Title

Essays on Excess Volatility and the Quality of Financial Markets

Author
Haghpanah, Farshad

Publication Date
2017

Peer reviewed|Thesis/dissertation

Essays on Excess Volatility and the Quality of Financial Markets

by

Farshad Haghpanah

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Business Administration

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Christine A. Parlour, Chair
Professor Dmitry Livdan
Professor Chris Shannon

Fall 2017

Essays on Excess Volatility and the Quality of Financial Markets

Abstract

Essays on Excess Volatility and the Quality of Financial Markets

by

Farshad Haghpanah

Doctor of Philosophy in Business Administration

University of California, Berkeley

Professor Christine A. Parlour, Chair

Many articles examine the quality of financial markets, and propose how to improve it. Most often, the goal is to improve the efficiency of the candidate market, specifically operational efficiency, i.e. the goal is to reduce the transaction costs. Although operational efficiency is one of the most important factors that contribute to the quality of financial markets (or at least the perception of it), several other factors play a role. Fairness, informational efficiency, and stability of financial markets are ultra important. My research exhibits the need to set a standard for quality assessment of financial markets. This purpose demands the illustration of the multidimensional nature of the quality of financial markets and the inadequacy of commonly used measures of market quality to capture it.

When exploring the quality of financial markets, contrary to common practice, I do not limit my attention to the sell-side. I have looked at both the sell-side and the buy-side, the impact of market design changes and new policies on their behavior, and the quality implications of it. However, since there is a lot to say on this enormous subject, I have chosen a principle of selection. I have selected three scenarios in which market quality is adversely affected by endogenous excess volatility induced by market participants' rational behavior.

Excess volatility is defined by Shiller (1981) as the movements in real stock prices that cannot be explained by new information about subsequent real dividends, and by LeRoy and Porter (1981) as the fluctuations in asset prices that are more than is consistent with present value models. They both found excess volatility using very different variance-bounds tests. Since then, however, their methods have been subject to many criticisms; from having little to none statistical significance, to having econometric problems, to their hypothesis requiring the risk neutrality assumption. At the same time, a lot has been done trying to reconcile financial models to the puzzling levels of volatility observed in financial markets.

I approach excess volatility in the three essays from a different angle. I define it as the movements in asset prices in excess of the changes in the fundamental value of the asset, and since it cannot be measured empirically, I have chosen a theoretical approach so I can examine which frictions or behaviors can cause excess volatility. Also, I have focused my

attention on volatility inducing frictions or behaviors for which we can find remedies in the form of market design change or policy. The implications of the theories presented can then be tested by implementing pilot programs in any exchange. One such program, the Securities and Exchange Commission's tick size pilot program, is already underway.

In the first essay, I examine the impact of inventory pressure on a single market maker. I present a continuous-time model of liquidity provision with long-lived information and endogenous inventory control. I show that an $(l, \underline{I}, \bar{I}, u)$ inventory control strategy is optimal. The optimal price depends on the inventory level. Furthermore, the instantaneous cost of holding a position (either long or short) to the market maker and the excess volatility in prices are in a direct relationship. The instantaneous cost can be interpreted as adverse selection cost (risk) of holding any non-zero position, cost of capital for holding long positions, and short selling cost for holding short positions. My result suggests reducing short selling costs to market makers, e.g. reducing borrowing cost or allowing naked short selling, decreases the excess volatility they induce as a result of their inventory control strategy.

My main focus in the second essay is on the characteristics of limit-order markets and the impact of market design changes on the incentives, actions, and payoffs of different kinds of liquidity providers. I show that agents with the lowest latency in the market have an incentive to front-run upcoming market orders. This behavior causes mismatch between expected execution prices and realized execution prices, excess volatility in spreads, and excess volatility in execution prices. In this model, the incentive depends on the latency of the fastest traders (not frequency) relative to the rest of the market and the tick size. Then, I examine several possible solutions; from changes to market latency or tick size to "taxing" front-running behavior through pricing of co-location based on order cancellations.

In the third essay, I switch my attention to the buy-side, and present a continuous-time model of trading on private information with uncertainty about the timing of information events. This uncertainty prevents partially-informed traders from knowing the "newness" of their private information. Their trades can cause the price to systematically diverge from fundamentals even when market participants are rational, there are no persistent exogenous demand (or supply) shocks, and there are no restrictions on trade. My result links the behavior of informed rational traders, i.e. the "smart money", to the seemingly manic episodes of price behavior, and suggests policy advice on the importance of transparency in maintaining informational efficiency and stability of financial markets.

Finally, I present recommendations aimed at standardizing a set of measures capable of capturing the multidimensional nature of the quality of financial markets. This is a first attempt to address this enormous subject, and it has not been my purpose to provide a "sufficient statistic" for market quality. I have aimed at providing a first draft to encourage further conversation and investigation. I have included only so much theory as I thought necessary, and have omitted altogether topics, although important, that did not seem to me to help with the comprehension of the problem at hand. Also, I have recorded seemingly unimportant details when I considered them illustrative of the nature of the problem.

JEL Classification: G14

To you, who showed me critical thinking.

Contents

List of Figures

Acknowledgments

I am eternally grateful to Christine A. Parlour for her enthusiastic support and resolute guidance. I am also immensely grateful to members of my Dissertation Committee—Dmitry Livdan and Chris Shannon. Furthermore, I would like to express my gratitude to William Fuchs, Nicolae Gârleanu, Brett Green, Terrence Hendershott, Amir Kermani, Martin Lettau, Ryan Liu, Gustavo Manso, Adair Morse, David Sraer, Richard Stanton, Annette Vissing-Jørgensen, and Johan Walden for providing helpful comments. Finally, I thank the authors of the research cited in the text for their insights and apologize for any misrepresentations in my discussion of their work.

Chapter 1

Naked Short Selling, Market Makers' Optimal Inventory Control Strategy, and Excess Volatility in Prices

"Lean is a way of thinking, not a list of things to do."

— Shigeo Shingo

1.1 Introduction

On main street, there has been much hand-wringing about the dangers of naked short selling.[1] In particular, market makers have been recently vilified for engaging in it.[2] In recent years, under immense pressure from market participants, the regulators of many countries have introduced extensive rules against naked short selling, tightened existing regulation, or expanded them by removing exceptions. This, however, developed without a consensus of opinion among academics on the subject.

Although the case for benefits of covered short selling is easy to defend with our understanding of the price discovery process and the importance of informationally efficient prices, the jury is still out on whether the benefits of allowing market participants to short naked outweigh its potential harms. The few papers published on the subject, however, seem to suggest that naked short selling is not as damaging as portrayed in the media. For instance, Fotak, Raman, and Yadav (2014) find that the positive impact of short sales on market quality does not depend on whether the short sale is covered. In this paper, I limit

[1] Naked short selling, i.e. uncovered short selling, is the practice of selling an asset short without having borrowed or located it in order to be able to make delivery.

[2] Some critics of naked short selling have even described it as counterfeiting shares.

my attention to one particular group of market participants, namely market makers, and examine the impacts of short selling costs on their behavior and consequently on prices.

This paper is related to the theory of inventory control, continuous-time inventory optimization in particular. Continuous-time inventory control literature, arguably, started with the work of Bather (1966). Following Bather (1966), many papers in this literature seek to show the optimality of the (s, S) restocking policy, in which when the inventory level falls below reorder point s, an order is placed to bring the inventory level up to S.[3] In this paper, I show that an $(l, \underline{I}, \bar{I}, u)$ inventory control strategy is optimal, in which the primary market maker places a request for quotation (RFQ) to enforce an upward jump to $\underline{I}$ whenever inventory level l is hit, and a downward jump to $\bar{I}$ whenever inventory level u is hit.

Restocking policy (also called re-balancing strategy), however, is not the only tool in the market makers' tool kit. They can also manage their inventory risk by controlling the future order flow through the price of liquidity, i.e. the spread, and the price of the asset, i.e. mid price. The former, i.e. the impact of market makers' inventory on the bid-ask spread and the other aspects of liquidity, has been studied extensively.[4] In this paper, I investigate the later and show that inventory pressure can cause the price to fluctuate in excess of the natural fluctuations of the underlying value of the risky asset. In other words, the price of an asset can deviate from the expected value of it as a result of the cost associated with holding a large (either positive or negative) position by the sell-side. Similar to Amihud and Mendelson (1980), my results are consistent with the existance of a prefered inventory position; a control band in this model. Surprisingly, however, the price is not necessarily downward monotonic here. The dynamics of the price is studied in Section 1.3.

The remainder of this chapter is organized as follows. First, I present the model in its general form and find an upper bound for the (expected) long-run average profit of the market maker. Then, the inventory control strategy of interest is defined, and it is shown that it can attain the upper bound established earlier. Finally, I examine the optimal pricing rule and the impact of inventory holding costs on the price of the asset and the pricing error. All proofs are in the Appendix.

1.2 The Model

I consider an infinite horizon, continuous-time model of liquidity provision with a primary market maker (PMM), a responding market maker (RMM), and liquidity takers with price-dependent cumulative order flow in the spirit of continuous-review inventory control models of Harrison, Sellke, and Taylor (1983) and Chen and Simchi-Levi (2006). There are two assets in the economy: one risk-free and one risky. Without significant loss of generality, I normalize the risk-free rate to zero. The underlying value of the risky asset, denoted by v,

[3]See Constantinides and Richard (1978), Bodt and Graves (1985), Sulem (1986), Bartmann and Beckmann (1992), and Chen and Simchi-Levi (2006).

[4]See Liu and Wang (2016).

is drawn from a continuous distribution with compact support and zero mean. Assume that the inverse of the distribution function is well defined on the interval $(0, 1)$ and its second moment is finite.

Trading takes place continuously. At any time $t \geq 0$, liquidity takers submit market orders to the primary market maker, who sets the price and clears the market. She makes pricing and inventory decisions simultaneously. When her position (long or short) becomes too big (too costly), she submits a request for quotation (RFQ) to the responding market maker in order to offload part of her position. The responding market maker has the obligation to respond to RFQs. He charges the primary market maker a fixed cost, denoted by k, each time she submits a RFQ and a variable cost, denoted by f, per share executed.[5]

The liquidity takers' cumulative order flow, denoted by X_t, follows a Brownian motion with time-dependent drift:

$$dX_t = \mu_t dt + \sigma dW_t, \qquad t \geq 0, \tag{1.1}$$

where μ_t is the drift, σ is the diffusion coefficient, and W denotes a Wiener process (Standard Brownian motion).[6] Assume that the drift, $\mu_t = \mu(e_t) = \mu(v - P_t)$, is strictly increasing in pricing error (market bias) at the time, denoted by e_t, and twice continuously differentiable. Assume further that $\mu(0) = 0$.[7]

The primary market maker, knowing her own inventory level Y_t, observes the cumulative order flow X_t, and sets the price P_t to maximize her (expected) long-run average profit. She also chooses whether to submit a RFQ in order to buy from the responding market maker or sell to him. Her strategy, therefore, consists of a pricing rule, denoted by P, and an inventory control strategy. The inventory control strategy, itself, consists of two sequences of stopping times $\{t_n^b\}_{n=1,2,\ldots}$ and $\{t_n^s\}_{n=1,2,\ldots}$ and two sequences of random variables $\{q_n^b\}_{n=1,2,\ldots}$ and $\{q_n^s\}_{n=1,2,\ldots}$, where $t_n^b \geq 0$ denotes the time of the nth RFQ submitted by the primary market maker to buy shares from the responding market maker and q_n^b the quantity of shares requested.[8] Given 1.1 and the primary market maker's inventory control strategy, her inventory level is given by:

$$Y_t = Y_{0^-} - \int_0^t \mu_\tau d\tau - \sigma W_t + \sum_{n=1}^{N^b(t)} q_n^b - \sum_{n=1}^{N^s(t)} q_n^s, \qquad t \geq 0. \tag{1.2}$$

Note that the jump-diffusion process Y_t is right-continuous left-limits (RCLL or cadlag).

[5] The responding market maker's obligation to respond to RFQs guarantees that there will be no failure to deliver by the primary market maker.

[6] All random variables in the model are defined on $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, \mathbb{P})$, and process W_t is a standard Brownian motion on $\mathcal{F}_t$. Furthermore, only strategies adapted to $\mathcal{F}$ are considered.

[7] All the results of the paper can be proven without the assumption that $\mu(0) = 0$. However, this assumption affects the price level and makes it more consistent with the rest of the literature and also the other essays in this piece.

[8] Both q_n^b and q_n^s are unsigned quantities.

Condition 1.2.1. *Under an admissible inventory control strategy, the primary market maker's controlled inventory level, i.e.* Y_t*, is bounded.*

A bounded controlled inventory remains in a control band at all times. In other words, there exists finite real numbers l and u, where $l < u$, such that $l \leq Y_t \leq u$.

The primary market maker's liquidity provision revenue up to time t is $\int_0^t r dX_\tau$, where $r \geq 0$ is the maker per-share rebate in the maker-taker fee structure, which is adopted by many financial markets to incentivize market participants to provide liquidity. But, she has to bear the cost of holding an inventory $\int_0^t c(Y_\tau)d\tau$, where $c(Y_t)$ is the inventory cost per unit of time, and the occasional cost of request for quotations $k+fq$, where q is the order quantity. Finally, part of her profit/loss comes from her valuation of the risky asset, $\int_0^t (P_\tau - v)dX_\tau$. She, therefore, solves:

$$\begin{aligned} \sup_{(P,\{t_n^b\},\{q_n^b\},\{t_n^s\},\{q_n^s\})} \liminf_{t\to\infty} \frac{1}{t}\mathbb{E}\Bigg[&\int_0^t (P_\tau - v + r)dX_\tau - \int_0^t c(Y_\tau)d\tau \\ &- \sum_{n=1}^{N^b(t)}(k + fq_n^b) - \sum_{n=1}^{N^s(t)}(k+fq_n^s) \mid Y_{0^-} = y \Bigg], \end{aligned} \tag{1.3}$$

where $N^b(t)$ and $N^s(t)$ denote the number of RFQs to buy and sell, respectively, up to and including time t.[9]

The instantaneous cost function $c : \mathbb{R} \to [0, \infty)$ can be interpreted as adverse selection cost (risk) of holding any non-zero position, cost of capital for holding long positions, and short selling cost for holding short positions. Assume that cost is continuously differentiable, strictly increasing in the absolute value of inventory level, and polynomially bounded. Assume further that holding no position has no cost, i.e. $c(0) = 0$.

Suppose the primary market maker's (expected) long-run average profit has an upper bound, denoted by π. Define function $\delta : \mathbb{R} \to \mathbb{R}$ as follows:

$$\begin{aligned} \delta(y) &= \mathbb{E}\left[\int_0^T \pi d\tau - \int_0^T (P_\tau - v + r)dX_\tau + \int_0^T c(Y_\tau)d\tau \mid Y_{0^-} = y\right] \\ &= \mathbb{E}\left[\int_0^T \big(\pi + c(Y_\tau) - (P_\tau - v + r)\mu_\tau\big)d\tau \mid Y_{0^-} = y\right], \end{aligned} \tag{1.4}$$

where y is the initial inventory level and T is the time of the first request for quotation under an admissible strategy by the primary market maker. Function $\delta(y)$ denotes the expected difference between earning the upper bound (expected) long-run average profit π (per unit of time) and the profit associated with the aforementioned strategy, on the time interval $[0, T)$.

[9] $\{N^b(t); t > 0\}$ is the counting process for the buying-RFQ arrival process $0 < t_1^b < t_2^b < \ldots$, and the random variables $t_1^b, t_2^b, \ldots$ are called buying-RFQ arrival epochs.

Proposition 1.2.1. *Assume π, v, r, l, and u are finite real numbers, where $l < u$. Assume further that Y_t is a diffusion process with drift $-\mu(v-p)$, $p(.)$ a bounded real function, and $c(.)$ a continuously differentiable, strictly increasing, and polynomially bounded function. Define function $\delta : \mathbb{R} \to \mathbb{R}$ as follows:*

$$\delta(y) = \begin{cases} \mathbb{E}\left[\int_0^T \big(\pi + c(Y_\tau) - (p(Y_\tau) - v + r)\mu(v - p(Y_\tau))\big)d\tau \mid Y_{0^-} = y\right] & l < y < u \\ 0 & y \leq l, y \geq u \end{cases} \tag{1.5}$$

where T is the first time Y_t hits either l or u. Then, function $\delta(.)$ satisfies:

$$[(p(y) - v + r)\mu(v - p(y)) - (c(y) + \pi)] = \mu(v - p(y))\delta_y - \frac{1}{2}\sigma^2\delta_{yy}, \tag{1.6}$$

for $l < y < u$.

Note that in Proposition 1.2.1, there are no assumptions about the primary market maker's (expected) long-run average profit having an upper bound, and π is just an arbitrary constant. Next, I prove that π is the upper bound for the primary market maker's (expected) long-run average profit.

Proposition 1.2.2. *Assume there exists a bounded right-continuous left-limits jump-diffusion process Y_t that follows 1.2 and a real-valued function $\delta(.)$ with continuous derivatives $\delta_y(y)$ and $\delta_{yy}(y)$ for all $y \in \mathbb{R}$ that satisfies 1.6. If the following conditions hold*

$$\delta(y_2) - \delta(y_1) \leq k + f|y_2 - y_1| \quad \forall\, l < y_1, y_2 < u, \tag{1.7}$$

$$\delta(y) = 0 \quad \forall\, y \leq l, y \geq u \tag{1.8}$$

then we have:

$$\begin{aligned} \liminf_{t\to\infty} \frac{1}{t}\mathbb{E}\Bigg[& \int_0^t (P_\tau - v + r)dX_\tau - \int_0^t c(Y_\tau)d\tau \\ & - \sum_{n=1}^{N^b(t)} (k + fq_n^b) - \sum_{n=1}^{N^s(t)} (k + fq_n^s) \mid Y_{0^-} = y\Bigg] \leq \pi. \end{aligned} \tag{1.9}$$

1.3 Optimal Strategy

In this section, I introduce a class of inventory control strategies and show that the primary market maker can attain the upper bound for her (expected) long-run average profit by following it. In doing so, I also find the conditions that her pricing strategy needs to satisfy.

Inventory Control Strategy

Definition 1.3.1. *In an $(l, \underline{I}, \bar{I}, u)$ inventory control strategy, the market maker places a request for quotation (RFQ) to enforce an upward jump to $\underline{I}$ whenever inventory level l is hit, and a downward jump to $\bar{I}$ whenever inventory level u is hit.*

Note that following this inventory control strategy is similar to adopting two (s, S) restocking policies for when the inventory level is too high or too low. All that is left to be done is to show that the upper bound for profit can be attained by following this invemtory strategy together with some pricing rule.

Theorem 1.3.1. *Consider a simultaneous inventory control and pricing problem, where the market maker maximizes her (expected) long-run average profit. If the controlled inventory process Y_t is a right-continuous left-limits jump-diffusion process, given in 1.2, satisfying Condition 1.2.1 and the function $\delta(.)$ in 1.4 satisfies condition 1.7, then there exists finite real numbers $l < \underline{I} \leq \bar{I} < u$ and a pricing strategy $p(Y_t)$ such that inventory control strategy $(l, \underline{I}, \bar{I}, u)$ is optimal.*

I think condition 1.7 is noteworthy amongst all the assumptions and conditions in Theorem 1.3.1. Intuitively, one can see that if the cost of re-balancing inventory to the primary market maker, the fixed cost in particular, is small enough, then an $(l, \underline{I}, \bar{I}, u)$ restocking policy cannot be optimal. For example, consider the unrealistic case of zero-cost RFQs. In this case, if there exists an optimal inventory level, then the primary market maker actively keeps the inventory level at the optimal level.

