Title
Essays in Information and Financial Markets

Author
Aoyagi, Jun

Publication Date
2021

Peer reviewed|Thesis/dissertation

Essays in Information and Financial Markets

by

Jun Aoyagi

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Economics

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Nicolae Garleanu, Co-chair
Associate Professor David Sraer, Co-chair
Professor Christine Parlour
Professor Chris Shannon

Spring 2021

Essays in Information and Financial Markets

Abstract

Essays in Information and Financial Markets

by

Jun Aoyagi

Doctor of Philosophy in Economics

University of California, Berkeley

Professor Nicolae Garleanu, Co-chair

Associate Professor David Sraer, Co-chair

The landscape of the modern financial market is rapidly changing due to innovations in trading methods and information technologies. Do innovations in financial markets improve market quality? How do traders change their trading behavior and information acquisition? How should exchange platforms and policy makers react to financial innovations? To answer these questions, I first analyze strategic information acquisition by traders and the impact of imposing market structures with different informational environments. Chapter 1 considers the quality aspect of information (i.e., the precision of private information), and Chapter 2 considers the speed aspect of information (i.e., how quickly a trader can process information and act on it). Finally, Chapter 3 analyzes one of the latest innovations in financial markets: the blockchain technology and decentralization.

In Chapter 1, I extend the canonical model of Kyle (1985) and accommodate strategic information acquisition by an informed trader and a fair disclosure regulation. A fair disclosure policy tries to mitigate information asymmetry between informed and uninformed traders by disseminating material information to all market participants. The literature has suggested that such a policy should discourage information acquisition by a potential informed trader, as it diminishes the value of privately possessing information. However, my model shows that a trader may exhibit the opposite reaction. In particular, if the disclosure policy provides a more precise public signal about asset fundamentals, it can promote information acquisition by a potential informed trader. This effect is referred to as the crowding-in effect. The crowding-in effect competes against the intended effect of fair disclosure, leading to an ambiguous reaction of private information production by an informed trader.

Chapter 2 deals with financial innovations in the speed of trading and information processing. It analyzes high-frequency traders (HFTs) and intentional delays imposed by exchange platforms. HFTs are ultra-fast traders who exploit sophisticated information and communication technologies in order to acquire information and take short-term arbitrage opportunities. The speed advantage of HFTs imposes an adverse selection cost on other traders, making a market less liquid. A growing number of exchanges have adopted intentional delays to exogenously slow down HFTs and to protect liquidity providers

against latency arbitrage. However, analogous to Chapter 1, my model shows that intentional delays have the crowding-in effect on speed acquisition by HFTs. Even though an exchange tries to slow down HFTs by exogenously imposing a delay on their transactions, HFTs may try to process information more quickly and move faster. As the crowding-in effect competes against the intended effect of intentional delays, the reaction of equilibrium market quality to the imposition of delays becomes ambiguous.

Chapter 3 considers the recent innovations in blockchain and decentralized exchanges. A growing number of exchanges are built on a decentralized information management system of the Ethereum blockchain, and they have implemented a novel market-making algorithm called Constant Product Market Makers (CPMM) to execute transactions. I consider the coexistence of a centralized exchange with the traditional order-book mechanism and a decentralized exchange with the CPMM. Informed and uninformed traders are endogenously differentiated between the traditional and the new market platforms and re-configure the informativeness of order flow on each exchange. The model demonstrates that liquidity on a decentralized exchange with the CPMM is positively associated with that on a centralized exchange.

To my family.

Contents

List of Figures

List of Tables

Acknowledgments

In preparing this dissertation, I have received tremendous professional and personal support from my advisers, colleagues, family, and friends. It is my real pleasure and honor acknowledging their support and companionship.

I would like to express my deepest appreciation to my primary advisor Nicolae Garleanu for his continual support and invaluable guidance. His first-year course *Discrete Time Asset Pricing* led me to the field of Finance. He helped me in all stages of research with encouragement, patience, insights, and knowledge. Whenever I meet and talk to Nicolae, he gave me honest and invaluable advice and always let me understand what is good and what needs to be improved in my research, my writing, and my presentation.

I am also deeply indebted to Christine Parlour. My interests in information economics and market microstructure are definitely rooted in her second-year course on microstructure. The idea of my job market paper arises from her offhand remark in one of our meetings. She also gave me the chance to learn Economics of FinTech through our discussion, and it has become one of my specialized fields. Her support and advice definitely made my life as a graduate student a productive one.

I would also like to express my sincere appreciation to professors at Berkeley for constructive discussions, insightful comments, and priceless advice. My special gratitude goes out to Terry Hendershott and David Sraer for important advice on high-frequency trading in my job market paper. I am also grateful to Chris Shannon for her insightful comments on my work on ambiguity and financial markets. Moreover, I was lucky to have opportunities to listen to people from other institutions. Special thanks to Ohad Kadan, Xuelin Li, Katya Malinova, Sophie Moinas, Katrin Tinn, and Mao Ye for being a discussant of my papers at finance conferences. Their discussions have sharpened my thoughts and arguments.

It all started from the six years of my undergraduate and graduate studies at the University of Tokyo. I am very pleased with acknowledging Kosuke Aoki, Shin-ichi Fukuda, Tomohiro Hirano, and Kiyohiko Nishimura for their continued encouragement and supports over the past ten years.

I would also like to cordially acknowledge financial supports from the Economics Department and the Haas School of Business at UC Berkeley, the Nakajima Foundation, and the Japan-IMF Scholarship Program. The Japan-IMF Scholarship Program also provided me with an opportunity to participate in its summer internship program. I enjoyed staying in Washington DC and working with IMF economists. Moriguchi Prize at ISER Osaka University and Berkeley Blockchain Initiative Award also encouraged me to pursue my research on HFT and blockchain via their awards and financial supports.

Friendship with my colleagues at Berkeley is a precious asset for me. Many thanks to Yuki Ito and Yui Ito for all the unconditional support. They helped me get through this very tough and intense academic year during COVID-19 pandemic. I also thank ChaeWon Baek and Yingge Yan for their companionship throughout my Ph.D. life at Berkeley. I would also like to thank the graduate advisers, Patrick Allen, and Janene Vernard for their support for all of us in the Economics Ph.D. program.

Last, but certainly not the least, I would like to extend my deepest appreciation to my family. I am forever indebted to my parents and my wife for giving me the opportunities and experiences that have made me who I am.

Chapter 1

Strategic Information Acquisition

1.1 Introduction

Asymmetric information between traders is one of the key determinants of market quality. Uninformed market makers tend to bear an adverse selection cost of a trade when it is initiated by a trader with private information. To compensate for the cost, market makers charge a wider bid-ask spread, making the market less liquid. Policy makers have tried to mitigate asymmetric information by imposing several information regulations. In particular, this chapter focuses on a fair disclosure policy, such as the Regulation Fair Disclosure (RegFD) in 2000 and Sarbanes-Oxley Act (SOX) in 2002. They try to mitigate information asymmetry between traders by promoting full and fair disclosure of material information by issuers (e.g., firms). For example, RegFD prohibits corporate issuers from revealing material nonpublic information to specific entities and requires them to publicly release such information. Another example is provided by SOX, which aims to alleviate information asymmetry by "improving the accuracy and reliability of corporate disclosures" (Public Law, 107–204). Many other regulations and policies have been implemented in the expectation that disseminating precise information to all market participants improves liquidity and price efficiency by mitigating information asymmetry.

However, the existing studies on fair disclosure suggest that it may lead to some unintended consequences.[1] Specifically, the literature on market microstructure argues that publicly revealing information diminishes private information acquisition by potential informed traders because their informational advantage dwindles when somewhat correlated material news becomes public. The negative impact of information disclosure on private information acquisition is called the *crowding-out* effect (Goldstein and Yang, 2017). Even though the crowding-out effect mitigates asymmetric information problem between traders, it may reduce the amount of information available for the market. The

[1] There are some other unintended effects of fair disclosure, including erosion of trading opportunities with risk sharing motives (Hirshleifer, 1971) and an increase in non-fundamental volatility due to the beauty-contest type activity (Morris and Shin, 2002).

existing models cast a doubt on the effectiveness of fair disclosure, as they predict that the price can be less informative and less efficient due to the crowding-out effect on private information production.

This paper studies a one-period version of Kyle (1985) model, where I introduce a public signal (i.e., fair disclosure) and information acquisition by an informed trader. Even though the deviation from the original environment is kept minimal, it shows that the crowding-out effect in the existing models can be overturned due to strategic information acquisition. Instead, fair disclosure *crowds in* the information-acquisition activity by the informed trader. Thus, the unintended impact of fair disclosure can be opposite to the existing models. In other words, fair disclosure not only provides a more precise public signal to the market but it also encourages a potential informed trader to (privately) discover new information. Therefore, the crowding-in effect helps fair disclosure with promoting the market and price efficiency.

The key mechanism is hard-wired in the standard Kyle-model. The informed trader's expected trading profit is increasing in the precision of her private signal, as it represents her informational advantage. Since she is strategic, however, she knows that trading on her private information imposes an adverse selection cost on an uninformed market maker. Then, the price impact of her order flow increases, and liquidity deteriorates. In other words, she is aware that improving the precision of her private signal harms the market depth and increases the trading cost. An increase in the trading cost, in turn, is perceived as an *endogenous* cost of acquiring a more precise private signal for the strategic trader. In contrast to the exogenous cost of information, the endogenous cost is affected by fair disclosure, because it yields a new source of information to the market maker, alleviates the adverse selection problem, and changes the price impact of informed trading.

Due to fair disclosure, the market maker's pricing behavior depends on two sources of information: the order flow and the publicly disclosed signal. When the public signal becomes more precise due to a fair disclosure policy, the price impact of the order flow diminishes, as the market maker's inference depends more on public news and less on the order flow. This provides the informed trader with room for increasing the precision of private information. Even if the informed trader increases the precision of her private signal, it does not deteriorates liquidity because the price impact (Kyle's λ) is small. Thus, fair disclosure decreases the *endogenous marginal cost* of information acquisition for the informed trader, thereby generating the crowding-in effect.

The crowding-in effect in my model provides novel implications. Firstly, it competes against the traditional crowding-out effect of fair disclosure. I show that the crowding-in effect arises due to the strategic motive of the trader, while the crowding-out effect is driven by the exogenous cost of information acquisition. Therefore, my model predicts that fair disclosure hampers private information acquisition when searching for news takes a relatively large cost, e.g., corporate and industry information is hard to obtain or hard to interpret, or a trader needs to learn synthetic securities that involve complex return structures. Therefore, depending on the cost of private information acquisition, the informativeness of the order flow (such as PIN measure by Easley, Kiefer, and O'Hara, 1997) exhibits a non-monotonic reaction to a fair disclosure policy.

Moreover, my model demonstrates that fair disclosure improves market quality measured by liquidity (market depth) and price efficiency. A more precise public signal directly improves both measures of market quality. Responding to the change in market quality (i.e., the price impact), the potential informed trader tries to acquire more or less precise private information. Even though the crowding-in effect undermines the direct effect of the disclosure policy on market liquidity, it is the indirect effect of disclosure and cannot dominate the direct one. In the existing models on fair disclosure, disseminating a public signal unambiguously improves the liquidity, as the information asymmetry shrinks, but it crowds out the amount of information that the price can impound and harms the price efficiency. My model implies that the crowding-in effect does not dominate the direct impact of disclosure on asymmetric information. Thus, it not only improves the liquidity measure but also injects more information into the price, leading to an optimistic view on the effectiveness of fair disclosure.

A long list of studies have analyzed the impact of information regulation in financial markets (see Dye, 2001; Verrecchia, 2001; Kanodia and Sapra, 2016 for comprehensive reviews). The existence of fair disclosure's crowding-out effect is pointed out by several papers, such as Verrecchia (1982), Diamond (1985), Kim and Verrecchia (1994), Gao and Liang (2013), Colombo, Femminis, and Pavan (2014), and Goldstein and Yang (2017). In most papers, however, the crowding-out effect is analyzed in a competitive environment, such as the noisy rational expectations equilibrium model by Grossman and Stiglitz (1980). In contrast, my model studies a strategic trader who knows the price impact of her information-acquisition activity. The strategic nature imposes the endogenous cost on information acquisition and leads to the crowding-in effect of fair disclosure.[2] In the strategic environment, Banerjee and Breon-Drish (2018) consider a continuous-time Kyle model but their focus is on the dynamic nature of information acquisition and traders' entry decision. They find that more volatile public information enhances information production by informed traders, i.e., they find the crowding-out effect in a dynamic environment.[3]

A couple of studies argue for the crowding-in effect of a fair disclosure policy, but they rely on ad-hoc assumptions on information or trading environment. For example, Bertomeu, Beyer, and Dye (2011) and Cheynel and Levine (2020) consider the mosaic theory, in which having a more precise private signal allows a trader to better interpret and process public information. McNichols and Trueman (1994) argue for the crowding-in effect based on the possibility of multiple rounds of trading, i.e., trade before and after information disclosure. In their environment, an informed trader can exploit her private

[2]Huddart, Hughes, and Levine (2001) and Gong and Liu (2012) consider strategic trading but disclosure policy in their models is dissemination of information about the trading behavior of a privately informed trader, i.e., an informed trader must report her trading return or strategy after a trading session.

[3]The public signal in their model does not correspond to that in my model, because the signal reveals information about the distribution of the fundamental value process rather than the noisy signal about the realized value of asset.

information to predict the information content of disclosure and the price reaction to disclosed information. The informed trader can take an appropriate position at the pre-announcement round in order to earn from a post-announcement trade. Finally, Han, Tang, and Yang (2016) derive the crowding-out effect in a model à la Grossman and Stiglitz (1980) with endogenous liquidity traders. When liquidity traders endogenously participate in the market, a more precise public signal reduces the trading cost and attracts a larger set of liquidity traders, which, in turn, motivate the informed trader to acquire more information. In contrast to the above theories, the result in my model does not rely on specific assumptions on trading environment.

The empirical literature has reported somewhat ambiguous results on the impact of disclosure. On the one hand, Bushee, Matsumoto, and Miller (2004), Chiyachantana et al. (2004), Eleswarapu, Thompson, and Venkataraman (2004), and Gintschel and Markov (2004) find that the adverse selection component in bid-ask spread shrinks after fair disclosure policies, consistent with the crowding-out effect. Also Chen and Lu (2019) directly measure the information-acquisition activity after TRACE in corporate bond markets and find that TRACE crowds out the information production. On the other hand, Krinsky and Lee (1996), Coller and Yohn (1997), Straser (2002), and Sidhu et al. (2008) report the opposite results, i.e., the asymmetric information problem worsens due to disclosure policies. Since the existing models with crowding-out effect unambiguously suggest a decline in asymmetric information, the crowding-in effect in my model helps reconcile the above empirical findings.

1.2 Model

Consider a one-shot trading model à la Kyle (1985) with two additional stages. There are three types of participants: a potential informed trader, a competitive market maker, and a noise trader. The competitiveness in the market making sector is justified by considering multiple potential market makers who stay inactive on the equilibrium path. The model differs from the original Kyle model only in two aspects: it involves (i) information acquisition by the potential informed trader at $t = 0$ and (ii) the implementation of a fair disclosure policy at $t = 1$. At $t = 2$, a trade takes place as in the original Kyle model.

A single asset is traded, and its *ex-post* liquidation value, denoted as v, follows the normal distribution with mean p_0 and variance Σ_0, i.e., $v \sim \mathcal{N}(p_0, \Sigma_0)$. The liquidation value of the asset is not observable *per se*.

Fair disclosure policy. Following the literature (e.g., Goldstein and Yang, 2017), the disclosure policy is described by a public signal about v. The signal is denoted as s_{pub}, and I assume that it is linear in v:

$$s_{pub} = v + e_{pub}$$

with $e_{pub} \sim \mathcal{N}(0, \sigma_{pub}^2)$. I denote the precision of the public signal as $\tau_{pub} \equiv \sigma_{pub}^{-2}$. The signal is disseminated to all traders before the trading game starts.

Potential informed trader. The informed trader's behavior involves two steps. Prior to the disclosure of signal s_{pub}, she obtains private information s. It is also linear in v and given by

$$s = v + e$$

with $e \sim \mathcal{N}(0, \sigma_e^2)$. I denote the precision of the private signal as $\tau_e \equiv \sigma_e^{-2}$. In the information acquisition stage, τ_e is a choice variable of the informed trader. I assume that obtaining a precise private signal is costly. In particular, learning s with τ_e takes information acquisition cost $C(\tau_e)$ with $C'(\cdot) > 0$. The positive marginal cost of τ_e can be seen as costs of information processing and market monitoring.

At the trading stage, the informed trader chooses the optimal trading strategy (quantity) $x \in \mathbb{R}$. With p denoting the price of the asset set by the market maker, the informed trader's trading profit is $(v - p)x$. For later use, I denote the *ex-ante* expected trading profit of the informed trader as V:

$$V(\tau_e; \tau_{pub}) = \mathbb{E}[(v - p)x(s, s_{pub})], \tag{1.1}$$

where $x(s, s_{pub})$ is the optimal trading strategy conditional on the realized value of signals (derived in the following sections), and $\mathbb{E}$ denotes the unconditional expectation operator over the random variables.

Noise trader and market maker. The noise trader's activity is summarized by the random order flow which is independent of (expected) asset value and the price level. Specifically, she places a market order with quantity $u \sim \mathcal{N}(0, \sigma_u^2)$ in the trading stage. u is also independent of other random variables.

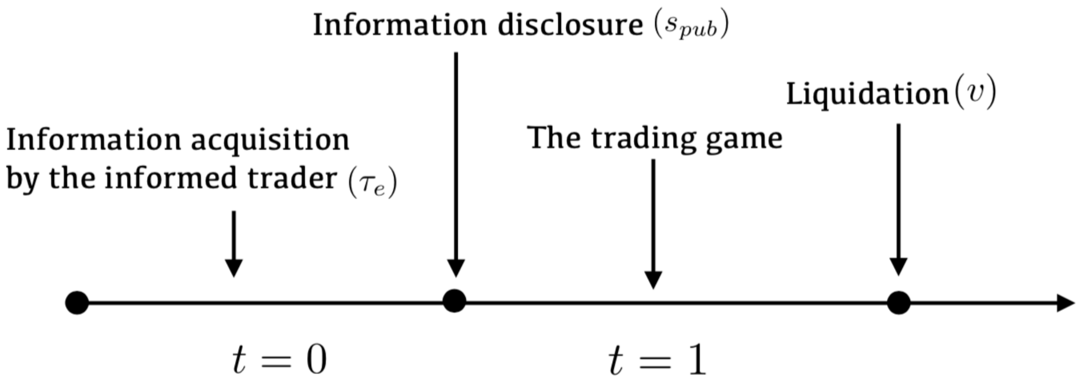
The market maker sets the competitive price given available information to clear the market. The set of available information for the market maker consists of the disclosed signal s_{pub} and the aggregate order flow $y \equiv x + u$. Since the competition leads to the efficient price, it holds that

$$p = \mathbb{E}[v | s_{pub}, y]. \tag{1.2}$$

Equilibrium. The equilibrium concept of the model is the subgame perfect equilibrium. The first stage involves the informed trader's decision on the precision (τ_e) of the private signal (s), and the second stage is the trading game. Figure 3.3 illustrates the timing of events.

