulator who updates the information about the National Best Bid and Offer (NBBO) and sends it back to the exchanges. NBBO in our model is based on the regulation enforced by the Security Exchange Commission (SEC) in the United States. In practice, the NBBO is an aggregate information with the price and name of the exchange offering the best bid and ask prices. The main purpose of the NBBO is to protect individual investors and traders who do not have the complete market information. Trader's orders are re-routed using the NBBO information from one exchange to another in order to get the best price possible.

The NBBO is a signal aggregated by the regulator across both exchanges, it is then made publicly available. The information which traders see is the market state after every clearing of the exchanges. In our model, we make use of the fact that slow traders often see the NBBO information lagged behind the real state of the market. This lagging then causes wrong order routing and arbitrage opportunities arise which in today's markets are mainly picked off by faster traders who can construct their own synthetic NBBO that is faster than the one provided by the regulator. This situation is explained in greater detail later.

Asset

There is one asset which is traded on both exchanges. The fundamental value of the asset is modeled in the same way as in LeBaron (2002), it is of a mean-reverting nature and follows a stationary auto-regressive process of order 1.

$$r_t = max(0, \mu + \alpha(r_{t-1} - \mu) + \epsilon_t) \tag{4.2}$$

In (4.2) r_t is publicly observable signal of the fundamental value of the asset at time t. For simplicity sake, we assume that the signal is perfectly correlated with the true fundamental value. The information is available to all traders with no delay. The current value of the signal depends on its past value r_{t-1} and on the mean of this process μ. In each period a shock is drawn from a normal distribution $\epsilon_t \sim \mathcal{N}(0, \sigma^2)$. The degree to which the asset reverses back to its mean is given by the mean-reversion parameter α. Given the initial values (specified later) and the characteristics of (4.2), it is unlikely that the fundamental value of the asset would drop below zero, but we want to make sure that it does not by taking the maximum of zero and the resulting value of the process.

4.2 Agents

4.2.1 Agent based modelling

We use the agent-based modeling (ABM) approach to investigate the aggregate behavior and market characteristics under various market regimes and conditions. In each simulation our two exchanges have the same properties - either both are trading continuously or using the batch auctions. We employ three main types of traders. Regular investors and low-frequency traders are be represented with a zero-intelligence trader. More sophisticated traders are modeled as market makers - only submitting passive limit orders on both sides of the market. Finally, we introduce the latency arbitrageur whose sole purpose is to profit off any arbitrage discrepancies.

4.2.2 Zero-intelligence traders

The core of our model is formed with Zero-Intelligence (ZI) traders as defined by Wah & Wellman (2017). The ZI traders were first introduced by Gode & Sunder (1993). Our agents do not have any initial endowment, they are restricted in terms of the asset's quantity constraint.

The arrival times of our ZI traders are given by a Poisson process of intensity λ. Upon arrival, the ZI trader is assigned with a probability of $p = 0.5$ to be either a buyer or a seller. In our environment of the two trading venues, each ZI trader are assigned to one of the trading exchanges by default. He basis his valuation of the asset on his private and public components. The public part is a reflection of the current fundamental value and the ZI trader estimates it using the following equation:

$$\hat{r}_t = \mu(1 - (1 - \alpha)^{T-t}) + r_t(1 - \alpha)^{T-t} \tag{4.3}$$

which can be also rewritten as:

$$\hat{r}_t = \mu + (r_t - \mu)(1 - \alpha)^{T-t}$$

where T is the length of the trading session, t is the current time, α is the mean-reversion factor and α is the mean of past asset values. In other words the estimate $\hat{r}_t$ is based on the current value of the asset r_t and on the mean of past asset values μ. As $t \to T$, the relative importance of latest asset value increases, once $t = T$ it is easy to see from the equation 4.3 that $\hat{r}_t = r_t$.

Apart from the public component, ZI trader is the only agent type who also has a private component for the valuation of the asset. The private part is based mainly on the number of assets that the trader already owns and it should reflect his preferences about his owning fewer or more of the additional assets. We impose a restriction on the number of units that the trader can be long or short at any time and we refer to this maximum (minimum) amount as q_{max}. Then ZI trader's preferences can be represented using the following vector:

$$\Phi = (\phi^{-q_{max}+1}, \ldots, \phi^{q}, \phi^{q+1}, \ldots, \phi^{q_{max}}) \tag{4.4}$$

Φ is a vector of length $2q_{max}$ of elements ϕ^q, where q is trader's current position restricted by the inequality $-q_{max} \leq q \leq q_{max}$. We construct the vector Φ by independently drawing $2q_{max}$ times from a normal distribution -

$\phi^q \sim \mathcal{N}(0, \sigma_q^2)$. We then sort ϕ^q in an non-increasing order, this way the vector Φ represents the diminishing marginal utility. Given a position q the private component is ϕ^{q+1} in case the ZI trader is a buyer or ϕ^q if he is a seller.

Putting the private and public components into one equation, we get the following cases for the asset valuation:

$$v_t = \begin{cases} \hat{r}_t + \phi^{q+1}, \text{if buying} \\ \hat{r}_t + \phi^q, \text{if selling} \end{cases} \tag{4.5}$$

We now have a mechanism for calculating the perceived value of the asset by any of the ZI traders. The price is derived by taking the perceived value and adjusting it for additional surplus which the ZI trader demands on top of the perceived value. The amount is drawn from a uniform distribution between the minimum and maximum amount that any ZI trader can demand. $s_t \sim U(S_{min}, S_{max})$. Combining the demanded surplus with (4.5) we obtain the following:

$$p_t \sim \begin{cases} U(v_t - S_{max}, v_t - S_{min}), \text{if buying} \\ U(v_t + S_{min}, v_t + S_{max}), \text{if selling} \end{cases} \tag{4.6}$$

Demanding additional surplus on one hand theoretically increases trader's potential surplus on a given trade, but it also decreases his chances of having the order executed - it represents the traditional trade-off between better execution time and better execution price.

Now let us look at the problem of order routing in case of ZI traders. First, if the trader had any resting limit order on any of the two exchanges, he cancels them in order to submit a new order. Therefore at any time each ZI trader only has up to one limit order in the order book. Our market consists of two trading venues - let e denote the number of a particular exchange, $e \in \{1, 2\}$. The trader is assigned to an exchange (denoted by X) X_e and by default routes all of his orders there. The only case at which the trader routes his orders to the other exchange is if the perceived (not actual) NBBO signal tells him that the other exchange offers an immediate and better execution than his original exchange.

In case our ZI trader is a buyer (seller), we want to see an ASK (BID) order on the second exchange to be in terms of the price below (above) the best ASK (BID) at the default exchange. Rewriting it formally, as e marks the default exchange let h denote the second exchange i.e. $e \neq h$, BID_i and ASK_i are the respective best bid and best ask price at venue i such that $i \in \{e, h, \text{NBBO}\}$ where NBBO represents the exchange which currently offers the best price of either a BID or an ASK side of the market. Let the ZI trader be a buyer, then his order is routed $\iff$ $(ASK_{\text{NBBO}} = ASK_h < ASK_e) \wedge (p_t \geq ASK_h)$. The first condition is that of a better execution price, the second implies an immediate execution. The routing is analogical in case the ZI trader is a seller.

We are not working with market orders, therefore technically all orders are sent as the limit orders with the price of p_t and are immediately matched against resting limit orders only if they are being submitted at price worse or equal to that of a resting limit order. In case the order is being rerouted and the expected resting limit order has already been executed or deleted the trader's new order is simply added to the order book at price p_t at its respective place in the queue.

4.2.3 Market makers

Market maker (MM) as an agent shares some of the characteristics of ZI traders. Our MM trader is based upon the model of Wah *et al.* (2017). Each MM is assigned to one of the exchanges by default, MM is also considered to be one of the slow traders - he is dependent on the NBBO provided by the regulator as he is not capable of constructing his own synthetic signal. MM too arrive at the market according to a Poisson process of intensity λ_{MM}.

Upon the arrival, MM deletes all of his resting limit orders from the order book and submits new ones. Market maker then is neutral when it comes to choosing a market direction. He does not submit only one order at one side of the market, instead, he submits a chain of orders on both sides of the market. There are n orders being submitted on each side of the market with δ ticks between them.

The center price of these orders is given by the estimate of the fundamental value of the asset, that is equation (4.3) which we have already defined. The market maker sets the default spread s_{MM} around this central price symmetrically, that is the chains of bid and ask orders start at prices $BID_{MM} = \hat{r}_t - \eta$ and $ASK_{MM} = \hat{r}_t + \eta$ for the bid and ask order respectively, given an estimate of the fundamental value $\hat{r}_t$. In other words, the size of the market maker's spread is given simply by the equation $s_{MM} = 2\eta$. Then the list of prices

ξ_t^j at which the n bid and ask orders are submitted by the market maker is represented with:

$$\xi_t^j \sim \begin{cases} (\hat{r}_t - \eta, \hat{r}_t - \eta - \delta, \ldots, \hat{r}_t - \eta - (n-1)\delta), j = \text{BID} \\ (\hat{r}_t + \eta, \hat{r}_t + \eta + \delta, \ldots, \hat{r}_t + \eta + (n-1)\delta), j = \text{ASK} \end{cases} \tag{4.7}$$

where $\hat{r}_t$ is the estimate of the asset's fundamental value, η is the initial offset from the estimate and δ marks the number of ticks between each order. The orders are spaced symmetrically on both sides of the market. Our MM model does not handle its current inventory (position) in any way, nor is he restricted to the number of contracts he can be long or short at any given time.

Unlike the ZI trader, the market maker does not route his orders in any way, all of the orders are processed at his default exchange. Similarly to the ZI trader, he is assigned by default to one of the exchanges. The market maker also works with the NBBO information - as his goal is to only send limit orders (orders which would not get executed right away), he filters out his own orders which would be executed immediately. By accounting for the NBBO information, he only keeps orders which fulfill the condition of not crossing the bid-ask spread on both exchanges - sell orders have to be strictly above the BID_{NBBO} and buy orders need to be strictly below the ASK_{NBBO}.

$$\hat{\xi}_t^j \sim \begin{cases} (B_t, B_t - \delta, \ldots, B_t - (n-k-1)\delta), j = \text{BID} \\ (A_t, A_t + \delta, \ldots, A_t + (n-k-1)\delta), j = \text{ASK} \end{cases} \tag{4.8}$$

where $\hat{\xi}_t^j$ is the filtered list of prices at which the MM trader submits his BID and ASK orders. B_t and A_t stand respectively for the first bid and ask price, at which the market maker can submit his order without having it immediately executed. The list of prices starts with A_t and B_t and is $n-k$ elements long, where k ($k \in \mathbb{N}, k \leq n$) is the number of orders which have been filtered out. The trader does not have to submit any orders on one side of the market if the condition of no bid-ask spread crossing is not fulfilled even for the order furthest from the mid. In such situation, the NBBO best bid (ask) price would be above (below) the MM's short (long) order which is furthest from the mid.

4.2.4 Latency arbitrageur

Latency arbitrageur (LA) is a slightly different agent compared to the previous two we have introduced. Her arrivals to the exchange are not proactive but rather reactive. There is no process which would dictate LA's arrivals, as she only trades if an arbitrage opportunity is present. In our context, an arbitrage is a situation which offers an immediate profit with zero risk. In a market with n exchanges such situation would be defined as $BID_i > ASK_j$, where $i, j \in (1, \ldots, n), i \neq j$. Formally we would define an arbitrage potential in time t as $\omega_t = max(BID_i - ASK_j, 0)$. The two prices have to be from two different exchanges as a situation with overlapping BID and ASK prices would never happen in a single market - the market would clear these orders instead. In case the bid and ask prices do not overlap, the arbitrage potential equals zero.

The arbitrageur in practice makes use of her superior speed to profit from the discrepancy by selling the asset on the exchange where it is overpriced and immediately buying it on the other exchange where it is underpriced. In our simulation, we work only with one arbitrageur, who is infinitely fast and she is the only trader capable of building a synthetic NBBO data feed. The infinite speed is the approximation of the fact that trade who is collocated (has servers extremely close to exchange's servers) can be orders of magnitude faster than other traders. She, therefore, knows with 0 latency what the best bid and best ask price is on all of the exchanges. And she can react to this information with zero latency as well.

Our LA agent trades at time t only when an arbitrage opportunity gives a profit potential larger than zero. She trades by submitting two orders at a price which is in the middle of the true NBBO's bid and ask prices.

$$p_t = BID_i - \frac{\omega_t}{2} = ASK_j + \frac{\omega_t}{2} \tag{4.9}$$

,

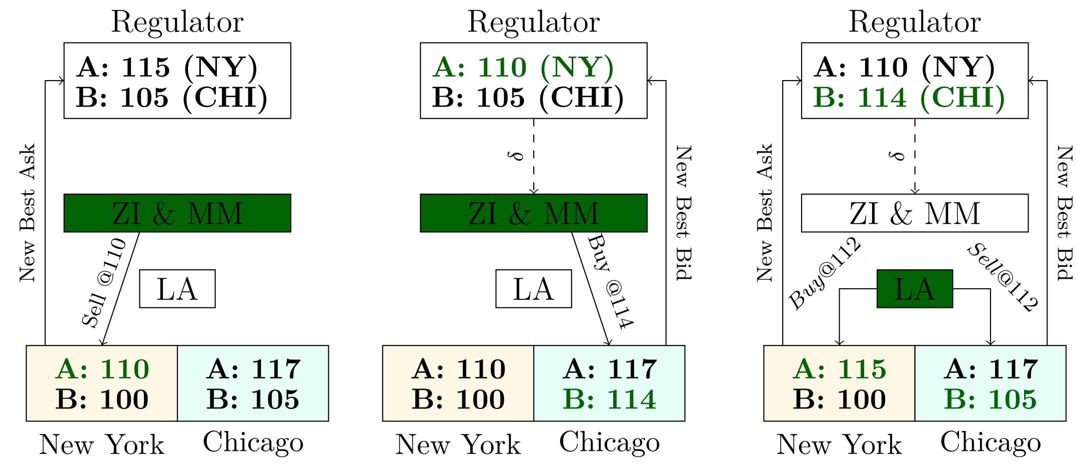
In the case mentioned above the LA trader sends the selling order to the exchange i and the buying order to the exchange j. In the ideal case both of the these orders get executed against the resting limit orders at prices BID_i and ASK_j. The profit π of the LA in this case would be equal to the full arbitrage potential ω.

4.2.5 Arbitrage opportunity

There are two ways in which an arbitrage opportunity can arise in our market settings - by lagging the NBBO information or by introducing the batch auction. Such opportunity emerges because of the order routing used by ZI traders. However, if we were to remove routing of the orders altogether and all ZI and MM traders would simply only look at the prices at their default trading venues, there would still be price discrepancies across exchanges even if all the participants had perfect NBBO information.