Pricing Rule

Theorem 1.3.1 doesn't pin down the pricing strategy to only one class of strategies. It, however, gives us conditions that the pricing strategy needs to satisfy. Note that for a given $(l, \underline{I}, \bar{I}, u)$ restocking policy, the optimal pricing strategy is not necessarily unique. Corollary 1.3.1 specifies one optimal pricing rule.

Corollary 1.3.1. *In Theorem 1.3.1, under an $(l, \underline{I}, \bar{I}, u)$ inventory control strategy, the pricing rule that satisfies*

$$P_t = p(Y_t) = \underset{p}{\operatorname{argmax}}\ (p - v + r - \delta_y(Y_t))\mu(v - p), \quad t \geq 0, \tag{1.10}$$

where δ_y is given in A.16, is optimal.

Consider a pricing strategy that satisfies 1.10. Since most of the finance literature agrees on the conjecture that the price of an asset is monotonically decreasing in the inventory level of the liquidity providers, it would be illuminating to examine it. The following proposition shows that, in fact, the optimal price is not necessarily monotonically decreasing in the inventory level.

Proposition 1.3.1. *Under an $(l, \underline{I}, \bar{I}, u)$ inventory control strategy, there exists an optimal pricing strategy in which the price is increasing in $[l, \alpha]$, decreasing in $[\alpha, \beta]$, and increasing in $[\beta, u]$ for $l \leq \alpha < \underline{I} \leq \bar{I} < \beta \leq u$.*

The reason for this price behavior is the costs that the primary market maker faces, the inventory instantaneous cost function in particular. When her position is large but not large enough that requires a request for quotation, she has two tools to bring down the inventory level: 1) reducing the price to incentivize future buy orders and disincentivize future sell orders and 2) doing the opposite to hit the cutoff point u sooner and submit a RFQ. The decision between these two depends on the inventory cost fuction. If the cost fuction is such that $l = \alpha$ and $\beta = u$ then the price is indeed monotonically decreasing in the inventory level. Meaning that the primary market maker is utilizing the first tool, i.e. price, to control her inventory as long as it is in the control band. However, in other cases, she proactively acts to catalyze the inevitable hitting event and interestingly she reduces her RFQ cost by doing so. Consider the state that the inventory level is very close to the cutoff point u. By increasing the price temporarily, instead of decreasing it to control the inventory level, she not only increases the probability of hitting u, which reduces the expected first hitting time and consequently the total cost of holding the position, but also makes the price more favorable for herself when submitting a request for quotation to sell (offload) part of her inventory to the responding market maker.

Comparative Dynamics

Having established the optimal strategies of the primary market maker, I illustrate the effects of the costs of her operation on her inventory control strategy and pricing rule. I will consider how changes in the cost function or RFQ charges may affect the width of the inventory control band and the impact of inventory level on the price.

The width of the inventory control band, i.e. $u - l$, depends on both the cost associated with holding an inventory and the fixed cost of RFQs. It is increasing in the fixed cost of RFQs and decreasing in the average cost of holding an inventory. Meaning, by increasing short selling costs, the primary market maker cannot afford to hold big negative positions. Next, we investigate whether this affects the informational efficiency of the price.

Corollary 1.3.2. *The higher the short selling costs of an asset, the higher the excess volatility of its price.*

The price of the risky asset, in this model, deviates from the true value of the asset as a result of the costs associated with holding an inventory. Corollary 1.3.1 shows how the pricing error (market bias), i.e. $e_t = v - P_t$, is linked to the inventory level through δ_y. Price is increasing in δ_y, and δ_y, given by A.16, is increasing in the level of function $c(y)$. In other words, $c(.)$ adds to the sensitivity of price to the inventory level. By increasing short selling costs, price will fluctuate around the underlying value more. That is, the higher the short selling costs of an asset, the higher the excess volatility of its price.

1.4 Conclusion

I examined the impact of inventory pressure on a primary market maker in a continuous-time model of liquidity provision with long-lived information and endogenous inventory control. I showed that an $(l, \underline{I}, \bar{I}, u)$ inventory control strategy is optimal, the optimal price depends on the inventory level, and the instantaneous cost of holding a position (either long or short) to the primary market maker and the excess volatility in prices are in a direct relationship. My results suggest that reducing short selling costs to market makers, e.g. reducing borrowing cost or allowing naked short selling, decreases the excess volatility they induce as a result of their inventory control strategy.

Chapter 2

Liquidity Mirages in Limit Order Markets and Market Stability

"A lack of transparency results in distrust and a deep sense of insecurity."

— Dalai Lama

2.1 Introduction

An open electronic limit order market is arguably the best way to organize trading activity in equity exchanges and many other secondary markets.[1] Not all limit order markets are created equal, however. They differ in organizational structure, matching principle, informational transparency, market latency, minimum execution latency, maximum trade frequency, minimum pricing increment (i.e. tick size),trading halts, maximum allowed intra-day volatility, trading curbs (and circuit breakers), fee (and rebate) schedule, and more.[2] Any change in market structure and design will have implications for the quality of limit order markets. Furthermore, any limit order market has many different participants and these participants are affected differentially by market design changes and new policies. Thus, to study limit order market design, we need to examine the impact of design changes or new policies on the incentives, actions, and payoffs of different kinds of market participants and then consider the impacts of their behavior on the quality of the market. For the last step, we need measures of market quality capable of capturing the complexity of limit order markets and

[1] See Parlour and Seppi (2008) for a survey on limit order markets.

[2] Examples of market organizational structure decisions include centralized or decentralized national markets and pure or hybrid limit order markets. Hybrid limit order markets add a specialist or an automatic execution system with guaranteed execution size to the limit order market design. See Seppi (1997) for more information on hybrid markets.

a general framework that shows which questions can be answered by using any specific set of measures.

In this paper, I limit my attention to a specific kind of market participant: participants with the lowest trading latency. In many exchanges, traders have the option to co-locate their servers next to the matching engine and hence low-latency traders are easily identified. They are often called high-frequency traders (HFTs).[3] Frequency of trades is not a bottleneck for any trader anymore, however, latency still is and will ever be. High-frequency trading has been publicized in the last decades and mostly negatively. Institutional investors have been the most vocal against the unnecessary complexity that high-frequency trading introduces to the trading process. Furthermore, there has been a lot of articles on its dangers and many in the media and finance industry blame HFTs for recent episodes of market instability, the flash crash in particular.

Academic literature, however, hasn't been so decisive about HFTs. On one hand, many papers argue high-frequency trading negatively impact different aspects of market quality. For example, Jarrow and Protter (2012) show that HFTs can create a mispricing and exploit it to the disadvantage of ordinary investors and Kirilenko et al. (2017) conclude that although HFTs did not cause the Flash Crash of 2010, they exacerbated market volatility. Furthermore, Budish, Cramton, and Shim (2015) have argued that the arms race between HFTs is socially wasteful and its cost is to investors. On the other hand, some papers have a favorable view of the HFTs. Brogaard, Hendershott, and Riordan (2014) show that HFTs improve price discovery, Angel and McCabe (2013) argue that because many HFT strategies are beneficial to other market participants, we cannot denounce the practice as unfair, and Brogaard et al. (2011) find that HFTs "tend to follow a price reversal strategy driven by order imbalances, add to the price efficiency, provide the best bids and offers for a significant portion of the trading day, and may dampen intraday volatility".

I examine the incentives of HFTs to take advantage of front-running opportunities built into limit order markets with price-time priority matching principle.[4] HFTs with access to the lowest latency in the market both take and make liquidity in anticipation of upcoming market orders. By taking liquidity at the same side of the limit order book to be hit by upcoming market orders, they not only prepare themselves to provide the liquidity later but also move the price in their favor temporarily. This behavior, a kind of predatory trading, is harmful in several ways. First, it increases the risk of submitting limit orders by slower market participants, i.e. the risk of providing longerlasting liquidity than what HFTs provide. Second, it induces excess volatility in quoted prices, and hence bid-ask spreads and executed prices. And, last but not least, the aforementioned excess volatility increases execution uncertainty of submitting market orders. This, in turn, exacerbates the breakdown

[3] Throughout this article, I use HFT to denote high-frequency trading and HFTs to denote high-frequency traders.

[4] The price-time priority principle is an order matching principle in which all orders are sorted by type, price and time of arrival (to the exchange). Market orders are given the highest priority for matching and are executed in the order in which they arrived. Limit orders are prioritized by price, and then by time of arrival.

between observed quotes and realized prices in limit order markets, amplifies the unfairness perceived by most market participants, and undermines investor confidence and public trust in the financial markets. This behavior is also evidence for the viewpoint that the distinction between liquidity makers and takers is blurred by the widespread adoption of algorithmic trading, high-frequency trading in particular, in limit order markets.

The remainder of this article is organized as follows. Section 2.2 presents the model in its most general form. In section 2.3, the optimal mid-period strategy of a "fast trader" is studied. Section 2.4 examines the regulatory and market-based solutions proposed to enhance market quality. Finally, section 2.5 concludes. All proofs are in the Appendix.

2.2 The Model

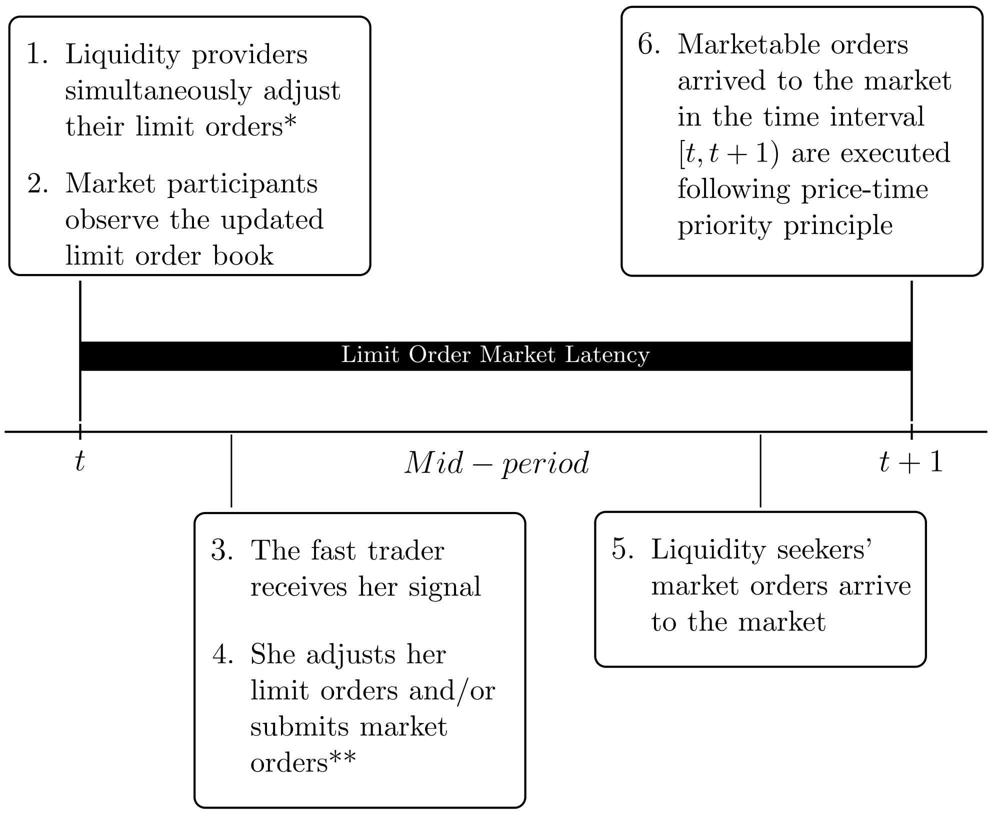
I examine a finite horizon, discrete-time trading model, inspired by Seppi (1997), with three kinds of traders: liquidity seekers, competitive risk-neutral liquidity providers, and a risk-neutral "fast trader" with a speed advantage over all other traders.[5] There are two assets in the economy: one risk-free and one risky. I normalize the risk-free rate to zero. The risky asset is traded at each time $t \in \{1, \ldots, T\}$ in an open, pure electronic limit order market. Prices and quotes are discrete, and the smallest increment by which they can move is the tick size, denoted by δ. In other words, the price grid is the discrete set $\mathcal{P} = \{p \mid \frac{p}{\delta} \in \mathbb{Z}\}$.

The timing and sequence of events in each period $[t, t+1)$, illustrated in Figure 2.1, are as follows: at time t, having observed the orders previously executed, liquidity providers and the fast trader simultaneously submit their limit orders, the limit order book is updated, and then all market participants observe the limit order book. Assume random and unobservable time priorities for all limit orders. Next, the fast trader receives a signal from which she can learn about the upcoming market orders from liquidity seekers. She then adjusts her limit orders and/or submits market orders strategically. Due to her speed advantage, her orders arrive to the market before liquidity seekers' orders. Then, market orders from liquidity seekers reach the market. And finally, all the marketable orders that arrived to the market in the time interval $[t, t+1)$ are executed at time $t+1$, following the price-time priority principle. The period time-length in this model is the limit order market latency—the reciprocal of the updating frequency of the limit order market information, the state of limit order book and order execution information in particular. Consequently, the number of time periods T is inversely proportional to the market latency.

In this model, only liquidity seekers and the fast trader demand liquidity by submitting market orders. Assume liquidity seekers only submit market orders, but the fast trader has a choice between market orders and limit orders. Since liquidity seekers have different trading latencies and frequencies that are not modeled here, their order flow is time indexed according to the orders' execution time. Their buy and sell market orders that arrive to the

[5] A trader co-located at the exchange can be considered an example of a fast trader. Co-located traders have the lowest latency in the market and hence speed advantage over all other traders.

Figure 2.1: Timing and Sequence of Events



** All liquidity providers; fast and slow.*
*** Fast trader's orders arrive to the market before liquidity seekers' orders.*

market in the time interval $[t, t+1)$, and hence get executed at time $t+1$, are denoted by unsigned quantities x^b_{t+1} and x^s_{t+1}. The net order flow from liquidity seekers is, therefore, $x_{t+1} = x^b_{t+1} - x^s_{t+1}$. Assume x^b_{t+1} and x^s_{t+1} are independent and identically distributed (i.i.d.) Poisson random variables with parameter $\frac{\mu}{T}$.[6]

Liquidity seekers' cumulative order imbalance over the trading horizon X_T is assumed to be informative about the terminal value of the risky asset. Specifically, the terminal value, which is publicly released at the end of the trading horizon T, is a non-decreasing function of the cumulative order imbalance from liquidity seekers, i.e.

$$v = \nu(X_T). \tag{2.1}$$

[6] Assume $0 < \frac{\mu}{T} < 1$.

Most of the results hold with this general terminal value function, but some of the resulting expressions are simpler in the following special case.

Assumption 2.2.1. *The terminal value of the risky asset v is linear in liquidity seekers' cumulative order imbalance over the trading horizon:*

$$v = \lambda X_T. \tag{2.2}$$

This assumption is closely related to the linear pricing rule in Kyle (1985). Furthermore, it is consistent with one of the primary results of the next chapter.

The fast trader demands liquidity if it is optimal given her private signal which carries information about the upcoming market orders that get executed at time $t+1$. The fast trader's market orders also get executed at time $t+1$, however they have higher time priority for matching due to the fact that they arrive to the market before liquidity seekers' market orders. Her buy and sell market orders are denoted by y^b_{t+1} and y^s_{t+1} and thus her net market order flow is $y_{t+1} = y^b_{t+1} - y^s_{t+1}$. Furthermore, the private signal is denoted by $\xi_t = (\xi^b_t, \xi^s_t)$, where ξ^b_t and ξ^s_t are drawn from discrete uniform distributions $\text{unif}(0, x^b_{t+1})$ and $\text{unif}(0, x^s_{t+1})$, respectively.

In order to simplify the notation, at any point in time, price levels on the price grid are indexed according to their location relative to the market valuation of the asset at that time. So if the valuation of the asset at time t is $u\delta \le v_t < (u+1)\delta$, then the price grid labels are $p^k_t = (u+k)\delta$.

The (state of) limit order book is the vector $q_t = (q^k_t)_{k\in\mathbb{Z}} = (..., q^{-1}_t, q^0_t, q^1_t, ...)$ where q^k_t is the signed quantity of limit orders available at price p^k_t. Positive (negative) quantities are outstanding buy (sell) limit orders. Given the state of the limit order book, we can determine depth at any price $|q^k_t|$, cumulative depths at or below any price higher than the current market valuation $Q^k_t = \sum_{i=1}^{k\times 1_{\{k>0\}}} |q^i_t|$, cumulative depths at or above any price lower than the current market valuation $Q^k_t = \sum_{i=k\times 1_{\{k\le 0\}}}^{0} |q^i_t|$, best bid price $b_t = \max\{p^k_t | q^k_t > 0\}$, best ask price $a_t = \min\{p^k_t | q^k_t < 0\}$, and bid-ask spread $s_t = a_t - b_t$.

Market participants are charged a liquidity fee, f per share, for removing liquidity from the limit order book by trading against it—submitting marketable orders—and a cancellation fee, c per share, for removing liquidity from the limit order book by canceling their own limit orders. On the other hand, they will receive a rebate, r per share executed, if they provide liquidity. This schedule of fees (and rebates) nests the Maker-taker transaction pricing model.

Liquidity Provision

Since the focus of this model is on market quality, market participants involved in the liquidity provision game—namely liquidity providers and the fast trader—are modeled as strategic agents. I describe these agents' action space and the notation used to represent their strategies here, and discuss their equilibrium strategies next.

Liquidity providers observe the aggregate buy and sell market orders, $z^b_t = x^b_t + y^b_t$ and $z^s_t = x^s_t + y^s_t$, and hence the cumulative buy and sell orders, Z^b_t and Z^s_t. Then, they estimate (filter) the current liquidity seekers' cumulative order imbalance X_t and consequently the terminal value of the risky asset. Throughout this article, I call the liquidity providers' valuation of the risky asset the market valuation, denoted by v_t,

$$v_t = \mathbb{E}_t[v|Z_t^b, Z_t^s] = \mathbb{E}_t[\nu(X_T)|Z_t^b, Z_t^s]. \tag{2.3}$$

Similar to Seppi 1997, liquidity providers are modeled as competitive agents that submit "limit orders until the marginal expected profit from additional submissions at each price is driven to zero". Since order sizes are discrete rather than continuous, the following criterion replaces the zero-profit condition widely used in the literature for modeling the aggregate behavior of competitive liquidity providers.

Condition 2.2.1 (Free Entry Condition (FEC))**.** *At any time t, after the limit order book is updated, the marginal expected profit of submitting an extra limit order at any price level p_t^k is non-positive.*

Alternatively, this condition can be replaced with a subgame where inventory and liquidity constrained agents with rational expectations arrive to the market sequentially, observe the limit order book, and submit an extra limit order if the marginal expected profit from additional submissions is still positive.

The aggregate liquidity provision strategy of competitive liquidity providers is denoted by $\mathcal{L}$. Let $l_t = (l_t^k)_{k\in\mathbb{Z}} = (..., l_t^{-1}, l_t^0, l_t^1, ...)$ be the updated vector of liquidity providers' limit orders at time t, where $|l_t^k|$ is the depth provided by "slow" liquidity providers at price p_t^k. Furthermore, let L_t^k be the cumulative depth at or below (above) p_t^k for positive (non-positive) k.

The fast trader's objective is to choose her strategy—denoted by $(\mathcal{H}, \mathcal{Y}, \Theta)$, where $\mathcal{H}$ is her liquidity provision strategy, $\mathcal{Y}$ is her mid-period market order strategy, and Θ is her mid-period limit order adjustment strategy—to maximize her final wealth. Let $h_t = (h_t^k)_{k\in\mathbb{Z}} = (..., h_t^{-1}, h_t^0, h_t^1, ...)$ be the updated vector of fast trader's limit orders at time t, y_{t+1}^b and y_{t+1}^s her mid-period buy and sell market orders, and $\theta_{t+1} = (\theta_{t+1}^k)_{k\in\mathbb{Z}} = (..., \theta_{t+1}^{-1}, \theta_{t+1}^0, \theta_{t+1}^1, ...)$ her mid-period limit order adjustments.