Definition 1.1. *The equilibrium of the model is defined by the set of variables (τ_e, x, p) such that the following three conditions hold:*

Figure 1.1: Timeline of the game



(i) Profit maximization: For any alternate trading strategies, $\bar{x}$, and for any (s, s_{pub}), the informed trader does not have profitable deviation, i.e.,

$$\mathbb{E}[(v-p)x|s, s_{pub}] \geq \mathbb{E}[(v-p)\bar{x}|s, s_{pub}].$$

(ii) Market efficiency: The equilibrium price, p, satisfies equation (1.2).
(iii) Information acquisition: The precision of the private signal, τ_e, maximizes the ex-ante expected trading profit of the informed trader V in equation (1.1).

1.3 Solution

Following Kyle (1985), I focus on the linear equilibrium. Since the informed trader has two sources of information, s and s_{pub}, her optimal trading strategy takes the following form:

$$x = \alpha + \beta s + \gamma s_{pub} \tag{1.3}$$

with some equilibrium coefficients (α, β, γ).

The market maker observes the aggregate order flow, $y = x + u$, and the public signal s_{pub}. Note that observing these two signals is informationally equivalent to observing

$$(s_{pub}, y - \gamma s_{pub}) = (s_{pub}, \alpha + \beta s + u).$$

Thus, for notational simplicity, I denote $\hat{y} \equiv y - \gamma s_{pub}$. Given the linear trading strategy of the informed trader, the standard filtering problem implies that the market efficiency condition is given by

$$\begin{aligned} p &= \mathbb{E}[v|s_{pub}, \hat{y}] \\ &= \mu + \theta s_{pub} + \lambda \hat{y} \end{aligned} \tag{1.4}$$

with some equilibrium coefficients (μ, θ, λ).

Solution. Consider the informed trader's trading strategy given p in equation (1.4). It solves

$$\begin{aligned} x(s, s_{pub}) &= \arg\max_x \mathbb{E}[(v-p)x|s, s_{pub}] \\ &= \arg\max_x (\mathbb{E}[v|s, s_{pub}] - \mu - \theta s_{sub} - \lambda \hat{y})x. \end{aligned}$$

The FOC implies that

$$x(s, s_{pub}) = \frac{\mathbb{E}[v|s, s_{pub}] - \mu - (\theta - \gamma\lambda)s_{pub}}{2\lambda}, \tag{1.5}$$

where the standard filtering problem yields

$$\mathbb{E}[v|s, s_{pub}] = \frac{\Sigma_0^{-1}}{\Sigma_0^{-1} + \tau_e + \tau_{pub}} p_0 + \frac{\tau_e}{\Sigma_0^{-1} + \tau_e + \tau_{pub}} s + \frac{\tau_{pub}}{\Sigma_0^{-1} + \tau_e + \tau_{pub}} s_{pub}. \tag{1.6}$$

By defining $w_i \equiv \sigma_i/\sigma_{pub}$ for $i \in \{e, u\}$, equations (1.5) and (1.6) imply that the optimal trading strategy is characterized by equation (1.3) with

$$\alpha = -\frac{1}{2\lambda}\left(\mu - \frac{\tau_e^{-1}}{\tau_e^{-1} + \Sigma_0(1+w_e^2)} p_0\right), \tag{1.7}$$

$$\beta = \frac{1}{2\lambda} \frac{\Sigma_0}{\tau_e^{-1} + \Sigma_0(1+w_e^2)}, \tag{1.8}$$

$$\gamma = \frac{1}{\lambda}\left(\frac{\Sigma_0 w_e^2}{\tau_e^{-1} + \Sigma_0(1+w_e^2)} - \theta\right). \tag{1.9}$$

Next, consider the market efficiency condition given the linear trading strategy (1.3). Once again, the standard filtering problem leads to

$$\begin{aligned} p =& p_0 + \left[\frac{\Sigma_0(\beta^2 w_e^2 + w_u^2)}{\Sigma_0(w_u^2 + \sigma_u^2) + \beta^2[\Sigma_0(w_e^2+1) + \tau_e^{-1}]}(s_{pub} - p_0)\right. \\ &\left. + \frac{\beta\Sigma_0}{\Sigma_0(w_u^2 + \sigma_u^2) + \beta^2[\Sigma_0(w_e^2+1) + \tau_e^{-1}]}(\hat{y} - \alpha - \beta p_0)\right]. \end{aligned} \tag{1.10}$$

Thus, the efficient price is characterized by equation (1.4) with

$$\mu = \frac{\Sigma_0\sigma_u^2 + \beta^2\tau_e^{-1}}{\Sigma_0(w_u^2 + \sigma_u^2) + \beta^2[\Sigma_0(w_e^2+1) + \tau_e^{-1}]} p_0 - \lambda\alpha, \tag{1.11}$$

$$\lambda = \frac{\beta\Sigma_0}{\Sigma_0(w_u^2 + \sigma_u^2) + \beta^2[\Sigma_0(w_e^2+1) + \tau_e^{-1}]}, \tag{1.12}$$

$$\theta = \frac{\Sigma_0(\beta^2 w_e^2 + w_u^2)}{\Sigma_0(w_u^2 + \sigma_u^2) + \beta^2[\Sigma_0(w_e^2+1) + \tau_e^{-1}]}. \tag{1.13}$$

Now, the explicit solution for the equilibrium is given by solving equations (1.7)-(1.13):

Proposition 1.1. *There exists a unique linear equilibrium in the trading stage, in which the trading strategy of the informed trader and the asset price are given by (1.3) and (1.4) with*

$$\begin{aligned}
\alpha &= -\frac{1}{\Sigma_0 \tau_{pub} + 1} \beta p_0, \\
\beta &= \sqrt{\frac{\Sigma_0 w_u^2 + \sigma_u^2}{\Sigma_0 (1 + w_e^2) + \tau_e^{-1}}}, \\
\gamma &= -\beta \frac{\Sigma_0 w_u^2}{\Sigma_0 w_u^2 + \sigma_u^2}, \\
\mu &= \frac{\sigma_u^2}{\Sigma_0 w_u^2 + \sigma_u^2} p_0, \\
\lambda &= \frac{1}{2} \frac{\Sigma_0}{\sqrt{(\Sigma_0 w_u^2 + \sigma_u^2)(\Sigma_0 (1 + w_e^2) + \tau_e^{-1})}}, \\
\theta &= \frac{\Sigma_0}{2} \left(\frac{w_e^2}{\Sigma_0 (1 + w_e^2) + \tau_e^{-1}} + \frac{w_u^2}{\Sigma_0 w_u^2 + \sigma_u^2} \right).
\end{aligned}$$

Note that the above equations converge to the original one-period Kyle model if $\sigma_{pub} \to \infty$ and $\sigma_e \to 0$ (i.e., a model with a perfectly informed trader and no public disclosure).

1.3.1 Equilibrium in the trading stage

Liquidity, market depth, and price impact. Proposition 1.1 implies that λ is decreasing in the precision of the public signal, $\tau_{pub} = \sigma_{pub}^{-2}$, and increasing in the precision of the informed trader's private signal, $\tau_e = \sigma_e^{-2}$. That is, the price impact of order flow weakens when (i) the informed trader has a less precise private signal and (ii) the public signal is more precise. Intuition follows the standard Kyle-model, because high τ_e and low τ_{pub} both induce severe asymmetric information between the informed trader and the market maker. In such a situation, the market maker renders the market less liquid (i.e., increases λ) to compensate for the adverse selection cost that stems from trading with the informed trader (Bagehot, 1971; Glosten and Milgrom, 1985).

Behavior of the informed trader. The informed trader trades based on the private and the public signals, s and s_{pub}, where β and γ represent the weight on each signal. Firstly, the weight on the private signal β is increasing in its precision (τ_e), because the informed trader optimally acts more on the private signal. Since the weight on the public signal (γ) is the complement of β, γ is a decreasing function of τ_e, i.e., the informed trader takes the weight away from s_{pub} and reallocates it to s when τ_e increases.

Interestingly, however, the impact of the precision of disclosure (τ_{pub}) on β and γ is not trivial. Firstly, a more precise public signal makes the informed trader rely more heavily on it. Thus, $\tau_{pub} = \sigma_{pub}^{-2}$ has a negative impact on β and a positive impact on γ. By the same token, however, the market maker's behavior (i.e., the execution price) becomes more dependent on the public signal. Since trading on the same information is not profitable for the informed trader, she has an incentive to lean more toward the private signal, i.e., she reduces γ and increases β. Put differently, a more precise public signal alleviates the asymmetric information problem and reduces the price impact of order flow, λ, which generates room for the informed trader to act more on her private signal s.

Although τ_{pub} has the above-mentioned two competing effects, the following result suggests that the second channel is dominant:

Corollary 1.1. *β is increasing in τ_{pub}, and γ is decreasing in τ_{pub}.*

Proof. Taking the first-order derivative of β and γ with respect to τ_{pub} yields the result. □

Corollary 1.1 implies the following: when the fair disclosure policy provides a more precise signal on the asset value, the informed trader avoids trading on that public information. That is, the strategic trader prefers to trade based on the information that other traders (i.e., the market maker) do not know, and this behavior becomes strong when the market maker has more precise information. This result highlights the key ingredient in information acquisition: the informed trader relies more heavily on the private signal when the public signal is more precise.

1.3.2 Ex-ante expected profit

With the solution in Proposition 1.1, the ex-ante expected profit of the informed trader, gross of the cost of information, is given by

$$\begin{aligned} V &= \mathbb{E}[(v-p)x(s, s_{pub})] \\ &= \frac{1}{4\lambda}\mathbb{E}\left[(\mathbb{E}[v|s, s_{pub}] - \mu - (\theta+\gamma\lambda)s_{pub})^2\right]. \end{aligned} \tag{1.14}$$

By using the optimal conditions for x (equation [1.5]), the inside of the parentheses is rewritten as

$$\begin{aligned} \mathbb{E}[v|s, s_{pub}] - \mu - (\theta+\gamma\lambda)s_{pub} &= 2\lambda x(s, s_{sup}), \\ &= 2\lambda(\beta+\gamma)\left(v + \frac{\beta}{\beta+\gamma}e + \frac{\gamma}{\beta+\gamma}e_{pub} + \frac{\alpha}{\beta+\gamma}\right). \end{aligned}$$

From Proposition 1.1, it holds that

$$v + \frac{\beta}{\beta+\gamma}e + \frac{\gamma}{\beta+\gamma}e_{pub} + \frac{\alpha}{\beta+\gamma} \sim \mathcal{N}\left(0, \Sigma_0 + \left(\frac{\beta}{\beta+\gamma}\right)^2\sigma_e^2 + \left(\frac{\gamma}{\beta+\gamma}\right)^2\sigma_{pub}^2\right). \tag{1.15}$$

Therefore, V is explicitly calculated by taking the variance of (1.15). With the explicit solution for (α, β, γ) in Proposition 1.1, I obtain the following result:

Proposition 1.2. *With $b_{pub} \equiv (\Sigma_0 \tau_{pub} + 1)^{-1}$, the ex-ante expected profit of the informed trader, gross of the information acquisition cost, is given by*

$$V = \frac{b_{pub}}{2} \sqrt{\frac{\Sigma_0^2 \sigma_u^2}{\tau_e^{-1} + b_{pub}\Sigma_0}}. \tag{1.16}$$

Acquiring a more precise private signal (i.e., a higher τ_e) has two competing effects on V. On the one hand, it provides the informed trader with a larger informational advantage and a higher expected profit. On the other hand, it worsens the asymmetric information problem, liquidity diminishes, and the profit margin declines (i.e., the endogenous cost of τ_e). However, equation (1.16) implies that the first positive impact of a higher precision (τ_e) is dominant. This is because the illiquid market arises *as a consequence of* a better informed trader, i.e., the negative impact is the indirect effect of a higher τ_e and cannot dominate its positive direct impact.

1.3.3 Information acquisition

To make the problem well-defined, I introduce a linear exogenous cost of information acquisition, $C(\tau_e) = c\tau_e$. It can be thought of as the cost of collecting and processing a large amount of information or an increasing monitoring cost. It can be easily checked that a quadratic cost function yields the same results.

Then, the optimal precision of the private signal solves the following problem:

$$\tau_e^* \equiv \arg\max_{\tau_e \geq 0} [V(\tau_e) - C(\tau_e)] = \arg\max_{\tau_e \geq 0} \left(\frac{b_{pub}}{2} \sqrt{\frac{\Sigma_0^2 \sigma_u^2}{\tau_e^{-1} + b_{pub}\Sigma_0}} - c\tau_e \right). \tag{1.17}$$

By denoting the net profit as $\bar{V} \equiv V - C$, the FOC is given by

$$\frac{d\bar{V}(\tau_e)}{d\tau_e} = \frac{b_{pub}}{4} \sqrt{\Sigma_0^2 \sigma_u^2} (\tau_e^{-1} + b_{pub}\Sigma_0)^{-\frac{3}{2}} \tau_e^{-2} - c = 0. \tag{1.18}$$

The SOC is easy to check: $\frac{d\bar{V}(\tau_e)}{d\tau_e}$ is decreasing in τ_e. Since $\lim_{\tau_e \to 0} \frac{d\bar{V}(\tau_e)}{d\tau_e} = \infty$ and $\lim_{\tau_e \to \infty} \frac{d\bar{V}(\tau_e)}{d\tau_e} = -c < 0$, the FOC has a unique positive solution.

1.3.4 Crowding-in effect versus crowding-out effect

Consider the partial derivative of the FOC with respect to τ_{pub} to understand the effect of fair disclosure on τ_e^*. Firstly, note that the FOC contains τ_{pub} only via $b_{pub} = (\Sigma_0 \tau_{pub} + 1)^{-1}$. Since b_{pub} is a monotone transformation of τ_{pub}, I compute the partial derivative of $\frac{d\bar{V}}{d\tau_e}$ with respect of b_{pub} at the optimal τ_e:

$$\frac{\partial}{\partial b_{pub}} \left(\frac{d\bar{V}}{d\tau_e} \right) |_{\tau_e = \tau_e^*} = c \frac{\tau_e^{-1} - \frac{1}{2} b_{pub}\Sigma_0}{b_{pub}(\tau_e^{-1} + b_{pub}\Sigma_0)}. \tag{1.19}$$

Equation (1.19) implies that the impact of b_{pub} on τ_e^* can be both positive and negative.

Crowding-out effect. On the one hand, when the public signal becomes precise (σ_{pub}^2 and b_{pub} decline), the expected profit of the informed trader shrinks because the precise disclosure harms the informational advantage of the informed trader. The first coefficient b_{pub} in equation (1.16) captures this effect. Therefore, the informed trader knows that the return from obtaining a more precise private signal is now smaller due the precise public signal. Since increasing τ_e takes the exogenous cost $C(\tau_e)$, she is less willing to acquire a higher τ_e. This channel represents the traditional *crowding-out* effect of fair disclosure on information acquisition, which is analyzed in the literature (Verrecchia, 1982; Diamond, 1985; Kim and Verrecchia, 1994; and others). Even if the informed trader obtains private information, the public signal diminishes the informational advantage, and paying the cost for the private signal becomes less worthwhile.

Crowding-in effect. On the other hand, equation (1.19) shows that there also exists the *crowding-in* effect on information acquisition. When the public signal becomes more precise, the market maker's pricing behavior depends more heavily on it, diminishing the price impact of the order flow (λ). In other words, the price becomes *insensitive* to a change in the order flow $y = x + u$ because the market maker has another reliable source of information, s_{pub}, to predict v. Since a higher price impact λ can be seen as the endogenous cost of being more informed, a decline in λ means that the *marginal* endogenous cost of obtaining a higher τ_e diminishes. Namely, even if s becomes more precise, market liquidity and the price impact (λ) worsen only slightly. This reduction in the sensitivity of the market impact λ allows the informed trader to trade more intensively on her private information, generating additional room for her to exploit the informational advantage. As a result, the informed trader finds it optimal to increase the precision of the private signal, on which she can act and derive an additional profit.

Thus, the crowding-in effect emerges from the strategic motive of the informed trader, i.e., she is aware of the endogenous cost of acquiring a precise private signal. The introduction of fair disclosure and a precise public signal reduces the endogenous marginal cost of being better informed, thereby promoting the informed trader's information acquisition.

From equation (1.19), we can analyze when the crowding-in effect becomes dominant:

Proposition 1.3. *The optimal precision of the private signal τ_e^* takes a hump-shaped curve against the precision of fair disclosure τ_{pub}. The tipping point is given by*

$$\tau_{pub}^* \equiv \left[\frac{1}{6}\left(\frac{c}{\sigma_u}\right)^{-\frac{3}{2}} - \frac{1}{\Sigma_0}\right]^+, \tag{1.20}$$

where $n^+ \equiv \max\{0, n\}$.

Proof. From equation (1.19), τ_e^* is increasing in b_{pub} if, and only if,

$$\left(\frac{1}{2}b_{pub}\Sigma_0\right)^{-1} > \tau_e^*.$$

Since the optimal τ_e^* is the solution of (1.18), and the SOC is satisfied, the above inequality holds if and only if

$$\frac{b_{pub}}{4}\sqrt{\Sigma_0^2\sigma_u^2}\left[\left(\frac{1}{2}b_{pub}\Sigma_0\right)+b_{pub}\Sigma_0\right]^{-\frac{3}{2}}\left(\frac{1}{2}b_{pub}\Sigma_0\right)^2-c<0.$$

By rearranging this inequality in terms of τ_{pub}, it is equivalent to $\tau_{pub}<\tau_{pub}^*$ where τ_{pub}^* is given by (1.20). $\square$

Proposition 1.3 shows that the crowding-in effect of fair disclosure dominates (resp. is dominated by) the crowding-out effect when τ_{pub} is small (resp. large).

Consider a marginal increase in the precision of the public signal, denoted as $\Delta\tau_{pub}>0$. When the level of τ_{pub} is small, the impact of $\Delta\tau_{pub}$ on the market maker's belief updating is large, while it has only an incremental impact on the belief updating when she already has a precise public signal (i.e., τ_{pub} is already high). Therefore, the reduction in the marginal endogenous cost due to $\Delta\tau_{pub}>0$ is strong in a small-τ_{pub} region, making the crowding-in effect more salient.