Arbitrage - lagged NBBO

Figure 4.4 represents the first case - arbitrage due to lags in the NBBO signal. We assume the market is clearing in a continuous regime, the situation would be analogous under frequent auctions. In the diagram we refer to the two exchanges as New York and Chicago, each has its own set of best bid and ask prices. They are both under the surveillance of the regulator, who sees the publicly available information about best bid and offer and aggregates it into a public signal and delivers it to the public with some delay δ. Green color marks the trader who is submitting his order during the period.



a New order
b New order & NBBO update
c Arbitrage found

Figure 4.4: Arbitrage under lagged NBBO

In the first period, one of the slow traders submits a new selling order to the New York exchange. This information is sent to the regulator. During the second stage, the regulator has processed all the information from the last period and he is sending it to the slow traders with some delay. Before the information arrives, one of the slow traders submits a new order which is not routed to the New York exchange (where it would be filled immediately) but rather is sent to his default Chicago exchange. Even though the regulator already has the information, due to the delay δ this information is not reflected in trader's decision-making process.

In the final step, the latency arbitrageur detects a new arbitrage opportunity. She, therefore, submits two orders at price 112 and sends them to the respective exchanges. They are cleared immediately and she receives a profit of 4. The exchange's best bid and ask prices return to the initial state. At this point, if one of the slow traders were to trade, his orders are routed according to the lagged information. Market makers, for instance, would not submit new orders because they perceive that their limit orders could get executed under the lagged NBBO.

Arbitrage - batch auction

Under frequent batch auctions clearing, an arbitrage could arise even with perfect information about the NBBO. Let's assume that the traders arrived both during one auction and submitted the same orders as in Figure 4.4. As any order which is submitted during an auction is not visible and becomes public only after the market clears, the second trader does not know of the first limit order. Once the market clears and the orders are added to their respective order books, only then do both slow and fast traders know about the new state of the order book. The latency arbitrageur immediately submits the same two orders, this time she has to wait until the market clears to see whether her orders were executed or not.

In summary, an arbitrage opportunity can have two sources - lagged NBBO or market clearing using frequent-batch auctions. Under continuous trading and perfect information all orders are routed properly and therefore there is no arbitrage. Under continuous trading with lagged NBBO, the arbitrageur's position remains always zero as thanks to her infinite speed she is surely the one who gets both misrouted quotes filled. During a frequent batch auction, the opportunity arise even with perfect information, but they are corrected during the next auction - either by the arbitrageur or by one of the slow traders.

4.3 Market metrics

Our model revolves around the zero-intelligence trader, who represents the general investment and trading public. Because of that, the primary metric at which we focus is the ZI trader's surplus. It is measured as an aggregate of all ZI traders - their profit, final position at the end of the trading session and the sum of their trading components. The private benefit in a sense is the core of our model, as that is where the overall surplus is generated. Without the private component, the sum of the profit from realized trades and current position marked to market would equal zero once we sum across all traders.

Let's have a zero-intelligence trader, who has made n trades during the entire session. Then formally his total surplus equals

$$\pi = \begin{cases} -q_T * r_T, q_T > 0 \\ q_T * r_T, q_T < 0 \end{cases} + \sum_{i=1}^{n} \begin{cases} \phi_i - p_i, q_i > 0 \\ p_i - \phi_i, q_i < 0 \end{cases} \tag{4.10}$$

where π is the total surplus, q_T is the quantity the trader holds at the end of the trading session. r_T is the final value of the asset's fundamental process. Each trade i is characterized by the price at which it was opened p_i, its quantity q_i and trader's private benefit ϕ_i. The left part of the equation summarizes the valuation of the position at the end of the trading session. The right part sums over all of the trades and accounts for the price at which the trade was executed and trader's personal benefit from this trade.

In a similar manner, we can measure market maker's and arbitrageur's surplus. These traders lack the personal component we, therefore, refer to this value as profit rather than a surplus. It is calculated in the same way as in 4.10, but $q_i = 0$ for all trades. The market surplus is the sum of these three metrics.

Next, we keep track of the total number of trades. Trade in this context is a pairing of a bid and ask order. This metric represents the magnitude of the activity in our model.

The average bid-ask spread is a time-weighted measure of the liquidity and transactions costs that traders face. Narrower spread means that the cost of getting into and out of a position immediately is relatively low. Spread is usually tight in high volume, highly liquid contracts such as the crude oil or S&P 500 futures contracts.

Similarly, the mean of the execution speed can be considered as a measure of liquidity. Faster execution is generally a proxy for the more liquid market.

This statistic is the mean of the difference between the arrival time of the order and its execution time. We measure both limit orders which are added to the order book as well as orders which are executed right away. This should draw a more realistic picture of the true average execution time. Previous works were calculating only the execution time of limit orders which were first added to the order book. Orders which were not executed are ignored in computing this statistic.

Finally, we also observe the price volatility. In the markets lower volatility stands for higher stability of the security and in general, is preferred by the risk-averse traders and investors. Here we measure it as the mean of the logarithm of the standard deviation of the price series.

4.4 Code and Simulation

The code is written using the objective-oriented programming (OOP) approach, where we define each entity (trader/regulator/exchange) as a separate object. These objects then are then linked - i.e. the traders know of the regulator and his attributes (current asset price, NBBO information, etc..). Even in continuous trading, the time is treated discretely and the principles of discrete-event simulation (DES) are used. DES ensures that during every event the order of agent's actions keeps its priority - e.g. in continuous mode, the asset's public signal is calculated first, then slow traders trade, market clears and at the end of the event arbitrageur reacts with no latency and market clears again.

For reproducibility purposes, the entire codebase of this thesis is written using open-source software and it is publicly available on GitHub[1]. Python 3.6 forms the core of the model, the results are saved in a PostgreSQL database. The simulation ran for two weeks on multiple computers, mainly in the computational center of UTIA - Institute of Information Theory and Automation. Each simulation has a separate pseudorandom seed, generated by the *numpy* Python library. These seeds are used for instance in generating the arrival times of the agents, in estimating the private valuation of ZI agents or computing the process of the asset's fundamental value.

[1] https://github.com/OskarGottlieb/master-thesis

Chapter 5

Preliminary analysis

5.1 Our environment

Before doing robust analysis from which statistically significant conclusions can be made, we first simulate the environment for a few sets of parameters. Given that the code used to simulating this is quite complex, these initial runs serve as an integration test (making sure that the agents behave accordingly when interacting with one another). From this initial analysis, we can see that the code is running as expected and we can set an expectation as to what should be the result of the statistically significant tests.

This initial analysis works with 200 zero-intelligence and 2 market makers. We assign the equal number of agents (equal number by category) to each exchange. In this testing environment, we do not include the arbitrageur. The intensities $\lambda_{MM} = \lambda_{ZI} = 0.005$ for both traders. Initially ZI traders do not discount any additional surplus $S_{min} = S_{max} = 0$. ZI trader's private value is drawn from the normal distribution with $\sigma_q^2 = 500,000$. Market makers submit 5 orders on each side of the market, with first order being 250 ticks away from the fundamental value. The orders are 100 ticks apart from one another. Session length equals $10,000$ where one time unit represents one millisecond. By default, we use the continuous clearing approach with no delay of the NBBO information. Given the initial values of our environment's parameters we could include the arbitrageur and still should receive (after enough sampling) a very similar result, as arbitrageur would not interact with the market because there would not be any arbitrage opportunities.

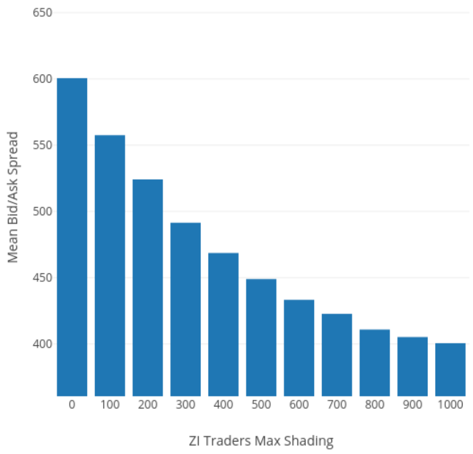
The asset's starting fundamental value is set to $10,000$ and its variance is $\sigma^2 = 500,000$. The mean-reversion parameter is set to $\alpha = 0.05$. The smallest possible change in the value of the asset (tick size) is set to 1. No trader can submit and trade against his order - no self-matching of orders is allowed.

5.2 Experiments

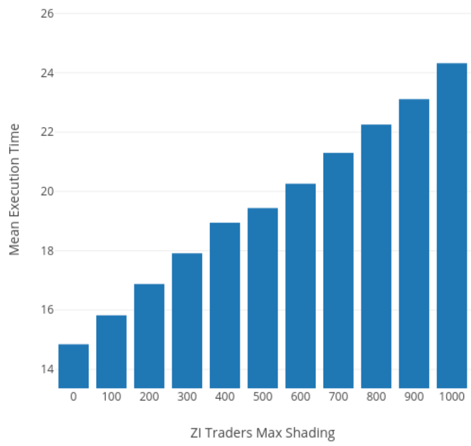
Each experiment will change up to three parameters and we will see what the implications of such change are for the market characteristics. Each experiment is the summary of two hundred repetitions unless explicitly specified otherwise. We plot the mean values over the samples, the mean along with its 95% confidence intervals are presented in their respective tables.

5.2.1 Changing ZI's additional surplus

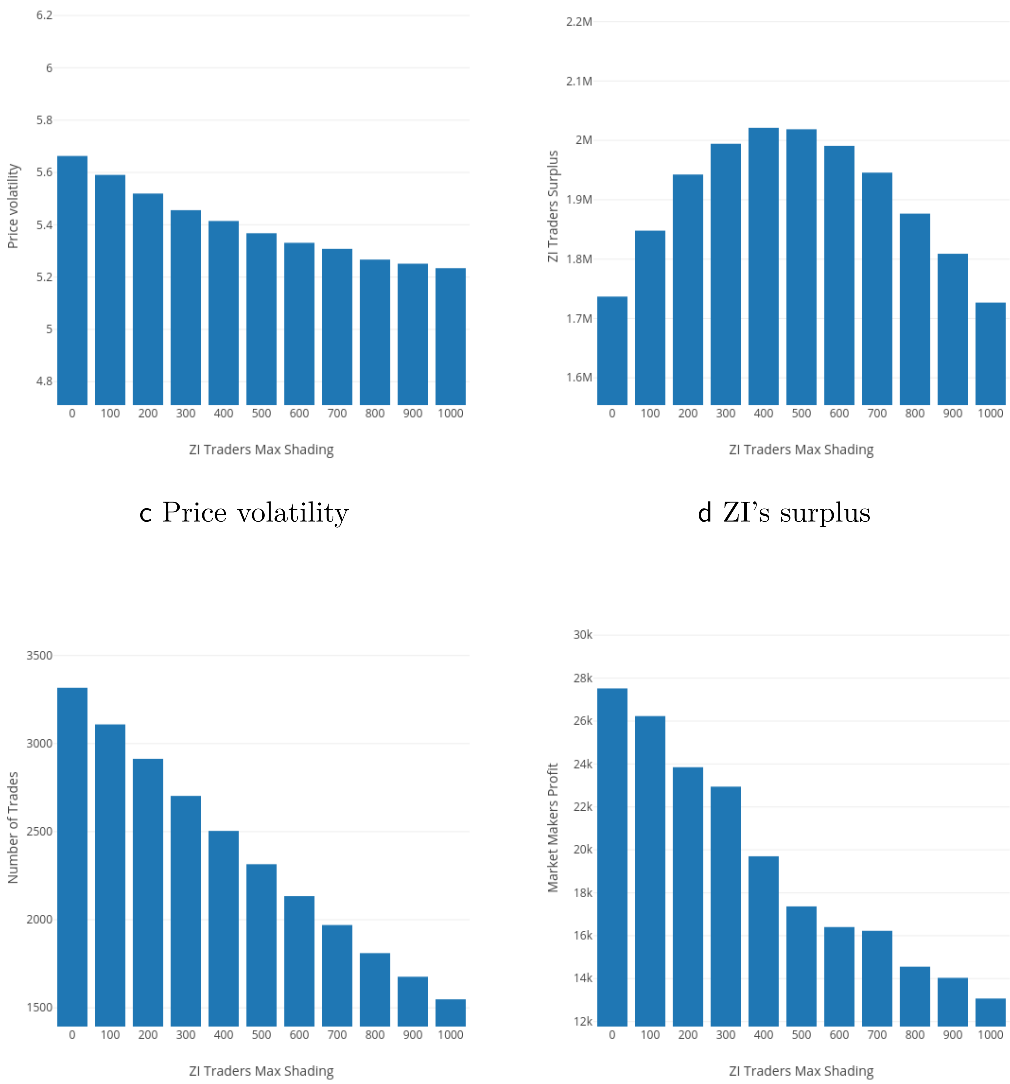
We keep the minimal additional surplus demanded at zero and only let the maximum move from zero to 1000 in increments of 100. The results are shown in Figure 5.1, Table 7.1 contains the detailed results.



a Mean bid-ask spread



b Mean execution time



c Price volatility

d ZI's surplus

e Number of trades

f MM profit

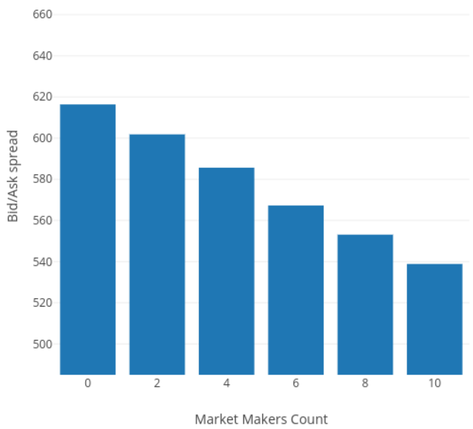
Figure 5.1: ZI's demands of additional surplus

Increasing the required amount has a direct effect on all of the metrics we observe. The increase should theoretically move ZI's bid and ask further away from the mid-price (the middle of the best bid and ask prices). Directly we see that the number of trades decreases linearly, with that the mean execution time increases. The decrease in the number of trades has a direct effect on market maker's profit, which decreases as well. This is in line with expectations, as fewer trades means fewer fills for market makers. Interestingly the mean bid-ask spread decreases, which should be a sign of increased liquidity.

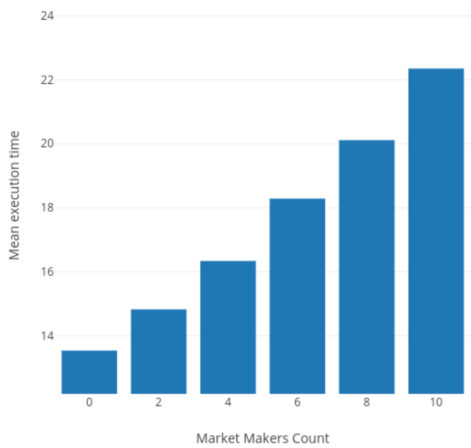
Similarly, due to fewer transactions in narrower bid-ask spread, we also observe a decrease in volatility. Finally, the surplus of ZI trader's illustrates the trade-off between having the trade executed and getting a better price. The resulting surplus has its peak around the values $400 - 500$. Even with fewer trades, the ZI trader can achieve a higher overall surplus. But demanding too much additional surplus could again decrease it.