Since the fast trader knows her own inventory in addition to the cumulative buy and sell orders, she can deduce the current market valuation and the cumulative buy and sell orders from liquidity seekers X_t^b and X_t^s. Her valuation which is not biased by her own orders, and consequently is the more accurate one, is

$$v_t^* = \mathbb{E}_t[\nu(X_T)|X_t^b, X_t^s]. \tag{2.4}$$

The market bias, here, is the difference between the current market valuation and the fast trader's valuation, i.e. $e_t = v_t - v_t^*$.

2.3 Optimal Strategy

In this paper, my purpose is not to solve this model in its general form. It is to show that the mid-period market order strategy and limit order adjustment strategy of the fast trader are uniquely pinned down by the state of the limit order book and her initial limit orders. Proposition 2.3.1 demonstrates this fact by providing an algorithm to find the optimal mid-period strategy of the fast trader, given any arbitrary limit order book that satisfies the following condition.

Condition 2.3.1 (No Predatory Pricing Condition)**.** *Market participants never submit buy (sell) limit orders above (below) the current market valuation of the risky asset, or equivalently*

$$\forall k \in \mathbb{Z} : k l_t^k \leq 0\ ,\ k h_t^k \leq 0.$$

This condition makes certain that the shape of the aggregate limit order book is well-defined. To satisfy it, we need to impose some restrictions on fees and rebate parameters. Intuitively, the maker rebates should not fully compensate the losses from submitting predatory limit orders. In addition to Condition 2.3.1, let us assume that the initial state of the limit order book satisfies $h_t^k = \alpha q_t^k$ for some $\alpha \in [0,1]$ and all $k \in \mathbb{Z}$. This means that the fast trader is providing a fixed fraction of the liquidity at every price level in the price grid. The following proposition shows that, in equilibrium, the fast trader's mid-period problem is reduced to choosing a vector of cut-off points η_{t+1} and a limit order adjustment vector θ_{t+1}.

Proposition 2.3.1. *Consider a limit order book $q_t = l_t + h_t$ that satisfies Condition 2.3.1 and $h_t^k = \alpha q_t^k$ for some $\alpha \in [0,1]$ and all $k \in \mathbb{Z}$. If $(1-\alpha)r < \alpha c + (1-\alpha) f$, then there exist vectors η_{t+1} and θ_{t+1} adapted to the fast trader's filtration (at time t after receiving her private signal), such that her optimal mid-period strategy is to submit market orders y_{t+1}^b and y_{t+1}^s and limit orders θ_{t+1}, where*

$$y_{t+1}^b = \begin{cases} 0 & \text{if } \mathbb{E}[x_{t+1}^b | \xi_t^b] < \eta_{t+1}^1 \\ |Q_t^i| & \text{otherwise,} \end{cases}$$

where $i = \underset{k>0}{\operatorname{argmax}}\, \mathbb{E}[x_{t+1}^b|\xi_t^b] \geq \eta_{t+1}^k$, and

$$y_{t+1}^s = \begin{cases} 0 & \text{if } \mathbb{E}[x_{t+1}^s | \xi_t^s] < \eta_{t+1}^0 \\ Q_t^j & \text{otherwise,} \end{cases}$$

where $j = \underset{k\leq 0}{\operatorname{argmin}}\, \mathbb{E}[x_{t+1}^s|\xi_t^s] \geq \eta_{t+1}^k$.

The vector of cut-off points η_{t+1} is found recursively in the Appendix. For $\mathbb{E}[x_{t+1}^b|\xi_t^b] \in [\eta_{t+1}^k, \eta_{t+1}^{k+1})$, where $k > 0$,

$$\theta_{t+1}^i = \begin{cases} 0 & \text{if } i > 0, i \neq k+1 \\ \max(0, x^s - |q_t^{k+1}|) & \text{if } i = k+1. \end{cases} \tag{2.5}$$

And, for $\mathbb{E}[x_{t+1}^s|\xi_t^s] \in [\eta_{t+1}^k, \eta_{t+1}^{k-1})$, where $k \leq 0$,

$$\theta_{t+1}^i = \begin{cases} 0 & \text{if } i \leq 0, i \neq k-1 \\ \max(0, x^s - q_t^{k-1}) & \text{if } i = k-1. \end{cases} \tag{2.6}$$

Note that the fast trader's optimal mid-period strategy is determined by her private signal, the limit order vectors l_t and h_t, and the fee schedule of the trading venue. If the maker rebate is high enough relative to the cost of removing liquidity from the book, i.e. an average of taker and cancellation fees, it acts as an incentive for the fast trader to front-run other traders' orders. Also, note that her strategy is not affected by her inventory in this model. I have deliberately shut down the inventory channel in this model, because it has been studied in the first essay.

As it is shown in Panels A and B of Figures 2.2 and 2.3, the mid-period strategy of the fast trader induces excess volatility in bid-ask spread and mid-price. Note that mid-price is particularly more volatile than the underlying value of the risky asset in short term. Also, since spread widens only when the fast trader receives large enough signals about the upcoming order flow and the event of receiving the signals by her is probabilistic, her behavior increases the uncertainty about the execution cost of market orders.

Comparative Dynamics

Having established the mid-period strategy of the fast trader, I illustrate the effects of the limit order market design on the incentive for her to front-run upcoming market orders.

Corollary 2.3.1. *For a given limit order book $q_t = l_t + h_t$, cut-off points, i.e. η_t^k for all $k \in \mathbb{Z}$, are decreasing in tick size δ and maker rebate r (relative to taker fee f and cancellation fee c).*

The lower the cut-off points, the higher the probability of front-running, given a fixed order size. In other words, larger tick size and higher maker rebate increases the fast trader's incentive to front-run upcoming orders. Owing to the fact that tick size and maker rebate are what the fast trader earns front-running market orders.

Corollary 2.3.2. *The execution price is convex in the order size.*

Although according to Assumption 2.2.1 the terminal value of the asset is linear in the cumulative demand for it, as a result of the mid-period strategy of the fast trader, the execution price is convex in the order size. This means that the fast trader's behavior is a volatility multiplier, i.e. introduces excess volatility to the price. Also, this result is consistent with the convexity of transaction costs in order size, well-established in empirical studies.[7]

2.4 Implications for Policy and Practice

Many proposals have been put forward in response to supposed dangers of high-frequency trading. Budish, Cramton, and Shim (2015) argue for changing the design of financial

[7] See Lillo, Farmer, and Mantegna (2003), Engle, Ferstenberg, and Russell (2008), and Breen, Hodrick, and Korajczyk (2002).

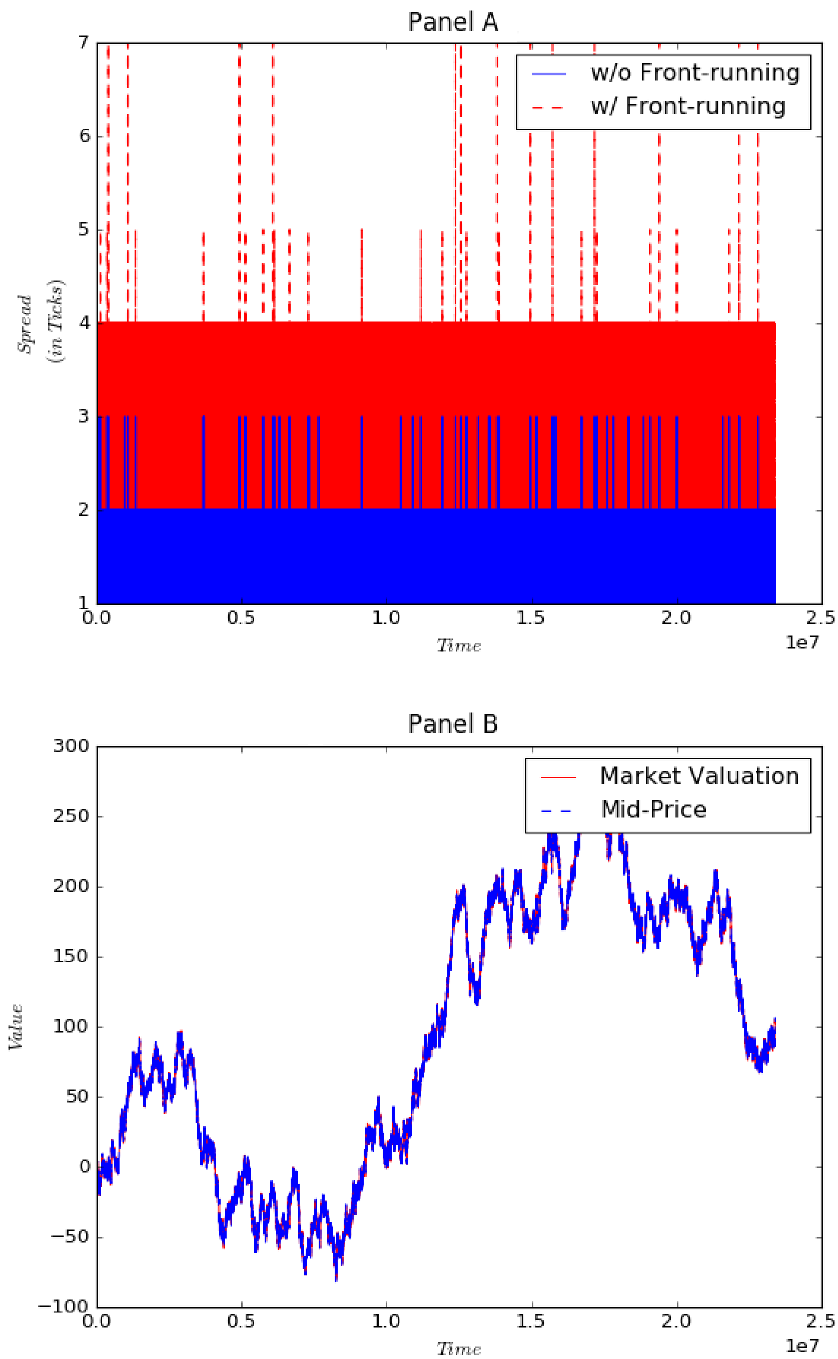


Figure 2.2: Bid-Ask Spread, Market Valuation, and Mid-Price Throughout a Trading Day. Panel A shows how the mid-period strategy of the fast trader affects bid-ask spreads, and Panel B illustrates its impact on the Mid-price.

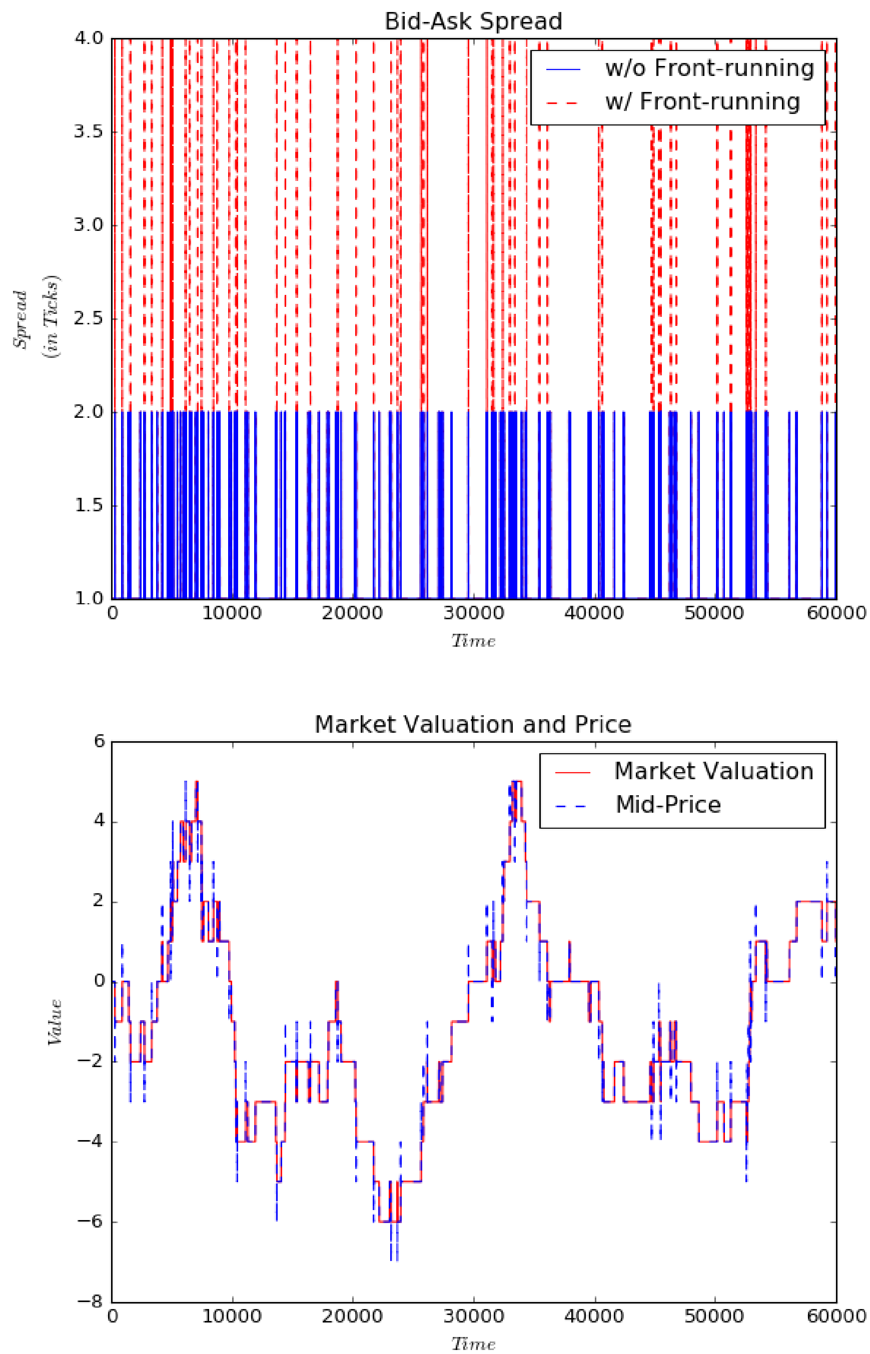


Figure 2.3: Bid-Ask Spread, Market Valuation, and Mid-Price Throughout a Minute. Panel A shows how the mid-period strategy of the fast trader affects bid-ask spreads, and Panel B illustrates its impact on the Mid-price.

exchanges from the continuous limit order book to frequent batch auctions. Brad Katsuyama, CEO of IEX stock exchange, has proposed a solution based on slowing the speed of trading.[8] In addition, there has been proposals for minimum resting time and message traffic tax. Minimum resting time policy makes it impossible for market participants to cancel their limit orders within a window of order submission. A message traffic tax is a tax on order-to-trade ratio (or cancel-to-fill ratio), disincentivizing canceling limit orders.[9]

Since almost every market design change or new policy will lead to a reallocation of resources among market participants. We need to ask ourselves a few questions before presenting a market design change or policy recommendation: would it lead to a Pareto improving outcome? If not, would the outcome be a Kaldor-Hicks improvement?[10] If so, which market participants are made better off, which are made worse off, and by how much? This paper is an attempt to shed some light on these issues.

First, we need to recognize that front-running is not the only strategy that HFTs follow. Many HFTs are actually providing liquidity at the best bid/offer. They can afford to do so by utilizing high-frequency trading to avoid being adversly selected by informed traders. This reduction in adverse selection risk faced by liquidity providers has led to lower quoted spreads. Also, excessive messaging and cancellation is not the problem, but the symptom. Furthermore, Proposition 2.3.1 shows that only large order flows are negatively impacted by the front-running activities of aggressive HFTs. Hence, proposals for minimum resting times and message traffic tax, although practicable in curbing some harmful behavior such as quote stuffing, would most likely benefit institutional traders with large order sizes rather than retail traders, if they are beneficial that is.[11]The reason for the uncertainty about beneficialness of these proposals is that they would also target those high-frequency liquidity providers who are not aggresively front-running market orders. Increasing the adverse selection risk of the liquidity providing HFTs would lead to higher quoted spreads, which makes the retail traders worse off. To make matters worse, in my opinion, these proposals would do nothing to prevent front-running; they just force the aggressive HFTs to use market orders instead of a mix of market and limit orders to front-run other market participants' order flows. Katsuyama's proposed solution has a similar problem. Since some HFTs are in the business of executing other market participants' orders, IEXs speed bump would also target those market participants. Furthermore, the front-running strategy studied in this essay would still work after implementing a speed bump.

Proposition 2.3.1 and Corollary 2.3.1 suggest simpler remedies for the problem at hand. Decreasing the tick size reduces the incentive for front-running upcoming orders and conse-

[8]IEX stock exchange slows down the speed of trading by adding 350 microseconds to the arrival and departure times of orders.

[9]Note that message traffic tax is not a tobin tax on transactions.

[10]see Kaldor 1939, Hicks 1939, and Scitovszky 1941 for more on Kaldor-Hicks efficiency.

[11]Quote stuffing is the scheme of entering and quickly canceling limit orders (quotes) to introduce excess noise in the real-time market data. It is used to make it costlier for competitors to process the real-time market data or camouflage real orders.

quently reduces the excess volatility introduced to the price by the fast trader's behavior. In addition, if we can reduce the expected market order size in each period, then the probability of front-running goes down. In this model, this can be achieved by reducing the limit order market latency. Limit order market latency is the reciprocal of the updating frequency of the limit order market information. Reducing it also improves the fairness perceived by market participants, since their execution price would be closer to the execution price estimated from the limit order book at the time of order submission.

In addition, a reasonable schedule of fees is crucial for having a well-functioning trading venue. The proof of Proposition 2.3.1 shows that maker rebates should not be too large relative to the taker and cancellation fees. Otherwise, market stability is undermined.

Another market design recommendation that could improve the stability of the market is to restrain excessively large market orders from destabilizing prices. This can be done by requiring market orders to be placed with price limits, in other words, by discarding market orders. In my opinion, reducing the number of order types on an exchange to one, i.e. only allowing submitting and canceling limit orders, would improve the stability of the market, particularly during volatile market conditions, without significant loss of functionality.

Finally, since there are so many unanswered questions about the effects of new policies and market design changes on the quality of our markets, implementing controlled experiments and pilot programs, like the Securities and Exchange Commission's tick size pilot program, could be illuminating. In particular, the idea of using frequent batch auctions instead of continuous limit order books deserves further investigation.

2.5 Conclusion

I showed that the "fast trader", the trader with the lowest trading latency, strategically submits limit orders that provide liquidity only to traders with small orders, i.e. traders least likely to be informed. She not only cancels her limit orders in anticipation of large orders but also front-runs those orders by taking liquidity from the limit order book. This results in tighter quoted spreads, but higher and more volatile realized spreads for large orders. In other words, as a result of the fast trader's behavior, spread widens and trades are executed at prices further from the mid-point of the best bid and offer, when the order flow is predictable. Also, I showed that, in this model, execution cost is convex in order size even when the terminal value of the asset is linear in the cumulative demand for it. This result suggests that the fast trader induces excess volatility in price. Finally, I examined several regulatory and market-based solutions proposed to enhance market quality.

Chapter 3

Private Timing of Information Events: A Theory of Momentum, Overshooting, and Correction

"Information is not knowledge."

— Albert Einstein

3.1 Introduction

Informational efficiency is one of the most important qualities of a market. Market participants contribute by collecting, processing, and transmitting (via trading) all the value-relevant information in the economy. This game of trading on information might nearly be a game of complete information. But almost surely, it is not a game of perfect information. Consider the following scenario: A friend of yours, an economist, informs you that she expects an increase in the federal funds rate at the next policy meeting. Assume she can predict the Fed rate increase with exceptional accuracy. Assume further that you know exactly how much of an impact the Fed rate increase would have on the value of an asset. Would you trade on it?