The tipping point τ_{pub}^* is decreasing in the marginal cost of private information c but increasing in the volatility of noise trading σ_u^2 and the initial uncertainty regarding the asset value Σ_0. A higher c magnifies the crowding-out effect of τ_{pub} because a large c implies that increasing the precision τ_e takes a larger cost. Knowing that some of the informational advantage is washed out by the public signal, the informed trader's incentive to increase τ_e declines more when its cost is large. In this case, it becomes more difficult for the crowding-in effect to be dominant. Thus, τ_e^* tends to be a decreasing function of τ_{pub}, and τ_{pub}^* becomes smaller.

In contrast, a more volatile noise trading σ_u and larger initial uncertainty Σ_0 promote the crowding-in effect. Both parameters make it harder for the market maker to infer v from the aggregate order flow. Then, the market maker becomes more dependent on the public signal. In this situation, the price p becomes more responsive to an increase in the precision of the public signal, leading to a larger decline in the marginal cost of increasing τ_e.

Proposition 1.3 implies that an attempt to create a level playing field by publicly and fairly disclosing information on v may *promote* the private information production by the informed trader. This result goes counter to the existing models on information disclosure, such as those cited by Goldstein and Yang (2017), because they only focus on the crowding-out effect. In my model, the key ingredient that drives the crowding-in effect is the strategic motive of the informed trader, which is absent in the existing studies that deal with competitive models. Moreover, my model shows that the crowding-in effect arises even in my basic model, as long as the trader is strategic. The existing models rely on some specific assumptions to derive the crowding-in effect, such as the mosaic theory with correlated private and public signals (Cheynel and Levine, 2020) and the possibility of resale after disclosure (McNichols and Trueman, 1994). In contrast, Proposition 1.3 depends solely on the "within-period" strategic motive with conditionally independent signals (s, s_{pub}).

1.4 Market quality

This subsection investigates how fair disclosure affects market quality, such as market depth (i.e., Kyle's λ) and price efficiency.

Price efficiency. Firstly, I derive the variance of v conditional on the price information. Since $p = \mu + \theta s_{pub} + \lambda(y - \gamma s_{pub})$, it becomes

$$\begin{aligned} Var(v|p) &= \Sigma_0 - \frac{(\lambda\beta + \theta)^2\Sigma_0^2}{(\lambda\beta + \theta)^2\Sigma_0 + \lambda^2\beta^2\sigma_e^2 + \lambda^2\sigma_u^2 + \theta^2\sigma_{pub}^2} \\ &= \frac{\Sigma_0}{1 + \Sigma_0\eta}, \end{aligned}$$

where

$$\eta \equiv \frac{(\lambda\beta + \theta)^2}{\lambda^2\beta^2\sigma_e^2 + \lambda^2\sigma_u^2 + \theta^2\sigma_{pub}^2} = \tau_{pub} + \frac{1}{2\tau_e^{-1} + b\Sigma_0}. \tag{1.21}$$

The second equality in (1.21) is derived by substituting equilibrium variables in Proposition 1.1.

Then, the price informativeness is measured by the signal-to-noise ratio of the price:

$$\Sigma \equiv \frac{Var(v)}{Var(v|p)} = 1 + \Sigma_0\eta,$$

Σ represents how much uncertainty in v is resolved by observing the equilibrium price.

Proposition 1.4. *The price informativeness, measured by Σ, is increasing in the precision of the public signal $\tau_{pub} = \sigma_{pub}^{-2}$.*

Proof. Firstly, the optimal τ_e^* solves the FOC in (1.18). Therefore, the implicit function theorem yields

$$\frac{d\tau_e^*}{db_{pub}} = \frac{\tau_e^*}{b_{pub}} \frac{\frac{1}{\tau_e^*} - \frac{1}{2}b_{pub}\Sigma_0}{\frac{1}{2}\frac{1}{\tau_e^*} + 2b_{pub}\Sigma_0}. \tag{1.22}$$

Secondly, we can rewrite

$$\eta = 2\frac{\tau_e^{*-1} + b_{pub}\Sigma_0}{b_{pub}\Sigma_0(2\tau_e^{*-1} + b_{pub}\Sigma_0)} - \Sigma_0^{-1}.$$

Thus, I analyze the behavior of $G(\tau_e, b_{pub}) \equiv \frac{\tau_e^{-1} + b_{pub}\Sigma_0}{b_{pub}\Sigma_0(2\tau_e^{-1} + b_{pub}\Sigma_0)}$ in the following. It holds that

$$\begin{aligned} \frac{dG(\tau_e^*, b)}{db} &= \frac{\partial G(\tau_e^*, b)}{\partial b} + \frac{d\tau_e^*}{db}\frac{\partial G(\tau_e^*, b)}{\partial \tau_e} \\ &\propto -\left(\frac{1}{2} + 2b_{pub}\Sigma_0\tau_e^*\right)\left[2\frac{1}{\tau_e^*}\left(\frac{1}{\tau_e^*} + b_{pub}\Sigma_0\right) + b_{pub}^2\Sigma_0^2\right] + b_{pub}\Sigma_0\left(\frac{1}{\tau_e^*} - \frac{1}{2}b_{pub}\Sigma_0\right) \\ &< 0. \end{aligned}$$

Therefore, η is an increasing function of b_{pub}, meaning that it decreases with τ_{pub}. $\square$

As suggested by (1.21), the precision of the public signal τ_{pub} directly improves the price informativeness, as it allows the market maker to know more about v by observing s_{pub}. Moreover, a higher τ_{pub} triggers the crowding-in and the crowding-out effects on the information-acquisition activity of the informed trader (Proposition 1.3), making the order flow y more or less informative. When the crowding-in effect is dominant, both the direct and indirect effects improve the price efficiency. When the crowding-out effect is dominant, the indirect effect reduces the amount of private information that the price can impound. However, even in this case, the crowding-out effect is indirect consequence of a change in τ_{pub} and cannot dominate the positive direct effect, leading to a more precise price information.

Note that the positive impact of τ_{pub} on Σ is robust due to the crowding-in effect. As suggested by Goldstein and Yang (2017), this result goes counter to most of the existing models, as they show that the crowding-out effect generates the ambiguous reaction of the price informativeness. In contrast, even though the crowding-out effect can be dominant in my model, there always exists the crowding-in effect, thereby (partially or completely) offsetting the crowding-out effect. Thus, the crowding-out effect cannot be sufficiently strong to harm price efficiency.

Market depth, liquidity, and the price impact. Consider the reaction of the price impact λ to an increase in the precision of the public signal τ_{pub}. As we have seen in Proposition 1.1, τ_{pub} reduces λ because it alleviates the asymmetric information problem between the market maker and the informed trader. However, it also affects the precision of the private signal τ_e: the informed trader becomes more or less informed, generating a non-trivial impact on λ.

From Proposition 1.1, and by using the definition of b_{pub}, the equilibrium λ is rewritten as

$$\lambda = \frac{\Sigma_0 b_{pub}}{2} \frac{1}{\sqrt{(\tau_e^{-1} + b_{pub}\Sigma_0)\sigma_u^2}}. \tag{1.23}$$

In (1.23), the negative impact of τ_{pub} on λ via $b_{pub} = \frac{1}{\Sigma_0 \tau_{pub}+1}$ represents the direct impact of τ_{pub} through a reduction in the informational problem. At the same time, λ is increasing in τ_e, as it worsens the asymmetric information problem. Thus, τ_{pub} may or may not increase λ by affecting τ_e at the equilibrium.

Once again, the following result shows that the indirect effect cannot dominate the direct effect of τ_{pub}, making the equilibrium λ a decreasing function of τ_{pub}.

Proposition 1.5. *Market liquidity, measured by $1/\lambda$, improves when the public signal becomes more precise (τ_{pub} increases).*

Proof. Note that it is sufficient to analyze the behavior of $\bar{\lambda} \equiv b_{pub}(\tau_e^{-1} + b_{pub}\Sigma_0)^{-\frac{1}{2}}$ at

$\tau_e = \tau_e^*$ to obtain the result. By using (1.22), it holds that

$$\frac{d\bar{\lambda}}{db_{pub}} \propto \frac{1}{\tau_e^*} + \frac{1}{2} b_{pub}\Sigma_0 + \frac{\frac{1}{\tau_e^*} - \frac{1}{2} b_{pub}\Sigma_0}{1 + 4 b_{pub}\tau_e^*\Sigma_0} > 0.$$

Therefore, λ is increasing in b_{pub}, meaning that it is decreasing in τ_{pub}. . □

Overall, Propositions 1.4 and 1.5 provides an optimistic prediction: both the market liquidity and the price efficiency improve when fair information disclosure provides a precise signal about the asset fundamentals. The existing studies have casted doubt on the effectiveness of fair disclosure, because disclosure unintendedly crowds out the private information production by the trader. In my model, fair disclosure has the other unintended impact that manifests in the opposite way—disclosure crowds in private information acquisition—and overturns the existing results on the reaction of market quality.

1.5 Conclusion

In this chapter, I study a model of strategic information acquisition and investigate whether a fair disclosure policy mitigates information asymmetry between traders. My model shows that disseminating material information to all traders in financial markets can promote private information acquisition by a potential informed traders, which I call the crowding-in effect. The existing studies on fair disclosure policies argue that public information should discourage such an information acquisition activity. Thus, my model proposes the possibility of the opposite result compared to the literature: a regulation that intends to discourage information acquisition may unintendedly encourage it.

The model in this chapter can be seen as the simplest building block to analyze the impact of innovations in information processing and regulations on information. That is, an informed trader may pay some cost to adopt new information technologies to acquire better information, while policy makers impose a regulation that reduces the marginal value of private information. My result indicates that a strategic trader may try to countervail the negative impact of the regulation by acquiring more information.

Prior to the recent innovations in ultra-fast information technologies, the above discussion has been enough because information acquisition is mostly about information quality, i.e., how precise information a trader can privately learn. However, high-frequency traders (HFT) add a new dimension to information acquisition: speed. In the next chapter, I provide a model of HFT to analyze the speed aspect of information acquisition and the impact of information regulations in terms of speed. The question is whether the crowding-in effect in this chapter survives even if we consider the speed aspect of information acquisition and trading, as well as whether market quality improves or deteriorates due to restrictions on traders' speed.

Chapter 2

The Dark Side of Delaying Order Execution

2.1 Introduction

Creating a level playing field is the centerpiece of financial regulation. Traditionally, it is designed to mitigate asymmetric information by requiring firms to disclose material news so that all traders obtain information of equal quality (e.g., Sarbanes–Oxley Act of 2002). In recent years, technological innovations have added a new dimension to traders' information acquisition—speed. Traders called high-frequency traders (HFTs) exploit sophisticated communication tools to learn information and act on it at the speed of light. On the one hand, their market-making behavior contributes to rendering markets more liquid. On the other hand, HFTs can quickly learn news and snipe standing limit orders at stale prices before liquidity providers update them based on new information. Thus, their speed advantage as liquidity takers exposes liquidity providers to adverse selection risk and tends to harm market liquidity.

Some exchange platforms try to slow down HFTs' liquidity-taking behavior by imposing delays on their order execution.[1] One of the most widely adopted forms of delays is so-called a speed bump. It imposes an intentional delay on the arrival or execution of trading orders at an exchange, aiming to protect liquidity providers against being sniped by HFTs. For example, Aequitas NEO and TSX Alpha, both Canadian exchanges, apply a few milliseconds of random delay to slow down liquidity-taking orders. As summarized by Table 2.1 (see Appendix A for more details), many other exchanges have also adopted or proposed adopting speed bumps. Reflecting their primary purpose, all speed bumps (except for those at IEX) are *asymmetric*, meaning that only liquidity-taking orders are delayed.

[1] Budish, Cramton, and Shim (2015) propose frequent batch auctions (FBA) as an alternative solution. However, adopting FBA needs considerable structural changes in the current continuous markets. In contrast, speed bumps are easier to adopt and has become a more popular tool to restrict HFT.

Table 2.1: Design of speed bumps

	Exchange	Date	Targets of delay	Length of delay
In operation	IEX	Octber 2013	All but pegged orders	350 microseconds
	Thomson Reuters*	June 2016	Non-cancellation	0-3 milliseconds
	Aequitas NEO*	March 2015	Liquidity takers	3-9 milliseconds
	TSX Alpha*	September 2015	Liquidity takers	1-3 milliseconds
	Eurex Exchange*	June 2019	Liquidity takers	1 or 3 milliseconds
	EBS Market*	July 2013	Liquidity takers	3-5 milliseconds
	ParFx*	March 2013	Liquidity takers	10-30 milliseconds
	Moscow Exchange*	April 2019	Liquidity takers	2-5 milliseconds
Proposed	CHX	Proposed	Liquidity takers	350 microseconds
	EDGA (Cboe)	Proposed	Liquidity takers	n/a
	NASDAQ OMX PHLX	Proposed	Liquidity takers	5 microseconds
	ICE Futures	Proposed	Liquidity takers	3 microseconds
	Interactive Brokers*	Proposed	Liquidity takers	10-200 milliseconds
	NYSE American**	July 2017	All but pegged orders	350 microseconds

Note: * indicates random speed bumps. **In December 2019, ICE announced removal of speed bumps from NYSE American based on their finding that speed bumps worsen liquidity and the trading share of the exchange. As of May 2020, exchanges with random speed bumps do not announce the distribution funcsion of random delays.

Although the intentional delays are growing popular, the study on the impact of delays is still in its infancy. This is the first study that considers it by incorporating strategic speed acquisition by HFTs. As the main contribution, this chapter shows that delaying execution of liquidity-taking orders (i.e., asymmetric delays) has a *crowding-in* effect on the HFTs' speed acquisition, meaning that asymmetric delays can endogenously increase the equilibrium speed of HFTs. The crowding-in effect undermines or even dominates the intended impact of delays. Thus, delaying order execution is not as effective as the literature has argued, and depending on market structures and parameters, it can worsen adverse selection and contribute to creating an "uneven playing field."

The first part of the chapter provides a simple benchmark model of a quote-driven market (à la Glosten and Milgrom, 1985) with a single HFT. In the model, an HFT as a liquidity taker acquires speed, and a competitive market maker sets a bid-ask spread that reflects the adverse selection cost. A speed-up by the HFT makes it easier for her to snipe a stale limit order and increases her expected profit. However, it also widens the bid-ask spread and reduces HFT's profit *margin* because a faster HFT imposes more severe adverse selection on the market maker (e.g., Biais, Foucault, and Moinas, 2015; Budish, Cramton, and Shim, 2015). Intentional delays in order execution affect the HFT's optimal speed level by altering the above tradeoff of being faster.

Delays in order execution bring about two positive impacts on the HFT's speed acquisition. Firstly, the intentional delays exogenously discount the arrival frequency of the HFT. Then, the market maker downplays the HFT's speed-up, making the equilibrium bid-ask spread less responsive to the HFT's speed acquisition. That is, the price impact of speed weakens, and it becomes easier for the HFT to exploit her speed advantage without adversely affecting the bid-ask spread. This effect can be seen as a decline in the endogenous marginal cost of being faster. Secondly, the intentional delays induce the market maker to tighten the spread by exogenously curbing the arrival of the HFT. It reduces the HFT's trading cost, boosts the sniping reward, and increases the marginal benefit of being faster. These are the *equilibrium* channels through which intentional delays strengthen the HFT's speed acquisition.

As the existing models argue, however, the intentional delays also have a negative equilibrium effect on speed acquisition, which I call a *crowding-out* effect. It arises because the intentional delays make it more difficult for the HFT to successfully snipe a stale limit order. As the chance of obtaining a reward declines, paying exogenous sunk costs for a speed-up becomes less worthwhile, making the HFT reluctant to increase her speed level.

I investigate when and why one of the crowding-in and crowding-out effects becomes dominant. A switch occurs depending on some observable parameters, such as the (exogenous) marginal cost of increasing speed, the expected length of delays, and the asset's volatility, measured by the arrival frequency of arbitrage opportunities and their size. For example, if the asset becomes more volatile, the HFT anticipates a more frequent arrival of sniping opportunities. It magnifies the expected sniping profit and reduces the *effective* exogenous cost of a speed-up (i.e., the cost per sniping opportunity), weakening the crowding-out effect. Thus, my model suggests that the imposition of delays tends to promote HFTs' trading speed when the volatility of assets is expected to be high (and vice versa).

In Section 2.3, I generalize the baseline model to accommodate multiple HFTs. It analyzes speed competition among HFTs —an "arms race"— and they serve both as snipers and high-frequency market makers. I show that an arms race can involve strategic complementarity, i.e., speed begets speed. An increase in market makers' speed has the same effect as an extension of order execution delays, as both of them reduce market makers' risk of being picked off. Therefore, a speed-up by an HFT as a market maker can promote other HFTs' speed acquisition, triggering a positive externality. The externality amplifies the original effect of order execution delays on speed acquisition and causes a substantial change in the equilibrium speed of HFTs.

Due to several competing effects, the impact of intentional delays on the adverse selection problem (i.e., the bid-ask spread) and market liquidity is ambiguous. Firstly, intentional delays *exogenously* reduce the bid-ask spread by directly hampering HFTs' arrival. Secondly, they *endogenously* affect the bid-ask spread through HFTs' equilibrium speed acquisition, the sign of which is ambiguous due to the crowding-in effect versus the crowding-out effect. Since an arms race involves the amplification effect, the endogenous impact of delays can even dominate their direct mitigating impact on the bid-ask spread when the crowding-in effect is dominant. Therefore, the attempt to slow down HFTs can not only encourage their speed acquisition but also harm market liquidity by imposing more severe adverse selection on market makers.

Section 2.4 analyzes welfare implications of the intentional delays, and I show that aggregate welfare can decrease due to the imposition of delays. Since the competitive market maker breaks even, the bid-ask spread is just a transfer between the HFT and a liquidity trader with exogenous trading motives. Thus, the intentional delays do not affect aggregate welfare via the bid-ask spread or the adverse selection cost for the market makers. The only remaining factor that affects aggregate welfare is the realized cost of speed investments that the HFT incurs. It is consistent with the literature and the real financial markets, where the most speed technologies are not general-purpose investments and distort social welfare. As the realized cost of speed is positively associated with the HFT's optimal speed, aggregate welfare declines when the crowding-in effect of order execution delays promotes the HFT's speed acquisition. Therefore, the crowding-in effect can be seen as a dark side of delaying order execution both in terms of the adverse selection problem (market liquidity) and aggregate welfare.