5.2.2 Market maker count

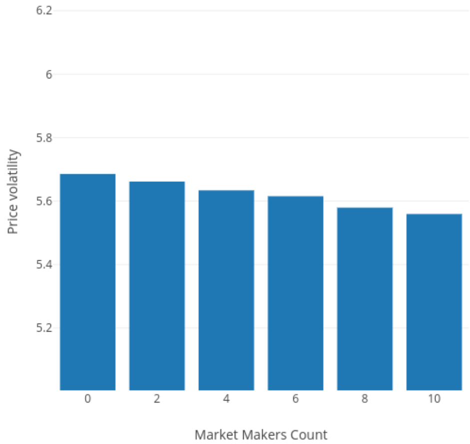
We keep market maker's parameters constant and only change the number of MM agents. Increments of two are selected as with each increment each trading venue gets exactly one new market maker. Figure 5.2 summarizes this behaviour, detailed results are in Table 7.2.



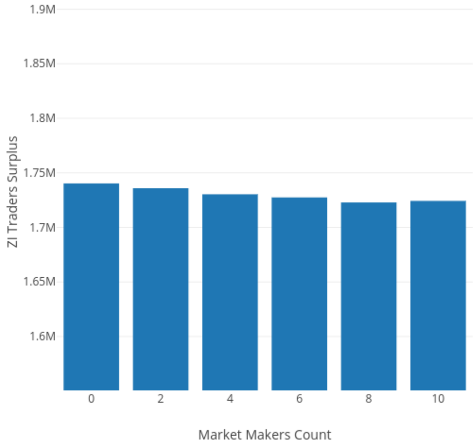
a Mean bid-ask spread



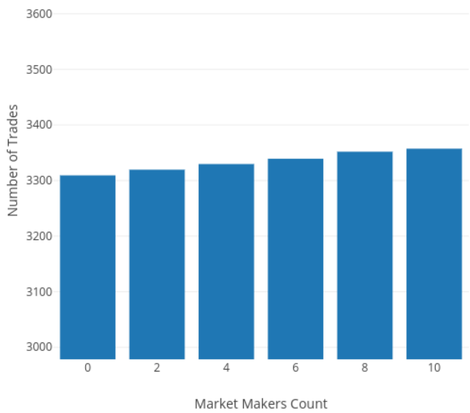
b Mean execution time



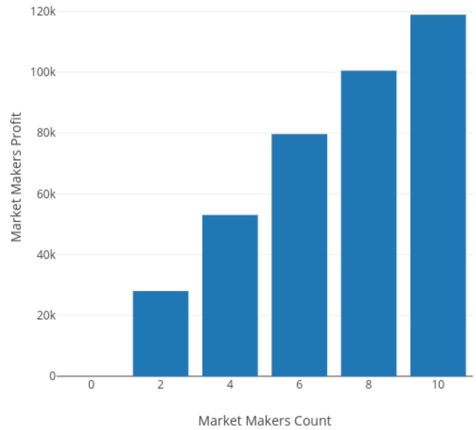
c Price volatility



d ZI's surplus



e Number of trades



f MM profit

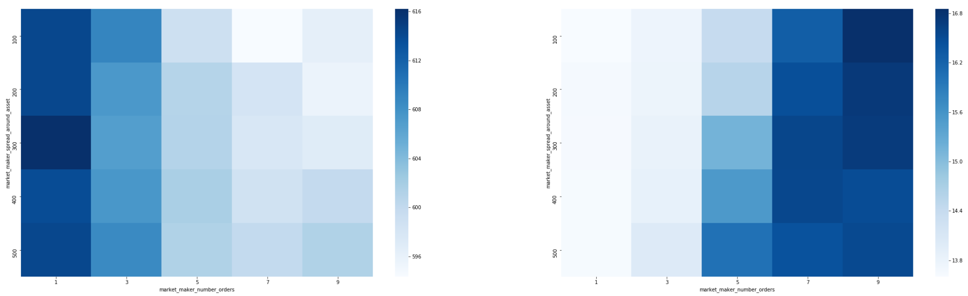
Figure 5.2: Number of MM agents

From the theoretical point of view, the role of market makers is to increase liquidity and market quality by absorbing some of the price risk. MM then profits from the bid-ask spread where he quotes passively. We can see that increasing the number of market makers increases their overall profit, but the marginal income with each additional MM decreases. This points to the fact that the market is getting more saturated with additional traders. The spread decreases and so does price volatility, both in line with the theoretical expectations. The number of trades increases only slightly.

The surprising part is that ZI trader's surplus decreases with additional market makers. This goes against the expected outcome as the market maker is considered generally beneficial and this would place him into the same category with the arbitrageur. However, the behavior of market makers has not been studied before in a similar model. In this settings with perfect information about the NBBO, the market maker only submits passive orders, accounting for the information from both exchanges. We could think of its actions as a form of front running of the resting limit orders. MM never directly takes the liquidity, but his profit could come from the pocket of the ZI trader - more specifically it would be the opportunity cost of ZI' trader's non-executed trades. Given that market maker only submits passive orders it is of no surprise that the mean execution time increases with additional traders.

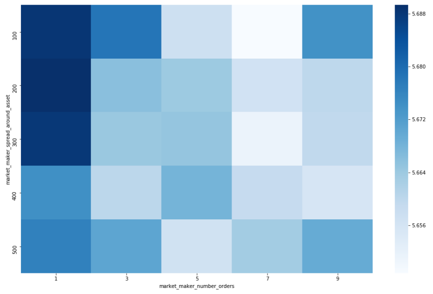
5.2.3 MM's number of orders & spread

We turn back to the default settings with two market makers and in this experiment, we change two of its attributes, the number of orders he submits on both sides of the market and the initial spread he chooses around the asset. The x-axis marks the number of orders which range from 1 to 9 in increments of two. On the y-axis, we plot the spread around the asset starting at 100 and increasing in increments of 100 until 500. The value is newly represented with a heatmap (Figure 5.3) where the deeper the color (blue) the larger the value of the market characteristic which we observe. The tables with the preliminary results are from now added in the attachment of this thesis.

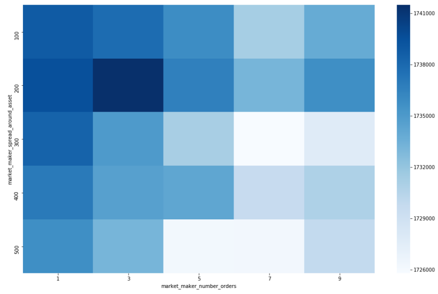


a Mean bid-ask spread

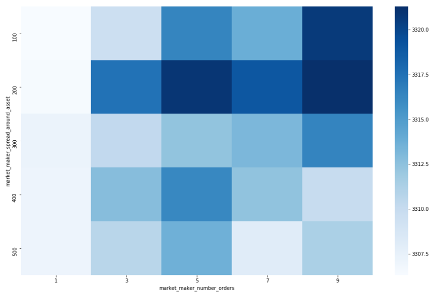
b Mean execution time



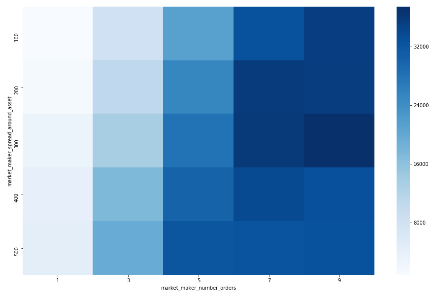
c Price volatility



d ZI's surplus



e Number of trades



f MM profit

Figure 5.3: Number of MM's orders and spread

The overall effect of the market makers in our settings is rather small, in terms of a couple of percentages for each of the observed variables. The exception being the profit of the market makers themselves. It increases with increasing number of orders submitted. The marginal increment is decreasing which is a similar trend we saw with the number of traders, although the underlying effect should be different. The additional orders are added further down the order book with a smaller probability of being executed, in the previous case we had more traders competing around similar price levels. The trend of the spread parameter does seem to be twofold. In case of fewer trades, it seems that larger initial spread yields a higher profit. The implication is inverse with 7 or 9 submitted orders.

The cause of this interaction lies in the way the market maker agent works. The MM first estimates the fundamental value of the asset. Then he submits the orders but cancels those which would be executed immediately. Submitting only a few orders with tight spread could result in very few orders actually being sent to the exchange because the market maker would cancel them as they would get executed immediately. As the spread increases the market maker cancels fewer orders and his profit increases. Once he is sending 7 or 9 orders, it could be more profitable to trade with a tighter spread as even though market maker cancels some of his orders, the rest of them remains in the order book. Choosing a large spread, on the other hand, could result in fewer orders being executed.

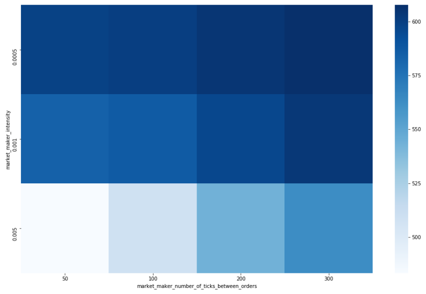
The fact that the probability that the orders deep in the order book will be executed is small also impacts the mean execution time which increases with the number of orders submitted. The initial spread parameter again has a twofold effect depending on the number of orders submitted. The explanation is analogical as in the case of MM's profit.

Submitting more orders causes an overall tighter spread. The effect of the market maker's initial spread on the market's spread is ambiguous. Similar conclusions can be made with price volatility which increases as the number MM's orders decreases and there is not a clear trend with spread's influence. The effect on the number of trades is negligible.

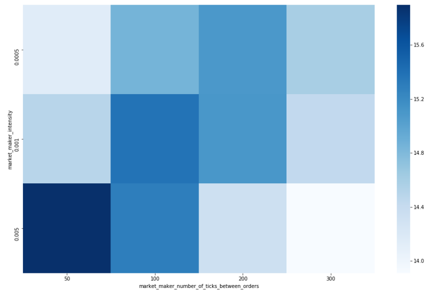
Interesting is the effect on zero-intelligence trader's surplus. Although small in magnitude, it decreases with the number of orders submitted - this is in line with our previous observation that more market makers have an adverse effect on ZI trader's surplus. The market maker's initial spread points to the fact that MM is more beneficial to ZI traders if he narrows the spread by quoting closer to the midpoint. ZI traders are then able to make trades with MM being their counterparty - those are trades which otherwise might not have been paired up.

5.2.4 MM's intensity of orders & order spacing

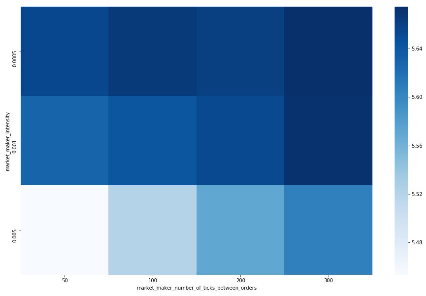
We look also at MM's second set of parameters - their intensity and number of ticks they place in between individual orders. The intensity approximates the speed of the trader as every arrival means that the trader updates his orders which might have been out-of-date. The x-axis represents the number of ticks (price units) between the bid and ask orders, the y-axis shows three intensities - 0.005, 0.01 and 0.05.



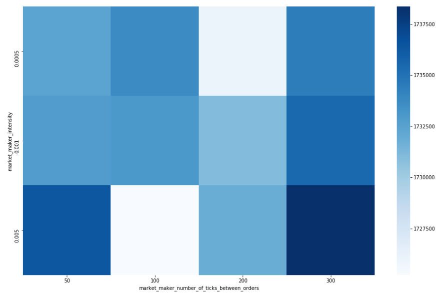
a Mean bid-ask spread



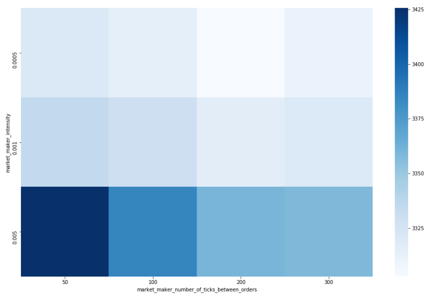
b Mean execution time



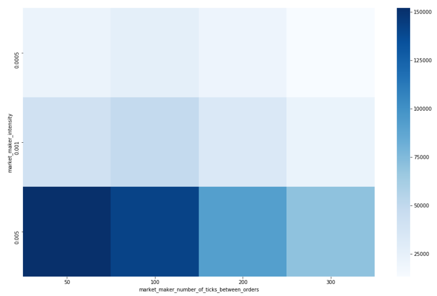
c Price volatility



d ZI's surplus



e Number of trades



f MM profit

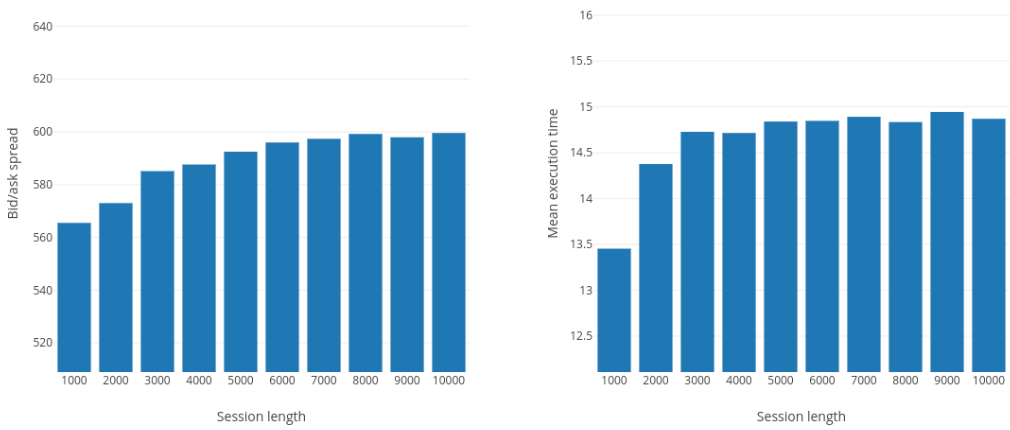
Figure 5.4: MM's intensity and spacing parameters

Visually we can see that the bid-ask spread exhibits similar behavior as price volatility. Similarly, the number of trades correlates with MM's profit. The largest difference in these figures is due to the intensity of the Poisson process and the fact that the y-axis is not linear. Increasing the intensity (trader's indirect speed) narrows the market's bid-ask spread and reduces the price volatility. The same effect can be achieved by narrowing the spacing between MM's orders.