The relevant question is: how much of the private information is already in the current price of the asset? This is the problem that economic agents face when they receive a signal carrying private information. To address it, we need to know who else knows about this "news", since when they have known it, and what their trading strategies are. Although markets with informational asymmetries have been studied in several strands of literature, the optimal trading strategy of partially-informed traders facing this type of uncertainty, i.e. uncertainty about the newness of their information, is yet to be characterized.

An extensive literature explores the impact of information on market prices.[1] In this

[1]See Madhavan (2000).

literature, the informational content of prices and trading strategies of informed agents are of particular interest. Kyle (1985) arguably started the literature discussing the equilibrium strategy of an insider acting as an intertemporal monopolist and the resulting price behavior. Back (1992) presented the continuous-time equilibrium for general distributions of the underlying value of the asset. Empirical evidence, however, suggests that often multiple agents trade on private information in any time period with proven insider trading.[2] Holden and Subrahmanyam (1992) studied the effects of competition among identically-informed insiders on informational efficiency of the market, trading strategies, and insiders' profits. They Provide a discussion on the behavior of insiders and the non-existance of equilibrium in the continuous-time limit. Following Foster and Viswanathan (1994), a significant effort has been made to address the inference problem faced by traders with different private information.[3] In particular, Ostrovsky (2012) studies dynamic markets with partially-informed traders and whether information gets aggregated in such markets. However, one question remains: what happens when partially-informed traders don't know whether others have already received the same private signal? After all, the date and time that an agent receives a private signal is itself private information.

This paper is a first attempt to address private timing of information events and the effects of the resulting uncertainty about the "newness" of partially-informed traders' private signals on informational efficiency of the market. I consider a continuous-time model of trading with one insider (similar to the one in Kyle (1985)) and multiple partially-informed traders. Partially-informed traders' signals and their arrival times are private information. They try to infer from the price both other agents' additional information and "newness" of their own signal. In this equilibrium Bayesian model of mis-valuation, price can systematically diverge from fundamentals. However, the more informed trader, i.e. the insider, gradually trades the mis-valuation away. Consequently, the price trajectory could exhibit a momentum-overshoot-correction cycle. Here, the intensity and the uncertainty of the inflow of partially-informed traders play a vital role in the divergance of beliefs and sustainability of mis-valuation. This result is consistent with the empirical findings in Xiong and Yu (2011).[4]

The remainder of this article is organized as follows. Section 3.2 presents the model in its general form. The equilibrium concept is defined and its existence and uniqueness are studied in section 3.3. In section 3.4, I examine the relationship between the uncertainty about the "newness" of private signals and prolonged mis-valuation in the market, followed by the implications of it for policy and practice in section 3.5. Finally, section 3.6 concludes. All proofs are in the Appendix.

---

[2]See Cornell and Sirri (1992).

[3]See He and Wang (1995), Foster and Viswanathan (1996), and Back, Cao, and Willard (2000).

[4]Xiong and Yu (2011) examine several theories of mis-valuation — "bubble theories" — using a sample of warrants traded in China from 2005 to 2008, and conclude that the joint effects of short-sales constraints and heterogeneous beliefs are more evident in driving mis-valuation than other theories considered. They also point to the importance of inflow of new investors in understanding prolonged mis-valuation.

3.2 The Model

I consider a finite horizon, continuous-time model of trading with competitive market makers, strategic informed traders, and noise traders in the spirit of Kyle (1985). There are two assets in the economy: one risk-free and one risky. Without significant loss of generality, I normalize the risk-free rate to zero. The underlying value of the risky asset, denoted by V, is assumed to be of the form:

$$V = \sum_{i=1}^{N} V_i, \tag{3.1}$$

where the vector $(V_1, \ldots, V_N)$ is drawn from a multivariate normal distribution with mean zero and covariance matrix $\Sigma = diag(\sigma_1^2, \sigma_2^2, \ldots, \sigma_N^2)$. The initial value of the asset is normalized to zero. Hence, each component V_i can be interpreted as a fundamental change in the underlying value of the risky asset. Furthermore, information on each V_i can be acquired independent of other components and at a "cost".

Trading takes place continuously during the time interval $[0, 1]$. At the end of the trading horizon — time 1 — the liquidation value of the risky asset, i.e. the underlying value v, is publicly released. At any time t in the interval $[0, 1)$, noise traders and risk-neutral Bayesian informed traders submit market orders simultaneously to competitive risk-neutral market makers, who set the price and clear the market. Following the literature, I assume noise traders' cumulative order flow, denoted by X_t^N, follows a standard Brownian motion with mean zero and variance σ^2 (per unit of time), independent of the underlying value of the risky asset.[5]

There are two classes of informed agents: one who is "fully" informed, i.e. the insider who knows the liquidation value v since time 0, and a random number of partially-informed ones. The insider doesn't observe the private information events at which partially-informed traders receive their signals. In other words, she doesn't know which part of her private information is leaked and when the leak has happened. Agents informed about the ith component of the value, i.e. v_i, arrive sequentially at the market according to a homogeneous Poisson point process with parameter θ_i, independent of the underlying value of the asset. The arrival processes of agents informed about different components of the value are also independent of each other. The arrival intensity θ_i measures not only the intensity of the inflow of partially-informed agents but also the uncertainty of it. The reciprocal of arrival intensity $1 \,/\, \theta_i$ can be interpreted as a measure of the cost associated with acquiring information about v_i. Furthermore, throughout the paper, I call $\eta_i = \sigma_i^2 \,/\, \sum_{j=1}^{N} \sigma_j^2$ the relative informativeness of the private information about the ith component of the value.

Let X_t^i denote the cumulative, aggregate order flow of all the traders who have received signal v_i at or before time t, and X_t^I denote the cumulative order flow of the insider. Market

[5]All random variables in the model are defined on $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, \mathbb{P})$, and process X_t^N is a standard Brownian motion on $\mathcal{F}_t$.

makers observe only the total order flow, denoted by X_t, where

$$X_t = X_t^N + X_t^I + \sum_{i=1}^{N} X_t^i. \tag{3.2}$$

Consequently, market makers' filtration is the natural filtration of $\mathcal{F}$ with respect to X, denoted by $\mathcal{F}_t^X$. The perfectly competitive nature of the liquidity provision side of the market and risk-neutrality of market makers make them set the price equal to their current expectation of the liquidation value of the risky asset,

$$P_t = \mathbb{E}\left[V \mid \mathcal{F}_t^X\right]. \tag{3.3}$$

As in Back, Cao, and Willard (2000), I study linear equilibria. Therefore, market makers' pricing rule is

$$dP_t = \lambda_t dX_t. \tag{3.4}$$

Agents who trade on information — the insider and the partially-informed traders — observe the price process P_t and maximize their expected final wealth given their signal. Assuming that λ_t is positive, all agents can infer the total order flow X from the price, because strict monotonicity of the price in the total order flow implies its invertibility. Thus, the filtration of the insider can be written as $\mathcal{F}_t^{X,X^I} \vee \sigma(V)$ and that of a partially-informed trader who arrives at time τ knowing v_i as $\mathcal{F}_t^{X,X^{i,\tau}} \vee \sigma(V_i)$ for $\tau \leq t < 1$, where $\sigma(.)$ denotes the sigma algebra generated by the corresponding random variable, and $X_t^{i,\tau}$ denotes the cumulative order flow of the agent.

Hence, the insider's objective is

$$\sup_{X^I} \mathbb{E}\left[\int_0^1 (V - P_{t-}) dX_t^I - [P, X^I]_1 \mid \mathcal{F}_t^{X,X^I} \vee \sigma(V)\right], \tag{3.5}$$

where

$$dX_t^I = (\beta_t v + \phi_t P_t)\, dt \qquad 0 \leq t < 1, \tag{3.6}$$

and for any $i \in \{1, \ldots, N\}$ and $\tau \in [0, 1)$, the agent who arrives at time τ knowing v_i maximizes

$$\sup_{X^{i,\tau}} \mathbb{E}\left[\int_\tau^1 (V - P_{t-}) dX_t^{i,\tau} - [P, X^{i,\tau}]_1 \mid \mathcal{F}_t^{X,X^{i,\tau}} \vee \sigma(V_i)\right], \tag{3.7}$$

where

$$dX_t^{i,\tau} = \left(\beta_t^i v_i + \phi_t^i P_t\right) dt \qquad \tau \leq t < 1. \tag{3.8}$$

To ensure that the problem is well-posed and to rule out doubling strategies, I limit the set of strategies to absolutely continuous functionals, denoted by χ, that satisfy:

$$\mathbb{E}\left[\int_0^1 (\lambda_t(\chi_t + X_t^N))^2 d_t\right] < \infty.^6$$

[6] See Back (1992).

Following Back, Cao, and Willard (2000), the trading strategies are not allowed to have a stochastic component.

Note that Kyle (1985) is a special case of this model by assuming no arrival of partially-informed agents, i.e. $\theta = (\theta_1, \ldots, \theta_N) = 0$.

3.3 The Equilibrium

The solution concept is linear Markovian rational-expectations equilibrium. To show existence and uniqueness, I show that each agent's equilibrium strategy, including market makers' pricing rule, is optimal given the strategies of others. First, I need to define linear Markovian rational-expectations equilibrium of this model.

Definition 3.3.1. *A 3-tuple* $\left(P, X^I, \{X^{i,\tau} : i \in \{1, \ldots, N\}, \tau \in [0,1)\}\right)$ *is an equilibrium if:*

1. *Given the trading strategies of the insider* X^I *and all partially-informed types* $\{X^{i,\tau} : i \in \{1, \ldots, N\}, \tau \in [0,1)\}$*, the pricing rule* P *satisfies 3.3.*
2. *Given* P *and* $\{X^{i,\tau} : i \in \{1, \ldots, N\}, \tau \in [0,1)\}$*,* X^I *solves the optimization problem 3.5 over the set of admissible strategies.*
3. *For all* $j \in \{1, \ldots, N\}$ *and* $s \in [0,1)$*, given* P*,* X^I*, and* $\{X^{i,\tau} : i \in \{1, \ldots, N\} \backslash \{j\}, \tau \in [0,1) \backslash \{s\}\}$*,* $X^{j,s}$ *solves the optimization problem 3.7 over the set of admissible strategies.*

If market makers are Bayesians with correct priors, they should not be able to distinguish the total order flow X from the noise traders' cumulative order flow in equilibrium, otherwise the informed order flow would be identified by the market makers, price — the market's estimate of liquidation value — would jump, and hence it wouldn't be revised according to 3.4.

Condition 3.3.1 (Undetectability)**.** *In any linear Markovian rational expectations equilibrium of this market, the total order flow* X_t *must be a standard Brownian motion with mean zero and variance* σ^2 *on its own filtration* $\mathcal{F}_t^X$*.*

The undetectability condition, Condition 3.3.1, and the presence of the insider, guarantee that the constant depth result of Kyle (1985) is robust to this generalization of the model.

Lemma 3.3.1. *Only a positive constant* λ_t *is consistent with equilibrium, i.e.*

$$dP_t = \lambda dX_t. \tag{3.9}$$

One important insight, here, is that the undetectability of informed order flow makes it more likely for the underlying logic of Kyle (1985) to carry over in richer settings. Now, we are in the position to state and prove the main theorem of this section.

Theorem 3.3.1. *The unique linear Markovian rational expectations equilibrium of this model is* $\left(P, X^I, \{X^{i,\tau} : i \in \{1, \ldots, N\}, \tau \in [0,1)\}\right)$, *where*

- *market makers revise the price according to:*

$$dP_t = \frac{\sqrt{\Sigma_0}}{\sigma} dX_t, \tag{3.10}$$

where $\Sigma_0 = \sum_{i=1}^N \sigma_i^2$,

- *the insider trades against market bias (mis-valuation):*

$$dX_t^I = \beta_t (v - P_t)\, dt \qquad 0 \le t < 1, \tag{3.11}$$

- *and at any time in the time interval* $[\tau, 1)$, *partially-informed trader who arrives at time* τ *knowing* v_i, *i.e. agent of type* (i, τ), *follows the following strategy:*

$$dX_t^{i,\tau} = \left(\beta_t^i v_i + \phi_t^i P_t\right) dt \qquad \tau \le t < 1, \tag{3.12}$$

where β_t, β_t^i, *and* ϕ_t^i *are given in A.49 and A.57 in the Appendix.*

In equilibrium, market depth is constant. The insider does not trade against her own information, destabilize the market, or "ride the bubble".[7] Her trades increase the informativeness of the price at any point in time. In fact, it is her presence that guarantees the convergence of the price to the fundamentals. Furthermore, the presence of partially-informed traders affects the price dynamics as shown in Figure 3.1. Most of them trade in the direction of their signal, i.e. buy when their private signal is good "news" for the value of the asset and sell when it is bad "news", when they first receive their private signals.

Corollary 3.3.1. *Upon receiving a private signal, the partially-informed traders, initially, trade in the direction of their signel, regardless of the price level, if their signal is not too informative. In other words,*

$$\forall P_\tau \in \mathbb{R}, \exists \epsilon > 0 : if\ \eta_i < \epsilon,\ then\ dX_\tau^{i,\tau} \times v_i \ge 0. \tag{3.13}$$

This is precisely the reason for the snowball effect of a piece of information on the price levels.

[7] Note that if the insider knew the timing of information events and the private information leaked to the partially-informed traders, she could manipulate the market. However, in this model, the timing of information events is itself private information.

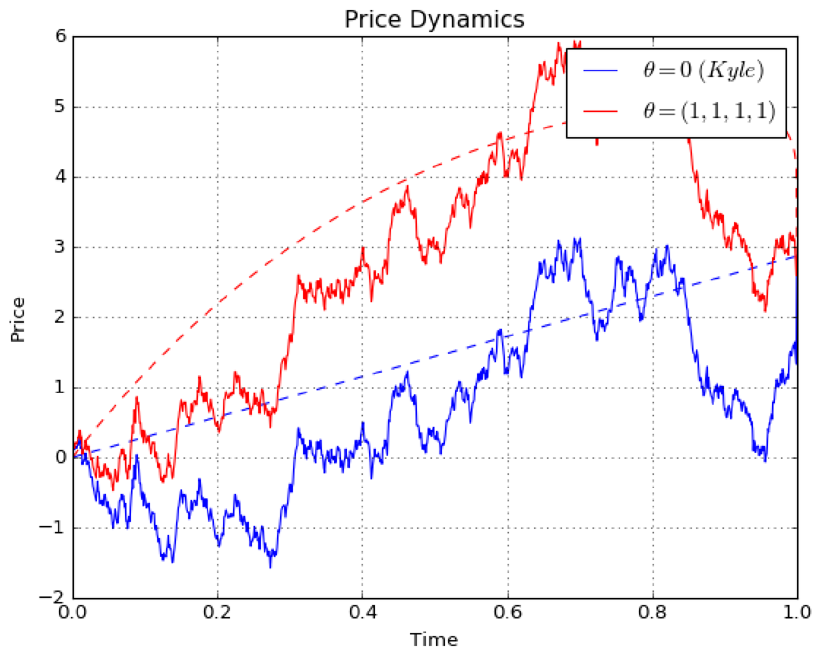


Figure 3.1: Price Dynamics. This figures shows the price dynamics in a market with only a monopolistic insider and a market with both the insider and partially-informed traders.

Comparative Dynamics

Having established the equilibrium strategies of market participants, I illustrate the effects of private timing of information events and the resulting uncertainties, in this section. I will consider how changes in the intensity and the uncertainty of the inflow of partially-informed traders, i.e. vector θ, may affect the strategies of different agents.

As it is shown in Panel A of Figure 3.2, the intensity of insider's trading is increasing in the intensity of the inflow of partially-informed traders. Compared to the monopolistic insider in Kyle (1985), i.e. special case of $\theta = 0$ in this model, she trades more aggressively on her information at any point in time, specially in the beginning of the game. It also depends on the relative informativeness of the signals. She trades more aggressively on her information when one group of partially-informed traders has most of the information. This is shown in Panel B of Figure 3.2.

The sensitivity of the partially-informed agents' trade to the market bias (mis-valuation) perceived by them depends on the relative informativeness of their signals and the intensity of the inflow of agents with the same information. This is shown in Figure 3.3. Increasing the intensity of the inflow of agents informed about a component of the underlying value results in more competition in that segment of the market. Consequently, agents who arrive early to the game trade more aggressively on the signal and those who arrive late barely trade on the signal. Unlike the insider, the partially-informed traders barely trade close to the public

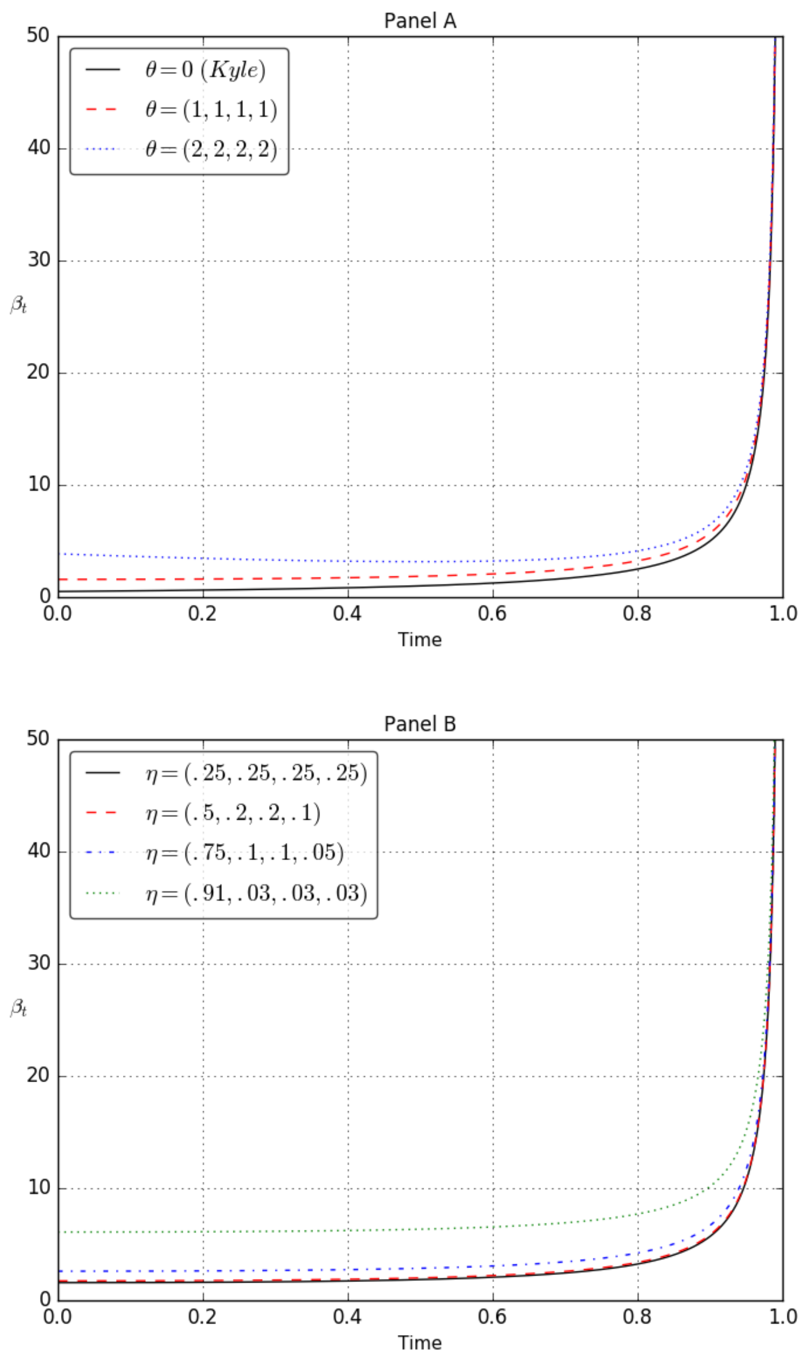


Figure 3.2: Intensity of Insider's trading. β_t is a measure of the sensitivity of the insider's trade to the market bias (mis-valuation). In both panels, $N = 4$, $\Sigma_0 = 4$, and $\sigma^2 = 1$. In Panel A, $\sigma_i^2 = 1$ for all $i \in \{1, \ldots, 4\}$ and the intensity (and the uncertainty) of the inflow of partially-informed traders is varied. In Panel B, $\theta = (1, 1, 1, 1)$ and the relative informativeness of the private signals is varied.