Section 2.5 provides extended discussions. For instance, it analyzes why an exchange platform has an incentive to introduce intentional delays when they have the crowding-in effect and can promote HFTs' speed acquisition. In reality, intentional delay is a self-imposed market structure that a *for-profit* exchange platform voluntarily adopts. Also, an exchange platform earns a large portion of profits by supplying speed services to HFTs (Budish, Lee, and Shim, 2018). Therefore, if the intentional delays have only the crowding-out effect, as argued by the literature, the adoption of delays cannot be explained by the profit-maximization behavior of an exchange: it reduces the demand for speed services and can harm exchange's profits. In contrast, the crowding-in effect of intentional delays promotes the demand for speed services and can increase exchange's profits. Hence, the intentional delays with the crowding-in effect can be consistent with the profit-maximization incentive of an exchange platform, allowing us to explain the adoption of delays in the real financial market.

This chapter contributes to the literature on HFT and market structure (see Jones, 2013; O'Hara, 2015; Menkveld, 2016 for reviews). The speed acquisition problem of HFT has been analyzed in the existing studies, such as Foucault, Roell, and Sandas (2003), Liu (2009), Foucault, Kadan, and Kandel (2013), Foucault, Kozhan, and Tham (2016).[2] However, the existing models study the problem in the absence of either a strategic motive of HFTs in speed acquisition or a slow market structure (i.e., order execution delays). My model integrates both factors and provides an insight into how intentional delays affect HFTs' speed acquisition and equilibrium market quality.

Moreover, my model shares the same interests as the studies on slow market structures, such as frequent batch auctions (Budish, Cramton, and Shim, 2015; Haas and Zoican, 2016) and speed bumps (Baldauf and Mollner, 2017; Brolley and Cimon, 2017; Aldrich and Friedman, 2018).[3] They highlight the benefits of slow market structures by focusing

[2] For the studies on high-frequency market making, see Aït-Sahalia and Saglam (2013), Han, Khapko, and Kyle (2014), Hoffmann (2014), and Conrad, Wahal, and Xiang (2015), Bongaerts and Van Achter (2016).

[3] The speed and frequency of order executions by a trading platform are also analyzed by Du and Zhu

on a particular design of delay that perfectly eliminates sniping by HFTs.[4] Also, they take speed of HFTs as given or a binary choice variable, while HFTs in my model strategically choose their speed with a continuous domain. Due to the above differences, my model uncovers a crowding-in effect of delays on speed acquisition.[5] It complements the literature by demonstrating that the impact of delays can be positive, negative, and ambiguous depending on the design of delays and other parameters.

The scope of the literature extends to empirical findings.[6] Hu (2018) and Chakrabarty, Huang, and Jain (2019) analyze the SEC approval of IEX with speed bumps as a National Securities Exchange (NSE) and find a net improvement in market quality measured by spreads. Shkilko and Sokolov (2016) exploit the interruptions of messaging caused by precipitation and find a reduction in quoted spreads.[7] In contrast, Chen, Foley, Goldstein, and Ruf (2017) analyze a speed bump at TSX Alpha and find an increase in quoted spreads. NYSE American (2019) also reports a decline in market quality and liquidity after the introduction of speed bumps and has decided to remove speed bumps from the exchange. Anderson, Andrews, Devani, Mueller, and Walton (2018) report that a change in the market-wide effective spread and the price impact is insignificant. Consistent with my model, a recent experimental study by Khapko and Zoican (2019) finds that a marginally longer speed bump stimulates the traders' investment in speed when the execution price is endogenous.[8]

2.2 Benchmark model

This section proposes a simple benchmark model to separate the main mechanism. Consider a one-shot exchange of an asset between three types of market participants: a single HFT, a competitive market maker, and a liquidity trader. As in the literature, the competitive market maker is justified by considering multiple potential market makers

(2017), Kyle and Lee (2017), Menkveld and Zoican (2017), and Pagnotta and Philippon (2018).

[4]Brolley and Cimon (2017) is an exception; they consider the possibility that a delay cannot eliminate sniping due to the randomness in sniping race and the length of a speed bump.

[5]Baldauf and Mollner (2017) consider information acquisition by investors by separating informed investors from HFTs, i.e., informed investors are assumed to be slow. Brolley and Cimon (2017) show that the equilibrium spread takes a hump-shaped reaction to a speed bump due to traders' migration between Slow and Fast exchanges. However, information acquisition in their model always weakens if a speed bump kicks in, and the speed of HFTs is an exogenous parameter.

[6]Other empirical studies on HFTs include, for example, Hendershott and Moulton (2011), Hasbrouck and Saar (2013), Riordan and Storkenmaier (2012),Ye, Yao, and Gai (2013), Frino, Mollica, and Webb (2014), Boehmer, Fong, and Wu (2015), and Brogaard et al. (2015).

[7]Although the interruption by precipitation may have a similar effect to a speed bump, this phenomenon is not paid much attention by financial institutions. In contrast, traders anticipate a speed bump and take it into their decision making.

[8]Khapko and Zoican (2019) find that a speed bump diminishes equilibrium investments into speed technologies when the price and trading profit for liquidity takers are exogenously fixed. This is also consistent with my model, as it suggests that an endogenous reaction of the spread to a speed bump contributes to the crowding-in effect.

who are inactive *on* the equilibrium path.[9] When the market opens at time $t = 0$, the asset has value $v = v_0 = 0$, which is common knowledge. $v = 0$ can be thought of as a situation where no mechanical arbitrage exists.

Innovations to the fundamental value arrive at a Poisson rate z. Conditional on a jump in the asset value, it becomes $v = v_0 \pm \sigma$ with equal probability. It can be thought of as a new arbitrage opportunity triggered by a jump in the asset's price that is not yet reflected by prices of other highly correlated assets (Budish, Cramton, and Shim, 2015).[10] I focus on the very short time interval in which innovations to v occurs at most once and denote the timing of the jump as $t_0 \sim \text{Exp}(z)$.

Following the convention of market microstructure (e.g., Glosten and Milgrom, 1985), I assume that a trade size is restricted to one unit.

2.2.1 Traders

Market maker. There is a risk-neutral competitive *market maker*. When the market opens, she posts a single-unit limit order (LO) by quoting ask and bid prices $(s, -s)$ and commits to trade at these prices. The market maker revises or cancels her initial quote if private information arrives or a liquidity taker takes her LO (defined below).[11] In what follows, price s is referred to as the (half) spread.

After posting a limit order, the market maker starts monitoring the market, anticipating a jump in the asset value. When the asset's value jumps at $t = t_0$, the market maker's learning process is put in motion, and she obtains news on true v with a Poisson process. Specifically, it takes stochastic time $T_M \sim \text{Exp}(\gamma)$ to learn v after the jump. T_M can be thought of as the shortest necessary time for the market maker to learn v and act on it, and γ represents the speed of the market maker.[12]

Upon privately learning v, the market maker re-prices her initial limit order by quoting new bid and ask prices (v, v) to avoid being picked off at stale prices. If her limit order is taken by a taker (as specified below), the market maker exits the market.

High-frequency trader. There is a risk-neutral *high-frequency trader (HFT)*. Before she enters the market, the HFT invests in a speed technology that provides speed $\phi \geq 0$. It could be colocation services, high-bandwidth connectivity, direct data feeds, or a

[9]Multiple market makers exist on the sidelines on-path. One of the inactive market makers may become active in an off-path event such that the market maker who is active on the equilibrium path deviates from the equilibrium spread.

[10]Innovations in the asset value could also stem from some public news, such as Fed announcements and the release of government statistics, or execution of Intermarket Sweep Orders (Baldauf and Mollner, 2017).

[11]The market maker does not have an incentive to revise her quote prior to the above events because the model is homogeneous in time due to Poisson distributions.

[12]Section 2.3 considers an endogenous choice of γ by high-frequency market makers.

sophisticated trading algorithm.[13] The HFT also monitors the market and, when the jump in v occurs, her learning process is put in motion. It takes stochastic time $T_H \sim \text{Exp}(\phi)$ for the HFT to learn v. Upon learning v, the HFT immediately submits a market order to "snipe" a limit order on the limit order book (LOB).[14]

The Poisson arrivals of traders are in line with the literature, such as Budish, Cramton, and Shim (2015) and Brolley and Zoican (2019). For example, the HFT's arrival intensity (ϕ) can be seen as the amount of computer power or the information processors that the HFT controls at a given point in time, making the learning/arrival rate of the HFT proportional to ϕ. Also, aside from the "speed," we can think of ϕ as the monitoring intensity, as in Foucault, Kozhan, and Tham (2016).

Liquidity trader. There is a *liquidity trader* (LT) who is exposed to a liquidity shock. The shock exogenously makes her submit a buy or sell market order with equal probability. Thus, her trading behavior conveys no information. The LT captures slow passive investors with no material information in the real world. Her behavior stems from some exogenous reasons, such as hedging motives. Let $T_L \sim \text{Exp}(\beta)$ be the stochastic timing that the liquidity shock hits and the LT places an market order. Parameter β captures the frequency of the liquidity shock, and I set $\beta > \frac{\gamma}{3}$ to make the model well defined.[15] Since her trading motive is independent of the asset value and its price, the LT is exposed to the shock from $t = 0$ onward.

In what follows, all random variables, $(t_0, v, \{T_i\}_{i=M,H,L})$, are assumed to be independent. Also, for simplicity, I suppose that there is no mechanical latency *per se* other than those mentioned above.

2.2.2 Asymmetric delay in order execution

A market imposes an *asymmetric* delay on order execution (e.g., speed bumps) and delays all market orders by the same length of time $\tilde{\delta} \geq 0$.[16] An asymmetric delay means that it is imposed only on liquidity takers, whereas liquidity-providing orders (i.e., limit

[13] The *ex-ante* choice of speed is in line with the literature, e.g., Foucault, Kadan, and Kandel (2013) and Brolley and Zoican (2019). Also, it is consistent with the real-world speed acquisition, as speed services are provided via pre-determined monthly subscriptions, and computer algorithm for fast trading must be set up before a trader starts trading.

[14] HFT does not intentionally delay the timing of the order submission: if she gets information at t, she immediately places an order at t. She might be motivated to put a time lag between information arrival and order placement, as it can reduce a spread and increase her sniping profit. However, without a commitment device, this cannot be an equilibrium since it is always optimal for the HFT to snipe immediately upon learning given a time lag she previously announces, i.e., there is a time inconsistency.

[15] This inequality implies that the liquidity trader arrives at the market at a sufficiently high frequency. If β becomes sufficiently small, the market maker anticipates a trade with the HFT almost surely and needs to charge a wide bid-ask spread, leading to market break downs.

[16] The model gives qualitatively the same (and simpler) results if delays are imposed only on the HFT. See also Internet Appendix B for the case that a delay is applied only to in an iid manner.

orders my the market maker in this model) are not restricted.[17] As shown by Table 2.1, asymmetric delays reflect the speed bumps in the real financial markets.

The model can describe both random and deterministic delays, but the primary focus of the model below is on the random ones. The case with a deterministic delay is analyzed in Section 2.5, where the results are qualitatively the same as those with random delays. When a delay is random, I assume that $\tilde{\delta}$ follows the exponential distribution with $\delta = \mathbb{E}[\tilde{\delta}]$ and is independent of other random variables.[18] Appendix B.7 shows that the following results are robust to generalizing the distribution of $\tilde{\delta}$ with some parameter restrictions.

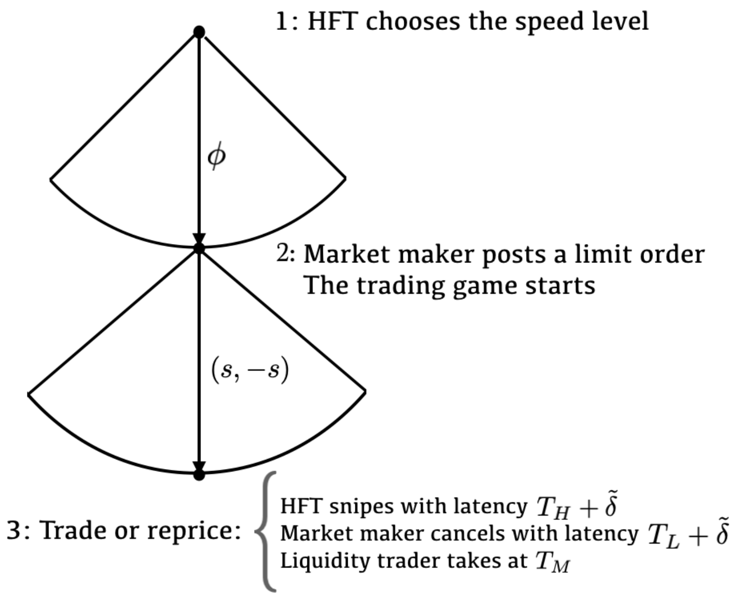
2.2.3 Equilibrium

The model is conceptualized as a sequential game with two stages, and the equilibrium concept is the subgame perfect equilibrium. Figure 2.1 visualizes the timeline of the game. In the first stage, the HFT chooses the level of speed ϕ. In the second stage, the market maker submits a limit order with competitive prices, and traders move as specified in Subsection 2.2.1. The end of the game is triggered by one of the three possible events: execution of a liquidity taking order by the HFT or the LT, or repricing by the market maker.

[17] My model is not an appropriate workplace to analyze symmetric speed bumps, such as those at IEX. It delays execution or arrival of all orders (both from takers and makers) with an exemption provided to a specific type of undisplayed pegged orders. As Appendix A explains in detail, the pegged orders cannot be analyzed by limit orders in my model.

[18] When a delay is random, the HFT is prohibited from sending multiple orders at any two points of time that are infinitely close, i.e., the HFT cannot send an order at time t and another order at t' for $|t - t'| < \epsilon$ with infinitely small ϵ. As in Baldauf and Mollner (2017), this restriction rules out the possibility that the HFT sends redundant orders, in the expectation that one of them can trade faster than others due to the randomness in a delay. It allows the model to avoid unnecessary complications.

Figure 2.1: Timeline of the game



2.2.4 Equilibrium behavior of a market maker

Unless otherwise mentioned, I consider a *random* delay. Without loss of generality, I analyze how the ask price (s) is determined when $v = +\sigma$.[19]

The expected profit of the market maker from posting s is given by

$$
\begin{aligned}
V_M(s) =&\mathbb{E}_{\tilde{\delta},t_0}\overbrace{\Bigg[s\int_{\tilde{\delta}}^{\infty}\beta e^{-\beta(t-\tilde{\delta})}e^{-\gamma(t-t_0)^+}e^{-\phi(t-t_0-\tilde{\delta})^+}dt}^{\text{liquidity trader takes LO}} \\
&+(s-\sigma)\underbrace{\int_{t_0+\tilde{\delta}}^{\infty}\phi e^{-\beta(t-\tilde{\delta})}e^{-\gamma(t-t_0)}e^{-\phi(t-t_0-\tilde{\delta})}dt\Bigg]}_{\text{HFT takes LO}}, \qquad (2.1)
\end{aligned}
$$

where $\mathbb{E}_{\tilde{\delta},t_0}$ denotes the expectation operator with respect to $\tilde{\delta}$ and t_0, and $X^+ \equiv \max\{X, 0\}$. The profit and the cost from market making stem from trading with the liquidity trader and the informed HFT.

The first line in (2.1) captures a trade with the liquidity trader, which happens from time $\tilde{\delta}$ and onward (given $\tilde{\delta}$). It takes place at t if the liquidity shock hits at $T_L = t - \tilde{\delta}$ but learning by the market maker and the HFT's trade do not happen until time t. Thus, given the liquidity trader arrives at t, she can fulfill her trading needs with probability

[19]The model is symmetric around zero, and the symmetric argument gives the results for the bid-side of the market.

density $e^{-\gamma(t-t_0)^+}e^{-\phi(t-t_0-\tilde{\delta})^+}$. For the market maker, the expected profit from this trade is $s - \mathbb{E}[v] = s$ since the liquidity trader does not bring new information.

In contrast, the second line in (2.1) represents a trade with the HFT, which happens in $t \geq t_0 + \tilde{\delta}$ (given $\tilde{\delta}$ and t_0). The integrand shows the density function for the HFT's order execution at time t, given that the market maker's learning and the liquidity trading do not happen before t. The market maker's expected profit from trading with the HFT is $s - \sigma$ because the HFT buys the asset if and only if $v = +\sigma$.

By incorporating the randomness in $\tilde{\delta}$ and t_0, the overall expected profit from market-making is given by equation (2.1). Note that an intentional delay ($\tilde{\delta}$) makes the market maker less likely to be picked off by the HFT given the innovation in the asset value, as the HFT's order execution is delayed and happens only after $t_0 + \tilde{\delta}$. This is the direct effect of an intentional delay in order execution that mitigates the adverse selection problem.

Since the market maker is competitive, her expected profit shrinks to zero, as in Glosten and Milgrom (1985). The break-even condition yields the following equilibrium spread.

Proposition 2.1. *(i) With* $\lambda(\delta) \equiv \frac{z^{-1}(1+\delta)+\delta}{1+z\delta}$*, the equilibrium (half) spread is given by*

$$s = \frac{\frac{\phi}{1+\lambda(\delta)(\phi+\beta+\gamma)}}{\frac{\phi}{1+\lambda(\delta)(\phi+\beta+\gamma)} + \beta}\sigma. \tag{2.2}$$

(ii) s *is monotonically increasing in* ϕ *and decreasing in* δ *(with* ϕ *fixed).*

Proof. See Appendix B.1 for point (i). Point (ii) is obvious from (2.2). □

Note that $\lambda(\delta)$ is monotonically increasing in the expected length of a delay, $\delta = \mathbb{E}[\tilde{\delta}]$. Thus, in equation (2.2), the direct effect of a delay appears in the form of *discount* on the arrival rate of the HFT (ϕ). This is because, conditional on a jump in v, a delay in order execution makes a trade less likely to happen in interval $[t_0 + \tilde{\delta}, \infty)$ before the market maker learns v, reducing the market maker's risk of being picked off by the HFT.[20]

We can think of the term $\frac{\phi}{1+\lambda(\delta)(\phi+\beta+\gamma)}$ as the effective arrival rate of the HFT. *Ceteris paribus*, the spread shrinks when the expected length of a delay becomes longer, as the delay mitigates adverse selection risk and reduces the HFT's effective arrival rate. However, this channel only captures the direct effect of a delay, and it is premature to conclude that it is effective. In particular, the HFT's speed (ϕ) is endogenous and depends on δ.

[20] An intentional delay also delays the liquidity trader's order execution, but equation (2.2) implies that it reduces the adverse selection cost more than it reduces the profit from liquidity trading. This is because the liquidity trader's trading motive is independent of the innovation in v. Given that an innovation in v occurs, the intentional delay reduces the arrival rate of the HFT and the liquidity trader by the same magnitude in expectation, thereby reducing the adverse selection cost and the profit from liquidity trading. However, even if a jump in v does not happen, the liquidity trader can be hit by the shock and trade with the market maker, in which case the HFT and market makers are inactive due to the absence of a jump in v. In this case, the intentional delay does not affect the market maker's expected profit from liquidity trading. As a result, the intentional delay mitigates adverse selection more than it reduces the profit from liquidity trading.