The larger intensity and smaller spacing also increase the total number of trades as well as MM's profit. The effect of the two parameters on mean-execution time is rather ambiguous. The most interesting part is yet again the effect on ZI's surplus. As we saw before increasing either the number of market makers or the number of orders that they send out had a negative effect on ZI trader's surplus. Here increasing the intensity does not yield the same result, quite the contrary. Increasing the number of trades through higher turnover of market makers could result in higher overall ZI's surplus.

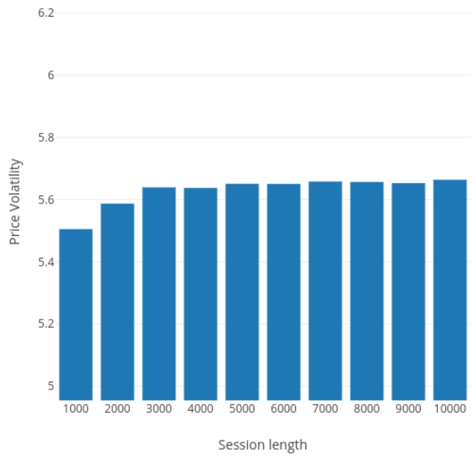
5.2.5 Session length

Next we alter the session length parameter and we let move it from $1,000$ to $10,000$ in increments of $1,000$. The details can be seen in Figure 5.5.

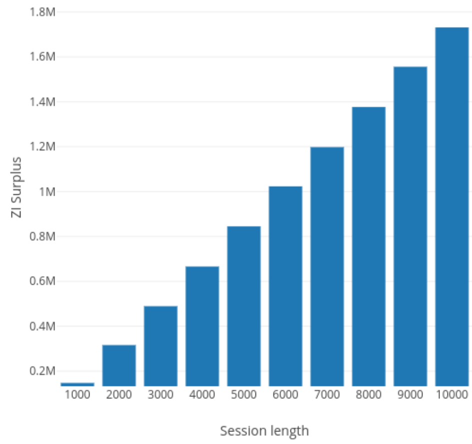


a Mean bid-ask spread

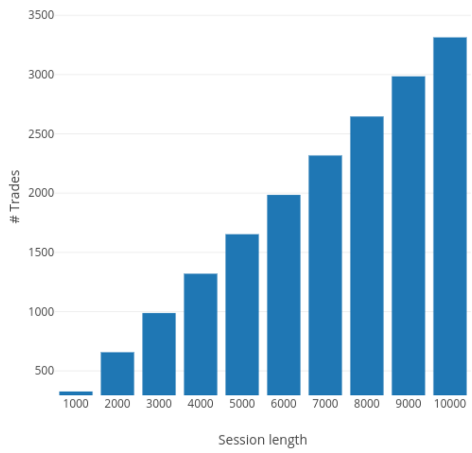
b Mean execution time



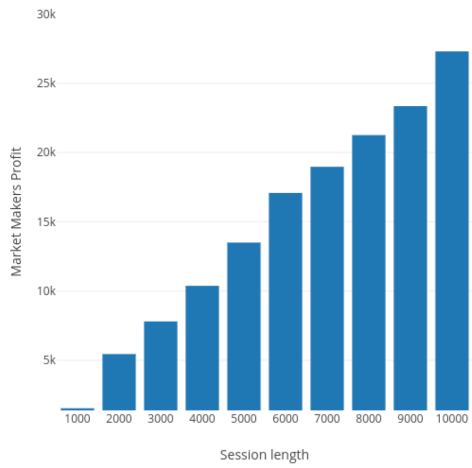
c Price volatility



d ZI's surplus



e Number of trades



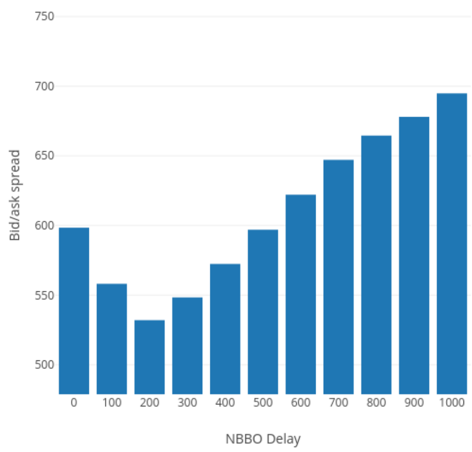
f MM profit

Figure 5.5: Session length

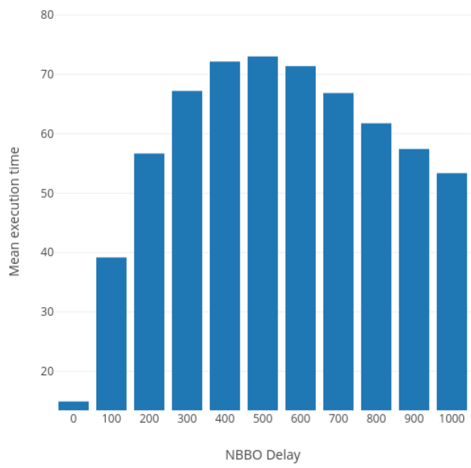
The effect is in line with the expectations - values which depend on the length of the session do exhibit a linear dependency - the number of trades as well as the profit of the market maker and zero-intelligence trader's surplus. On the other hand, the market state characteristics do stabilize rather quickly and do not change as much with increasing session length - the spread, execution time and price volatility. The reason for the initial shift in these characteristics is that in the beginning, the market does not have any orders and it takes a while for it to achieve a stable state in which all the traders either are actively trading or already have traded.

5.2.6 National best bid and offer

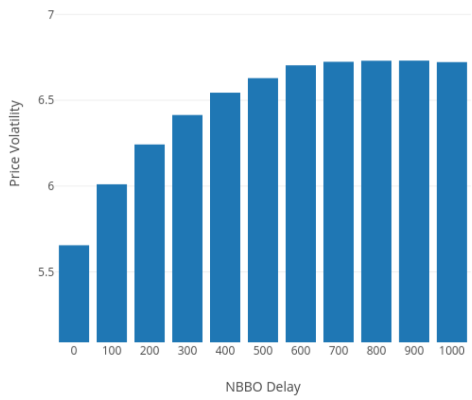
A delayed NBBO information allows for the creation of arbitrage opportunities. In this case, these opportunities are not immediately exploited as there is no arbitrageur. We lag the NBBO information in increments of 100 where zero lag means that all market participants have perfect information. Figure 5.6 summarizes our findings.



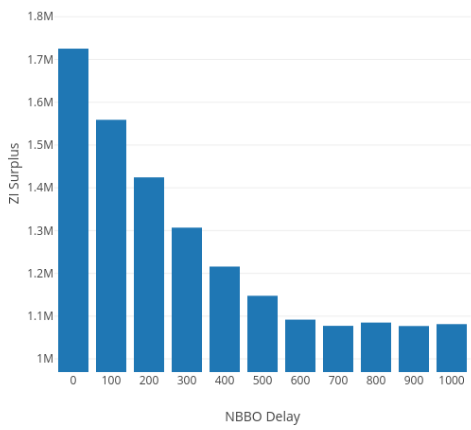
a Mean bid-ask spread



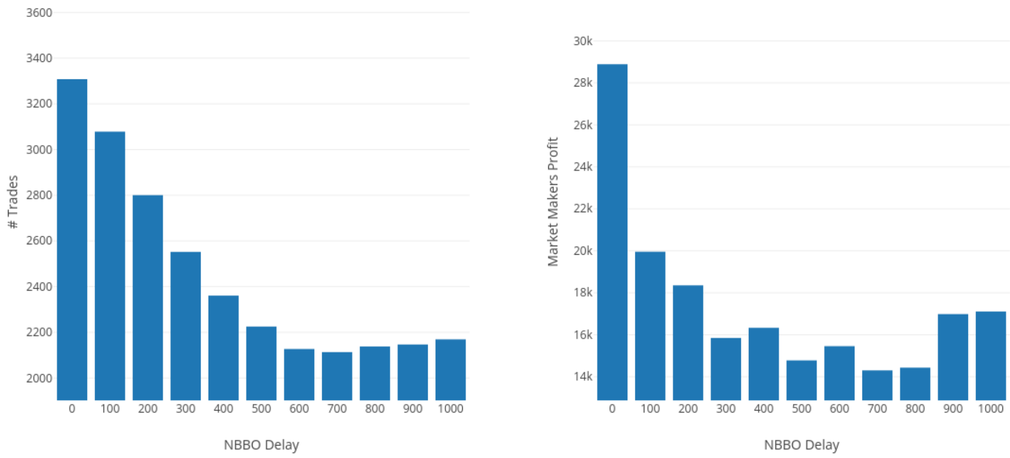
b Mean execution time



c Price volatility



d ZI's surplus



e Number of trades　　f MM profit

Figure 5.6: Delayed NBBO signal

The NBBO delay exhibits a similar effect on ZI Surplus, MM profit and the number of trades as the session's length - these value stabilize around the delay of 500. Market maker's first drop from no delay to delay of 100 yields a significant drop, in case of the number of trades and ZI surplus the decrease is more gradual. Every arbitrage opportunity can be thought of as a misallocation of individual orders. Therefore the delayed information causes that ZI traders submit orders which they perceive will be executed, but are added into the order book instead. Because of that, the mean execution time increases with increasing NBBO delay. The price volatility is also affected negatively by the lagged signal.

Finally, the bid-ask spread decreases until the delay value of 200, after which it increases with increasing NBBO lag. The interpretation of the bid-ask spread is unclear, the important note is that the value of 200 corresponds to the mean arrival time of both MM and ZI agents, as $\lambda_{MM} = \lambda_{ZI} = 0.005$. This information will be important in the next section in which we look at the relationship between the length of a batch auction and the delay of the NBBO information.

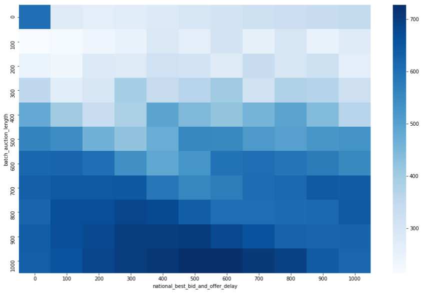
Chapter 6

Results

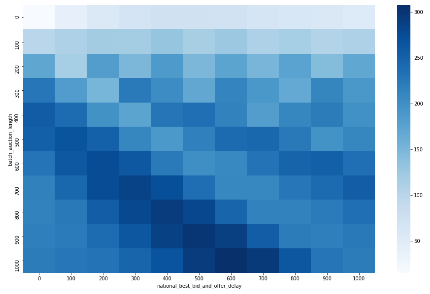
In this section the parameters of the environment are the same as in Chapter 5. In each experiment, it is explicitly specified which parameters differ from the default environment. As we are dealing with a simulation we use the Monte Carlo method in order to obtain statistically significant estimates of the market's characteristics.

6.1 Batch auction and lagged NBBO

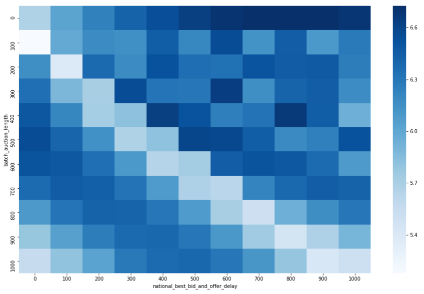
Now we have a solid understanding of the model, its agents and their influence on the market. In this experiment, we introduce the arbitrageur along with the batch auctions. Latency arbitrageur reacts to profit opportunities which are worth at least 1. Batch auction of length 0 is the standard continuous market clearing which we have used so far. Previous works only analyzed a batch auction where the length of the auction was set equal to the delayed NBBO information. In this experiment, we allow for asymmetric market settings, letting both parameters move from 0 to $1,000$ in increments of 100. Figures 6.1 and 6.2 represent the market settings without and with an arbitrageur respectively.



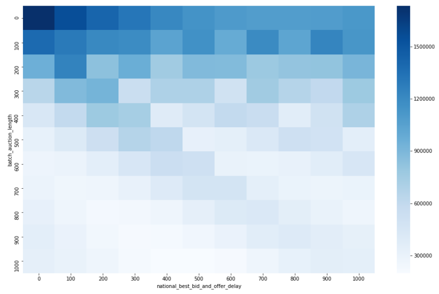
a Mean bid-ask spread



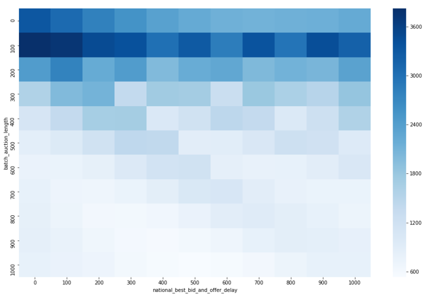
b Mean execution time



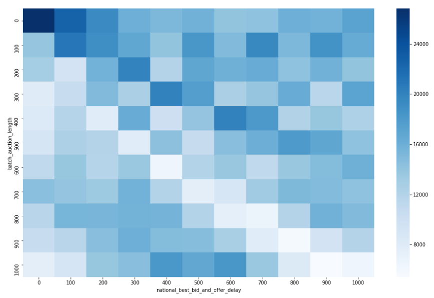
c Price volatility



d ZI's surplus

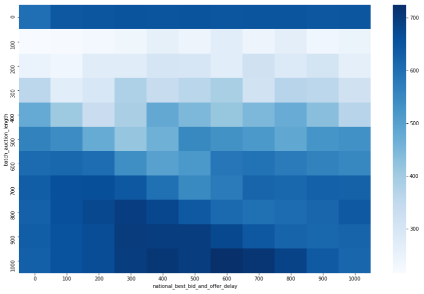


e Number of trades

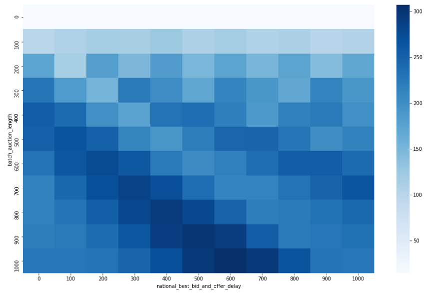


f MM profit

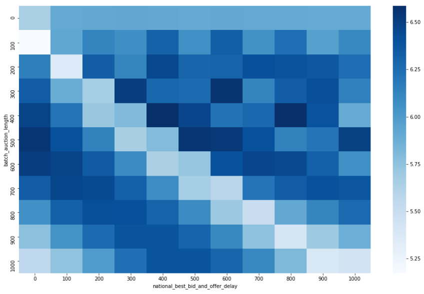
Figure 6.1: Batch auction with NBBO delay, no arbitrageur



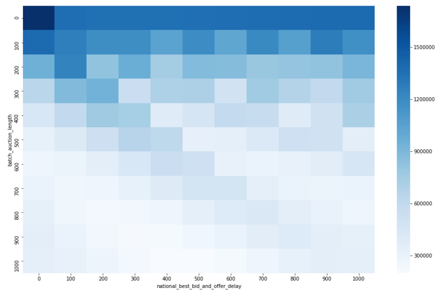
a Mean bid-ask spread



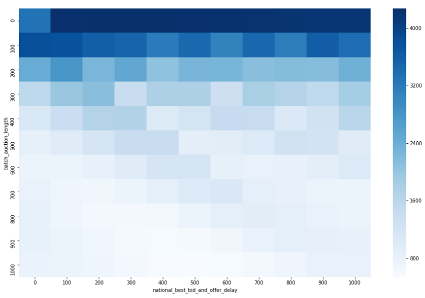
b Mean execution time



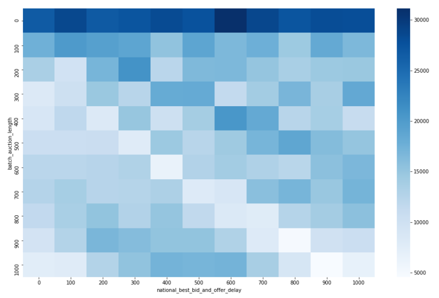
c Price volatility



d ZI's surplus



e Number of trades



f MM profit

Figure 6.2: Batch auction with NBBO delay, with arbitrageur

The two sets of plots are fairly similar, but there are a couple of important distinctions mainly when it comes to ZI trader's surplus and MM's profit. For each delay of the NBBO, we want to test the means of the market's metrics in the continuous auction against each length of the batch auction. The hypothesis is then set in the following way:

$$\begin{aligned} H_0 &: \mu_c^i = \mu_a^i \\ H_1 &: \mu_c^i > \mu_a^i \end{aligned} \tag{6.1}$$

where μ_c^i and μ_a^i are the means of the given variable i for the continuous trading and respective batch auction. In the same way we can set the alternative hypothesis to $\mu_c^i < \mu_a^i$.