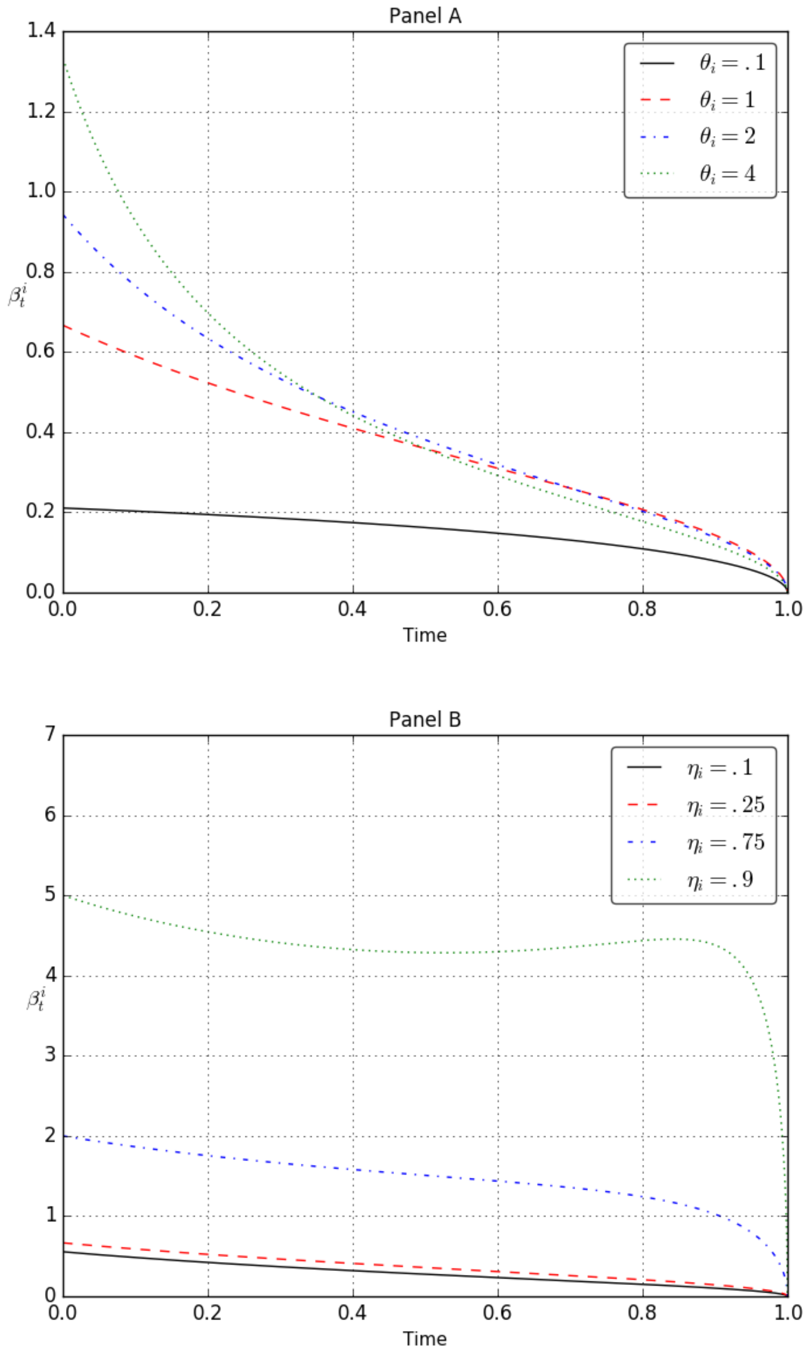


Figure 3.3: Intensity of Partially-Informed Agents' trading. β_t^i is a measure of the sensitivity of the partially-informed agents' trade to the market bias (mis-valuation) they perceive. In both panels, $N = 4$, $\Sigma_0 = 4$, and $\sigma^2 = 1$. In Panel A, $\sigma_i^2 = 1$ for all $i \in \{1, \ldots, 4\}$ and the intensity (and the uncertainty) of the inflow of partially-informed traders is varied. In Panel B, $\theta_i = 1$ and the relative informativeness of the private signal is varied.

release of the information. This is similar to the findings of Back, Cao, and Willard (2000), and consistent with the observed behavior of specialized institutional traders close to public announcements that carry information they are not fully knowledgeable about.

It is worth noting that the learning process of the partially-informed traders is affected by not only the relative informativeness of their signals but also the exact time that they receive them — the time of the agents' arrival in this model. The difference between the beliefs of two agents with the same private signal who arrive at the market at different times comes from their knowledge about their own presence. Also, the more informative their signal is about the liquidation value of the asset, the less the weight they put on the market price and their inventory in their learning process.

3.4 Overshoot in Market Valuation

Bubbles, market crashes, and financial crises provide us with fascinating bits of intuition and help us identify alternative explanations for different stages or the entirety of the misvaluation cycle — momentum, overshooting ("bubble"), and correction ("crash"). A bird's-eye view of all the related models in the literature demonstrates that proposed explanations are founded on one or a combination of the following assumptions: (1) the presence of irrational agents in the economy who not only fail to acknowledge that prices are highly inflated (deflated) and bear little to no relation to the intrinsic value of the assets but also might believe that the prices will keep rising (falling), (2) the existence of a persistent exogenous demand (or supply) shock in the economy, and (3) the existence of market frictions, short-sale constraints in particular.[8]

In this essay, I show prices can begin to diverge systematically from fundamentals in markets in which participants are rational, there are no persistent exogenous demand (or supply) shocks, and there are no restrictions on trade. In doing so, I aim to shed some light on the importance of transparency in markets and how private information propagates.

I define the fundamental value of a risky asset as the expectation of its liquidation value given all value relevant information, i.e. $\mathbb{E}[V \mid v_1, \ldots, v_N]$. The value relevant information is long-lived in this model, hence the fundamental value is time-invariant

$$\mathbb{E}[V \mid v_1, \ldots, v_N] = v.$$

The pricing error at any point in time, therefore, is $e_t = v - P_t$.

I examine the informational efficiency of this market using the following measures: (1) the unconditional variance of pricing error, (2) the conditional variance of the fundamental value given public information, denoted by Σ_t, (3) the expected price path and pricing error

[8] See Harrison and Kreps (1978), Allen, Morris, and Postlewaite (1993), Allen and Gorton (1993), Morris (1996), Shleifer and Vishny (1997), Hong and Stein (1999), Abreu and Brunnermeier (2003), Hong and Stein (2003), Ofek and Richardson (2003), Scheinkman and Xiong (2003), Brunnermeier and Nagel (2004), Hong, Scheinkman, and Xiong (2006), Brunnermeier and Oehmke (2013), and Shiller (2015).

given all value relevant information, (4) the expected conditional variance of pricing error given all value relevant information, and (5) the probability of overshoot.[9]

In this model, the conditional variance of the fundamental value given public information is equal to the unconditional variance of pricing error. From equations A.43 and A.44, we have

$$\Sigma_t = Var(V|\mathcal{F}_t^X) = \mathbb{E}\left[(V - P_t)^2\right] = \Sigma_0 - \Sigma_0 t. \tag{3.14}$$

This is similar to the results of Kyle (1985). In other words, the presence of the partially-informed traders doesn't change how much market makers expect to learn from the order flow. Intuitively, this is because noise traders provide camouflage for the informed order flow and market makers (the public) can not learn much from the order flow. All that market makers know is that regardless of the pricing rule they choose, the presence of the insider — the fully informed agent — guarantees the convergence of the price to the fundamental value of the risky asset without a jump at the date and time of the public release of information, and all they can do is to minimize their loss by minimizing the maximum profit that the insider can make, i.e. setting the price equal to their expectation of the fundamental value.

However, given all value relevant information, i.e. having observed v_i for all $i \in \{1, \ldots, N\}$, the price does not necessarily converge to the fundamental value monotonically, even in expectation. Consequently, the expected pricing error is not necessarily monotonically decreasing in time. This can be seen in Figure 3.1. In this model, the expected price path given all value relevant information is a function of the intensity and the uncertainty of the inflow of partially-informed traders, i.e. vector θ, and the realization of the value vector $(v_1, \ldots, v_N)$. Moreover, it can be infered that, in expectation, the price can overshoot. To examine the probability of overshoot, however, we need an accurate definition of the term.

Definition 3.4.1. *For any real number u, consider the first hitting time*

$$\gamma(u) = \begin{cases} inf\{t \in [0,1) : \frac{P_t}{(1+u)} = g(t)\} & \textit{if such } t \textit{ exists,} \\ 1 & \textit{otherwise,} \end{cases}$$

where $g(t)$ is some measure of fundamental value at time t, e.g. $g(t) = v$ in this model. The probability of overshoot with respect to g is $Prob(\gamma(u) < 1)$.

Proposition 3.4.1. *The probability of overshoot with respect to $g(t) = v$ is monotonically increasing in the intensity and the uncertainty of the inflow of partially-informed traders, i.e. θ_i's.*

[9] In control theory, overshoot refers to the response of a system — the market valuation (price) — exceeding its final, steady-state value — the fundamental value — and it is argued that the amplitude of the transient response, its time duration, and the settling time of any system are of considerable importance, because they must be kept within tolerable limits. For more information, see Ogata (2010) and Golnaraghi and Kuo (2009).

3.5 Implications for Policy and Practice

O'Hara (1995) defines the degree of market transparency as the ability of market participants to observe the information available in the trading process. In this model, since the price process is continuously observable by all market participants and the aggregate order flow is inferable from prices, market transparency can only be improved by providing market participants with additional information about the sources of the order flow or by decreasing the latency of public releases of information about the underlying value of the asset.

A decrease in the latency of public announcements about the value of the asset, in this model, means playing a shorter game each time. It makes the insider trade on her information more aggressively which results in a more informative price process. It also decreases the uncertainty about the number of agents informed about each component of the value between two public announcements which translates to lower probability of overshoot, again resulting in a more informative price process. This is in favor of legalizing trading on private information with timely disclosure requirements. The shorter the period of time insiders have to disclose their trades the more informationally efficient the market.

Providing market participants with additional information about the sources of the order flow can be done in many different ways. Here, I examine an easily implementable example. Consider increasing market transparency by disclosing summary statistics for the net order flow of different "types" of traders at the end of each trading day. For instance, if the risky asset in question is a firm then the requirement would be disclosing the net order imbalance of market makers (as a whole), corporate insiders, members of Congress, macro hedge funds (as a whole), other institutional traders (as a whole), and retail investors (as a whole). By doing so, we make it harder for informed traders, who arrive early to the game, to hide their information from the market makers and also from those who arrive later. We also provide partially-informed traders with more information about orders likely to be motivated by the same private signal as theirs which makes them less likely to "overreact" to their signal. Consequently, not only the informational efficiency of the market is increased but also the probability of overshoot is decreased. In other words, increasing the transparency of the market by disclosing more information post-trade increases price informativeness and decreases the instability of the market.

Since specialized financial institutions fit this model's description of partially-informed agents, the opacity of their trading strategies is another source of uncertainty about suitability of any piece of information for use in investment decision making. Therefore, releasing information about the inputs to their investment decision making process would facilitate the implementation of this recommendation.

3.6 Conclusion

I have studied the equilibrium behavior of the price of a risky asset in the presence of partially-informed traders facing uncertainty about the timing of information events. Although private timing of information events adds to the uncertainties faced by all market participants, partially-informed traders are taxed by these uncertainties the most. The resulting uncertainty about the "newness" of their otherwise "noiseless" signal about the value of the asset translates to lack of knowledge on their part about the suitability of their signal for use in investment decision making. I showed that the presence of partially-informed traders does affect the price process. In particular, the price trajectory could exhibit a momentum-overshoot-correction cycle. This is a new argument for how prices begin to diverge from the fundamental value of the assets. I also showed that the intensity and the uncertainty of the inflow of partially-informed traders play a vital role in the occurrence of prolonged mis-valuation. Since institutional traders fit this model's description of partially-informed agents, the results of the model link the prevalence of specialized financial institutions and the lack of transparency about their strategies to the probability of observing manic episodes of price behavior.

I also showed that the insider, i.e. the more informed trader, provides the market with information about the fundamental value of the risky asset through her trading, which, slowly but surely, negates the mis-pricing resulted from other traders' order flows. In other words, the presence of the insider is beneficial to the market, in the sense that it leads to higher informational efficiency of the equilibrium price process. However, note that I haven't studied strategic leaking of information by the insider in a market where some participants actively seek value relevant information and trade on it. The setup and structure of my model can be used to examine this behavior and the possibility and profitability of information-based market manipulation. Moreover, by adding the costs associated with acquiring different kinds of information to this model, one can examine the specialization of financial institutions and its impacts on informational efficiency of our markets.

Chapter 4

Conclusion, Discussion, and Recommendations for Future Research

"That which is measured improves. That which is measured and reported improves exponentially."

— Karl Pearson

Many measures of market quality exist. The purpose of this text was not to survey all the existing measures, compare their properties, or specify the relationships among them. It was to illustrate the inadequacy of commonly used measures of market quality to capture the multidimensional nature of the quality of financial markets. In my opinion, this is of paramount importance in today's fragmented financial markets. Because, not only trading venues' marketing departments are guilty of cherry-picking the most favorable performance measures in their reports to exaggerate their competitive advantage, but also regulators and academics occasionally refer to only one specific measure of market quality as a justification for their policy recommendations. These circumstances call for setting a standard for quality assessment of financial markets, especially markets easily accessible to the general public, like exchanges.

A brief look at the roles of financial markets in the economy and the functions that these roles entail can be of great benefit to our understanding of market quality and to our attempt to pick measures capable of capturing different aspects of it. First and foremost, secondary markets bring investors' preferences over investment horizon into harmony with the capital needs of businesses by unlocking trade; providing investors with resale option to smooth liquidity shocks and to realize gains-from-trade when they appear. Thus, together with primary markets, they enable public to participate in the production side of the economy by providing liquidity, which in turn provides businesses with access to more capital, improves risk sharing, and promotes saving. Second, they facilitate the aggregation of information dispersed in the economy, the mechanism for price discovery, and provide market statistics. These statistics, price in particular, can help all market participants in their resource allocation decisions and, hence, affect social welfare. Thus, financial markets form the feedback loop for businesses, industries, and the whole economy that helps managers and policy makers to evaluate their decisions and strategies.

Clearly, the functions of financial markets are interdependent, and naturally, different measures of market quality are highly correlated in normal times. For instance, the more operationally efficient a market, all else equal, the easier it is for capital to move and consequently the more informationally efficient the market will be. This, together with the fact that no widely accepted standards exist for measuring market quality, has led to the cherry picking problem discussed earlier. A remedy for this problem is to develop a general framework for evaluating market quality. And, a first step in doing so is to shed light on the limitations of commonly used measures of market quality.

Before going into details, a general characterization of what we call quality when it comes to financial markets may prove helpful. Let us cover the many dimensions of the quality of financial markets by reminding ourselves of what an investor wants in a market. First and foremost, she wants to be able to trade in a timely manner when she needs to. Some measures of liquidity, e.g. market depth, capture the first moment of this aspect of market quality. The other moments also play a role, as does the trading Latency. Second, she prefers lower transaction costs. The bid-ask spread and many other measures of operational efficiency capture that. Again, we need to be more mindful of the variations of transaction costs over time. Third, she wants to know that she will be able to offload her position (buy or sell) at a fair price in the future. Fair price, however, needs to be defined. If we agree that the fair price of an asset is the true underlying value of it, then measures of price efficiency, i.e. informational efficiency, are needed to express this aspect of market quality. In addition, she would deem the market "unfair", if her trades would always get executed at worse prices than she had observed when submitting her market orders.

It follows that if we want to compare the quality of trading venues, study the impacts of market design changes, or recommend new policies, we need to utilize multiple measures of market quality. In my opinion, we need to measure, at the minimum, the following: 1) the quoted execution cost of orders with different sizes and its volatility calculated from limit order book data, 2) the effective (realized) execution cost of orders with different sizes and its volatility calculated from trade and quote data, 3) the difference between realized and quoted execution costs given the state of limit order book at the time of order submission, and 4) informational efficiency of the prices.

The quoted execution cost can be measured by quoted spread for very small orders. Quoted spread is calculated by subtracting the best bid from the best offer, at every unit of time, and then averaging those values for the period of time under study. Clearly, the tighter the quoted spread, the more favorable the market to small retail investors. For larger orders, however, we need measures of market depth or better yet measures of effective (realized) execution costs. Measures of market depth, calculated from limit order book data, are not good proxies for the liquidity available to large investors. The liquidity has to be available to them for a period of time long enough for them to use it. Effective spread is designed to be a more adequate measure of market liquidity by using executed marketable orders information in conjunction with displayed quotations' mid-point. It is calculated by finding the weighted average of double the absolute difference between the execution price and the midpoint of best bid and best offer at the time of arrival of a marketable order to the limit order market, where the weights are the order sizes. Quoted spread and effective spread, together, capture a lot of what we want to capture and many existing measures of liquidity don't add much to them. They have their own weaknesses, however. Quoted spread is sensitive to sampling frequency and if the sampling times are known it can be manipulated. This could be remedied by using random sampling times. In addition, quoted spread doesn't take into account hidden and midpoint liquidity available in the market. Effective spread, on the other hand, is path dependent.[1] The same market can behave very differently given the sequence in which buyers and sellers arrive in the market and the size of their orders. That makes effective spread hard to sell if our purpose is to use it to make any comparisons, including comparing different trading venues' performance and the same venue's performance at different periods of time. My suggestion is to filter the trade data by order sizes and report multiple conditional effective spread measures.

Tighter spreads and more depth are not sufficient for concluding that a market is better, however. The uncertainty about execution costs needs to be measured, as well. The time dimension of this uncertainty can be captured by calculating the volatilities of the aforementioned measures.[2] The other dimension of this uncertainty is the difference between realized and quoted execution costs given the state of the limit order book at the time of order submission. The pre-trade execution cost calculated given the state of the limit order book, most often, doesn't match the realized cost of a large trade. Note that effective spread or effective/quoted spread ratio can not capture this, because effective spread is calculated at the time of arrival of a marketable order to the limit order market.[3] The breakdown between observed quotes and realized prices in limit order markets can amplify the perceived unfairness and consequently public distrust of the financial markets, which has the potential to impact the distribution of financial risk within the economy. Yet, commonly used measures of market quality can not capture it. This inadequacy is evident from the response of the media and Wall Street regarding claims of financial markets being improved. Although the spreads are thighter now than ever before and almost all measures of liquidity show improvements, institutional traders are not happy with the trading venues and they often claim that a large order that needed only tens of executions to manage the execution costs in the late 1990's needs hundreds or thousands of executions in recent years.

Last but not least, measuring informational efficiency of the prices is in the interests

[1] Execution prices in limit order markets exhibit path dependencies.

[2] These volatilities can also be used as imperfect proxies for the stability of financial markets.