2.2.5 Speed acquisition by an HFT

Consider the HFT's optimal speed acquisition problem to investigate how ϕ depends on the expected length of a delay, δ.

Firstly, the following lemma provides the *ex-ante* expected probability that the HFT can snipe a standing limit order:

Lemma 2.1. *When the HFT has speed ϕ, and the expected length of a delay is δ, the HFT's expected sniping probability (as of $t = 0$) is given by*

$$\pi(\phi, \delta) = \left(\underbrace{\frac{\phi}{\phi + \beta + \gamma}}_{\equiv \pi_{ND}} \times \underbrace{\frac{1}{1 + \gamma\delta}}_{\equiv h} \right) \frac{1}{1 + \frac{\beta}{z}} \tag{2.3}$$

Proof. With $\tilde{\delta}$ and t_0 fixed, the HFT can snipe the standing LO at time $t + \tilde{\delta}$ if she learns v and places her order at time $T_H = t$, while the market maker does no learn v or exits the market prior to t. These events happen with density $\pi_t(\phi, \tilde{\delta}) = \phi e^{-\phi(t-t_0)} e^{-\beta t} e^{-\gamma(t-t_0+\tilde{\delta})}$ for $t > t_0$. Then, by taking the expectation with respect to $\tilde{\delta}$ and t_0, the expected sniping probability is given by computing $\pi(\phi, \delta) = \mathbb{E}_{\tilde{\delta}, t_0} \left[\int_{t_0}^{\infty} \pi_t(\phi, \tilde{\delta}) dt \right]$. □

Since the learning process of the HFT (T_H) is independent of a random delay ($\tilde{\delta}$), the sniping probability in (2.3) can be decomposed into three factors. The first component, denoted as $\pi_{ND}(\phi)$ in equation (2.3), is the *no-delay sniping probability.* It represents the probability of sniping (conditional on the occurrence of the jump in v before the liquidity trading) when no intentional delays are imposed in expectation. The second component, denoted as $h(\delta)$, is the *delay effect* that discounts π, i.e., sniping the standing LO becomes harder if the expected length of a delay grows longer. The last coefficient, $\frac{1}{1+\frac{\beta}{z}}$, in (2.3) represents the discount of the expected sniping probability that emanates from the possibility that the liquidity trader takes the LO before the asset value jumps

The HFT chooses the optimal speed ϕ^* to maximize the following *ex-ante* expected profit net of the cost of speed:

$$W_{HFT}(\phi) = \frac{1}{1 + \beta z^{-1}} \pi_{ND}(\phi) h(\delta)(\sigma - s(\phi)) - C(\phi), \tag{2.4}$$

$$\text{s.t. } s = \frac{\frac{\phi}{1+\lambda(\delta)(\phi+\beta+\gamma)}}{\beta + \frac{\phi}{1+\lambda(\delta)(\phi+\beta+\gamma)}} \sigma.$$

In equation (2.4), $C(\phi)$ denotes the cost of investing into the speed technology. Although I assume that C is weakly increasing in ϕ ($C' \geq 0$), I do not bring a specific form of C unless otherwise mentioned.

The FOC of (2.4) is given by

$$0 = \frac{dW_{HFT}(\phi)}{d\phi} = \frac{\sigma}{1 + \beta z^{-1}} h(\delta) \left[\left(1 - \frac{s(\phi, \delta)}{\sigma}\right) \frac{d\pi_{ND}}{d\phi} - \pi_{ND} \frac{d}{d\phi} \frac{s(\phi, \delta)}{\sigma} \right] - C'(\phi). \tag{2.5}$$

Firstly, ϕ increases W_{HFT} because the HFT with faster speed technologies is more likely to snipe, i.e., $\frac{d\pi_{ND}}{d\phi} > 0$. Secondly, a speed-up reduces the HFT's profit margin because trading cost s increases, i.e., $\frac{d(1-\frac{s}{\sigma})}{d\phi} < 0$. Put differently, the equilibrium spread involves the positive price impact and can be seen as an *endogenous* cost of being faster,[21] with $\frac{ds}{d\phi} > 0$ measuring the *endogenous* marginal cost of a speed-up. In what follows, I denote the solution of the FOC as ϕ^* and call it the optimal speed for the HFT.

2.2.6 Impact of delay in order execution

I first analyze how a longer expected delay (a larger δ) affects the optimal speed of the HFT. As equation (2.5) suggests, δ affects the FOC and ϕ^* via three channels: the level of the spread (s), the slope of the spread ($\frac{ds}{d\phi}$), and the delay effect in the sniping probability (h). Regarding the second channel, the following lemma holds:

Lemma 2.2. *In the equilibrium, the price impact of speed is decreasing in δ, i.e., $\frac{\partial}{\partial\delta}\left(\frac{ds}{d\phi}\right) < 0$.*

Proof. The result is derived by evaluating $\frac{\partial}{\partial\delta}\left(\frac{ds}{d\phi}\right)$ at the equilibrium speed in equation (2.7) below. □

Lemma 2.2 implies that a longer expected delay renders the spread less sensitive to the HFT's speed-up. A longer delay is more likely to protect the market maker against sniping by the HFT. In other words, the market maker's risk of being picked off grows insensitive to ϕ. Consequently, even if the HFT becomes faster, the market maker widens the spread only slightly. A decline in the price impact of speed makes it easier for the HFT to increase her speed without reducing her profit margin.

Now, consider the partial derivative of $\frac{dW_{HFT}}{d\phi}$ with respect to the expected length of a delay (δ) to understand the impact of δ on the optimal speed. From equation (2.5), it holds that

$$\frac{\partial}{\partial\delta}\left(\frac{dW_{HFT}}{d\phi}\right)|_{\phi=\phi^*} = \frac{\sigma}{1+\beta z^{-1}} h(\delta)\left[\overbrace{\frac{\partial(1-\frac{s}{\sigma})}{\partial\delta}\frac{d\pi_{ND}(\phi^*)}{d\phi}}^{\text{effect (i)}>0} + \overbrace{\pi_{ND}(\phi)\frac{\partial}{\partial\delta}\left(\frac{d(1-\frac{s}{\sigma})}{d\phi}\right)}^{\text{effect (ii)}>0}\right] + \frac{C'(\phi^*)}{h(\delta)}\underbrace{\frac{dh(\delta)}{d\delta}}_{\text{effect (iii)}<0}. \tag{2.6}$$

[21] Aquilina, Budish, and O'Neill (2020) show that more than 80% of sniping races are played by top 6 high-frequency trading firms. Thus, incorporating the strategic motive of HFTs is natural.

Increasing the expected length of a delay (δ) has two positive effects and one negative effect on ϕ^*, where the last term becomes proportional to C' due to the FOC (i.e., the envelope condition).

1. An intentional delay exogenously mitigates adverse selection, reduces the spread, and increases the (normalized) profit margin $1 - \frac{s}{\sigma}$ (Proposition 2.1). As a result, being faster and increasing the sniping probability become more worthwhile, as effect (i) in equation (2.6) shows.

2. The spread becomes less responsive to the speed-up by the HFT, as Lemma 2.2 attests. This reduces the marginal cost of being faster for the HFT. Effect (ii) in (2.6) captures this channel.

3. A longer delay makes it difficult for the HFT to snipe, and $h(\delta)$ declines. It reduces the marginal benefit of being faster, as shown by effect (iii) in equation (2.6).

If $C' = 0$ (i.e., investments in speed technologies take only a fixed cost), effect (iii) disappears. As a result, the first two positive effects become dominant, and a longer expected delay (δ) unambiguously encourages the HFT to invest in the speed technology (ϕ) rather than thwarting it. If $C' > 0$, in contrast, effect (iii) disincentivizes speed acquisition. Intuitively, increasing the speed level involves the exogenous marginal cost C', but it pays out only if the HFT can snipe an LO. Since a longer delay makes it difficult to snipe, it becomes less worthwhile to increase the speed. Effect (iii) competes with effects (i) and (ii), and the result becomes ambiguous.

Proposition 2.2 (Crowding-in effect). *(i) If the exogenous cost of speed is constant ($C' = 0$), the HFT's optimal speed is given by*

$$\phi^* = \begin{cases} \frac{\sqrt{\beta(\beta+\gamma)}(1+\lambda(\delta)(\beta+\gamma))}{1-\lambda(\delta)\sqrt{\beta(\beta+\gamma)}} & \text{if } \lambda < \frac{1}{\sqrt{\beta(\beta+\gamma)}} \\ \infty & \text{otherwise.} \end{cases} \tag{2.7}$$

Moreover, ϕ^* is increasing in δ (when ϕ^* is bounded).
(ii) When the exogenous marginal cost of speed is positive ($C' > 0$), the optimal speed is increasing in δ if C' and β are small, and z and σ are large (and vice versa).

Proof. Solving $W'_{HFT} = 0$ with $C' = 0$ yields ϕ^*. It is increasing in δ because $\lambda'(\delta) > 0$. See Appendix B.2 for point (ii). □

Proposition 2.2 demonstrates the existence of a *crowding-in* effect, i.e., an intentional delay can increase the HFT's equilibrium speed. The only negative effect of δ, which I call a *crowding-out* effect, stems from effect (iii) in equation (2.6) and is effective only if $C' > 0$, i.e., investments in speed technologies take an increasing variable cost.

Whether the crowding-in effect dominates the crowding-out effect (negatively) depends on the *effective* marginal cost of speed, i.e., the cost per sniping opportunity. First of all, $C' > 0$ is the driving force of the crowding-out effect (see effect [iii] in equation [2.6]). However, the marginal cost of speed, C', is amortized over the expected trading opportunities. Thus, a more frequent jump in v (parameterized by z) and a larger jump size (σ) dwarf the marginal cost of speed. They weaken the crowding-out effect, and the optimal speed of the HFT tends to be an increasing function of the expected length of a delay. The liquidity trader's arrives frequency (β) also matters, as less frequent liquidity trading makes sniping easier for the HFT, magnifying the crowding-in effect.

Overall, the baseline model demonstrates that the attempt to slow down high-frequency trading by imposing an intentional delay can encourage HFT's endogenous speed acquisition. The next subsection analyzes the reaction of the adverse selection problem to the imposition of a delay by incorporating the crowding-in and the crowding-out effects.

2.2.7 Adverse selection and liquidity

Hereafter, I use the equilibrium half spread (s) as a measure of adverse selection and market liquidity.[22] An order execution delay directly mitigates adverse selection by delaying the HFT's arrival (Proposition 2.1) but it can promote speed acquisition by the HFT (Proposition 2.2), thereby worsening adverse selection.

Proposition 2.3. *(i) If the cost of speed investments is constant ($C' = 0$), the effective arrival rate of the HFT is constant in the equilibrium, i.e.,*

$$\textit{Effective arrival rate of HFT} \equiv \frac{\phi^*}{1 + \lambda(\delta)(\phi^* + \gamma + \beta)} = \sqrt{\beta(\beta + \gamma)}. \tag{2.8}$$

Moreover, the equilibrium spread is independent of the expected length of a delay, i.e., $\frac{ds(\phi^*,\delta)}{d\delta} = 0$.
(ii) If the cost of speed investments is increasing in ϕ ($C' > 0$), the equilibrium spread is decreasing in the expected length of a delay, i.e., $\frac{ds}{d\delta} < 0$.

Proof. See Appendix B.3 for (i) and Appendix B.2 for (ii). □

Firstly, if the marginal cost of speed is absent, the expected length of a delay is irrelevant to the HFT's effective arrival rate and the equilibrium adverse selection problem. Although the crowding-in effect (Proposition 2.2) fuels speed acquisition, the strategic HFT knows that a higher ϕ exacerbates adverse selection and reduces her profit margin. Since she is a monopolistic HFT facing no externalities, and there is no exogenous cost of adjusting ϕ, she can freely choose the level of ϕ such that her effective arrival rate stays constant at the optimal level in equation (2.8) regardless of δ. Since that the

[22] Since my model does not have trading fees and commissions, the bid-ask spread corresponds to the price impact measure and reflects the adverse selection cost for the market maker.

spread in equation (2.2) depends on the effective arrival rate of the HFT, it also becomes independent of the expected length of a delay.[23]

However, as suggested by result (ii), the irrelevance result holds only with $C' = 0$ in the benchmark model. More precisely, it rests on the environment with a single HFT and no adjustment costs of speed ($C' = 0$). By applying the chain rule to the equilibrium spread, we have

$$\frac{ds^*(\phi^*, \delta)}{d\delta} = \underbrace{\frac{\partial s^*}{\partial \delta}}_{<0} + \underbrace{\frac{d\phi^*}{d\delta}\frac{\partial s^*}{\partial \phi}}_{\gtrless 0}. \tag{2.9}$$

The first term is the negative direct impact of a delay on the spread, while the second term is its impact through HFT's speed acquisition. When the cost of speed is constant ($C' = 0$), the second term is unambiguously positive due to the crowding-in effect and completely offsets the negative first effect. In contrast, $C' > 0$ generates the crowding-out effect and weakens the second channel, i.e., $\frac{d\phi^*}{d\delta}|_{c=0} > \frac{d\phi^*}{d\delta}|_{c>0}$. Since C' does not directly affect $\frac{s}{\sigma}$, the expression in (2.9) becomes negative.

Overall, the baseline model with a single HFT proposes the crowding-in effect of delaying order execution, but delays are still effective in reducing the spread, as long as increasing ϕ takes some exogenous costs with $C' > 0$. However, the following section with an arms race generalizes this result, showing that delays can backfire in terms of the spread in some parameter regions.

2.3 Multiple HFTs and high-frequency market makers

In reality, both liquidity snipers and liquidity providers adopt high-frequency trading technologies. To accommodate this fact with minimal deviations from the benchmark model, I consider the case with $N = 3$ HFTs, indexed by $i = 1, 2, 3$.[24]

As in the baseline model, HFTs choose the speed level ϕ_i before the market opens. The flow of the subsequent events for HFTs in the generalized model are given by the following.

[23] Proposition 2.3 resembles the result in Kyle (1985) that shows that an increase in the volatility of noise trading does not affect the market depth. In Kyle (1985), the monopolistic insider knows that more volatile noise trading reduces the price impact of the order flow (Kyle's lambda), and she tries to exploit her informational advantage more intensively. At the same time, however, she knows more intensive informed trading increases the price impact. Thus, she picks the trading intensity so that more aggressive informed trading just offsets the impact of more volatile noise trading.

[24] In reality, the number of trading firms (N) that participate in a sniping race tends to be small. For example, Aquilina, Budish, and O'Neill (2020) show that more than 80% of sniping races are played by the top 6 HFT firms, meaning that a race is highly concentrated (see also Table **??** in Internet Appendix **??**). Thus, considering a relatively small N reasonably reflects a sniping race in the real financial markets.

1. Upon entering the market, one of the HFTs is randomly selected as a high-frequency sniper (HFS; $i = S$),[25] while the other two HFTs become high-frequency market makers (HFMs). The random selection of role by nature can be thought of as a situation where HFTs are "registered (or designated) market makers" at an exchange.[26]

2. At $t = 0$ (when the market opens), both HFMs post a single-unit limit order with (half) spread s_i. At this stage, they engage in *strategic* competition for liquidity provision à la Dennert (1993).

3. Conditional on a jump in asset value, HFT i learns v with stochastic latency $T_i \sim \text{Exp}(\phi_i)$. If HFT i is a sniper, she instantly places a market order to snipe *all* available limit orders (i.e., two units) provided by HFMs. If HFT i is an HFM, she reprices her limit order so that bid and ask prices become (v, v). Also, at time $T_L \sim \text{Exp}(\beta)$, the liquidity shock hits a liquidity trader, and she posts a single unit buy or sell market order to fulfill her trading needs at the best price.

In order to describe strategic competition between high-frequency market makers, I follow Dennert (1993) and Baruch and Glosten (2019) and relax the assumption on a unit trading size for an HFS: she takes all available liquidity on the LOB (i.e., two units) as long as they are profitable, while a liquidity trader trades only one unit due to an exogenous capacity constraint. This is consistent with the literature (e.g., Easley and O'hara, 1987) that shows that informed traders tend to trade larger quantities than other traders. Other structures of the game, including an order execution delay, stay the same.

2.3.1 Strategic liquidity provision

Suppose that HFT i and j become HFMs. Since HFMs are strategic, posting the competitive spread is no longer an equilibrium.

HFM i knows that an HFS tries to snipe her limit order as long as $s_i \leq \sigma$, as the sniper takes all available liquidity. In contrast, a liquidity trader tries to trade one unit at the *best* price. Thus, HFM i can compensate for the loss from adverse selection only if she posts a better price than her rival (HFM j). In the case of a tie ($s_i = s_j$), I assume that HFM i trades with a liquidity trader with probability $g_i \in (0, 1)$.[27]

After posting her initial limit order, there are three possible events that trigger further action by HFM i. Firstly, if the HFS takes liquidity, then both HFMs exit the market, as the HFS takes both limit orders. Secondly, if the liquidity trader takes HFM i's limit

[25]Internet Appendix of Aoyagi (2020) considers more than one snipers.

[26]If a financial institution is a registered (or designated) market maker, she is assigned a particular stock that she must make markets. Thus, if a jump occurs in a security price which HFT i must make markets, she serves as an HFM. If a jump occurs in other assets' prices, then HFT i may serve as a sniper. NYSE is the most famous exchanges that adopt designated market making, and NASDAQ is another example of exchanges that adopt registered market maker ("NASDAQ member" firms). For example, GETCO was one of the HFTs serving as a designated market maker on NYSE.

[27]The tie-breaking rule, g_i, does not affect the following results.

order, HFM i exits the market. If, in contrast, the liquidity trader trades with HFM j, then HFM i can realize the arrival of the liquidity trader by looking at the LOB. This is because a single unit limit order disappears from one side of the LOB only if the LO is taken by the liquidity trader. Then, HFM i updates her quote to $(ask, bid) = (\sigma, -\sigma)$ because the liquidity trader no longer arrives. Finally, if one of the HFMs obtains her private news and updates her quote to (v, v), then the other HFM also learns v by looking at her rival's new quote, i.e., the LOB is observable.