In case of the ZI traders, both with and without the arbitrageur we set the alternative hypothesis to the mean of their surplus being lower in batch auctions than it is continuous trading. We set the hypothesis in the same way for the arbitrageur and market maker's profit. The p-values are presented in Tables 7.3, 7.4 7.5, 7.6 and 7.7. The x-axis of each table represents the delay of the NBBO information, y-axis stands for the length of the auction. The auctions of length zero is not in the table, as in each column it stands for the continuous clearing which we test against.

In case of the ZI traders' surplus we can reject the null hypothesis in almost all the cases, although there are a couple of instances where we can not reject it. Market maker's profit is significantly lower for all lags of the NBBO and all lengths of the auctions when there is an arbitrageur present. Without the arbitrageur, the profit is still significantly lower in most cases under NBBO lag smaller than 200. With larger lags, the situation does get more complicated as there are areas with small and areas with high p-values. Even though they look randomly scattered at first, we can see one pattern which can be seen i.e. on multiple plots in Figure 6.2. The p-values are very low along the line where the length of the auction is slightly above the delay of the NBBO information. This is a pattern which can be seen throughout other experiments which we present in next section. The second explanation for the scattered p-values is the fact that there are only two market makers in this settings, which is a relatively low number compared to the number of ZI traders.

Finally, the profit of the arbitrageur is smaller in the batch-auction regime with the exception of the leftmost column where the NBBO signal is not lagged. This corresponds to the theoretical model, where the arbitrageur is able to profit in the continuous regime where other market participants react slowly to the public lagged NBBO signal. Once the auction is introduced the arbitrageur loses his edge of absolute certainty of getting the full arbitrage potential, which results in lower profit than in the continuous regime. Similarly, we can see that he is active in batch auctions even when the NBBO information is perfectly known to all agents. However here he does not have any comparative advantage as not only he is not clearing his orders instantly but also his competition has the exact same information.

The mean execution time increases with the introduction of batch auctions, which is in line with the expected outcome as generally an increase in the time interval it takes to process the order increases the average time an order rests in the order book. The execution time does seem lower along the same diagonal of values where the NBBO delay is slightly smaller than the length of the batch auction. The bid-ask spread exhibits strong improvements mainly in situations with the latency arbitrageur. This is a direct consequence of the fact that there are more spots in the market without arbitrageur, where misrouted orders decrease the average spread.

The number of trades in both markets decreases with increasing length of the auction. The exception is the shortest batch auction of length 100. Generally, the main reason for this decrease is the fact that there are fewer spots for market clearing. Every trader can trade (submit orders) only once during an auction. Any future orders replace the ones which were submitted before. In continuous markets this is not the case as any arrival of a trader means a new opportunity for a trade.

The introduction of batch auctions of this length has brought in overall lower surplus for all market participants. Also interestingly the presence of the arbitrageur seems to increase the benefit of the zero-intelligence traders in the continuous trading regime. These conclusions contradict the results of Wah & Wellman (2017) and we investigate them in the next section.

6.2 Arbitrageur's effect

For each lag of the NBBO and each length of the auction (or continuous clearing), we test the two means of the ZI trader's surplus with the alternative hypothesis being that ZI trader's surplus with arbitrageur is greater than without him. The p-values are shown in Table 7.8.

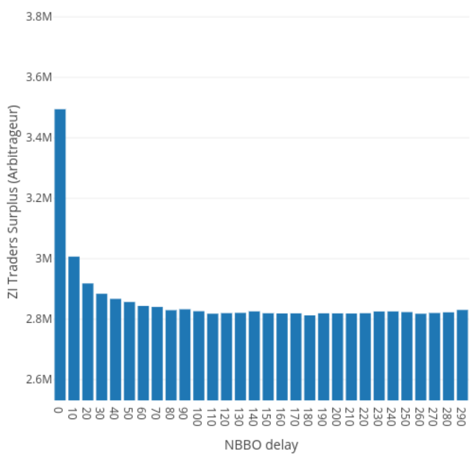
We can reject the null hypothesis in the continuous regime with the NBBO's lag being greater than or equal to 300. Some of the p-values of the shortest batch auction are also significant. Although counter-intuitive at first, this result does make sense in our market settings where the intensity of both ZI traders and MMs is $\lambda = 0.005$. Therefore the mean arrival time of all the traders (with exception of arbitrageur) is 200 time units. Once we lag the NBBO information above this threshold it seems that the routing of the orders is random and the traders react to market with a delay which is longer than their respective mean arrival time.

The arbitrageur under these conditions can be actually beneficial, as he would execute (for a cost - his profit) the misrouted trades which would be canceled in a market without arbitrageur. These orders are in themselves valuable even when they are not executed at the best possible price. Once the delay is above the threshold of 200 before this information is known to all market participants, chances are that the order has already been canceled by the trader, it, therefore, represents an opportunity cost for the ZI traders and MMs.

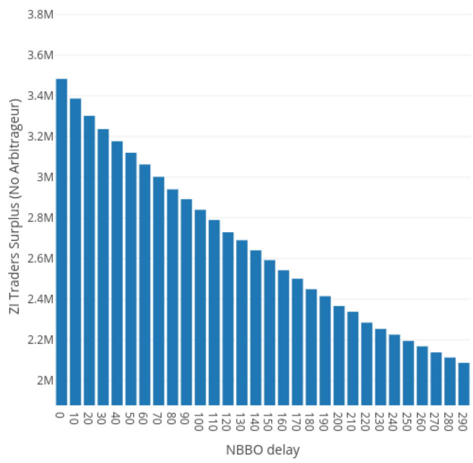
6.3 NBBO delay and Poisson intensity

We conduct one more experiment by testing our previous hypothesis with a slightly new environment. We test now only the continuous trading with delayed NBBO. The NBBO delay ranges from 0 to 290 in increments of 10. We also changed the intensity of the Poisson process of both ZI traders and MMs to $\lambda = 0.01$. The p-values from testing the means of zero intelligence trader's surplus are summarized in Table 7.9. Each NBBO delay has been sampled 200 times.

From the results, we can conclude that the arbitrageur is again beneficial once the delay of the NBBO information increases past the mean arrival time of the traders, in our case $\frac{1}{0.01} = 100$. Figure 6.3 shows the relationship between the surplus and the delay of the signal.



a Delayed NBBO with arbitrageur



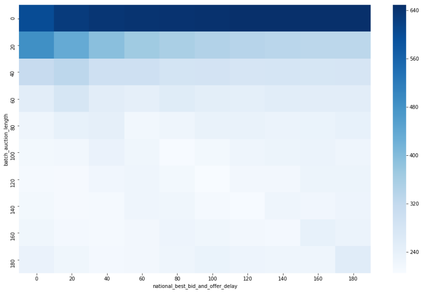
b Delayed NBBO without arbitrageur

Figure 6.3: Continuous trading with and without arbitrageur, $\lambda = 0.01$

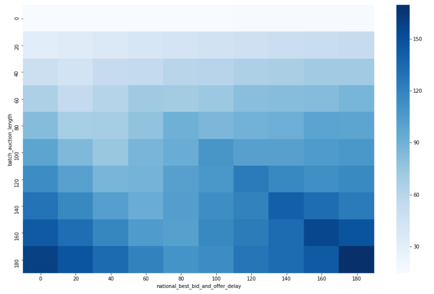
Under the arbitrageur, there is a sudden drop in the surplus, which stabilizes and remains roughly constant for larger delays. The market without an arbitrageur exhibits a more gradual decline, which seems to be proportional to the length of the delay. The decline also most likely stabilizes once the delay is large enough - we saw similar behavior in the plot of the figure 5.6. The two bar plots intersect roughly around the delay of 100, which corresponds to the point where the difference in means is statistically significant.

6.4 Batch auction and lagged NBBO in smaller increments

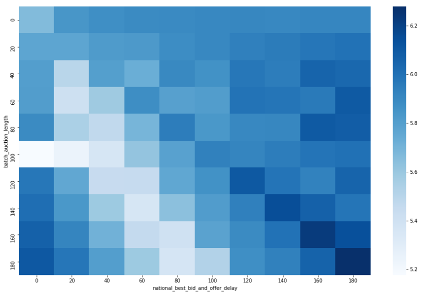
We now look at the same experiment as in 6.1 but we change the clearing intervals and NBBO delays. They now range from 0 to 180 in increments of 20. The Poisson intensity is back at its original level of $\lambda = 0.005$, which means that we investigate the area within the mean arrival time of our agents. Figures 6.4 and 6.5 summarize these measured metrics in a market settings with and without an arbitrageur respectively. The NBBO signal delay is again on the x-axis, auction length is on the y-axis.



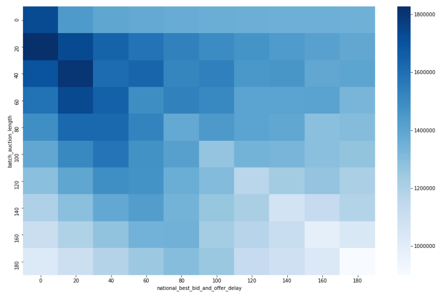
a Mean bid-ask spread



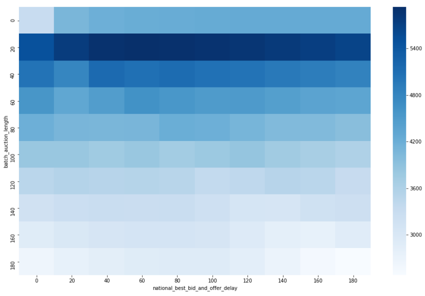
b Mean execution time



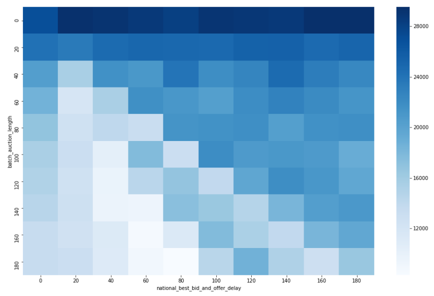
c Price volatility



d ZI's surplus

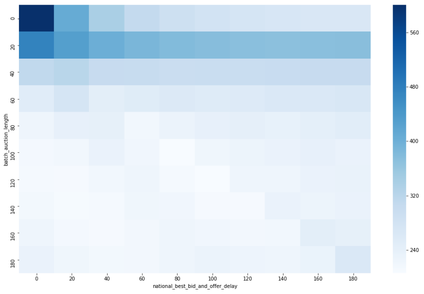


e Number of trades

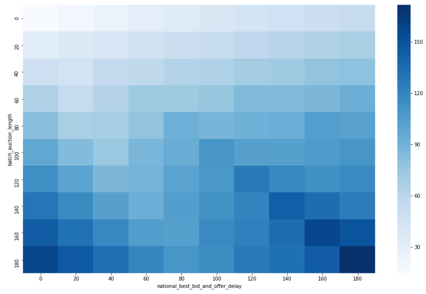


f MM profit

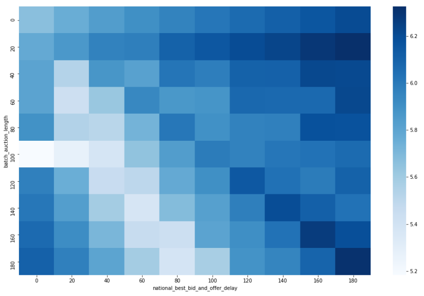
Figure 6.4: Batch auction with NBBO delay, with arbitrageur - (Delay up to 180)



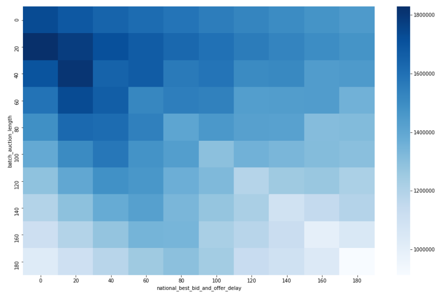
a Mean bid-ask spread



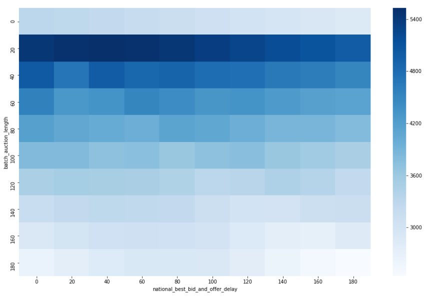
b Mean execution time



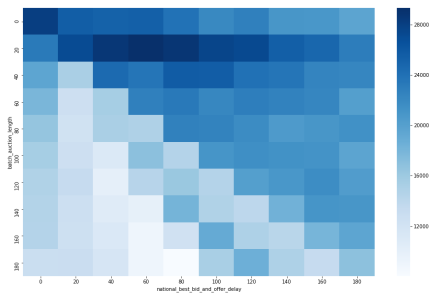
c Price volatility



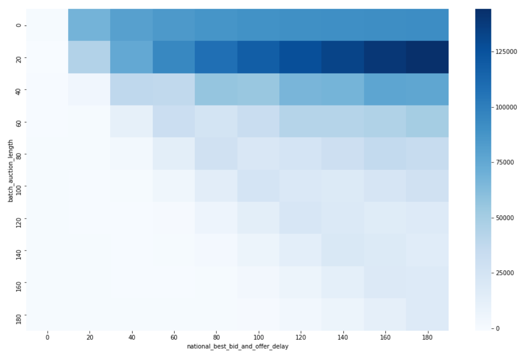
d ZI's surplus



e Number of trades



f MM profit



g Arbitrageur profit

Figure 6.5: Batch auction with NBBO delay, no arbitrageur - (Delay up to 180)

The tests of the means of ZI trader's surplus in batch-auctions against continuous trading with and without arbitrageur are summarized in tables 7.10 and 7.11 respectively. The alternative hypothesis in these tests is that the mean of the surplus in batch auctions is greater than the surplus in continuous trading. In both cases, there is a threshold beyond which switching to batch auctions is not beneficial to ZI traders. Market settings with an arbitrageur have this threshold around a greater length of the auction than the ones without the arbitrageur. The interpretation of this is that in a market setting with an arbitrageur, changing the clearing to batch auctions has a stronger impact on ZI trader's surplus as the general public is partially protected from the actions of the arbitrageur. In the first column where the NBBO signal is not lagged, we can see that with the exception of the shortest auction switching to batch auctions would not be beneficial to ZI traders. On average an increase in the length of the auction means a lower number of trades, hence lower overall benefit.