[3] Effective/quoted spread ratio is advertised as a measure of price improvement in a trading venue. An effective/quoted spread ratio of less than 1 indicates that, on average, traders have paid less to execute their orders than what quoted spread suggests.

of everyone in the economy. Bounded information asymmetry is a desirable property of any socio-technical system. An informationally efficient market not only creates value by providing all market participants with reliable information but also reduces the risk of adverse selection and as a result the fear of not knowing the right prices which is one of the factors that catalyzes limited participation of households in the capital markets. In spite of that, I am yet to find a good measure of informational efficiency. Note that our measures of price impact are not good proxies for the informational efficiency of the market. The fact that limiting the market impact of large trades has become much more difficult in todays markets does not necessarily translate to a more or less efficient market. In a quality market, the per-share impact of trades on the price is close to the information content of the order. However, a common misconception is that the per-share impact of trades on the price has to be as small as possible.

At this point, it must be clear to the trained eye why I chose the three essays with excess volatility as the recurring theme. In the first essay, under the pressure from a costly-to-hold inventory, the sell-side introduces excess volatility to the price, reducing the informational efficiency of the market. In the second one, the harmful front-running strategies of the fastest traders in the market add to the execution uncertainty in the market, increasing the difference between expected and realized execution costs. These strategies also exacerbate the excess volatility problem in the market by increasing both the price volatility and the spread volatility. Finally, the third essay presents the idea that lack of pre- and post-trade transparency can cause buy-side induced excess volatility by increasing the uncertainties that partially-informed traders face. In other words, although transparency of a market by itself is not a measure of quality, it helps fairness and informational efficiency of the market.[4]

The three essays exhibit the inadequacy of available measures of market quality, the need for a proxy for informational efficieny of a market, and the need for better data. They, however, barely scratched the surface. A lot of work has to be done both to develop a good proxy for price efficieny and to examine ways to minimize the excess volatility in prices. The former, I expect, can be a lucrative research agenda, mainly because the true value of an asset is not observable and consequently we will need either model based measures or new data-driven techniques. The latter, although unfashionable and maybe unpublishable, is of great importance, inasmuch as eliminating excess volatility would improve all aspects of market quality discussed earlier and enhance the public confidence in the financial system.[5]

---

[4]Throughout this text, for lack of a better term, I used fairness to capture the notions of control and expectation, meaning the extent to which market participants are able to determine the execution cost of their trades, given the information available to them.

[5]Note that eliminating or reducing the natural volatility associated with a well-functioning financial market does not necessarily improve all aspects of market quality. When the fundamental value of an asset changes, if the price doesn't move as well, it would be detrimental to the informational efficiency of the market.

Appendix A

Appendix. Proofs

Proof of Proposition 1.2.1. This proposition follows from Equation C (3.3) in Karlin and Taylor (1981), p. 193 and the proof can be found in pages 193 to 197. □

Proof of Proposition 1.2.2. Consider the process $Z_t = \delta(Y_t)$, where $\delta(y)$ is a real-valued function with continuous derivatives $\delta_y(y)$ and $\delta_{yy}(y)$ for all $y \in \mathbb{R}$. Then, applying Ito's Lemma, we have:

$$\begin{aligned}\delta(Y_t) - \delta(Y_{0^-}) = &- \int_0^t \left[\delta_y(Y_{\tau^-})\mu_{\tau^-} - \frac{1}{2}\delta_{yy}(Y_{\tau^-})\sigma^2\right] d\tau - \int_0^t [\delta_y(Y_{\tau^-})\sigma] dW_\tau \\ &+ \sum_{n=1, t_n^b \leq t}^{N^b(t)} [\delta(Y_{t_n^{b-}} + q_n^b) - \delta(Y_{t_n^{b-}})] \\ &+ \sum_{n=1, t_n^s \leq t}^{N^s(t)} [\delta(Y_{t_n^{s-}} - q_n^s) - \delta(Y_{t_n^{s-}})], \qquad a.s.\end{aligned} \tag{A.1}$$

Since $\delta(y_2) - \delta(y_1) \leq k + f|y_2 - y_1|$, we have: $\delta(Y_{t_n^{b-}} + q_n^b) - \delta(Y_{t_n^{b-}}) \leq k + fq_n^b$ and $\delta(Y_{t_n^{s-}} - q_n^s) - \delta(Y_{t_n^{s-}}) \leq k + fq_n^s$. Thus,

$$\begin{aligned}\delta(Y_t) - \delta(Y_{0^-}) \leq &- \int_0^t \left[\delta_y(Y_{\tau^-})\mu_{\tau^-} - \frac{1}{2}\delta_{yy}(Y_{\tau^-})\sigma^2\right] d\tau \\ &- \int_0^t [\delta_y(Y_{\tau^-})\sigma] dW_\tau + \sum_{n=1}^{N^b(t)} [k + fq_n^b] + \sum_{n=1}^{N^s(t)} [k + fq_n^s].\end{aligned} \tag{A.2}$$

Knowing $Y_{0^-} = y$, taking expectations and dividing by t yield:

$$\begin{aligned}\frac{1}{t}\left(\mathbb{E}[\delta(Y_t)] - \delta(y)\right) \leq \frac{1}{t}\mathbb{E}\Bigg[&- \int_0^t \left[\delta_y(Y_{\tau^-})\mu_{\tau^-} - \frac{1}{2}\delta_{yy}(Y_{\tau^-})\sigma^2\right] d\tau \\ &+ \sum_{n=1}^{N^b(t)} [k + fq_n^b] + \sum_{n=1}^{N^s(t)} [k + fq_n^s] \mid Y_{0^-} = y\Bigg].\end{aligned} \tag{A.3}$$

After plugging in 1.6 from Proposition 1.2.1 and rearranging, we get:

$$\begin{aligned}\frac{1}{t}\mathbb{E}\left[\int_0^t [(P_\tau - v + r)\mu_\tau - c(Y_\tau)]\, d\tau - \sum_{n=1}^{N^b(t)}[k + fq_n^b] - \sum_{n=1}^{N^s(t)}[k + fq_n^s] \mid Y_{0^-} = y\right] \\ \leq \pi - \frac{1}{t}\left(\mathbb{E}[\delta(Y_t)] - \delta(y)\right).\end{aligned} \tag{A.4}$$

All we need to prove that π is the upper bound for the primary market maker's expected long-run average profit is to show that: $\liminf_{t\to\infty} -\frac{1}{t}\left(\mathbb{E}[\delta(Y_t)] - \delta(y)\right) \leq 0$. Then we have:

$$\liminf_{t\to\infty}\frac{1}{t}\mathbb{E}\left[\int_0^t [(P_\tau - v + r)\mu_\tau - c(Y_\tau)]\, d\tau - \sum_{n=1}^{N^b(t)}(k + fq_n^b) - \sum_{n=1}^{N^s(t)}(k + fq_n^s)\right] \leq \pi. \tag{A.5}$$

Note that the left-hand side is the expected long-run average profit in the maximization problem 1.9.

Now, I prove that $\liminf_{t\to\infty} -\frac{1}{t}\left(\mathbb{E}[\delta(Y_t)] - \delta(y)\right) \leq 0$.

*C*ase 1. $l < y < u$: If $y = Y_{0^-}$ is in the control band, Y_t almost surely hits Y_{0^-} again and we have $\liminf_{t\to\infty} -\frac{1}{t}\left(\mathbb{E}[\delta(Y_t)] - \delta(y)\right) \leq 0$.

*C*ase 2. $y \leq l$ or $y \geq u$: If $y = Y_{0^-}$ is out of the control band, from condition 1.8, $\delta(y) = 0$. Thus, we need to show $\liminf_{t\to\infty} -\frac{1}{t}\mathbb{E}[\delta(Y_t)] \leq 0$.

*P*roof by contradiction: Assume $\liminf_{t\to\infty} -\frac{1}{t}\mathbb{E}[\delta(Y_t)] > 0$. This means $\delta(Y_t) < 0, \forall Y_t$. However, we already know that $\delta(Y_t)$ can be zero. Contradiction! □

Proof of Theorem 1.3.1. Consider an arbitrary $(l, \underline{I}, \bar{I}, u)$ inventory control strategy defined in 1.3.1, where $l < \underline{I} \leq \bar{I} < u \in \mathbb{R}$. The primary market maker's controlled inventory level, given in 1.2, can be written as follows:

$$Y_t = Y_{0^-} - \int_0^t \mu_\tau d\tau - \sigma W_t + \sum_{n=1}^{N^b(t)}(\underline{I} - l) - \sum_{n=1}^{N^s(t)}(u - \bar{I}), \qquad t \geq 0. \tag{A.6}$$

Since the jump-diffusion process Y_t is a bounded right-continuous left-limits process and the function $\delta(.)$, given by 1.4, satisfies conditions 1.6, 1.7, and 1.8, we can follow a similar proof as in Proposition 1.2.2. Let $Z_t = \delta(Y_t)$, applying Ito's Lemma, we have:

$$\begin{aligned}\delta(Y_t) - \delta(Y_{0^-}) = &- \int_0^t \left[\delta_y(Y_{\tau^-})\mu_{\tau^-} - \frac{1}{2}\delta_{yy}(Y_{\tau^-})\sigma^2\right] d\tau - \int_0^t [\delta_y(Y_{\tau^-})\sigma] dW_\tau \\ &+ \sum_{n=1}^{N^b(t)}[\delta(\underline{I}) - \delta(l)] + \sum_{n=1}^{N^s(t)}[\delta(\bar{I}) - \delta(u)], \qquad a.s.\end{aligned} \tag{A.7}$$

To find the highest profit attainable under an $(l, \underline{I}, \bar{I}, u)$ inventory control strategy, we need to maximize the utility of a jump. Therefore, for given l and u, we find $\underline{I}$ and $\bar{I}$ to maximize

$$\delta(\underline{I}) - \delta(l) - k - f(\underline{I} - l) \tag{A.8}$$

and

$$\delta(\bar{I}) - \delta(u) - k - f(u - \bar{I}), \tag{A.9}$$

respectively. Therefore, we have $\delta_y(\underline{I}) = f$ and $\delta_y(\bar{I}) = -f$. Similarly, for given $\underline{I}$ and $\bar{I}$, we find l and u to maximize A.8 and A.9, respectively. Hence, we get $\delta_y(l) = f$ and $\delta_y(u)In = -f$. Furthermore, from 1.7 and 1.8, we know $\delta(l) = \delta(u) = 0$ and consequently $\delta(\underline{I}) = k + f(\underline{I} - l)$ and $\delta(\bar{I}) = k + f(u - \bar{I})$.Thus,

$$\begin{aligned}\delta(Y_t) - \delta(Y_{0^-}) = &- \int_0^t \left[\delta_y(Y_{\tau^-})\mu_{\tau^-} - \frac{1}{2}\delta_{yy}(Y_{\tau^-})\sigma^2\right] d\tau - \int_0^t [\delta_y(Y_{\tau^-})\sigma] dW_\tau \\ &+ \sum_{n=1}^{N^b(t)} [k + f(\underline{I} - l)] + \sum_{n=1}^{N^s(t)} [k + f(u - \bar{I})].\end{aligned} \tag{A.10}$$

Knowing $Y_{0^-} = y$, taking expectations and dividing by t yield:

$$\begin{aligned}\frac{1}{t}\left(\mathbb{E}[\delta(Y_t)] - \delta(y)\right) = \frac{1}{t}\mathbb{E}\Bigg[&- \int_0^t \left[\delta_y(Y_{\tau^-})\mu_{\tau^-} - \frac{1}{2}\delta_{yy}(Y_{\tau^-})\sigma^2\right] d\tau \\ &+ \sum_{n=1}^{N^b(t)} [k + f(\underline{I} - l)] + \sum_{n=1}^{N^s(t)} [k + f(u - \bar{I})] \mid Y_{0^-} = y\Bigg].\end{aligned} \tag{A.11}$$

After plugging in 1.6 from Proposition 1.2.1 and rearranging, we get:

$$\begin{aligned}\frac{1}{t}\mathbb{E}\Bigg[&\int_0^t [(P_\tau - v + r)\mu_\tau - c(Y_\tau)]\, d\tau - \sum_{n=1}^{N^b(t)} [k + f(\underline{I} - l)] \\ &- \sum_{n=1}^{N^s(t)} [k + f(u - \bar{I})] \mid Y_{0^-} = y\Bigg] = \pi - \frac{1}{t}\left(\mathbb{E}[\delta(Y_t)] - \delta(y)\right).\end{aligned} \tag{A.12}$$

All we need to prove that π is attainable under this policy is to show that:

$$\liminf_{t\to\infty} -\frac{1}{t}\left(\mathbb{E}[\delta(Y_t)] - \delta(y)\right) = 0.$$

Then we have:

$$\liminf_{t\to\infty} \frac{1}{t}\mathbb{E}\left[\int_0^t [(P_\tau - v + r)\mu_\tau - c(Y_\tau)]\, d\tau - \sum_{n=1}^{N^b(t)} (k + fq_n^b) - \sum_{n=1}^{N^s(t)} (k + fq_n^s)\right] = \pi. \tag{A.13}$$

Now, I prove that $\liminf_{t\to\infty} -\frac{1}{t}\left(\mathbb{E}[\delta(Y_t)] - \delta(y)\right) = 0$.

*C*ase 1. $l < y < u$: If $y = Y_{0^-}$ is in the control band, then from 1.7, we have

$$\delta(Y_t) - \delta(y) \leq k + f|Y_t - y| \quad \forall\, l < Y_t, y < u, \tag{A.14}$$

This holds for any pair $l < Y_t, y < u$ and consequently implies that $\delta(Y_t) - \delta(y)$ is bounded, and $\liminf_{t\to\infty} -\frac{1}{t}\left(\mathbb{E}[\delta(Y_t)] - \delta(y)\right) = 0$.

*C*ase 2. $y \le l$ or $y \ge u$: If $y = Y_{0^-}$ is out of the control band, by definition, $\delta(y) = 0$. Thus, we need to show $\liminf_{t\to\infty} -\frac{1}{t}\mathbb{E}[\delta(Y_t)] = 0$. Since $\delta(l) = 0$, $\delta(Y_t) = \delta(Y_t) - \delta(l) \le k + f|Y_t - l|$. Hence from Case 1, we know that $\delta(Y_t)$ is bounded, and $\liminf_{t\to\infty} -\frac{1}{t}\mathbb{E}[\delta(Y_t)] = 0$. $\square$

Notes on Corollary 1.3.1. Let us rewrite 1.6 in first-order differential equation form

$$\frac{d\delta_y}{dy} - \frac{2\mu(v - p(y))}{\sigma^2}\delta_y = \frac{(p(y) - v + r)\mu(v - p(y)) - c(y) - \pi}{-\frac{1}{2}\sigma^2}. \tag{A.15}$$

Solving for δ_y

$$\delta_y = \frac{\int exp(\int \frac{\mu(v-p(y))}{-\frac{1}{2}\sigma^2}dy)\frac{(p(y)-v+r)\mu(v-p(y))-c(y)-\pi}{-\frac{1}{2}\sigma^2}dy + constant}{exp(\int \frac{\mu(v-p(y))}{-\frac{1}{2}\sigma^2}dy)}, \tag{A.16}$$

and plugging in 1.10, we have:

$$P_t = \operatorname*{argmax}_p \, (p + r - \frac{\int exp(\int \frac{\mu(-p)}{-\frac{1}{2}\sigma^2}dy)\frac{(p-v+r)\mu(-p)-c(y)-\pi}{-\frac{1}{2}\sigma^2}dy + constant}{exp(\int \frac{\mu(v-p)}{-\frac{1}{2}\sigma^2}dy)})\mu(-p), \quad t \ge 0. \tag{A.17}$$

$\square$

Proof of Proposition 1.3.1. Consider the optimal pricing rule $p(Y_t)$ and let

$$y_{min} = \operatorname*{argmin}_{y\in[l,u]} p(y) \tag{A.18}$$

and

$$y_{max} = \operatorname*{argmax}_{y\in[l,u]} p(y). \tag{A.19}$$

First, I prove that y_{min} and y_{max} are unique.

*P*roof by contradiction: Assume $y^1_{min} < y^2_{min}$ are both arguments of the minima of the optimal pricing strategy, i.e function $p(y^1_{min}) = p(y^2_{min}) = p_{min}$. Then,

$$\left[(p(y^1_{min}) - v + r)\mu(v - p(y^1_{min})) - (c(y^1_{min}) + \pi)\right] = \mu(v - p(y^1_{min}))\delta_y(y^1_{min}) - \frac{1}{2}\sigma^2\delta_{yy}(y^1_{min}) \tag{A.20}$$

and

$$\left[(p(y^2_{min}) - v + r)\mu(v - p(y^2_{min})) - (c(y^2_{min}) + \pi)\right] = \mu(v - p(y^2_{min}))\delta_y(y^2_{min}) - \frac{1}{2}\sigma^2\delta_{yy}(y^2_{min}). \tag{A.21}$$

Now, I find π from A.20 and plug it in A.21,

$$\begin{aligned}&\big[(p_{min} - v + r)\mu(v - p_{min}) - (c(y^2_{min}) + \mu(v - p_{min})\delta_y(y^1_{min}) - \frac{1}{2}\sigma^2\delta_{yy}(y^1_{min})\\ &- [(p_{min} - v + r)\mu(v - p_{min}) - c(y^1_{min})]\big)\big] = \mu(v - p_{min})\delta_y(y^2_{min}) - \frac{1}{2}\sigma^2\delta_{yy}(y^2_{min}),\end{aligned} \tag{A.22}$$

yielding

$$c(y^1_{min}) - (c(y^2_{min}) + \mu(v - p_{min})(\delta_y(y^1_{min}) - \delta_y(y^2_{min})) - \frac{1}{2}\sigma^2(\delta_{yy}(y^1_{min}) - \delta_{yy}(y^2_{min})) = 0. \tag{A.23}$$

Since both y^1_{min} and y^2_{min} are arguments of the minima, $\frac{dp(y)}{dy}|_{y=y^1_{min}} = \frac{dp(y)}{dy}|_{y=y^2_{min}} = 0$ and consequently $\frac{d\delta_y(y)}{dy}|_{y=y^1_{min}} = \frac{d\delta_y(y)}{dy}|_{y=y^2_{min}} = 0$. Also, p_{min} is an optimal price, so it has to be the argument of the maximum for function $(p + r - \delta_y(y))\mu(v - p)$ at both inventory levels. This implies that $\delta_y(y^1_{min}) = \delta_y(y^2_{min})$. Thus, we have:

$$c(y^1_{min}) - c(y^2_{min}) = 0. \tag{A.24}$$

Now, $c(y^1_{min})$ and $c(y^2_{min})$ can not be equal if $y^1_{min} < y^2_{min} < 0$ or $0 < y^1_{min} < y^2_{min}$, because function $c(.)$ is strictly increasing in the absolute value of y. $c(y^1_{min}) = c(y^2_{min})$ is only possible for $y^1_{min} \leq 0 < y^2_{min}$.

From 1.3.1, we know that any optimal price satisfies:

$$(1 - \delta_{yp})\mu(v - p) + (p - v + r - \delta_y)\mu_p(v - p) = 0. \tag{A.25}$$

Taking the derivative of this equation with respect to y, and pluging in y^1_{min} and y^2_{min}, we have

$$\delta_{yyp}(y)\mu(v - p(y))|_{y\in\{y^1_{min},y^2_{min}\}} = 0. \tag{A.26}$$

Now, either $\mu(v - p(y))|_{y\in\{y^1_{min},y^2_{min}\}} = 0$ or $\delta_{yyp}(y)|_{y\in\{y^1_{min},y^2_{min}\}} = 0$. The former together with A.20 and A.21 yields: $\pi = -c(y)|_{y\in\{y^1_{min},y^2_{min}\}}$. Contradiction! The latter can be written as:

$$\frac{-2}{\sigma^2}[(1 - \delta_{yp})\mu(v - p) + (p - v + r - \delta_y)\mu_p(v - p) - c_p(y)]|_{y\in\{y^1_{min},y^2_{min}\}} = 0, \tag{A.27}$$

which implies $c_y(y^1_{min}) = c_y(y^2_{min})$. Function $c(.)$, however, is decreasing at $y = y^1_{min}$ and increasing at $y = y^2_{min}$. Contradiction!