Due to the last point, learning by the HFMs takes place if one of them learns and updates her LO. By exploiting the property of the exponential distribution, each HFM learns v with intensity $\phi_M \equiv \phi_i + \phi_j$. This property simplifies the competition between HFMs, as they become homogeneous regarding the learning process.[28]

2.3.1.1 HFMs' expected profit

In the following, I denote the speed level of the HFS as ϕ_S. Since the structure of the trading game is the same as the benchmark model, HFM i's expected profit from market-making, given her rival's quote s_j, is analogous to equation (2.1):

$$\begin{aligned} V_{M,i}(s_i, s_j) =& \mathbb{E}_{\tilde{\delta}, t_0} \left[\theta(s_i, s_j) s_i \int_{\delta}^{\infty} \beta e^{-\beta(t-\tilde{\delta})} e^{-\phi_M (t-t_0)^+} e^{-\phi_S (t-t_0-\tilde{\delta})^+} dt \right. \\ & \left. + (s-\sigma) \int_{t_0+\tilde{\delta}}^{\infty} \phi_S e^{-\beta(t-\tilde{\delta})} e^{-\phi_S(t-t_0-\tilde{\delta})} e^{-\phi_M (t-t_0)} dt \right] \qquad (2.10) \\ \propto & \frac{\phi_S}{1+\lambda(\delta)(\phi_M+\phi_S+\beta)} (s_i - \sigma) + \theta(s_i, s_j) \beta s_i \qquad (2.11) \end{aligned}$$

with

$$\theta(s_i, s_j) \equiv \mathbb{I}_{\{s_i < s_j\}} + g \mathbb{I}_{\{s_i = s_j\}}.$$

$\theta(s_i, s_j)$ represents the probability that HFM i can trade with the liquidity trader when the posted prices are s_i and s_j, conditional on the liquidity shock.

Also, $\lambda(\delta) = \frac{z^{-1}(1+\delta)+\delta}{1+z\delta}$ is the same as that in the baseline model. Thus, equation (2.11) indicates that the arrival rate of the HFS (ϕ_S) is discounted by the impact of a delay, λ, while the probability of trading with the liquidity trader is also discounted by θ due to competition for liquidity provision.

To solve the model, I introduce the following variable:

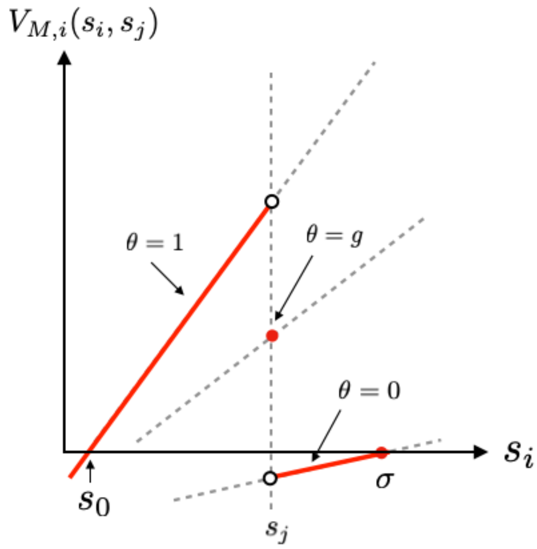
$$s_0 \equiv \frac{\frac{\phi_S}{1+\lambda(\delta)(\phi_M+\phi_S+\beta)}}{\frac{\phi_S}{1+\lambda(\delta)(\phi_M+\phi_S+\beta)} + \beta} \sigma. \qquad (2.12)$$

s_0 represents the break-even spread for HFMs when $\theta(\phi_k, \phi_{-k}) = 1$ for $k = i, j$.

Due to the strategic liquidity provision, the following result holds.

[28] The simplification relies on the assumption that limit orders are "displayed." When the LOB is undisplayed, one HFM cannot learn v from her rival's update. Although this modification complicates the equilibrium equations, we can show that the following results stay (qualitatively) the same.

Figure 2.2: HFM i's profits



Note: This figure illustrates the expected profit of HFM i as a function of her quote s_i.

Lemma 2.3. *In the liquidity-provision stage, there exists no equilibrium in pure strategies.*

Proof. See discussions below. □

The result is analogous to Dennert (1993) and Baruch and Glosten (2013, 2019), and the mechanism is explained by Figures 2.2 and 2.3. Figure 2.2 draws HFM i's profit from market-making, $V_{M,i}$, as a function of her strategy, s_i. Figure 2.3 shows the best-response function of HFM k ($k = i, j$), denoted as s_k^*, to her rival's quote.

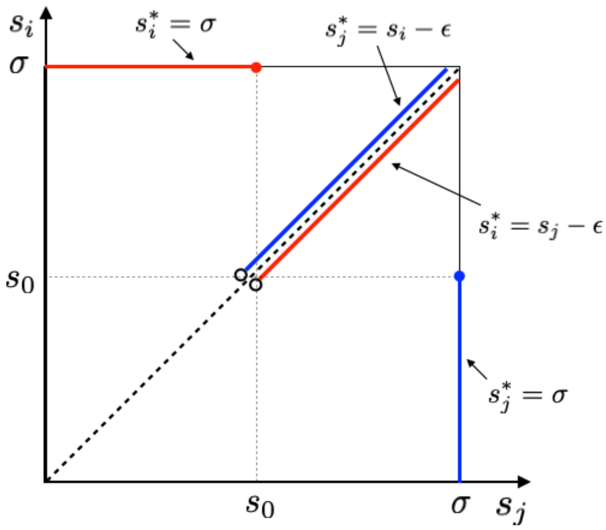
Figure 2.2 shows that it is optimal for HFM i to slightly undercut s_j as long as $s_j > s_0$ because a better price attracts a profitable order flow from the liquidity trader. Thus, in Figure 2.3, the best-response price of HFM i for $s_j > s_0$ is $s_i^* = s_j - \epsilon$ with $\epsilon \to +0$. Once s_j hits s_0, however, quoting $s_i \in [0, \sigma)$ generates negative profits. Since placing $s_i = \sigma$ always grantees zero profit, HFM i's best response price jumps to $s_i^* = \sigma$. Symmetric arguments provide the best response of HFM j, denoted as s_j^* in Figure 2.3.

Figure 2.3 shows that price competition between strategic HFMs does not result in equilibrium in pure strategies. This is because HFMs comprehend how prices (s_i, s_j) affect their profit and try to exploit discontinuity at $s_i = s_j$.

Hence, I turn to mixed strategy equilibrium.

Proposition 2.4. *(i) There is a unique mixed strategy equilibrium, in which HFM $k(= i, j)$ randomizes her quote s_k over the support $[s_0, \sigma]$ according to the following cumulative*

Figure 2.3: Best response and equilibrium



Note: This figure illustrates the optimal quote of each HFM as a function of her rival's quote.

distribution function.

$$F(s) \equiv \Pr(s_k \leq s) = 1 - \frac{1}{\beta}\frac{\phi_S}{1+\lambda(\delta)(\phi_M+\phi_S+\beta)}\frac{\sigma - s}{s}, \tag{2.13}$$

where ϕ_S is the sniper's speed, and s_0 is given by (2.12).
(ii) In the equilibrium, HFMs earn zero expected profits: $\mathbb{E}[V_{M,k}(s_k, s_{-k})] = 0$.
(iii) The expected spread posted by HFM k is given by

$$\bar{s}(\phi_S, \phi_M) = \mathbb{E}[s_k] = -\sigma\frac{1}{\beta}\frac{\phi_S}{1+\lambda(\delta)(\phi_M+\phi_S+\beta)}\log s_0. \tag{2.14}$$

(iv) The expected spread (2.14) is monotonically increasing and concave in ϕ_S and is decreasing in δ (with ϕ_S fixed).

Proof. (Derivation of the mixed strategy): Suppose that HFM i earns expected profit $V^*_{M,i}$ when HFM j randomizes quote s_j over $[\underline{s}_j, \sigma]$ with distribution F_j. The goal is to find three unknowns $(V^*_{M,i}, \underline{s}_j, F_j)$ that constitute a mixed strategy equilibrium. Firstly, θ in (2.11) is replaced by

$$\theta(s_i, F_j) = 1 - F_j(s_i) \tag{2.15}$$

leading to

$$V^*_{M,i} \propto \frac{\phi_S}{1+\lambda(\delta)(\phi_M+\phi_S+\beta)}(s_i - \sigma) + \underbrace{\beta\,(1 - F_j(s_i))}_{\text{Prob(trade w/ liq.trader)}}\; s_i.$$

For the mixed strategy to be an equilibrium, $V^*_{M,i}$ must be independent of s_i. Since $s_i = \sigma$ is feasible, we have $V^*_{M,i} = 0$. By solving the above equation with $V^*_{M,i} = 0$ for F_j, we obtain (2.13). Finally, the lower bound of the distribution $\underline{s}_j$ is derived from $F_j(\underline{s}_j) = 0$, leading to $\underline{s}_j = s_0$.

For uniqueness and other properties, see Appendix B.4 $\square$

Due to the mixed strategy, posted spreads are random. Thus, the average spread $\bar{s}$, which is increasing in sniper's speed (ϕ_S), works as the expected trading cost for HFTs in the speed acquisition stage.

2.3.2 Speed acquisition

Consider HFT i's speed acquisition problem before her role is determined. Since the indifference condition drives her expected profit from market-making zero, the *ex-ante* expected gains come only from sniping. Thus, in the symmetric equilibrium, she chooses ϕ_i for the case that she becomes an HFS and the other HFTs serve as HFMs with symmetric speed $\phi_j = \phi_l$ (for $j, l \neq i$). Accordingly, I denote $\phi_{-i} \equiv \phi_M = 2\phi_j$.

HFT i solves

$$\max_{\phi_i} W_i = \frac{2}{3}\pi(\phi_i, \phi_{-i}, \delta)\left(1 - \frac{\bar{s}(\phi_i, \phi_{-i})}{\sigma}\right) - C(\phi_i), \tag{2.16}$$

subject to $\bar{s}(\phi_i, \phi_{-i})$ in equation (2.14) and the sniping probability that is given by

$$\pi(\phi_i,\phi_{-i},\delta) = \left(\underbrace{\frac{1}{1+\delta\phi_{-i}}}_{=h(\delta,\phi_{-i})} \times \underbrace{\frac{\phi_i}{\phi_i + \phi_{-i} + \beta}}_{=\pi_{ND}(\phi_i,\phi_{-i})}\right)\frac{\sigma}{1+\frac{\beta}{z}}. \tag{2.17}$$

Note that the expected sniping probability is almost the same as that in the baseline model, while it contains the rival's speed ϕ_{-i} instead of γ (the exogenous speed of the market maker). Also, the profits from market making do not show up, as $\mathbb{E}\left[V_{M,i}(s_i, s_k)\right] = 0$ from Proposition 2.4. Since the coefficient does not qualitatively affect the result, I ignore $2/3$ on W_i in the following.

The FOC is given by

$$0 = \frac{dW_i}{d\phi_i} = \frac{\sigma}{1+\frac{\beta}{z}}h(\delta,\phi_{-i})\left(\frac{d\pi_{ND}}{d\phi_i}(1-\bar{s}) - \pi_{ND}\frac{d\bar{s}(\phi_i,\phi_{-i})}{d\phi_i}\right) - C'(\phi_i). \tag{2.18}$$

It captures the marginal benefit and cost of being faster, as in the benchmark model (see [2.5]). That is, a speed-up increases HFT i's profit by improving the probability of sniping, but it also harms her profit because it reduces her profit margin by widening the spread and by increasing the marginal cost of speed. I denote the solution of (2.18) as $\phi^*_i = BR_i(\phi_{-i})$ and call it the best response speed of HFT i to her rivals' speed ϕ_{-i}.

2.3.3 Crowding-in effect

As we have established in the previous section, the exogenous marginal cost C' drives the crowding-out effect of delays on speed acquisition: if $C' = 0$, only the crowding-in effect is at play. As the first step, the following discussion considers the case with $C' = 0$ to highlight the crowding-in effect and to obtain analytical solutions. Afterwards, I retrieve the positive marginal cost to obtain full characterizations.

Optimal speed Although the explicit formula for $\phi_i^* = BR_i$ is hard to obtain, the following proposition analytically shows that the crowding-in effect in the benchmark model is robust. Moreover, it uncovers the strategic nature of an arms race.

Proposition 2.5. *If the exogenous cost is constant ($C' = 0$);*
(i) An order execution delay has a positive impact on the best response function, i.e., $\frac{\partial BR_i(\phi_{-i})}{\partial \delta} > 0$.
(ii) The arms race exhibits strategic complementarity, i.e., $\frac{dBR_i(\phi_{-i})}{d\phi_{-i}} > 0$.

Proof. See Appendix B.4. □

Intuition for point (i) is the same as the crowding-in effect in the benchmark model (i.e., equation [2.6]). Moreover, point (ii) indicates that the rivals' speed also has the crowding-in effect on HFT i's speed acquisition. Intuitively, an increase in ϕ_{-i} makes HFMs less likely to be picked off by HFT i, leading to a narrower and less sensitive average spread. Thus, a speed-up by HFMs and an extension of a delay have (almost) the same impact on HFT i's speed acquisition, i.e., the crowding-in effect. As a result, the speed arms race involves strategic complementarity.[29]

2.3.3.1 Equilibrium speed

The solution of $\phi^* = BR(\phi^*)$ determines the speed level in the symmetric equilibrium, in which all HFTs take the same speed level, $\phi_i = \phi_{-i} = \phi^*$.

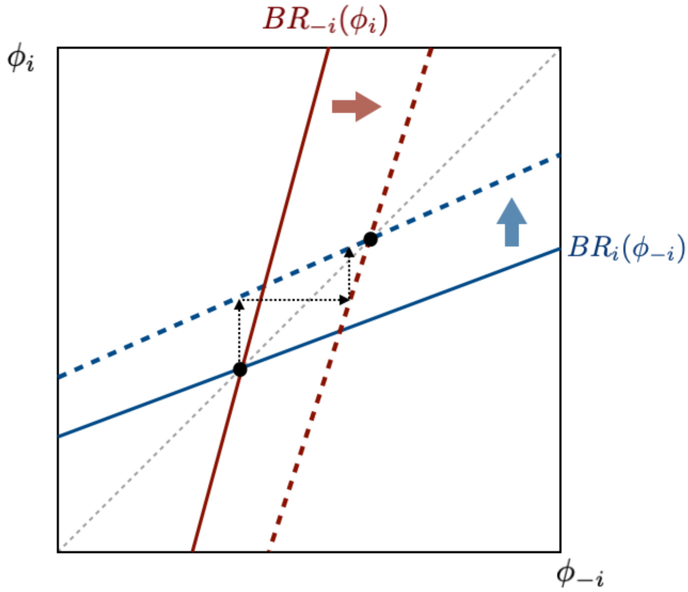
Proposition 2.6. *If the exogenous cost is constant ($C' = 0$), the equilibrium speed ϕ^* is monotonically increasing in the expected length of a delay, δ.*

Proof. See Appendix B.6. □

Figure 2.4 illustrates the symmetric equilibrium and the impact of increasing δ. As in the benchmark model, the crowding-in effect of a delay shifts BR_i upward. On top of that, the effect is amplified by the strategic complementarity, which is captured by the upward-sloping best response functions, generating a positive feedback loop, i.e., speed begets speed.

[29] See Appendix B.5 for more rigorous discussions and intuition regarding point (ii).

Figure 2.4: Impact of an extention of delays on BR



Note: This figure illustrates the best-response speed of HFTs. The dashed line in the middle represents the 45-degree line. The shifts from the solid lines to the upward-sloping dashed lines show the impact of increasing $\delta = \mathbb{E}[\tilde{\delta}]$. The equilibrium evolves from the lower black dot to the upper black dot according to the dotted arrows.

2.3.3.2 Market quality

Adverse selection. In the benchmark model with constant C, the irrelevance result holds ($\frac{ds}{d\delta} = 0$) because the single HFT can exploit her monopolistic power to set ϕ^*. By contrast, an arms race between multiple HFTs generates positive externality that amplifies the crowding-in effect. Thus, the equilibrium speed, ϕ^*, becomes substantially high and dominates the direct impact of δ on the spread.

Proposition 2.7. *If the effective exogenous cost is constant ($C' = 0$), an order execution delay widens the average spread, i.e., $\frac{d\bar{s}(\phi^*)}{d\delta} > 0$.*

Proof. See Appendix B.6. □

The introduction of an order execution delay or a longer expected delay can backfire, increasing the equilibrium speed of HFTs and worsening adverse selection. In the equilibrium, not only a sniper but also market makers increase their speed, responding to a longer delay. However, the average spread deteriorates because there is a liquidity trader whose speed is unchanged.[30]

This result highlights the fact that fast informed trading by liquidity takers harms liquidity provision whether liquidity providers are fast or slow. Since, in reality, the primary purpose of delaying order execution is to mitigate adverse selection, Proposition 2.7 indicates that it may lead to unfavorable consequences due to the crowding-in effect.

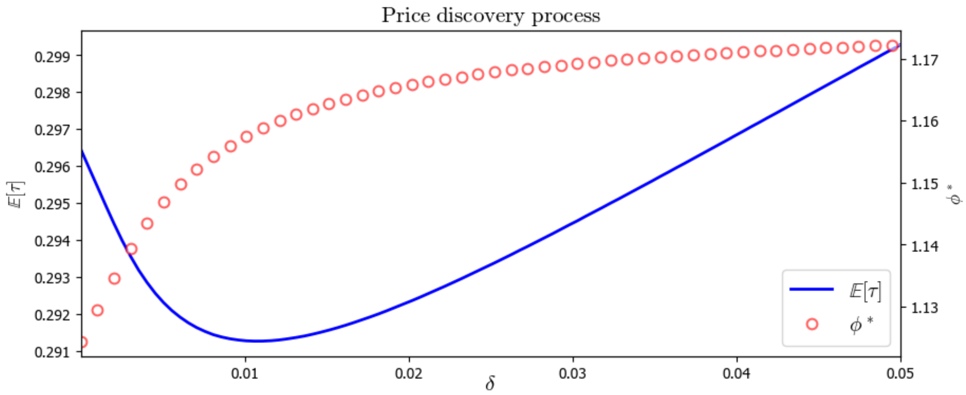
Moreover, what matters to the crowding-in effect and amplification is the fact that the market making is done by HFTs with endogenous ϕ_i. Also, if their role is fixed (e.g., HFT i always serves as an HFM), the amplification effect does not arise. Thus, as in the real markets, the fact that HFTs are serving both as snipers and market makers is another important factor to derive the above results.

Price discovery. How fast does the equilibrium price incorporate new information? For information to be reflected by the price, a stale quote must be removed from the market or repriced by a taker or a maker. It is triggered at stochastic time $\tau \equiv \min\{T^* + \tilde{\delta}, T^*, T^*\}$ with $T^* \sim \text{Exp}(\phi^*)$ where ϕ^* denotes the symmetric equilibrium speed. By exploiting the property of the exponential distribution, the expected length of time until price discovery from a jump in v is computed as

$$\mathbb{E}[\tau] = \frac{1}{3\phi^*}\left(1 + \frac{\delta\phi^*}{1 + 2\delta\phi^*}\right). \tag{2.19}$$

[30] The easiest way to understand why the existence of a liquidity trader (or a slow trader, in general) leads to $\frac{d\bar{s}}{d\varepsilon} > 0$ is to consider two extreme cases. One the one hand, when $\phi^* \to 0$ (i.e., both the HFS and HFMs do not invest in speed), HFMs are almost sure that an arriving market order is sent from the liquidity trader, so that they post $s^* = 0$. On the other hand, when $\phi^* \to \infty$ (i.e., both the HFS and HFMs have very fast speed technologies), market makers are almost sure that the liquidity trader falls behind the HFS, leading to a strictly positive spread. Since the impact of ϕ^* on the spread is monotonic, the global argument above can be applied to analyze the impact of increasing ϕ^* for $\phi^* \in (0, \infty)$.