Next, we compare the relative difference between ZI traders in markets with and without the arbitrageur for all combinations of parameters. The results are in Table 7.12. In the first column, none of the results are significant as that is where the arbitrageur does not have an edge. Once we introduce the delayed signal we obtain a diagonal above which ZI trader's surplus is larger when the arbitrageur is not present - we can see the degree to which an arbitrageur actually decreases ZI trader's surplus. The implication that an arbitrageur can be malicious is opposite to what we found in a market setting with longer auctions and NBBO signal delays.

The MM's results were not as significant and they are studied in a slightly different market settings in more detail in the next section. Arbitrageur's profit is summarized in the Table 7.13. Compared to the continuous trading the batch auction on average has the profit. The exception is the shortest auction with larger NBBO delays, where the arbitrageur's profits are higher than in continuous regime.

6.5 Changing the number of MMs

In the last two experiments, we test for the influence of the batch auctions on the market where the number of ZI traders or MMs changes. We use the same grid as in the previous experiment but we restrict ourselves only to a market with an arbitrageur. First, we compare ZI trader's surplus in markets with 8

MMs against markets with 4 and 0 MMs. The results are summarized in tables 7.14 and 7.15 respectively. There are 500 observations for each market settings.

We have used the same parameters for the MMs as in Section 5. We keep the number of ZI traders at 200 and we set the alternative hypothesis to ZI traders' surplus with 8 MMs being lower than one with fewer MMs. In both tests, we can reject the null hypothesis (that the means equal) in most of the frequent batch-auctions and continuous trading with no latency. An exception is the bottom center part of our table, where batch auctions are longest but it is impossible to infer any sort of trend from this discrepancy. In continuous trading, with latency, more market makers seem to have a positive effect on the zero intelligence traders' surplus.

Tables 7.16 and 7.17 test for the difference in means of arbitrageur's profit for the same market setting. In both tests there seem to be a threshold on the diagonal where the NBBO delay is about the same as the length of the batch auction. Above this threshold, the arbitrageur's profit actually gets smaller when we include more market makers. Below the threshold, the difference in profit between the two setups is about zero and mostly non-significant.

There is also a significant jump between delayed continuous market and frequent batch auctions - with more MMs arbitrageur's profit decreases more in continuous markets (with a delayed NBBO signal) than it does in frequent auctions. We could say that arbitrageur is better off without any market makers. From these results, it seems that the market makers used in our model act in a similar way as the arbitrageur - they compete with the arbitrageur in all market settings and are helpful towards ZI traders only in continuous trading with lagged NBBO.

We also tested for settings with same possible numbers of market makers but a different number of ZI traders (50 and 100). In both cases, the results were less significant but very similar in nature. Of course with increasing number of MMs, their profit increased as well.

6.6 Changing the number of ZI traders

Finally, we compare markets with 50 ZI agents against markets with 100 and 200 agents. The number of MMs is set to 8. We first test the means of the MM's profit, setting the alternative hypothesis to them being higher in markets with more ZI traders. The results are summarized in tables 7.18 and 7.19 respectively.

Once we increase the number of ZI traders from 50 to 100 the MM's profit increases in a majority of the grid. The exceptions are the shortest auction and a steep diagonal (slope of this diagonal could be expressed as "NBBO"$= -3$ "auction length"). In both cases, the difference in means oscillates around zero. With the second increase of ZI agents to 200, the diagonal starts to look as a threshold above which the profit increases and below which decreases with more agents. The shortest auction remains neutral in its effect on MM's profit.

The effect on arbitrageur's profit for the same market settings is shown in Tables 7.20 and 7.21. The results of the two experiments are almost identical, they only differ in the magnitude of the difference. A threshold can be drawn again dividing the grid into two parts. Below this diagonal threshold, the difference in both cases is on average zero. We could say that the arbitrageur hits a certain limit which does not increase once we introduce more ZI traders. Above the diagonal the profit of the arbitrageur increases with more ZI traders.

The diagonal is very similar to the one we saw in the previous example - it affects mainly the arbitrageur and it limits his profit potential. The sets of parameters below this diagonal represent situations where the length of the auction is significantly larger than the delay of the NBBO signal. Then a slow trader who arrives throughout the auction could already have the correct signal about the state of the market and he could base his decision upon the same information as the arbitrageur. This then means that the relative advantage of the arbitrageur decreases as there is a chance that a new trader submits an order at a better price and gets his order executed "at the expense" of the arbitrageur.

6.7 Discussion

The results show the benefits of the proposed batch-auctions for the general trading and investment public. As we saw in previous sections, given the number of parameters the model produces non-trivial results. For instance, the arbitrageur as an agent could be of benefit if the NBBO signal is delayed significantly compared to the mean arrival times of the agents. An almost completely opposite outcome can be achieved if we change the intensities of ZI's and MM's processes.

Similarly, the market maker's influence on ZI's surplus is not in line with both the theory and previous experiments. When we compared MM's behavior in both market regimes we only tested for one set of MM's parameters. Using multiple sets of parameters for both regimes would be extremely computationally demanding, we have therefore skipped these calculations in this thesis. Yet we did see a different effect the MMs had on ZI traders when we changed e.g. the intensity of the MM's process (ZI trader's intensity remained the same). This points to the possibility that even the simple MMs could be beneficial when appropriate parameters of the MM are picked. The interpretation of the results is therefore not straightforward, we have to take into account the relationship between various system attributes and the ways in which they interact with one another.

We also use a rather simple models of MMs, one which does not handle the inventory in any way and is in fact more primitive than our ZI traders. MMs tend to be more sophisticated traders, so in the future a different model could be more appropriate. Another option is that the MM could be included along with the arbitrageur as the fast trader, as market making is today mainly done through HFT.

We assumed that the behavior of our agents remains constant and does not change with the change of the market regime. In continuous trading, the arbitrageur has the ultimate advantage if the NBBO signal is delayed. He, therefore, submits two orders which are in the middle of the overlapping best bid and best ask prices and he has the certainty of getting the trade. This certainty is lost in batch-auctions yet the arbitrageur keeps on trading the same way. He still does have an information edge but does not use it fully - he could, for instance, submit orders very close to the best bid and best ask levels - increasing the probability of getting the trade even with a smaller profit.

Also, the switch to batch auctions gives a clear advantage to those with resting limit orders. Latency arbitrageur's profit is cut almost immediately by half, due to the fact that the resting limit orders get a better clearing price. This is of course under the assumption that there is only this trade during the auction and the clearing price is set only by it.

Chapter 7

Conclusion

High-frequency trading is still a relatively new concept with fairly unclear impact on the market quality. We focus on one aspect of HFT - the high-speed latency arbitrage which has been criticized by many for its social non-optimality. From various solutions which have been suggested in the literature, we pick and study the change of the market clearing - the effect of a switch from continuous to frequent-batch auctions. This change does not impact the general principles of the exchanges, it rather focuses on the micro-level of market clearing and how this change could impact the surplus of the general public.

The basis of our model comes from the fragmented market model used by Wah & Wellman (2017). Compared to the previous work, this thesis introduces a new agent (market maker) and allows for an NBBO signal and the clearing interval to be of different length. This allows for an interesting study of relationships between the two parameters and its impact on the overall market quality. The tests show the existence of multiple thresholds/lines where the market characteristics differ substantially.

We find that for a short enough auction the ZI traders experience a higher surplus in batch-auctions than in continuous regime. This is mainly due to the fact that in batch auctions the arbitrageur's edge decreases significantly. We also conclude that the arbitrageur can increase ZI surplus under longer delays of NBBO signals, as he pairs ZI traders' orders which would have been canceled otherwise. Contrary to previous research the market maker in our environment has not been of a benefit to ZI traders in most of the situations.

The latency arbitrage can be interpreted in the economic theory as a market failure, one that most likely can not be solved with a market solution. The proposed change to frequent-batch auctions does have a significant drawback for the exchange - lower number of trades. The exchange's business model is easily scalable and relies on the volume of securities traded. Choosing an approach which yields a lower revenue is not in line with exchange's optimal way of running their business. Therefore, the most likely way of implementing this policy would be a state-wide or international regulation, as the incentives of the agents (traders, exchanges) do not align.

The optimal policy would depend on three factors - the regulator's goal (maximizing trader's surplus, or minimizing abitrageur's profit), the length of the NBBO delay and the length of the auction. The first is selected by the regulator, second is observable empirically and third would have to be selected appropriately. The regulator would have to choose the length of the auction carefuly, as on one hand it does improve ZI trader's surplus, but it then decreases once we increase the length of the auction significantly. It is therefore possible that these two offseting effects do not produce a state in which batch-auctions are beneficial - simply because the NBBO is delayed by a significant amount.

	ZI Surplus	MM Profit	Bid/Ask
	1736890.7	27520.21	600.59
0	(1731897.65, 1741883.75)	(25910.45, 29129.96)	(598.34, 602.84)
	1847959.5	26230.04	557.54
100	(1842928.73, 1852990.27)	(24846.99, 27613.09)	(555.65, 559.43)
	1942544.7	23847.81	524.18
200	(1937369.71, 1947719.69)	(22399.94, 25295.69)	(522.45, 525.91)
	1994291.1	22943.62	491.47
300	(1988723.58, 1999858.62)	(21453.86, 24433.38)	(489.7, 493.24)
	2021207.4	19699.18	468.65
400	(2015400.18, 2027014.62)	(18152.18, 21246.18)	(467.06, 470.23)
	2018753.81	17363.68	448.96
500	(2013158.32, 2024349.29)	(16136.58, 18590.77)	(447.32, 450.6)
	1990797.55	16401.87	433.27
600	(1985070.68, 1996524.41)	(15165.23, 17638.51)	(431.69, 434.86)
	1945695.5	16226.2	422.72
700	(1939764.76, 1951626.24)	(14946.63, 17505.77)	(421.19, 424.26)
	1876604.55	14556.11	410.87
800	(1870228.9, 1882980.2)	(12865.98, 16246.24)	(409.38, 412.36)
	1809020.9	14035.75	405.19
900	(1802979.4, 1815062.4)	(12699.9, 15371.59)	(403.69, 406.69)
	1726788.15	13072.67	400.55
1000	(1720691.89, 1732884.41)	(11852.2, 14293.13)	(398.94, 402.17)

	# Trades	Volatility	Time
	3316.82	5.66	14.85
0	(3310.36, 3323.29)	(5.65, 5.67)	(14.73, 14.96)
	3108.73	5.59	15.82
100	(3103.03, 3114.43)	(5.58, 5.6)	(15.72, 15.93)
	2913.32	5.52	16.88
200	(2907.33, 2919.3)	(5.51, 5.53)	(16.75, 17.01)
	2702.95	5.46	17.92
300	(2696.58, 2709.32)	(5.44, 5.47)	(17.77, 18.06)
	2504.26	5.41	18.95
400	(2498.48, 2510.04)	(5.4, 5.43)	(18.78, 19.11)
	2315.29	5.37	19.44
500	(2309.79, 2320.78)	(5.36, 5.38)	(19.26, 19.62)
	2134.17	5.33	20.26
600	(2129, 2139.35)	(5.32, 5.34)	(20.09, 20.43)
	1969.99	5.31	21.3
700	(1964.68, 1975.31)	(5.3, 5.32)	(21.12, 21.48)
	1810.53	5.27	22.25
800	(1805.3, 1815.75)	(5.26, 5.28)	(22.04, 22.46)
	1676.52	5.25	23.11
900	(1671.65, 1681.39)	(5.24, 5.26)	(22.89, 23.34)
	1548.39	5.23	24.32
1000	(1543.47, 1553.32)	(5.22, 5.25)	(24.07, 24.58)

Table 7.1: Preliminary results - ZI's additional surplus.

	ZI Surplus	MM Profit	Bid/Ask
	1740301.57	0	616.38
0	(1735611.87, 1744991.27)	(NaN, NaN)	(614.45, 618.31)
	1735913.43	28037.64	601.84
2	(1731011.57, 1740815.29)	(26548.49, 29526.79)	(599.89, 603.78)
	1730364.9	53065.6	585.64
4	(1724990.9, 1735738.9)	(50957.95, 55173.24)	(583.29, 587.99)
	1727467.89	79666.66	567.28
6	(1723486.06, 1731449.72)	(77659.91, 81673.41)	(565.71, 568.85)
	1722878.9	100521.63	553.1
8	(1717221.49, 1728536.31)	(97316.92, 103726.35)	(551.06, 555.14)
	1724289.52	118938.73	538.88
10	(1719075.45, 1729503.6)	(115874.95, 122002.51)	(536.83, 540.92)

	# Trades	Volatility	Time
	3309.24	5.69	13.53
0	(3303.32, 3315.15)	(5.67, 5.7)	(13.45, 13.62)
	3319.53	5.66	14.82
2	(3313.15, 3325.92)	(5.65, 5.67)	(14.71, 14.93)
	3329.64	5.63	16.33
4	(3322.71, 3336.56)	(5.62, 5.65)	(16.2, 16.47)
	3339.18	5.62	18.28
6	(3334.17, 3344.19)	(5.61, 5.62)	(18.16, 18.41)
	3351.88	5.58	20.11
8	(3344.57, 3359.18)	(5.57, 5.59)	(19.91, 20.32)
	3357.25	5.56	22.35
10	(3351.12, 3363.39)	(5.55, 5.57)	(22.11, 22.59)

Table 7.2: Preliminary results - MM's count.