So, y_{min} is unique. Similarly, it can be shown that y_{max} is unique. I conjecture that $y_{max} < 0 < y_{min}$. Intuitively, the primary market maker chooses the minimum price when she wants to sell, that is when her inventory is large enough. Let $\alpha = y_{max}$ and $\beta = y_{min}$. Next, I show that the price is increasing in $[l, \alpha]$, decreasing in $[\alpha, \beta]$, and increasing in $[\beta, u]$ for $l \leq \alpha < \underline{I} \leq \bar{I} < \beta \leq u$.

*P*art 1. Price is decreasing in $[\alpha, \beta]$. *P*roof by contradiction: Assume there exists $\alpha < y^1 < 0 < y^2 < \beta$ such that $p(y^1) < p(y^2)$. Since α is the argument of the maximum, there exists $y^0 \in [\alpha, y^1]$ such that $p(y^0) = p(y^2)$. Similarly, since β is the argument of the minimum, there exists $y^3 \in [y^2, \beta]$ such that $p(y^1) = p(y^3)$. From 1.6, we have

$$c(y^2) - c(y^0) + \mu(v - p(y^0))[\delta_y(y^2) - \delta_y(y^0)] = \frac{1}{2}\sigma^2(\delta_{yy}(y^2) - \delta_{yy}(y^0)) \tag{A.28}$$

and

$$c(y^3) - c(y^1) + \mu(v - p(y^1))[\delta_y(y^3) - \delta_y(y^1)] = \frac{1}{2}\sigma^2(\delta_{yy}(y^3) - \delta_{yy}(y^1)). \tag{A.29}$$

Thus,

$$\begin{aligned} &c(y^3) - c(y^2) - c(y^1) + c(y^0) + \mu(v - p(y^1))[\delta_y(y^3) - \delta_y(y^1)] - \mu(v - p(y^0))[\delta_y(y^2) - \delta_y(y^0)] \\ &-\frac{1}{2}\sigma^2(\delta_{yy}(y^3) - \delta_{yy}(y^2) - \delta_{yy}(y^1) + \delta_{yy}(y^0)) = 0. \end{aligned} \tag{A.30}$$

The fact that $p(y^1) < p(y^2)$ implies that there exists $y^{\#} \in [y^1, y^2]$ such that $\frac{dp(y)}{dy}|_{y=y^{\#}} > 0$. This, in turn, implies that there exists a local minimum at $\underline{y} \in [\alpha, y^{\#}]$ and a local maximum at $\bar{y} \in [y^{\#}, \beta]$. All that we showed for y^1 and y^2 also holds for $\underline{y}$ and $\bar{y}$. In addition, $\frac{dp(y)}{dy}|_{y \in \{\underline{y}, \bar{y}\}} = 0$ means that $\delta_{yy}(y)|_{y \in \{\underline{y}, \bar{y}\}} = 0$. Hence,

$$\begin{aligned} &c(y^3) - c(\bar{y}) - c(\underline{y}) + c(y^0) + \mu(v - p(\underline{y}))[\delta_y(y^3) - \delta_y(\underline{y})] - \mu(v - p(\bar{y}))[\delta_y(\bar{y}) - \delta_y(y^0)] \\ &-\frac{1}{2}\sigma^2(\delta_{yy}(y^3) + \delta_{yy}(y^0)) = 0. \end{aligned} \tag{A.31}$$

We know $c(y^3) - c(\bar{y}) - c(\underline{y}) + c(y^0) > 0$ and $-\frac{1}{2}\sigma^2(\delta_{yy}(y^3) + \delta_{yy}(y^0)) > 0$, thus

$$\mu(v - p(\underline{y}))[\delta_y(y^3) - \delta_y(\underline{y})] < \mu(v - p(\bar{y}))[\delta_y(\bar{y}) - \delta_y(y^0)]. \tag{A.32}$$

In addition, $\mu(v - p(\underline{y})) < 0$, $\mu(v - p(\bar{y})) > 0$, and $\delta_y(y^3) - \delta_y(\underline{y}) < 0$, hence

$$\delta_y(\bar{y}) - \delta_y(y^0) > 0. \tag{A.33}$$

Contradiction!

Parts 2 and 3 of the proof are similar to part 1 and hence omitted. □

Proof of Proposition 2.3.1. Suppose limit order book $q_t = l_t + h_t$ satisfies Condition 2.3.1 and $h_t^k = \alpha q_t^k$ for some $\alpha \in [0, 1]$ and all $k \in \mathbb{Z}$, and that the fast trader has received a private signal (ξ_t^b, ξ_t^s) about the upcoming market orders arriving at time $t + 1$. Her strategy $(y_{t+1}^b, y_{t+1}^s, \theta_{t+1})$ consists of submitting market orders y_{t+1}^b and y_{t+1}^s and limit orders

θ_{t+1}. Consider the difference between the fast trader's profit with and without a mid-period adjustment. In a general setting, mid-period adjustments could have two benefits: a short-term benefit resulting from front-running the upcoming market orders in this period and a longer-term benefit resulting from limit order adjustments before all other liquidity providers observe the trade executions and adjust their limit orders, and hence gaining time prority in different price levels of the limit order book over time. However, since we are assuming random and unobservable time priorities for all limit orders, all she has to maximize is the short-term benefit of the mid-period adjustments. Having received a private signal (ξ_t^b, ξ_t^s) about the upcoming market orders, the fast trader solves:

$$\begin{aligned}
\max_{(y^b_{t+1}, y^s_{t+1}, \theta_{t+1})} \mathbb{E}\Bigg[& \sum_{k=1}^{i} \big[f l_t^k + c h_t^k - (v_t^* + \lambda x_{t+1} - p_t^k) Q_t^k \big] \\
& + (y^b_{t+1} + Q_t^i)\big(- f \frac{l_t^{i+1}}{q_t^{i+1}} - c \frac{h_t^{i+1}}{q_t^{i+1}} + v_t^* + \lambda x_{t+1} - p_t^{i+1} \big) \\
& + \sum_{k=j}^{0} \big[- f l_t^k - c h_t^k - (v_t^* + \lambda x_{t+1} - p_t^k) Q_t^k \big] \\
& + (y^s_{t+1} - Q_t^j)(- f \frac{l_t^{j-1}}{q_t^{j-1}} - c \frac{h_t^{j-1}}{q_t^{j-1}} - v_t^* - \lambda x_{t+1} + p_t^{j-1}) \\
& + \pi(y^b_{t+1}, y^s_{t+1}, \theta_{t+1}) - \pi(\mathbf{0}) \Bigg]
\end{aligned} \tag{A.34}$$

where $i = \operatorname{argmax}_{k>0} y^b_{t+1} \geq Q_t^k$, $j = \operatorname{argmin}_{k\leq 0} y^s_{t+1} \geq Q_t^k$, and $\pi(.)$ is the fast trader's payoff from providing liquidity.

First, If $(1-\alpha)r \geq \alpha c + (1-\alpha)f$, then since removing one unit of liquidity and replacing it by a limit order at the same price level in the grid is profitable, the fast trader always front-runs all upcoming market orders by removing all the liquidity available in the book. She, then, sets bids at the lowest price possible and offers at the highest price possible. In other words, without exogenous limits on price movements, there will be no equilibrium.

If $(1-\alpha)r < \alpha c + (1-\alpha)f$, then removing one unit of liquidity and replacing it by a limit order at the same price level in the grid has negative impact on the fast trader's profit, if the price impact of the upcoming order flow is not in favor of removing the liquidity. However, removing all the liquidity at a price level and submitting a bid at a lower price could be profitable. This means that if the fee schedule of a market is reasonable, i.e. $(1-\alpha)r < \alpha c + (1-\alpha)f$, then two factors can incentivize the fast trader to front-run upcoming market orders: 1) large enough price impact of the upcoming order flow and 2) large enough tick size. It follows that $y^b_{t+1} \in \{0, |Q_t^1|, |Q_t^2|, \dots\}$ and $y^s_{t+1} \in \{0, |Q_t^0|, |Q_t^{-1}|, \dots\}$. All that is left to do is to provide the algorithm to find the cut-off points vector η_{t+1} and limit order adjustment vector θ_{t+1}, given the state of the limit order book and the fast trader's signal (ξ_t^b, ξ_t^s). I will only describe the algorithm to find the bids side of the vectors η_{t+1} and θ_{t+1}, because the other side is found by a similar algorithm.

If $\mathbb{E}[x^s_{t+1}|\xi^s_t] = 0$, then $y^s_{t+1} = 0$. Since the incentive to front-run an order flow is increasing in the size of the order, the algorithm to find the cut-off points is a recursive one. Let us find η^0_{t+1} first. We start with small order flows. For $0 < x^s = \mathbb{E}[x^s_{t+1}|\xi^s_t] \leq q^0_t$, if

$$\begin{aligned}
&- q^0_t[\alpha c + (1-\alpha)f + (1-\alpha)(p^0_t - v^*_t - \lambda(x^b - x^s))] \\
&+ \left\{x^s\alpha[r + v^*_t + \lambda(x^b - x^s) - p^0_t + \delta]\right\}\mathbb{1}_{\{x^s < q^{-1}_t\}} \\
&+ \left\{(q^{-1}_t\alpha + x^s - q^{-1}_t)[r + v^*_t + \lambda(x^b - x^s) - p^0_t + \delta]\right\}\mathbb{1}_{\{x^s > q^{-1}_t\}} \\
&\geq x^s\alpha[r + (v^*_t + \lambda(x^b - x^s) - p^0_t)]
\end{aligned} \tag{A.35}$$

for some $x^s \in (0, q^0_t]$, then the value of x^s for which the equality holds is η^0_{t+1}. If no $x^s \in (0, q^0_t]$ satisfies A.35, we need to check the following condition for $x^s \in (q^0_t, Q^{-1}_t]$:

$$\begin{aligned}
&- q^0_t[\alpha c + (1-\alpha)f + (1-\alpha)(p^0_t - v^*_t - \lambda(x^b - x^s))] \\
&+ \left\{x^s\alpha[r + v^*_t + \lambda(x^b - x^s) - p^0_t + \delta]\right\}\mathbb{1}_{\{x^s < q^{-1}_t\}} \\
&+ \left\{(q^{-1}_t\alpha + x^s - q^{-1}_t)[r + v^*_t + \lambda(x^b - x^s) - p^0_t + \delta]\right\}\mathbb{1}_{\{x^s > q^{-1}_t\}} \\
&\geq q^0_t\alpha[r + (v^*_t + \lambda(x^b - x^s) - p^0_t)] \\
&+ (x^s - q^0_t)\alpha[r + (v^*_t + \lambda(x^b - x^s) - p^0_t) + \delta].
\end{aligned} \tag{A.36}$$

If no $x^s \in (q^0_t, Q^{-1}_t]$ satisfies A.36, we need to repeat this until we find a $x^s \in (Q^k_t, Q^{k-1}_t]$, where $k \leq -1$, that satisfies:

$$\begin{aligned}
&- q^0_t[\alpha c + (1-\alpha)f + (1-\alpha)(p^0_t - v^*_t - \lambda(x^b - x^s))] \\
&+ \left\{(q^{-1}_t\alpha + x^s - q^{-1}_t)[r + v^*_t + \lambda(x^b - x^s) - p^0_t + \delta]\right\} \\
&\geq q^0_t\alpha[r + (v^*_t + \lambda(x^b - x^s) - p^0_t)] + \sum_{i=k}^{-1} q^i_t\alpha[r + (v^*_t + \lambda(x^b - x^s) - p^0_t) - i\delta] \\
&+ (x^s - Q^k_t)\alpha[r + (v^*_t + \lambda(x^b - x^s) - p^0_t) - (k-1)\delta].
\end{aligned} \tag{A.37}$$

Now, assume η^k_{t+1} is known, η^{k-1}_{t+1} can be found as follows. If we find the first $x^s > \eta^k_{t+1}$ that satisfies the following condition

$$\begin{aligned}
&\sum_{i=k-1}^{0} -q^i_t[\alpha c + (1-\alpha)f + (1-\alpha)(p^0_t - v^*_t - \lambda(x^b - x^s)) + i\delta] \\
&+ \left\{x^s\alpha[r + v^*_t + \lambda(x^b - x^s) - p^0_t - (k-2)\delta]\right\}\mathbb{1}_{\{x^s < q^{k-2}_t\}} \\
&+ \left\{(q^{k-2}_t\alpha + x^s - q^{k-2}_t)[r + v^*_t + \lambda(x^b - x^s) - p^0_t - (k-2)\delta]\right\}\mathbb{1}_{\{x^s > q^{k-2}_t\}} \\
&\geq \sum_{i=k}^{0} -q^i_t[\alpha c + (1-\alpha)f + (1-\alpha)(p^0_t - v^*_t - \lambda(x^b - x^s)) + i\delta] \\
&+ \left\{x^s\alpha[r + v^*_t + \lambda(x^b - x^s) - p^0_t - (k-1)\delta]\right\}\mathbb{1}_{\{x^s < q^{k-1}_t\}} \\
&+ \left\{(q^{k-1}_t\alpha + x^s - q^{k-1}_t)[r + v^*_t + \lambda(x^b - x^s) - p^0_t - (k-1)\delta]\right\}\mathbb{1}_{\{x^s > q^{k-1}_t\}},
\end{aligned} \tag{A.38}$$

then we have found η_{t+1}^{k-1}.

Knowing η_{t+1}, for $x^s \in [\eta_{t+1}^k, \eta_{t+1}^{k-1})$, where $k \leq 0$, we have

$$\theta_{t+1}^i = \begin{cases} 0 & \text{if } i \leq 0, i \neq k-1 \\ \max(0, x^s - q_t^{k-1}) & \text{if } i = k-1. \end{cases} \tag{A.39}$$

□

Proof of Lemma 3.3.1. Assume that equations 3.4, 3.6, and 3.8 are the laws of motion and that λ_t, β_t, ϕ_t, $\{\beta_t^i : i \in \{1, \ldots, N\}\}$, and $\{\phi_t^i : i \in \{1, \ldots, N\}\}$ are continuous functions on $(0, 1)$. Assume further that the pricing rule satisfies 3.3 and X^I solves the optimization problem 3.5. Take an arbitrary period of time (t_1, t_2), where $0 \leq t_1 < t_2 < 1$.

*P*art 1. In equilibrium, λ_t is increasing (nondecreasing) in (t_1, t_2). Considering that with positive probability ($\mathrm{e}^{-t_2 \sum_{i=1}^N \theta_i}$) the monopolistic insider is the only informed trader in the market at and before time t_2, this part of the proof follows from Kyle 1985, p. 1328-1329.

*P*art 2. In equilibrium, λ_t is decreasing (nonincreasing) in (t_1, t_2). The proof of this part is similar to Kyle 1985, p. 1329 and hence omitted. □

Proof of Theorem 3.3.1. I show that the equilibrium is given by 3.10, 3.11, and 3.12 in three steps.

*P*art 1. Given 3.11 and 3.12, 3.10 satisfies 3.3. Assume that the insider's strategy is 3.11, and all partially informed agents trade according to 3.12 upon arrival. The change in the total order flow from the point of view of the market makers is of the form:

$$dX_t = \left[\beta_t(V - P_t) + \sum_{i=1}^{N} \left(\beta_t^i V_i \theta_i t + \phi_t^i P_t \theta_i t \right) \right] dt + \sigma dW_t, \tag{A.40}$$

where W_t is a Wiener process and β_t, β_t^i, and ϕ_t^i are given continuous functions on $(0, 1)$. Applying Itô's lemma with $f(X) = X^2$ and taking expectation gives

$$\begin{aligned} m(t) = \mathbb{E}\left[X_t^2\right] = \sigma^2 t &+ \int_0^t 2\beta_s \left(\mathbb{E}[X_s V] - \mathbb{E}[X_s P_s]\right) ds \\ &+ \sum_{i=1}^{N} \left\{ \int_0^t 2\theta_i \Big(\beta_s^i s \mathbb{E}[X_s V_i] + \phi_s^i s \mathbb{E}[X_s P_s] \Big) ds \right\}. \end{aligned} \tag{A.41}$$

From Lemma 3.3.1, we know that λ_t is constant over time. The solution to this ODE for an arbitrary $\lambda > 0$ is

$$m(t) = \mathbb{E}\left[X_t^2\right] = \sigma^2 t. \tag{A.42}$$

Equation 3.3 implies that the conditional variance of V given public information is the variance of pricing error

$$\Sigma_t = Var(V | \mathcal{F}_t^X) = \mathbb{E}\left[(V - P_t)^2\right]. \tag{A.43}$$

Plugging A.42 into A.43, we obtain

$$\Sigma_t = \sum_{i=1}^{N} \sigma_i^2 - \lambda^2 \sigma^2 t. \tag{A.44}$$

Σ_t is widely used as a measure of price informativeness. The smaller it gets, the more informative is the price. Given that insider's trading strategy is optimal, no money should be left on the table at the end of the game, and we have

$$\lambda = \frac{\sqrt{\sum_{i=1}^{N} \sigma_i^2}}{\sigma}. \tag{A.45}$$

*P*art 2. Given 3.10 and 3.12, 3.11 solves 3.5. Assume 3.10 is the pricing rule, and all partially informed agents trade according to 3.12 upon arrival. The change in the total order flow is of the form:

$$dX_t = \left[\beta_t (V - P_t) + \sum_{i=1}^{N} \left(\beta_t^i V_i \theta_i t + \phi_t^i P_t \theta_i t \right) \right] dt + \sigma dW_t, \tag{A.46}$$

where W_t is a Wiener process and β_t^i is a given continuous function on $(0, 1)$. From Lemma 1 in Back, Cao, and Willard (2000), we know that Σ_t, defined by

$$\frac{1}{\Sigma_t} = \int_0^t \left[\beta_u^2 + \sum_{i=1}^{N} (\beta_u^i \theta_i \eta_i u)^2 \right] du + \frac{1}{\sum_{i=1}^{N} \sigma_i^2}, \tag{A.47}$$

is the conditional variance of V given $\mathcal{F}_t^X$, and that the market makers' estimate of the value, and hence price, is revised according to $dP_t = (\beta_t + \sum_{i=1}^{N} \beta_t^i \theta_i \eta_i t) \Sigma_t dX_t$. Since from 3.10, we have

$$dP_t = \lambda \left\{ \left[\beta_t (V - P_t) + \sum_{i=1}^{N} \left(\beta_t^i V_i \theta_i t + \phi_t^i P_t \theta_i t \right) \right] dt + \sigma dW_t \right\}, \tag{A.48}$$

given β_t^i, we can find β_t from

$$\beta_t + \sum_{i=1}^{N} \beta_t^i \theta_i \eta_i t = \lambda \int_0^t \left[\beta_u^2 + \sum_{i=1}^{N} (\beta_u^i \theta_i \eta_i u)^2 \right] du + \frac{\lambda}{\sum_{i=1}^{N} \sigma_i^2}. \tag{A.49}$$

All that is left to do in part 2 of the proof is to show $\phi_t = -\beta_t$. To do so, I turn to the insider's objective. From Lemma 6 in Back, Cao, and Willard (2000), we know $\phi_t = -\beta_t \eta_{Insider}$. Since the insider knows V, $\eta_{Insider} = 1$, and thus $\phi_t = -\beta_t$.