Figure 2.5: Expected time until price discovery



Note: This figure illustrates the price discovery measure against the length of delays by using parameters value $\beta = z = 1.00$.

Figure 2.5 plots $\mathbb{E}[\tau]$ against $\delta = \mathbb{E}[\tilde{\delta}]$ in comparison with ϕ^*. Initially, a longer delay remarkably expedites the price discovery process, while it starts to slow down as δ becomes longer. On the one hand, a delay crowds in the speed level (ϕ^*). Therefore, the price can impound information more swiftly, leading to a lower $\mathbb{E}[\tau]$. On the other hand, it hampers the price discovery process because an intentional delay directly restricts the sniper's trading behavior. As a result, $\mathbb{E}[\tau]$ takes a U-shaped curve.

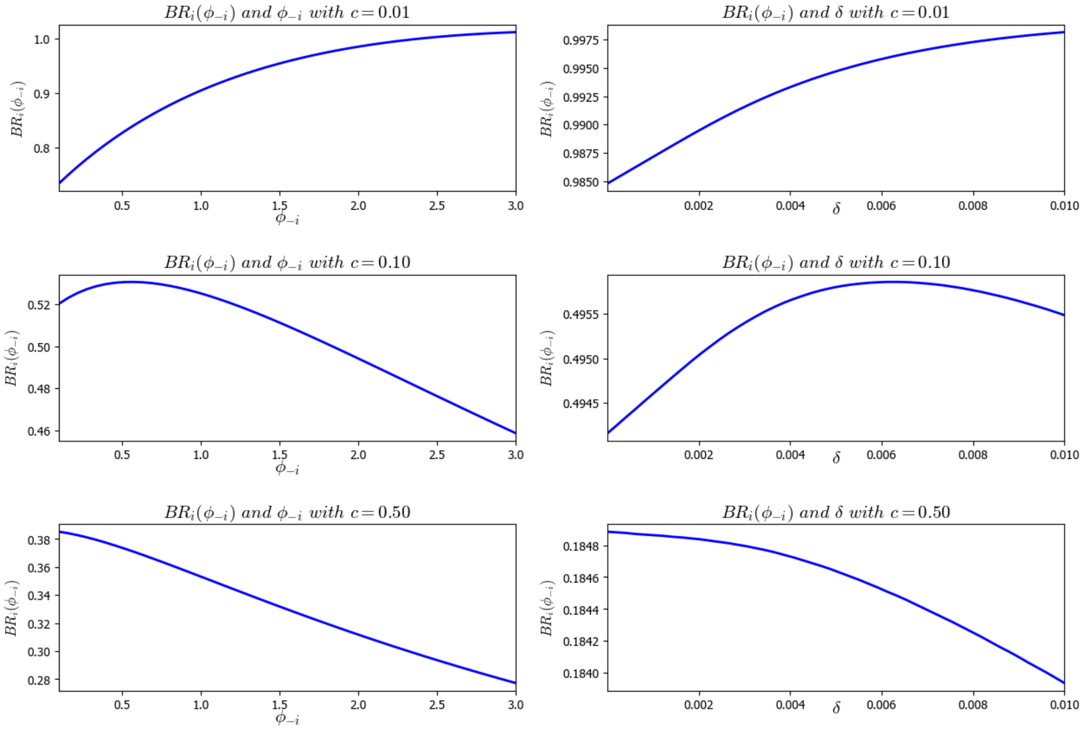
2.3.4 Introducing crowding-out effect

As in the benchmark model, $C' > 0$ is the source of the crowding-out effect (see [2.6]). Thus, I retrieve the increasing cost with $C(\phi) = \frac{c}{2}\phi^2$ for illustrative purposes. In the following, I deal with different values of c to describe the crowding-in versus the crowding-out effects, but the similar results hold when I change other parameters, such as the frequency of jumps in asset value (z), the size of a jump (σ), and the arrival rate of the LT (β).

Figure 2.6 illustrates the best response function (BR_i) with different c. The left column shows that an arms race involves strategic complementarity ($\frac{dBR_i}{d\phi_{-i}} > 0$) when c is small, while strategic substitution arises as c increases. The best response function becomes hump-shaped if c is intermediate, as these two competing effects offset each other. The right column establishes that analogous results hold for the reaction of BR_i to the expected length of delays, $\delta = \mathbb{E}[\tilde{\delta}]$. Figures exhibit the crowding-in effect versus the crowding-out effect, where an increase in c magnifies the crowding-out effect relative to the crowding-in effect.

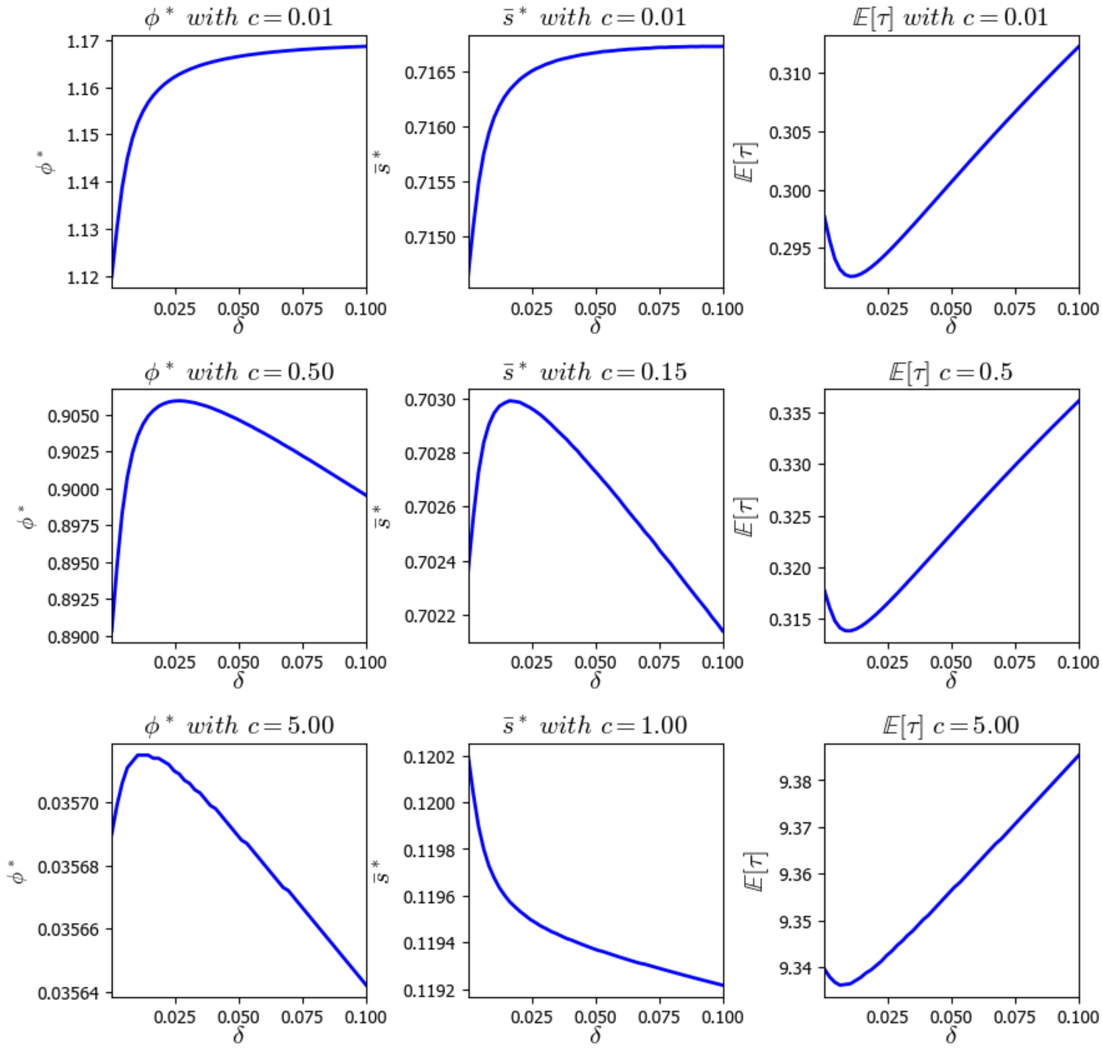
Figure 2.7 shows the impact of a longer expected delay ($\delta = \mathbb{E}[\tilde{\delta}]$) on the equilibrium speed ϕ^* (the left column), the expected spread $\bar{s}^* = \bar{s}(\phi^*, \phi^*)$ (the middle column), and

Figure 2.6: Best response function: $BR_i(\phi_{-i})$



Note: This figure is illustrated by using the following parameter value. $\beta = z = 1.0$, $\delta = 10^{-3}$ (the left column), and $\phi_{-i} = 1.0$ (the right column).

Figure 2.7: Equilibrium speed, spread, and price discovery



Note: This figure is illustrated by using the following parameter value: $\beta = z = 1.0$.

the price discovery process $\mathbb{E}[\tau]$ defined by (2.19) (the right column).

The case with small c follows the analytical result in Propositions 2.6 and 2.7, i.e., ϕ^* and $\bar{s}^*$ are increasing in δ. If c is relatively large, a higher δ reduces the sniping probability π, and the exogenous cost C stands out more. Thus, it tries to hamper the HFTs' incentive to be faster, leading to competition between the crowding-out effect and the crowding-in effect. With intermediate c, the reaction of the equilibrium speed becomes hump-shaped.

The reaction of the expected spread ($\bar{s}^*$) to the expected length of delay (δ) in the middle column shows that a higher ϕ^* does not necessarily promote $\bar{s}^*$ because the negative direct impact of δ on $\bar{s}^*$ is also at play (see equation [2.9]). Thus, the spread becomes hump-shaped with intermediate c and monotonically decreasing in δ if c becomes sufficiently large. Finally, the right column indicates that the U-shaped configuration of the price discovery process, measured by $\mathbb{E}[\tau]$, survives even if the exogenous cost is introduced.

Overall, considering the crowding-in effect with multiple HFTs generates the non-monotonic reaction of equilibrium variables to the mean length of order execution delays due to the crowding-in effect versus the crowding-out effect. The expected spread can deteriorate because the arms race between multiple HFTs involves complementarity and amplifies the original effect of order execution delays.

2.4 Welfare

To conduct welfare analyses, I define the utility of the liquidity trader (LT). Remember that the liquidity shock makes the LT need to buy or sell one unit of asset. I assume that the LT obtains the following net utility if she trades at prices $p = (ask, bid)$.

$$u(v,p) = \begin{cases} \alpha + v - ask & \text{if buys at ask price } ask \\ \alpha + bid - v & \text{if sells at bid price } bid. \end{cases} \tag{2.20}$$

α represents net private utility from trading. It could be some exogenous private value of trading that stems from hedging motives, margin constrains, borrowing/lending needs, and so on. In what follows, I set $\alpha > \sigma$ to avoid market breakdowns.

Since the arrival frequency of the LT and her trading size are the same as the previous section, so are the equilibrium spread and the optimal speed of the HFTs. In the following, I derive LT welfare by using the multiple-HFT model in Section 2.3 but the results for the single-HFT model can be derived by the symmetric argument. Once again, I focus on the case with $v = +\sigma$ (i.e., a positive jump in asset value).

2.4.1 Ex-ante expected utility of liquidity trader

Suppose that HFTs i and j are serving as HFMs. I denote a set of the *best* ask and bid prices on the top of the LOB at time t as $p_t = (s_{min}, -s_{min})$. Before the HFMs update their quote (either by learning private news or trading with the HFS), the best ask price is $s_{min} = \min\{s_i, s_j\}$, where s_i is the random spread that follows cdf F in equation (2.13). After the HFMs learn v, it becomes $p_t = (v, v)$.

The expected trading utility of the liquidity trader (when HFTs i and j are HFMs) is given by

$$W_{LT}^{i,j} = \mathbb{E}_Y \left[\int_0^\infty \beta e^{-\beta T_L} u(v, p_{T_L+\tilde{\delta}}) dT_L \right]. \tag{2.21}$$

$\mathbb{E}_Y$ denotes the expectation operator regarding a set of random variables $Y = (T_S, T_M, t_0, \tilde{\delta}, p)$.

Firstly, if the liquidity shock hits before the asset's value jumps ($T_L < t_0$), the prices on the LOB at T_L are not updated yet. However, due to a delay, the order is executed either at $p_{T_L+\delta} = (s_{min}, -s_{min})$ or (v, v), depending on the timing of market maker's learning. By incorporating this uncertainty, it holds that

$$\begin{aligned}
\mathbb{E}_Y \left[\int_0^{t_0} \beta e^{-\beta T_L} u(p_{T_L+\tilde{\delta}}) dT_L \right] &= (\alpha - \bar{s}_{min}) \overbrace{\mathbb{E}_{t_0,\tilde{\delta}} \left[\int_0^{t_0} \beta e^{-\beta T_L} e^{-\phi_M (T_L + \tilde{\delta} - t_0) \mathbb{I}_{\{\tilde{\delta} > t_0 - T_L\}}} dT_L \right]}^{\text{Pr(trade before MM learns } v)} \\
&\quad + \alpha \overbrace{\mathbb{E}_{t_0,\tilde{\delta}} \left[\int_0^{t_0} \beta e^{-\beta T_L} (1 - e^{-\phi_M (T_L + \tilde{\delta} - t_0) \mathbb{I}_{\{\tilde{\delta} > t_0 - T_L\}}}) dT_L \right]}^{\text{Pr(trade after MM learns } v)} \\
&= \frac{\beta}{z + \beta} \left[\alpha - \bar{s}_{min} \left(1 - \frac{\phi_M \delta}{1 + \phi_M \delta} \frac{\delta}{\delta + z^{-1}} \right) \right],
\end{aligned} \tag{2.22}$$

where $\bar{s}_{min} = \mathbb{E}_F \left[\min\{s_i, s_j\} \right]$ denotes the expected best ask price. The coefficient of the brackets amounts to the probability of $T_L < t_0$. Inside the brackets, the coefficient on $\bar{s}_{min}$ represents the probability that the LT's order is executed at prices $(s_{min}, -s_{min})$ before the market maker learns v.

As the second case, suppose that the liquidity shock hits after the jump in asset value ($T_L \geq t_0$). In this case, there are two scenarios: (i) the LOB at time T_L still displays $p_{T_L} = p_{min} \equiv (s_{min}, -s_{min})$ or (ii) it has already updated to $p_{T_L} = p_v \equiv (v, v)$. In case (i), the expected utility is either $\alpha - s_{top}$ or α, while case (ii) yields α for sure. Incorporating these possibilities, the expected utility is

$$\begin{aligned}
\mathbb{E}_Y \left[\int_{t_0}^\infty \beta e^{-\beta T_L} u(p_{T_L+\tilde{\delta}}) dT_L \right] =& \mathbb{E}_{t_0,\tilde{\delta}} \left[\int_{t_0}^\infty \beta e^{-\beta T_L} \Pr(p_{T_L} = p_{min}) \right. \\
& \left. \times \left(\alpha - \bar{s}_{min} \Pr(p_{T_L+\tilde{\delta}} = p_{min} | p_{T_L} = p_{min}) \right) \right] \qquad (2.23) \\
& + \alpha \mathbb{E}_{t_0,\tilde{\delta}} \left[\Pr(p_{T_L+\tilde{\delta}} = p_v) \right] \\
=& \frac{z}{z + \beta} \left(\alpha - \bar{s}_{min} \frac{1}{1 + \phi_M \delta} \frac{\beta}{\beta + \phi_M + \phi_S} \right). \qquad (2.24)
\end{aligned}$$

The first coefficient represents the probability that $T_L \geq t_0$ holds. As the first term in parentheses shows, the LT obtains α in any cases. The second term shows the case that the LT observes $p_{T_L} = p_{min} = (s_{min}, -s_{min})$ upon hit by the shock and executes her trade at these prices.

Since the selection of HFTs' role is random, equations (2.22) and (2.24) imply the following result.

Proposition 2.8. *Let ϕ^* denote the symmetric equilibrium speed of HFTs. LT welfare is given by*

$$W_{LT} \equiv \mathbb{E}_{i,j}[W_{LT}^{i,j}] = \alpha - w_s(\delta, \phi^*)\bar{s}_{min}, \tag{2.25}$$

where

$$\bar{s}_{min} = \mathbb{E}_F\left[\min\{s_i, s_j\}\right],$$

the random spread is such that $s_k \overset{iid}{\sim} F$ in (2.13) with $\phi_S = \frac{1}{2}\phi_M = \phi^*$, and

$$w_s(\delta, \phi^*) = \frac{\beta}{z+\beta}\left(1 - \frac{2\phi^*\delta}{1+2\phi^*\delta}\frac{\delta}{\delta+z^{-1}}\right) + \frac{z}{z+\beta}\frac{\beta}{(1+2\phi^*\delta)(\beta+3\phi^*)}.$$

w_s represents the probability that the LT trades at pre-learning prices $(s_{min}, -s_{min})$ and pays the positive spread $s_{min} > 0$. From (2.25), the best spread has a negative direct impact on LT welfare because it increases the trading cost for the LT. In contrast, all else equal, a faster HFT (ϕ^*) and a longer delay (δ) improve the LT's utility by reducing w_s. A faster HFT promotes the price discovery process and allows the LT to trade at updated prices with zero spread. The direct impact of an intentional delay is also positive because it delays the LT's order execution and makes it more likely to fall behind the market maker's learning.

In the single-HFT model, the following analytical result holds.