	0	100	200	300	400	500	600	700	800	900	1000
100	351 ***	94 ***	173 ***	178 ***	311 ***	176 ***	343 ***	166 ***	307 ***	104 ***	224 ***
200	782 ***	117 ***	528 ***	387 ***	607 ***	483 ***	497 ***	573 ***	560 ***	559 ***	480 ***
300	1,090 ***	482 ***	417 ***	812 ***	661 ***	656 ***	869 ***	613 ***	704 ***	788 ***	621 ***
400	1,294 ***	770 ***	582 ***	618 ***	971 ***	894 ***	775 ***	800 ***	991 ***	884 ***	679 ***
500	1,412 ***	958 ***	826 ***	696 ***	738 ***	1,037 ***	1,017 ***	955 ***	853 ***	886 ***	1,034 ***
600	1,461 ***	1,074 ***	1,002 ***	913 ***	811 ***	843 ***	1,048 ***	1,073 ***	1,059 ***	1,019 ***	944 ***
700	1,434 ***	1,107 ***	1,084 ***	1,035 ***	963 ***	885 ***	885 ***	1,030 ***	1,078 ***	1,102 ***	1,092 ***
800	1,404 ***	1,103 ***	1,120 ***	1,113 ***	1,084 ***	1,024 ***	964 ***	949 ***	1,020 ***	1,074 ***	1,113 ***
900	1,383 ***	1,060 ***	1,107 ***	1,137 ***	1,134 ***	1,108 ***	1,063 ***	1,005 ***	976 ***	1,028 ***	1,055 ***
1000	1,395 ***	1,049 ***	1,075 ***	1,123 ***	1,158 ***	1,163 ***	1,141 ***	1,104 ***	1,056 ***	1,026 ***	1,044 ***

Table 7.3: Difference in means (in thousands) of the ZI's surplus in continuous regime against batch-auctions (with arbitrageur)

	0	100	200	300	400	500	600	700	800	900	1000
100	347 ***	260 ***	213 ***	123 ***	174 ***	-9	113 ***	-114	52 ***	-147	-9
200	778 ***	315 ***	584 ***	349 ***	458 ***	267 ***	224 ***	300 ***	269 ***	269 ***	205 ***
300	1,088 ***	675 ***	495 ***	770 ***	517 ***	449 ***	602 ***	325 ***	419 ***	485 ***	345 ***
400	1,297 ***	962 ***	647 ***	580 ***	826 ***	676 ***	503 ***	521 ***	713 ***	589 ***	419 ***
500	1,413 ***	1,152 ***	892 ***	656 ***	598 ***	824 ***	754 ***	664 ***	565 ***	596 ***	769 ***
600	1,457 ***	1,268 ***	1,068 ***	872 ***	672 ***	629 ***	790 ***	788 ***	768 ***	722 ***	673 ***
700	1,442 ***	1,299 ***	1,153 ***	999 ***	816 ***	675 ***	622 ***	740 ***	790 ***	807 ***	823 ***
800	1,410 ***	1,294 ***	1,189 ***	1,076 ***	941 ***	813 ***	693 ***	661 ***	727 ***	781 ***	846 ***
900	1,384 ***	1,254 ***	1,170 ***	1,095 ***	989 ***	896 ***	794 ***	718 ***	688 ***	727 ***	790 ***
1000	1,397 ***	1,247 ***	1,148 ***	1,093 ***	1,015 ***	952 ***	879 ***	817 ***	761 ***	732 ***	777 ***

Table 7.4: P-values of ZI surplus test (No arbitrageur)

	0	100	200	300	400	500	600	700	800	900	1000
100	0	129 ***	121 ***	129 ***	105 ***	138 ***	102 ***	139 ***	111 ***	146 ***	124 ***
200	0	179 ***	152 ***	165 ***	142 ***	154 ***	159 ***	141 ***	151 ***	135 ***	159 ***
300	0	179 ***	184 ***	162 ***	163 ***	170 ***	150 ***	166 ***	157 ***	158 ***	159 ***
400	0	179 ***	186 ***	186 ***	166 ***	167 ***	164 ***	167 ***	158 ***	157 ***	166 ***
500	0	179 ***	186 ***	187 ***	184 ***	171 ***	165 ***	166 ***	167 ***	162 ***	159 ***
600	0	179 ***	186 ***	187 ***	185 ***	185 ***	171 ***	170 ***	167 ***	162 ***	163 ***
700	0	179 ***	186 ***	187 ***	185 ***	185 ***	178 ***	174 ***	171 ***	164 ***	163 ***
800	0	179 ***	186 ***	187 ***	185 ***	185 ***	179 ***	179 ***	174 ***	167 ***	166 ***
900	0	179 ***	186 ***	187 ***	185 ***	185 ***	179 ***	180 ***	177 ***	169 ***	168 ***
1000	0	179 ***	186 ***	187 ***	185 ***	185 ***	179 ***	180 ***	178 ***	171 ***	170 ***

Table 7.5: Difference in means of the arbitrageur's profit in continuous regime against batch-auctions

	0	100	200	300	400	500	600	700	800	900	1000
100	9,121 ***	8,462 ***	7,330 ***	8,343 ***	12,989 ***	8,705 ***	14,470 ***	11,152 ***	12,789 ***	9,773 ***	11,372 ***
200	13,223 ***	18,996 ***	9,792 ***	6,301 ***	16,185 ***	11,095 ***	14,428 ***	13,728 ***	13,683 ***	13,460 ***	13,247 ***
300	18,517 ***	18,382 ***	12,125 ***	14,909 ***	10,190 ***	9,344 ***	19,684 ***	14,638 ***	10,368 ***	14,460 ***	9,514 ***
400	17,891 ***	16,866 ***	18,619 ***	12,334 ***	18,153 ***	13,749 ***	10,566 ***	10,422 ***	14,992 ***	14,273 ***	17,032 ***
500	16,261 ***	18,097 ***	16,086 ***	19,480 ***	13,685 ***	15,225 ***	16,601 ***	11,720 ***	8,547 ***	11,914 ***	12,940 ***
600	14,518 ***	16,490 ***	14,355 ***	14,272 ***	21,865 ***	14,780 ***	17,008 ***	15,653 ***	15,054 ***	12,424 ***	11,254 ***
700	14,083 ***	14,771 ***	14,334 ***	14,673 ***	15,029 ***	19,639 ***	22,256 ***	13,010 ***	10,291 ***	13,301 ***	10,860 ***
800	15,128 ***	15,027 ***	11,562 ***	14,368 ***	13,325 ***	15,969 ***	22,725 ***	21,015 ***	14,798 ***	14,062 ***	12,301 ***
900	17,260 ***	15,788 ***	9,954 ***	11,181 ***	13,051 ***	12,418 ***	18,076 ***	20,736 ***	22,447 ***	18,254 ***	17,487 ***
1000	19,187 ***	20,778 ***	14,030 ***	11,921 ***	11,106 ***	10,643 ***	13,890 ***	14,998 ***	18,351 ***	23,602 ***	21,336 ***

Table 7.6: Difference in means of the MM's profit in continuous regime against batch-auctions (with arbitrageur)

	0	100	200	300	400	500	600	700	800	900	1000
100	12,894 ***	1,572	319	-871	1,047	-2,505	-888	-5,246	850	-2,913	832
200	13,797 ***	13,172 ***	3,567 **	-4,007	3,140 *	-1,046	-1,780	-2,080	1,734	-143	3,167 *
300	18,865 ***	11,955 ***	4,364 ***	3,431 **	-4,963	-1,855	1,521 +	328	-367	4,133 ***	81
400	18,459 ***	10,674 ***	11,447 ***	-308	5,070 ***	1,811 +	-5,963	-3,818	3,858 **	1,900 *	4,944 ***
500	17,751 ***	10,079 ***	7,360 ***	8,125 ***	763	5,069 ***	-482	-1,865	-2,025	-1,187	2,966 *
600	15,691 ***	8,769 ***	7,446 ***	2,341 *	8,707 ***	3,736 ***	370	3,069 **	2,260 *	871	1,308
700	12,322 ***	8,623 ***	5,838 ***	226	3,127 **	8,259 ***	5,106 ***	976	865	778	2,955 **
800	15,114 ***	7,073 ***	3,960 **	278	-430	3,797 **	6,636 ***	7,481 ***	3,975 ***	-111	2,212 *
900	16,292 ***	10,875 ***	4,791 ***	-44	330	1,152	1,397 +	6,360 ***	10,188 ***	6,348 ***	5,276 ***
1000	19,149 ***	13,457 ***	5,501 ***	1,445	-3,094	-912	-4,222	617	7,681 ***	10,290 ***	10,751 ***

Table 7.7: Difference in means of the MM's profit in continuous regime against batch-auctions (without arbitrageur)

	0	100	200	300	400	500	600	700	800	900	1000
0	628	-191,818	-66,617	40,210 ***	142,903 ***	212,625 ***	266,018 ***	288,154 ***	291,834 ***	297,080 ***	269,210 ***
100	-3,907	-25,868	-26,929	-15,161	5,544	27,865 *	36,548 ***	7,542	36,891 **	46,743 ***	35,994 ***
200	-3,043	5,850 *	-10,308	1,951	-5,747	-3,957	-7,182	14,671 **	1,238	6,733	-5,811
300	-1,422	1,272	11,597 **	-2,617	-798	6,004 *	-1,319	484	6,679 +	-5,931	-6,595
400	3,459 +	504	-958	1,291	-2,918	-5,352	-6,263	9,384 **	14,027 *	1,909	9,333
500	1,501	3,166 +	-112	161	3,192	-507	3,294	-3,388	4,179 +	6,988 *	4,452
600	-3,369	2,719	-288	-1,075	3,318 +	-2,088	7,940 *	2,466	741	-13	-1,528
700	9,196 *	622	1,773	3,912 *	-3,261	2,709 +	2,828	-2,049	3,960	2,052	-129
800	6,690 +	-957	1,943	3,322 +	-583	1,175	-4,691	-408	-1,992	3,498	2,472
900	2,256	2,142	-2,899	-1,420	-2,126	456	-2,285	963	4,152 *	-3,072	4,887
1000	2,713	6,004 +	5,733 +	10,385 **	183	1,730	4,413 *	1,397	-3,704	3,976 *	2,202

Table 7.8: Difference in means of ZI's surplus in all market settings with and without the arbitrageur

0	10	20	30	40	50	60	70	80	90	100	110	120	130	140
11 ***	-380	-383	-353	-311	-263	-219	-161	-110	-58	-12	31 ***	94 ***	131 ***	185 ***
150	160	170	180	190	200	210	220	230	240	250	260	270	280	290
226 ***	277 ***	319 ***	364 ***	402 ***	453 ***	482 ***	535 ***	572 ***	600 ***	628 ***	650 ***	682 ***	710 ***	744 ***

Table 7.9: Difference in means of ZI's surplus in continuous regime with and without the arbitrageur, $\lambda_{MM} = \lambda_{ZI} = 0.01$

	0	20	40	60	80	100	120	140	160	180
20	95 ***	288 ***	248 ***	207 ***	167 ***	135 ***	109 ***	82 ***	58 ***	37 ***
40	-27	360 ***	215 ***	256 ***	151 ***	176 ***	99 ***	104 ***	40 ***	48 ***
60	-139	292 ***	254 ***	111 ***	168 ***	146 ***	48 ***	49 ***	58 ***	-22
80	-243	182 ***	229 ***	150 ***	9 ***	83 ***	49 ***	36 ***	-64	-50
100	-341	63 ***	186 ***	95 ***	48 ***	-92	-16	-27	-63	-78
120	-439	-47	92 ***	91 ***	-10	-53	-190	-121	-84	-146
140	-532	-151	-11	50 ***	-29	-106	-149	-283	-218	-167
160	-619	-249	-119	-36	-21	-134	-185	-242	-369	-314
180	-703	-341	-212	-128	-74	-110	-225	-262	-321	-456

Table 7.10: Difference in means (in thousands) of the ZI's surplus in continuous regime against batch-auctions (with arbitrageur) - delay up to 180

	0	20	40	60	80	100	120	140	160	180
20	91 ***	92 ***	68 ***	58 ***	44 ***	38 ***	32 ***	34 ***	24 ***	22 ***
40	-32	123 ***	1	56 ***	-14	32 ***	-25	13 ***	-33	0
60	-142	50 ***	18 ***	-91	-31	-18	-94	-59	-32	-94
80	-242	-58	-29	-64	-180	-100	-103	-76	-165	-136
100	-344	-176	-71	-132	-150	-264	-174	-164	-162	-153
120	-442	-291	-161	-149	-215	-234	-337	-245	-206	-234
140	-536	-393	-264	-186	-249	-286	-308	-406	-328	-260
160	-625	-491	-370	-269	-241	-330	-349	-372	-477	-403
180	-708	-581	-461	-361	-290	-313	-401	-398	-435	-540

Table 7.11: Difference in means (in thousands) of the ZI's surplus in continuous regime against batch-auctions (with arbitrageur) - delay up to 180

	0	20	40	60	80	100	120	140	160	180
0	3	240 ***	250 ***	232 ***	212 ***	193 ***	171 ***	141 ***	122 ***	98 ***
20	0	45 ***	70 ***	83 ***	89 ***	96 ***	94 ***	94 ***	88 ***	83 ***
40	-2	4 +	36 ***	32 ***	47 ***	49 ***	47 ***	51 ***	49 ***	49 ***
60	0	-2	14 ***	30 ***	13 ***	29 ***	29 ***	34 ***	31 ***	26 ***
80	4 *	0	-8	18 ***	23 ***	11 ***	19 ***	30 ***	21 ***	13 ***
100	0	1	-7	5 *	13 ***	21 ***	14 ***	4 *	23 ***	24 ***
120	0	-4	-3	-8	7 ***	13 ***	24 ***	16 ***	0	10 ***
140	0	-3	-3	-4	-8	13 ***	13 ***	18 ***	12 ***	5 **
160	-2	-2	-1	-1	-9	-2	8 ***	12 ***	14 ***	10 ***
180	-1	0	1	0	-4	-9	-5	6 **	8 ***	13 ***

Table 7.12: Difference in means (in thousands) of ZI's surplus in all market settings without and with the arbitrageur (delay up to 180)

	0	20	40	60	80	100	120	140	160	180
20	2 ***	47 ***	11 ***	-22	-41	-58	-74	-85	-94	-104
40	1 **	127 ***	85 ***	94 ***	63 ***	67 ***	48 ***	46 ***	30 ***	30 ***
60	0	135 ***	137 ***	107 ***	125 ***	112 ***	96 ***	95 ***	94 ***	83 ***
80	0	136 ***	156 ***	141 ***	120 ***	136 ***	132 ***	120 ***	112 ***	116 ***
100	0	137 ***	160 ***	160 ***	147 ***	129 ***	141 ***	145 ***	137 ***	129 ***
120	0	136 ***	161 ***	167 ***	164 ***	151 ***	136 ***	143 ***	152 ***	149 ***
140	0	136 ***	161 ***	169 ***	172 ***	165 ***	154 ***	140 ***	147 ***	153 ***
160	0	136 ***	161 ***	169 ***	175 ***	173 ***	167 ***	157 ***	146 ***	148 ***
180	0	136 ***	160 ***	169 ***	176 ***	178 ***	174 ***	167 ***	160 ***	148 ***