*P*art 3. Given 3.10, 3.11, and 3.12, no partially informed agent has an incentive to deviate. Consider an agent who arrives at time τ^* knowing v_j. Assume that price is revised according to 3.10, the insider's strategy is 3.11, and all the other partially informed agents trade according to 3.12. Assume further that he follows an arbitrary strategy $dX_t^{j,\tau^*} = \left(b_t^j v_j + f_t^j P_t\right) dt$, where $\tau^* \le t < 1$. The change in the total order flow for $t \in [\tau^*, 1)$, from his point of view, is of the form

$$dX_t = \left[\beta_t(V - P_t) + \sum_{i=1}^{N} \left(\beta_t^i V_i \theta_i t + \phi_t^i P_t \theta_i t\right)\right] dt + \sigma dW_t + \left(b_t^j v_j + f_t^j P_t\right) dt. \tag{A.50}$$

He tries to infer other agents' additional information from the difference between the actual total order flow X and the expected total order flow. Note the similarity between his filtering problem and market makers' filtering problem. It can be shown via the Kalman filter that this agent maximizes

$$\sup_{\alpha(t)} \mathbb{E}\left[\int_{\tau^*}^{1} (V - P_{t-}) dX_t^{j,\tau^*} - [P, X^{j,\tau^*}]_1 \mid \mathcal{F}_t^{X, X^{j,\tau^*}} \vee \sigma(V_j)\right], \tag{A.51}$$

where

$$dX_t^{j,\tau^*} = \left(b_t^j v_j + f_t^j P_t\right) dt, \tag{A.52}$$

$$b_t^j = \alpha(t)\left(1 - \lambda\sqrt{1 - \eta_j} \int_0^t \beta_s ds - \lambda\theta_j\sqrt{1 - \eta_j} \int_0^t \beta_s^j s ds\right), \tag{A.53}$$

$$f_t^j P_t = \alpha(t)\left(P_t\left(\sqrt{1 - \eta_j} - 1\right) - \lambda\theta_j\sqrt{1 - \eta_j} \int_0^t \phi_s^j P_s s ds\right), \tag{A.54}$$

and $\alpha(t)$ is a continuous function on $(0, 1)$. Also, if $\eta_j \ne 1$, then the conditional variance of V given his filtration is Σ_t^j,

$$\frac{1}{\Sigma_t^j} = \int_0^t \left[\beta_u^2 + \sum_{i=1}^{N} (\beta_u^i \theta_i \eta_i u)^2\right] du + \frac{1}{(1 - \eta_j)\sum_{i=1}^N \sigma_i^2}. \tag{A.55}$$

Clearly, Σ_t^j is decreasing over time. Following Back, Cao, and Willard (2000), define

$$\delta_t^j = 1 - \frac{\Sigma_t^j}{\Sigma_t}, \tag{A.56}$$

then from Lemma 6 in Back, Cao, and Willard (2000), we have

$$\begin{aligned} \beta_t^i &= \frac{\beta_t}{1 - \sum_{i=1}^N \theta_i \eta_i t}, \\ \phi_t^i &= -\eta_i \beta_t^i. \end{aligned} \tag{A.57}$$

□

Proof of Proposition 3.4.1. The proof of Proposition 3.4.1 is straightforward, given the fact that the presence of partially-informed traders induces excess volatility in price, and hence omitted. □

Bibliography

Abreu, Dilip and Markus K. Brunnermeier (2003). “Bubbles and Crashes”. In: *Econometrica* 71.1, pp. 173–204. ISSN: 1468-0262. DOI: 10.1111/1468-0262.00393. URL: http://dx.doi.org/10.1111/1468-0262.00393.

Allen, F., S. Morris, and A. Postlewaite (1993). “Finite Bubbles with Short Sale Constraints and Asymmetric Information”. In: *Journal of Economic Theory* 61.2, pp. 206–229. DOI: 10.1006/jeth.1993.1067.

Allen, Franklin and Gary Gorton (1993). “Churning Bubbles”. In: *The Review of Economic Studies* 60.4, pp. 813–836. ISSN: 00346527, 1467937X. URL: http://www.jstor.org/stable/2298101.

Amihud, Yakov and Haim Mendelson (1980). “Dealership market: Market-making with inventory”. In: *Journal of Financial Economics* 8.1, pp. 31–53. URL: https://EconPapers.repec.org/RePEc:eee:jfinec:v:8:y:1980:i:1:p:31-53.

Angel, James J. and Douglas McCabe (2013). “Fairness in Financial Markets: The Case of High Frequency Trading”. In: *Journal of Business Ethics* 112.4, pp. 585–595. ISSN: 01674544, 15730697. URL: http://www.jstor.org/stable/23433669.

Back, Kerry (1992). “Insider Trading in Continuous Time”. English. In: *The Review of Financial Studies* 5.3, pp. 387–409. ISSN: 08939454. URL: http://www.jstor.org/stable/2962132.

Back, Kerry, C. Henry Cao, and Gregory A. Willard (2000). “Imperfect Competition among Informed Traders”. In: *The Journal of Finance* 55.5, pp. 2117–2155. ISSN: 00221082, 15406261. URL: http://www.jstor.org/stable/222485.

Bartmann, D. and M.J. Beckmann (1992). *Inventory control: models and methods*. Lecture notes in economics and mathematical systems. Springer-Verlag. ISBN: 9783540558200. URL: https://books.google.com/books?id=cYxtAAAAIAAJ.

Bather, J. A. (1966). “A Continuous Time Inventory Model”. In: *Journal of Applied Probability* 3.2, pp. 538–549. ISSN: 00219002. URL: http://www.jstor.org/stable/3212137.

Biais, Bruno, Pierre Hillion, and Chester Spatt (1999). “Price Discovery and Learning during the Preopening Period in the Paris Bourse”. English. In: *Journal of Political Economy* 107.6, pp. 1218–1248. ISSN: 00223808. URL: http://www.jstor.org/stable/10.1086/250095.

Bodt, Marc A. De and Stephen C. Graves (1985). “Continuous-Review Policies for a Multi-Echelon Inventory Problem with Stochastic Demand”. In: *Management Science* 31.10, pp. 1286–1299. ISSN: 00251909, 15265501. URL: http://www.jstor.org/stable/2631717.

Breen, William J., Laurie Simon Hodrick, and Robert A. Korajczyk (2002). "Predicting Equity Liquidity". In: *Management Science* 48.4, pp. 470–483. ISSN: 00251909, 15265501. DOI: 10.2307/822546. URL: http://www.jstor.org/stable/822546.

Brogaard, Jonathan, Terrence Hendershott, and Ryan Riordan (2014). "High-Frequency Trading and Price Discovery". In: *The Review of Financial Studies* 27.8, pp. 2267–2306. DOI: 10.1093/rfs/hhu032. eprint: /oup/backfile/content_public/journal/rfs/27/8/10.1093_rfs_hhu032/3/hhu032.pdf. URL: +http://dx.doi.org/10.1093/rfs/hhu032.

Brogaard, Jonathan et al. (2011). "High frequency trading and its impact on market quality". In:

Brunnermeier, Markus K. and Stefan Nagel (2004). "Hedge Funds and the Technology Bubble". In: *The Journal of Finance* 59.5, pp. 2013–2040. ISSN: 00221082, 15406261. URL: http://www.jstor.org/stable/3694815.

Brunnermeier, Markus K. and Martin Oehmke (2013). "Bubbles, Financial Crises, and Systemic Risk". In: *Handbook of the Economics of Finance*. Amsterdam: Elsevier.

Budish, Eric, Peter Cramton, and John Shim (2015). "The High-Frequency Trading Arms Race: Frequent Batch Auctions as a Market Design Response *". In: *The Quarterly Journal of Economics* 130.4, pp. 1547–1621. DOI: 10.1093/qje/qjv027. eprint: /oup/backfile/content_public/journal/qje/130/4/10.1093_qje_qjv027/4/qjv027.pdf. URL: +http://dx.doi.org/10.1093/qje/qjv027.

Chen, Xin and David Simchi-Levi (2006). "Coordinating inventory control and pricing strategies: The continuous review model". In: *Operations Research Letters* 34.3, pp. 323 –332. ISSN: 0167-6377. DOI: https://doi.org/10.1016/j.orl.2005.04.012. URL: http://www.sciencedirect.com/science/article/pii/S0167637705000635.

Constantinides, George M. and Scott F. Richard (1978). "Existence of Optimal Simple Policies for Discounted-Cost Inventory and Cash Management in Continuous Time". In: *Operations Research* 26.4, pp. 620–636. DOI: 10.1287/opre.26.4.620. eprint: https://doi.org/10.1287/opre.26.4.620. URL: https://doi.org/10.1287/opre.26.4.620.

Cornell, Bradford and Erik R. Sirri (1992). "The Reaction of Investors and Stock Prices to Insider Trading". In: *The Journal of Finance* 47.3, pp. 1031–1059. ISSN: 00221082, 15406261. URL: http://www.jstor.org/stable/2328976.

Engle, Robert F., Robert Ferstenberg, and Jeffrey R. Russell (2008). "Measuring and Modeling Execution Cost and Risk". In: *Chicago GSB Research Paper Series* 08-09. DOI: 10.2139/ssrn.1211162. URL: https://ssrn.com/abstract=1211162.

Foster, F. Douglas and S. Viswanathan (1994). "Strategic Trading with Asymmetrically Informed Traders and Long-Lived Information". English. In: *The Journal of Financial and Quantitative Analysis* 29.4, pp. 499–518. ISSN: 00221090. URL: http://www.jstor.org/stable/2331107.

Foster, F. Douglas and S. Viswanathan (1996). “Strategic Trading When Agents Forecast the Forecasts of Others”. English. In: *The Journal of Finance* 51.4, pp. 1437–1478. ISSN: 00221082. URL: http://www.jstor.org/stable/2329400.

Fotak, Veljko, Vikas Raman, and Pradeep K. Yadav (2014). “Fails-to-deliver, short selling, and market quality”. In: *Journal of Financial Economics* 114.3, pp. 493 –516. ISSN: 0304-405X. DOI: https://doi.org/10.1016/j.jfineco.2014.07.012. URL: http://www.sciencedirect.com/science/article/pii/S0304405X14001597.

Glosten, Lawrence R. (1994). “Is the Electronic Open Limit Order Book Inevitable?” English. In: *The Journal of Finance* 49.4, pp. 1127–1161. ISSN: 00221082. URL: http://www.jstor.org/stable/2329182.

Goettler, Ronald L., Christine A. Parlour, and Uday Rajan (2005). “Equilibrium in a Dynamic Limit Order Market”. English. In: *The Journal of Finance* 60.5, pp. 2149–2192. ISSN: 00221082. URL: http://www.jstor.org/stable/3694745.

— (2009). “Informed traders and limit order markets”. In: *Journal of Financial Economics* 93.1, pp. 67 –87. ISSN: 0304-405X. DOI: http://dx.doi.org/10.1016/j.jfineco.2008.08.002. URL: http://www.sciencedirect.com/science/article/pii/S0304405X09000567.

Golnaraghi, Farid and Benjamin C. Kuo (2009). *Automatic Control Systems*. 9th. Wiley. ISBN: 978-0-470-04896-2.

Harrison, J. Michael and David M. Kreps (1978). “Speculative Investor Behavior in a Stock Market with Heterogeneous Expectations”. In: *The Quarterly Journal of Economics* 92.2, pp. 323–336. ISSN: 00335533, 15314650. URL: http://www.jstor.org/stable/1884166.

Harrison, J. Michael, Thomas M. Sellke, and Allison J. Taylor (1983). “Impulse Control of Brownian Motion”. In: *Mathematics of Operations Research* 8.3, pp. 454–466. DOI: 10.1287/moor.8.3.454. eprint: https://doi.org/10.1287/moor.8.3.454. URL: https://doi.org/10.1287/moor.8.3.454.

He, Hua and Jiang Wang (1995). “Differential Information and Dynamic Behavior of Stock Trading Volume”. In: *The Review of Financial Studies* 8.4, pp. 919–972. ISSN: 08939454, 14657368. URL: http://www.jstor.org/stable/2962295.

Hicks, J. R. (1939). “The Foundations of Welfare Economics”. English. In: *The Economic Journal* 49.196, pp. 696–712. ISSN: 00130133. URL: http://www.jstor.org/stable/2225023.

Holden, Craig W. and Avanidhar Subrahmanyam (1992). “Long-Lived Private Information and Imperfect Competition”. In: *The Journal of Finance* 47.1, pp. 247–270. ISSN: 1540-6261. DOI: 10.1111/j.1540-6261.1992.tb03985.x. URL: http://dx.doi.org/10.1111/j.1540-6261.1992.tb03985.x.

Hong, Harrison, Jos Scheinkman, and Wei Xiong (2006). “Asset Float and Speculative Bubbles”. In: *The Journal of Finance* 61.3, pp. 1073–1117. ISSN: 00221082, 15406261. URL: http://www.jstor.org/stable/3699318.

Hong, Harrison and Jeremy C. Stein (1999). “A Unified Theory of Underreaction, Momentum Trading, and Overreaction in Asset Markets”. English. In: *The Journal of Finance* 54.6, pp. 2143–2184. ISSN: 00221082. URL: http://www.jstor.org/stable/797990.

Hong, Harrison and Jeremy C. Stein (2003). "Differences of Opinion, Short-Sales Constraints, and Market Crashes". In: *Review of Financial Studies* 16.2, pp. 487–525. DOI: `10.1093/rfs/hhg006`. eprint: `http://rfs.oxfordjournals.org/content/16/2/487.full.pdf+html`. URL: `http://rfs.oxfordjournals.org/content/16/2/487.abstract`.

Jarrow, Robert A. and Philip Protter (2012). "A DYSFUNCTIONAL ROLE OF HIGH FREQUENCY TRADING IN ELECTRONIC MARKETS". In: *International Journal of Theoretical and Applied Finance* 15.03, p. 1250022. DOI: `10.1142/S0219024912500227`. eprint: `http://www.worldscientific.com/doi/pdf/10.1142/S0219024912500227`. URL: `http://www.worldscientific.com/doi/abs/10.1142/S0219024912500227`.

Kaldor, Nicholas (1939). "Welfare Propositions of Economics and Interpersonal Comparisons of Utility". English. In: *The Economic Journal* 49.195, pp. 549–552. ISSN: 00130133. URL: `http://www.jstor.org/stable/2224835`.

Karlin, Samuel and Howard M. Taylor (1981). *A Second Course in Stochastic Processes*. Academic Press, pp. 193 –197. ISBN: 0123986508.

Kirilenko, Andrei A. et al. (2017). "The Flash Crash: High-Frequency Trading in an Electronic Market". In: *The Journal of Finance* 72.3, pp. 967–998. ISSN: 1540-6261. DOI: `10.1111/jofi.12498`. URL: `http://dx.doi.org/10.1111/jofi.12498`.

Kyle, Albert S. (1985). "Continuous Auctions and Insider Trading". English. In: *Econometrica* 53.6, pp. 1315–1335. ISSN: 00129682. URL: `http://www.jstor.org/stable/1913210`.

LeRoy, Stephen F. and Richard D. Porter (1981). "The Present-Value Relation: Tests Based on Implied Variance Bounds". In: *Econometrica* 49.3, pp. 555–574. ISSN: 00129682, 14680262. URL: `http://www.jstor.org/stable/1911512`.

Lillo, Fabrizio, J. Doyne Farmer, and Rosario N. Mantegna (2003). "Econophysics: Master curve for price-impact function". In: *Nature* 421. DOI: `10.1038/421129a`. URL: `http://dx.doi.org/10.1038/421129a`.

Liu, Hong and Yajun Wang (2016). "Market making with asymmetric information and inventory risk". In: *Journal of Economic Theory* 163.Supplement C, pp. 73 –109. ISSN: 0022-0531. DOI: `https://doi.org/10.1016/j.jet.2016.01.005`. URL: `http://www.sciencedirect.com/science/article/pii/S0022053116000065`.

Madhavan, Ananth (2000). "Market microstructure: A survey". In: *Journal of financial markets* 3.3, pp. 205–258.

Morris, Stephen (1996). "Speculative Investor Behavior and Learning". In: *The Quarterly Journal of Economics* 111.4, pp. 1111–1133. DOI: `10.2307/2946709`. eprint: `http://qje.oxfordjournals.org/content/111/4/1111.full.pdf+html`. URL: `http://qje.oxfordjournals.org/content/111/4/1111.abstract`.

Ofek, Eli and Matthew Richardson (2003). "DotCom Mania: The Rise and Fall of Internet Stock Prices". In: *The Journal of Finance* 58.3, pp. 1113–1137. ISSN: 1540-6261. DOI: `10.1111/1540-6261.00560`. URL: `http://dx.doi.org/10.1111/1540-6261.00560`.

Ogata, K. (2010). *Modern Control Engineering*. Instrumentation and controls series. Prentice Hall. ISBN: 9780136156734. URL: `https://books.google.com/books?id=Wu5GpNAelzkC`.

O'Hara, Maureen (1995). *Market Microstructure Theory*. English. Cambridge, Mass. : Blackwell Business. ISBN: 1-55786-443-8.

Ostrovsky, Michael (2012). "Information Aggregation in Dynamic Markets With Strategic Traders". English. In: *Econometrica* 80.6, pp. 2595–2647. ISSN: 00129682. URL: http://www.jstor.org/stable/23357236.

Parlour, Christine A. and Duane J. Seppi (2008). "Limit Order Markets: A Survey". In: *Handbook of Financial Intermediation and Banking*. Ed. by Anjan V. Thakor and Arnoud W.A. Boot. Handbooks in Finance. San Diego: Elsevier, pp. 63 –96. DOI: http://dx.doi.org/10.1016/B978-044451558-2.50007-6. URL: http://www.sciencedirect.com/science/article/pii/B9780444515582500076.

Ranaldo, Angelo (2004). "Order aggressiveness in limit order book markets". In: *Journal of Financial Markets* 7, pp. 53–74. ISSN: 13864181. DOI: 10.1016/S1386-4181(02)00069-1.

Scheinkman, Jos A. and Wei Xiong (2003). "Overconfidence and Speculative Bubbles". In: *Journal of Political Economy* 111.6, pp. 1183–1220. ISSN: 00223808, 1537534X. URL: http://www.jstor.org/stable/10.1086/378531.

Scitovszky, T. De (1941). "A Note on Welfare Propositions in Economics". English. In: *The Review of Economic Studies* 9.1, pp. 77–88. ISSN: 00346527. URL: http://www.jstor.org/stable/2967640.

Seppi, Duane J. (1997). "Liquidity Provision with Limit Orders and Strategic Specialist". English. In: *The Review of Financial Studies* 10.1, pp. 103–150. ISSN: 08939454. URL: http://www.jstor.org/stable/2962258.

Shiller, Robert J. (1981). "Do Stock Prices Move Too Much to be Justified by Subsequent Changes in Dividends?" In: *The American Economic Review* 71.3, pp. 421–436. ISSN: 00028282. URL: http://www.jstor.org/stable/1802789.

— (2015). *Irrational Exuberance*. Third. Princeton University Press. ISBN: 9780691166261. URL: http://www.jstor.org/stable/j.ctt1287kz5.

Shleifer, Andrei and Robert W. Vishny (1997). "The Limits of Arbitrage". In: *The Journal of Finance* 52.1, pp. 35–55. ISSN: 00221082, 15406261. URL: http://www.jstor.org/stable/2329555.

Sulem, Agns (1986). "A Solvable One-Dimensional Model of a Diffusion Inventory System". In: *Mathematics of Operations Research* 11.1, pp. 125–133. ISSN: 0364765X, 15265471. URL: http://www.jstor.org/stable/3690058.

Xiong, Wei and Jialin Yu (2011). "The Chinese Warrants Bubble". In: *The American Economic Review* 101.6, pp. 2723–2753. ISSN: 00028282. URL: http://www.jstor.org/stable/23045656.