Table 7.13: Difference in means of the arbitrageur's profit in continuous regime against batch-auctions

	0	20	40	60	80	100	120	140	160	180
0	6 ***	-11	-13	-14	-15	-13	-14	-12	-13	-12
20	17 ***	11 ***	11 ***	7 ***	8 ***	9 ***	7 ***	5 **	6 **	4 *
40	17 ***	6 **	13 ***	12 ***	11 ***	9 ***	9 ***	6 ***	11 ***	0
60	12 ***	5 **	6 **	9 ***	8 ***	4 *	9 ***	10 ***	9 ***	8 ***
80	14 ***	7 ***	6 ***	5 **	13 ***	11 ***	8 ***	8 ***	7 *	7 **
100	10 ***	6 ***	4 *	6 ***	7 ***	13 ***	9 ***	7 ***	10 ***	7 ***
120	10 ***	7 ***	7 ***	1	9 ***	6 ***	12 ***	13 ***	7 ***	5 **
140	8 ***	10 ***	5 **	1	3 +	4 *	6 ***	11 ***	14 ***	7 ***
160	8 ***	4 **	5 **	5 **	0	3 +	8 ***	3 *	11 ***	10 ***
180	8 ***	8 ***	7 ***	4 **	5 **	0	3 *	3 +	8 ***	8 ***

Table 7.14: Difference in means (in thousands) of ZI's surplus in all market settings with 8 and 4 MMs

	0	20	40	60	80	100	120	140	160	180
0	11 ***	-22	-24	-27	-26	-24	-29	-26	-27	-23
20	36 ***	27 ***	23 ***	16 ***	19 ***	23 ***	19 ***	14 ***	16 ***	12 ***
40	34 ***	20 ***	28 ***	21 ***	25 ***	20 ***	23 ***	18 ***	19 ***	11 ***
60	31 ***	14 ***	21 ***	27 ***	19 ***	17 ***	25 ***	22 ***	15 ***	16 ***
80	29 ***	20 ***	10 ***	16 ***	29 ***	23 ***	19 ***	21 ***	16 ***	11 ***
100	26 ***	16 ***	12 ***	16 ***	15 ***	28 ***	21 ***	11 ***	19 ***	16 ***
120	22 ***	17 ***	13 ***	5 **	15 ***	13 ***	25 ***	24 ***	14 ***	13 ***
140	20 ***	20 ***	13 ***	6 ***	6 **	13 ***	15 ***	26 ***	24 ***	15 ***
160	18 ***	16 ***	15 ***	8 ***	3 *	5 **	13 ***	13 ***	21 ***	20 ***
180	20 ***	17 ***	14 ***	12 ***	9 ***	-1	5 **	9 ***	12 ***	18 ***

Table 7.15: Difference in means (in thousands) of ZI's surplus in all market settings with 8 and 0 MMs

	0	20	40	60	80	100	120	140	160	180
0	0	12 ***	15 ***	16 ***	16 ***	17 ***	16 ***	16 ***	16 ***	16 ***
20	-1	4 ***	5 ***	6 ***	6 ***	8 ***	6 ***	7 ***	8 ***	9 ***
40	0	0	2 ***	3 ***	3 ***	5 ***	5 ***	5 ***	5 ***	7 ***
60	0	0	0	2 ***	2 ***	2 ***	3 ***	4 ***	5 ***	4 ***
80	0	0	-1	1 **	2 ***	1 *	3 ***	2 ***	3 **	3 ***
100	0	0	0	0	1 ***	2 ***	2 ***	2 ***	2 ***	3 ***
120	0	0	0	-1	0	1 ***	2 ***	2 ***	0	1 ***
140	0	0	0	0	0	1 ***	1 ***	2 ***	1 ***	1 **
160	0 *	0	0 *	0	0	0	0 *	1 ***	1 ***	2 ***
180	0 *	0	0	0	0	0	0	0 *	1 ***	2 ***

Table 7.16: Difference in means (in thousands) of arbitrageur's profit in all market settings with 8 and 4 MMs

	0	20	40	60	80	100	120	140	160	180
0	0	26 ***	31 ***	32 ***	32 ***	34 ***	33 ***	33 ***	34 ***	33 ***
20	-1	7 ***	11 ***	13 ***	13 ***	15 ***	16 ***	15 ***	16 ***	18 ***
40	-1	-1	5 ***	8 ***	7 ***	9 ***	10 ***	11 ***	11 ***	12 ***
60	0	0	1 **	3 ***	4 ***	5 ***	5 ***	8 ***	8 ***	9 ***
80	0	0	0	2 ***	4 ***	3 ***	5 ***	4 ***	6 ***	7 ***
100	0	0	0	1 +	1 ***	4 ***	3 ***	4 ***	5 ***	5 ***
120	0	0	0	-1	1 ***	2 ***	4 ***	4 ***	3 ***	3 ***
140	0	0	0	0	1 **	1 ***	2 ***	4 ***	3 ***	3 ***
160	0	0 *	0 **	0	0	1 *	1 ***	2 ***	3 ***	4 ***
180	0	0	0	0 +	0	1 *	1 **	1 ***	2 ***	3 ***

Table 7.17: Difference in means (in thousands) of arbitrageur's profit in all market settings with 8 and 0 MMs

	0	20	40	60	80	100	120	140	160	180
0	23 ***	25 ***	19 ***	18 ***	19 ***	17 ***	17 ***	17 ***	19 ***	17 ***
20	4 ***	1	2 *	3 **	2 *	-2	-1	-4	-3	-7
40	11 ***	-3	6 ***	3 **	9 ***	7 ***	7 ***	8 ***	11 ***	10 ***
60	13 ***	-2	2 **	12 ***	8 ***	7 ***	13 ***	13 ***	13 ***	17 ***
80	13 ***	0	0	5 ***	13 ***	11 ***	13 ***	13 ***	18 ***	18 ***
100	10 ***	3 ***	-1	7 ***	7 ***	15 ***	14 ***	15 ***	17 ***	13 ***
120	8 ***	6 ***	1 +	2 *	12 ***	9 ***	17 ***	17 ***	16 ***	19 ***
140	7 ***	6 ***	2 **	1	7 ***	13 ***	6 ***	18 ***	17 ***	16 ***
160	5 ***	4 ***	4 ***	0	3 ***	10 ***	14 ***	9 ***	16 ***	17 ***
180	4 ***	6 ***	4 ***	2 *	0	3 ***	14 ***	15 ***	8 ***	14 ***

Table 7.18: Difference in means (in thousands) of MM's profit in all market settings with 50 and 100 ZI traders

	0	20	40	60	80	100	120	140	160	180
0	36 ***	35 ***	31 ***	28 ***	28 ***	28 ***	26 ***	27 ***	27 ***	25 ***
20	4 ***	-1	-2	-2	1 +	1	4 ***	3 **	6 ***	4 ***
40	11 ***	-10	10 ***	8 ***	18 ***	15 ***	22 ***	21 ***	23 ***	19 ***
60	6 ***	-10	2 *	19 ***	16 ***	15 ***	22 ***	26 ***	24 ***	24 ***
80	5 ***	-6	3 ***	1	25 ***	20 ***	28 ***	18 ***	22 ***	26 ***
100	1	-3	-5	15 ***	1	24 ***	23 ***	29 ***	28 ***	18 ***
120	-2	-2	-3	6 ***	19 ***	3 **	24 ***	25 ***	28 ***	32 ***
140	-3	1	0	-4	20 ***	17 ***	4 ***	23 ***	26 ***	27 ***
160	-5	-3	-1	-5	3 ***	29 ***	17 ***	5 ***	19 ***	26 ***
180	-4	-1	-1	-1	-3	12 ***	33 ***	14 ***	5 ***	14 ***

Table 7.19: Difference in means (in thousands) of MM's profit in all market settings with 50 and 200 ZI traders

	0	20	40	60	80	100	120	140	160	180
0	0	32 ***	39 ***	42 ***	43 ***	44 ***	44 ***	45 ***	45 ***	45 ***
20	-2	21 ***	42 ***	61 ***	78 ***	93 ***	104 ***	114 ***	121 ***	130 ***
40	0	3 ***	31 ***	29 ***	48 ***	48 ***	58 ***	58 ***	64 ***	64 ***
60	1 +	1 +	6 ***	26 ***	21 ***	27 ***	36 ***	32 ***	35 ***	39 ***
80	1 *	0	2 ***	11 ***	21 ***	15 ***	17 ***	22 ***	26 ***	22 ***
100	0	0	0	3 ***	12 ***	17 ***	13 ***	12 ***	15 ***	18 ***
120	0	0	-1	1 *	4 ***	12 ***	14 ***	11 ***	8 ***	9 ***
140	0 +	0	-1	0	1 *	5 ***	11 ***	13 ***	10 ***	8 ***
160	0	0 *	0	0	0	2 ***	6 ***	10 ***	11 ***	10 ***
180	0	0	0	0	0	0	2 ***	6 ***	9 ***	11 ***

Table 7.20: Difference in means (in thousands) of arbitrageur's profit in all market settings with 50 and 100 ZI traders

	0	20	40	60	80	100	120	140	160	180
0	0	104 ***	115 ***	120 ***	121 ***	123 ***	124 ***	125 ***	124 ***	125 ***
20	-2	77 ***	129 ***	161 ***	184 ***	199 ***	213 ***	223 ***	230 ***	239 ***
40	0	7 ***	62 ***	54 ***	82 ***	80 ***	93 ***	92 ***	104 ***	101 ***
60	1 *	0	17 ***	44 ***	32 ***	44 ***	52 ***	51 ***	52 ***	55 ***
80	1 ***	0	2 ***	22 ***	34 ***	23 ***	27 ***	35 ***	37 ***	33 ***
100	0 +	0	0	6 ***	20 ***	29 ***	20 ***	18 ***	25 ***	29 ***
120	0	0	-1	1 ***	8 ***	18 ***	23 ***	18 ***	14 ***	18 ***
140	0 *	0	-1	0	1 **	9 ***	17 ***	21 ***	18 ***	13 ***
160	0	0 *	0	0	0	3 ***	10 ***	15 ***	19 ***	17 ***
180	0	0	0 +	0	0	0	4 ***	9 ***	14 ***	18 ***

Table 7.21: Difference in means (in thousands) of arbitrageur's profit in all market settings with 50 and 200 ZI traders

Bibliography

Baron, M., J. Brogaard, & A. A. Kirilenko (2016): "Risk and return in high frequency trading." *SSRN Electronic Journal* pp. –.

Biais, B., T. Foucault, & S. Moinas (2015): "Equilibrium fast trading." *Journal of Financial Economics* **116(2)**: pp. 292–313.

Biais, B., T. Foucault *et al.* (2014): "Hft and market quality." *Bankers, Markets & Investors* **128(1)**: pp. 5–19.

Bishop, A. (2017): "The evolution of the crumbling quote signal." *SSRN Electronic Journal* **Q2 | 2017(43)**: pp. –.

Boehmer, E., K. Fong, & J. Wu (2015): "International evidence on algorithmic trading." .

Brogaard, J., A. Carrion, T. Moyaert, R. Riordan, A. Shkilko, & K. Sokolov (2017): "High-frequency trading and extreme price movements."

Brogaard, J., T. Hendershott, & R. Riordan (2014): "High-frequency trading and price discovery." *The Review of Financial Studies* **27(8)**: pp. 2267–2306.

Brown, A. & F. Yang (2016): "Adverse selection, speed bumps and asset market quality." *SSRN Electronic Journal* pp. –.

Budish, E., P. Cramton, & J. Shim (2015): "The high-frequency trading arms race: Frequent batch auctions as a market design response." *The Quarterly Journal of Economics* **130(4)**: pp. 1547–1621.

Cliff, D., J. Bruten *et al.* (1997): "Zero not enough: On the lower limit of agent intelligence for continuous double auction markets." *HP Laboratories Technical Report HPL* .

Ding, S., J. Hanna, & T. Hendershott (2014): "How slow is the nbbo? a comparison with direct exchange feeds." *Financial Review* **49(2)**: pp. 313–332.

Duffy, J. & M. U. Ünver (2006): "Asset price bubbles and crashes with near-zero-intelligence traders." *Economic Theory* **27(3)**: pp. 537–563.

Farmer, J. D., P. Patelli, & I. I. Zovko (2005): "The predictive power of zero intelligence in financial markets." *Proceedings of the national academy of sciences of the united states of america* **102(6)**: pp. 2254–2259.

Gode, D. K. & S. Sunder (1993): "Allocative efficiency of markets with zero-intelligence traders: Market as a partial substitute for individual rationality." *Journal of political economy* **101(1)**: pp. 119–137.

Harris, L. (2013): "What to do about high-frequency trading."

Hoffmann, P. (2014): "A dynamic limit order market with fast and slow traders." *Journal of Financial Economics* **113(1)**: pp. 156–169.

Hunsader, E. S. (2013): "Einstein and the great fed robbery." `http://www.nanex.net/aqck2/4436.html`. Accessed: 2010-09-30.

Kirilenko, A., A. S. Kyle, M. Samadi, & T. Tuzun (2011): "The flash crash: The impact of high frequency trading on an electronic market." *Available at SSRN* **1686004**.

Kyle, A. S. & J. Lee (2017): "Toward a fully continuous exchange." *Oxford Review of Economic Policy* **33(4)**: pp. 650–675.

LeBaron, B. (2002): "Building the santa fe artificial stock market." *Physica A* .

Lewis, M. & D. Baker (2014): *Flash boys.* WW Norton New York.

Rossi, M., G. Deis, J. Roche, & K. Przywara (2015): "Recent civil and criminal enforcement action involving high frequency trading." *Journal of Investment Compliance* **16(1)**: pp. 5–12.

Sar, M. (2017): "Dodd-frank and the spoofing prohibition in commodities markets." *Fordham J. Corp. & Fin. L.* **22**: p. 383.

Tong, L. (2015): "A blessing or a curse? the impact of high frequency trading on institutional investors." *Available at SSRN* .

Wah, E., D. Hurd, & M. P. Wellman (2016): "Strategic market choice: Frequent call markets vs. continuous double auctions for fast and slow traders." *EAI Endorsed Trans. Serious Games* **3(10)**: p. e1.

Wah, E. & M. P. Wellman (2017): "Latency arbitrage in fragmented markets." *Algorithmic Finance* **5(3-4)**: pp. 69–93.

Wah, E., M. Wright, & M. P. Wellman (2017): "Welfare effects of market making in continuous double auctions." *Journal of Artificial Intelligence Research* **59**: pp. 613–650.

Walia, V., A. Byde, & D. Cliff (2003): "Evolving market design in zero-intelligence trader markets." In "E-Commerce, 2003. CEC 2003. IEEE International Conference on," pp. 157–164. IEEE.