Corollary 2.1. *(i) In the single-HFT model, $W_{LT} = \alpha - w_s s$ where*

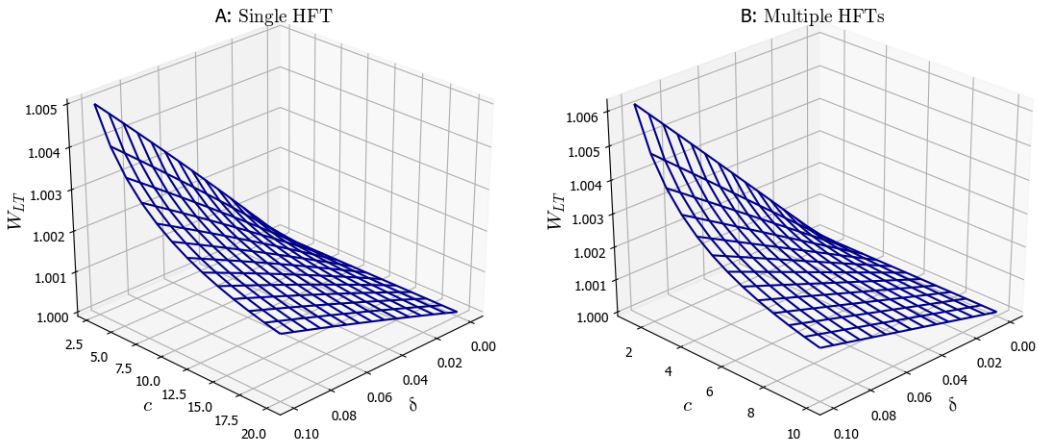
$$w_s = \frac{\beta}{z+\beta}\left(1 - \frac{\gamma\delta}{1+\gamma\delta}\frac{\delta}{\delta+z^{-1}}\right) + \frac{z\beta}{(z+\beta)(1+\gamma\delta)(\beta+\gamma+\phi^*)},$$

and s is the equilibrium spread in (2.2).
(ii) LT welfare is increasing in the expected length of a delay when C' and β are small, and z and σ are large.

When the *effective* marginal cost is small, the crowding-in effect of delays dominates the crowding-out effect. Thus, the HFT's speed is an increasing function of the length of delays. We also know from Proposition 2.3 that the spread is decreasing in δ. Therefore, the imposition of intentional delays improves LT welfare.

Figure 2.8 plots the reaction of LT welfare to δ for a wide range of c, showing that the result in Corollary 2.1 is robust in the multiple-HFT model. The figures establish that the positive impact of a delay on W_{LT} is robust even if the crowding-out effect is dominant. Also, the multiple-HFT model makes the average spread increasing in δ when c is relatively small (Proposition 2.7). Even though a wider average spread increases the trading cost for the LT, it cannot decrease W_{LT} (the right panel). This is because faster

Figure 2.8: Impact of delays on LT welfare



Note: The figures are illustrated with $\beta = \gamma = 0.5, z = 0.1, a = \sigma = 1.0$

HFTs make the LT more likely to trade at the updated prices (v, v) rather than the pre-learning best prices $(s_{min}, -s_{min})$, i.e., an HFS's sniping and HFMs' learning are more likely to occur before the LT's arrival. Thus, the negative impact of a wider spread on W_{LT} is limited.

Overall, the reaction of the LT's indirect utility to the imposition of order execution delays can be analyzed through the lens of the tradeoff between adverse selection and price discovery. The tradeoff has been a central issue and an open question in the context of informed trading, market quality, and trader welfare (see Glosten and Putnins, 2016; Rosu, 2018). The above discussions show that the benefit of a faster price discovery process dominates the negative impact of adverse selection in terms of LT welfare.

2.4.2 Aggregate welfare

Aggregate welfare is defined by

$$W = 3W_{HFT} + W_{LT},$$

where W_{HFT} represents individual HFT welfare, i.e.,

$$W_{HFT} = \frac{2}{3}\frac{1}{1+\frac{\beta}{z}}\underbrace{\frac{\phi^*}{3\phi^*+\beta}\frac{1}{1+2\delta\phi^*}}_{=\pi(\phi^*,\delta)}(\sigma - \bar{s}) - C(\phi^*). \tag{2.26}$$

W_{HFT} is identical to equation (2.16) in Section 2.3.

Note that the equilibrium spread s_i with the mixed strategy F satisfies the break-even condition in expectation. Specifically, if HFT i becomes an HFM, her expected profit in the liquidity provision stage (given s_i) is rewritten as

$$\begin{aligned} 0 &= \mathbb{E}_{s_j}[V_{M,i}(s_i, s_j)] \\ &= \frac{1}{1+\frac{\beta}{z}}\pi(\phi^*, \delta)(s_i - \sigma) + \mathbb{E}_{s_j}[\mathbb{I}_{s_i<s_j} s_i] w_s(\delta, \phi^*). \end{aligned}$$

By taking average over s_i and summing up to the symmetric equation for HFT j, it holds that

$$\begin{aligned} 0 &= 2\frac{1}{1+\frac{\beta}{z}}\pi(\phi^*, \delta)(\bar{s} - \sigma) + \mathbb{E}_F[s_{min}] w_s(\delta, \phi^*) \\ &= -3\left[W_{HFT} + C(\phi^*)\right] + (\alpha - W_{LT}). \end{aligned} \tag{2.27}$$

The first two terms in the brackets show the adverse selection cost that market makers bear due to the informed trading, and it corresponds to the (gross) sniping profit of the HFS. The last term is market makers' (positive) profit from trading with the LT before they learn v. This term captures the LT's trading cost when her order is executed at the pre-learning best spread s_{min}. The break-even condition implies that market makers set the spread s so that the adverse selection cost is perfectly compensated by the revenue from trading with the LT.

From Lemma 2.8 and equation (2.27), we have the following result:

Proposition 2.9. *(i) In the multiple-HFTs case, aggregate trader welfare is given by*

$$W = \alpha - 3C(\phi^*). \tag{2.28}$$

(ii) When the crowding-in (resp. crowding-out) effect of order execution delays on ϕ^* is dominant, aggregate welfare is decreasing (resp. increasing) in the length of delays.

Firstly, equation (2.28) shows that the equilibrium (pre-learning) spread s has no impact on aggregate trader welfare. This is due to the break-even condition of market makers: the spread is just a transfer of money from the LT to an HFS. What remains after the transfer is the LT's private utility from trading (α) and the exogenous realized cost of the speed investment (C). As a result, the behavior of W is perfectly traced by analyzing the HFTs' optimal speed, as suggested by point (ii) in Proposition 2.9. When the parameter values make the crowding-in effect dominant, order execution delays can harm aggregate trader welfare, and vice versa. Also, the same result as Proposition 2.9 holds in the single-HFT model:

Corollary 2.2. *In the single-HFT model, the result in Proposition 2.9 holds with $W = \alpha - C(\phi^*)$.*

Table 2.2: Impact of imposing a longer delay

	Varibles	Effective marginal cost small	intermediate	large
Single HFT	HFT speed	increasing*	ambiguous*	decreasing*
	Adverse selection (spread)	mitigated*	mitigated*	mitigated*
	Price discovery	faster	ambiguous	slower
	LT welfare	increasing*	increasing*	increasing
	Aggregate welfare	decreasing*	ambiguous*	increasing*
Multiple HFTs	HFT speed	increasing*	ambiguous*	decreasing*
	Adverse selection (spread)	worse*	ambiguous	mitigated
	Price discovery	faster	ambiguous	slower
	LT welfare	increasing	increasing	increasing
	Aggregate welfare	decreasing*	ambiguous*	increasing*

* Note: This table tabulates the reaction of equilibrium variables to the introduction of order execution delays or their extension. * represents analytical results. Variables with no * report numerical results.

2.4.3 Welfare implication of order execution delay

The literature on the slow market structure has sought a solution for the problem of the optimal length of delays. However, there is no consensus on the objective of the imposition of delays. For example, Brolley and Cimon (2017) analyze speed bumps that maximize aggregate trader welfare or trading volume.[31] Baldauf and Mollner (2017) take the efficient frontier of information production and LT welfare as their objective function, where LT welfare is perfectly determined by the equilibrium spread (i.e., $w_s = 1$ in my model). Moreover, in reality, most exchanges with speed bumps and other delays claim that their primary purpose is to protect liquidity providers, meaning that the adverse selection cost for market makers is the target variable.

Table 2.2 summarizes the impact of a longer expected delay on different target variables when the *effective* marginal cost of speed is large, intermediate, and small. The top half of the table shows the results with a single HFT, and the bottom half shows those with multiple HFTs. In the literature, delaying order execution of high-frequency traders always discourages their speed acquisition (i.e., only the crowding-out effect has been highlighted). If the crowding-out effect is the only mechanism at play, the impact of delays is fully analyzed by focusing on the right column of the table (relatively large

[31]Brolley and Cimon (2017) argue that trading volume is the objective function of an exchange, as it can be seen as the proxy for the exchange's fee revenue. In contrast, Budish, Lee, and Shim (2018) analyze the impact of FBA on HFTs' demand for speed because exchanges earn a large portion of profits from fees for speed services.

effective marginal costs). In this situation, adopting order execution delays (or extending the length of delay) may improve most of the target variables. For example, a speed bump slows down HFTs, reduces the investment into speed technologies, improves both LT and aggregate welfare, and protects liquidity providers by mitigating the adverse selection problem.

In contrast, my model points out that delays in order execution have the crowding-in effect on HFTs' speed acquisition. Due to this new and opposing effect, the first two columns (small and intermediate effective costs) must be added to the table. In particular, when the effective marginal cost is relatively small, a longer delay increases the speed of HFTs, reduces aggregate welfare, and worsens the adverse selection problem. The only variable that shows a positive reaction is LT welfare, causing a tradeoff between these objective functions.

Therefore, due to the crowding-in effect of delays, normative discussions on the optimal length of delays become more nuanced. Not only does it require policy makers or exchanges to tradeoff between the above objective variables, but it also makes measuring parameter values for the effective speed cost important. In conclusion, the crowding-in effect can be seen as the dark side of delaying order execution because my model suggests that it can harm liquidity providers (i.e., the adverse selection problem), in contrast to its intended purpose, as well as aggregate trader welfare.

2.5 Discussion

2.5.1 Deterministic delay in order execution

In reality, asymmetric speed bumps in operation are all random, while several exchanges have proposed asymmetric and deterministic speed bumps. A deterministic delay can be analyzed by slightly modifying the benchmark model with a random delay.

As in the benchmark model, a deterministic delay of length $\tilde{\delta}$ is imposed on all market orders, but it is constant $\tilde{\delta} = \bar{\delta} \geq 0$. By replacing a random delay in the benchmark model by constant $\bar{\delta}$, the break-even spread (denoted as s_{det}) is

$$s_{det} = \frac{\frac{\phi}{1+\lambda(\bar{\delta})(\phi+\beta+\gamma)}}{\frac{\phi}{1+\lambda(\bar{\delta})(\phi+\beta+\gamma)} + \beta}\sigma$$

with

$$\lambda(\bar{\delta}) = \frac{z}{z-\gamma}\left(z - \gamma e^{-(z-\gamma)\bar{\delta}}\right). \tag{2.29}$$

Note that λ is monotonically increasing in $\bar{\delta}$. Moreover, the objective function of the HFT is

$$\phi^* = \arg\max_{\phi} e^{-\gamma\bar{\delta}} \frac{\phi}{\phi+\beta+\gamma}(1 - \frac{s_{det}}{\sigma}) - C(\phi).$$

Thus, by denoting $h(\bar{\delta}) = e^{-\gamma\bar{\delta}}$ and using λ in (2.29), the model becomes identical to the benchmark model with a random delay. The identity confirms the robustness of my main arguments in the previous sections.[32]

2.5.2 Will for-profit exchanges introduce delays in order execution?

One of the normative discussions is related to the question "Will the market fix the market?" put forth by Budish, Lee, and Shim (2018) in the context of frequent batch auctions (FBA). They analyze whether competing exchanges have an incentive to introduce an FBA that aims to mitigate adverse selection caused by HFTs.

Exchange platforms, in reality, earn a large portion of their profits from fees for fast data/exchange access, such as direct data feeds. Budish, Lee, and Shim (2018) argue that exchanges do not have an incentive to introduce FBA.[33] Since their model does not have the crowding-in effect, introducing FBA always invalidates the speed advantage of HFTs. Then, HFTs stop purchasing fast data/connection services from exchanges, undermining exchanges' profits from speed fees. This pessimistic result holds because the previous studies focus on the crowding-out effect of FBA on speed acquisition and its beneficial impact on adverse selection.

Although adopting multiple exchanges and competition between them is beyond the scope of my model, incorporating the crowding-in effect of an intentional delay into the above question can provide a new implication. Specifically, the following claim is derived from my model:

Claim 2.1. *With appropriate design of delays (e.g., value of $\delta = \mathbb{E}[\tilde{\delta}]$) and the schedule of a speed fee (C'), order execution delays promote speed acquisition by HFTs and mitigate the adverse selection problem at the same time.*

One of the examples comes from Propositions 2.2 and 2.3 in the single-HFT model that shows

$$\exists(\delta, c) \text{ s.t., } \frac{d\phi^*}{d\delta} > 0 \text{ and } \frac{ds^*}{d\delta} < 0.$$

There is a region of δ and c such that the crowding-in effect dominates the crowding-out effect, leading to a faster HFT ($\frac{d\phi^*}{d\delta} > 0$), whereas the direct effect of a delay on the spread

[32] Appendix A discusses an advantage of random speed bumps over deterministic ones. Briefly, a random delay makes it harder to synchronize the arrival of "sprayed" orders across multiple exchanges, allowing market makers to learn from execution of a part of sprayed orders in other exchanges.

[33] Budish, Lee, and Shim (2018) show that, in equilibrium, exchanges charge positive fees on speed but no fees on transaction services. Such an equilibrium arises due to the trading rule in the US, such as RegNMS and UTP. On the one hand, trading services are homogeneous goods across exchanges, meaning that exchanges are fungible in terms of trading. On the other hand, traders need to purchase exchange specific speed technologies (ESST), making an exchange a monopolistic supplier of her ESST. In this situation, even if an exchange who deviates and adopts FBA earns profits by increasing trading fees, the profit margin easily disappears since other competing exchanges can imitate the FBA structure, resulting in lower profits compared to the status-quo structure.

dominates the speed-up by the HFT ($\frac{ds^*}{d\delta} < 0$). Such parameter regions also exist in the general models with multiple HFTs (e.g., see Figure 2.7 in Section 2.3).

As long as parameters (δ, c) satisfy the condition in Claim 1, an exchange can charge a higher price for speed by delaying order execution, as HFTs try to compensate for the loss from a delay and demand more speed services. Simultaneously, the exchange can achieve its intended goal, i.e., alleviating the adverse selection problem. Of course, we need to formalize the environment and the objective function of exchanges, but the above claim leads to the following prediction:

Conjecture 2.1. *Suppose that an exchange's objective function consists of (i) maximizing fee revenues paid by HFTs for fast access to the exchange and (ii) alleviating the adverse selection problem. Then, an exchange has an incentive to adopt order execution delays with appropriate design.*

My model's crowding-in effect shows that a slow market structure does not always diminish HFTs' demand for speed. It can be seen as a dark side of delaying order execution *per se* but proposes a more optimistic prediction once we consider exchanges' incentive: they will introduce a new market design that contributes to creating a level playing field.

2.5.3 Cost of speed and asset volatility

Discussions so far indicate that the *effective* marginal exogenous cost of speed $C'(\phi)$ is one of the important determinants of the effectiveness of delays. In reality, typical high-frequency trading firms pay separate (monthly) subscription fees for physical and logical connectivity to an exchange, an additional connectivity charge to receive market data through acquired connectivity, and for the content of market data itself.

My model provides some testable predictions, as the impact of a delay can be substantially different depending on the marginal cost of speed.

Conjecture 2.2. *All else equal, if the exogenous marginal cost of increasing speed is low, introduction of a delay or its extension is accompanied by an increase in spread (and vice versa).*

However, high-frequency trading firms are notoriously secretive regarding their costs, and measuring the marginal cost of speed is not easy. Toward a remedy of this issue, considering the *effective* cost may provide an alternative prediction.

My model restricts HFTs' trading volume and assumes that the asset's volatility, σ, and the frequency of the Poisson jumps in asset value, z, are fixed. These parameters have been measured in empirical literature, such as Aquilina, Budish, and O'Neill (2020). Variations in these parameters shrink or magnify the cost of speed per sniping opportunity, leading to different consequences.

For example, Virtu Financial, Inc., the one of the largest HFT firms, reported spending approximately \$209 million on communication and data processing and \$383 million on employee compensation and payroll taxes in 2019.[34] Although the aggregate cost looks large, it represents the *annual* amount. In contrast, according to the analysis by Laughlin (2014) and Bloomberg, Virtu conducts 2.5-3.5 million trades *per day* across all venues and all asset classes in the United States. The large number of transactions dwarfs the above operational cost, making the cost per transaction amount to \$0.67-\$0.94. Of course, the above back-of-envelope calculation cannot separate the cost for sniping from that for market making. Yet, what matters is the fact that a more frequent arrival of arbitrage opportunities reduces the cost per sniping and leads to a strong crowding-in effect (and vice versa). This discussion is translated into a change in z (i.e., the frequency of a Poisson jump in asset value) that negatively affects the effective cost of speed.

Conjecture 2.3. *All else equal, if the volatility of asset (measured by the frequency of a jump in asset value z and its size σ) is high, delaying order execution is accompanied by an increase in spreads (and vice versa).*

2.5.4 Application to general information design

My result can fit into the discussion on more general information regulation in financial markets. For example, one of the closest market structures to a speed bump is "Last Look" in foreign exchange markets (FX): after a liquidity taker trades on a market maker's posted price, the market maker obtains a small "latency buffer" in which she has an option to reject the trade. The design of the Last Look options is custom-tailored and differs across all Electronic Communication Networks and OTC (Over-the-Counter) markets. Thus, it is not as standardized as speed bumps. However, the primary goal is to mitigate adverse selection and other structural frictions for market makers that stem from the speed/informational advantage of liquidity takers in the highly fragmented FX markets. A couple of theoretical papers have studied the Last Look feature (e.g., Oomen, 2017; Cartea, Jaimungal, and Walton, 2019), but the speed acquisition by liquidity takers is not taken into account.

In general, financial markets have been experienced many forms of informational regulation that try to mitigate information asymmetry between traders, such as Regulation Fair Disclosure (2000), Sarbanes–Oxley Act (2002), and Dodd-Franc Act (2010). The theoretical literature has argued that attempts to mitigate asymmetric information by disseminating public information can hammer private information acquisition by potential informed traders (i.e., the crowding-out effect; Verrecchia, 1982).[35] If the crowding-out effect is the only force at play, fair disclosure policies unambiguously mitigate asymmetric information and tighten the spread. However, "the empirical evidence [...] is still inconclusive" (Beyer et al., 2010). Even though my model focuses on speed acquisition, the mechanism for the crowding-in effect should survive in a general regulatory framework.

[34]The expenditure in 2019 is the largest in these past 8 years. Virtu's SEC filings are available at https://ir.virtu.com/financials-and-filings/sec-filings/default.aspx.

[35]The studies that propose the crowding-out effect include Diamond (1985), Gao and Liang (2013), Colombo, Femminis, and Pavan (2014), and Banerjee, Davis, and Gondhi (2018).

For example, in the benchmark model, parameter ϕ governs the learning process of the HFT. In the context of fair disclosure, we can re-interpret ϕ (or $\phi/(\phi+\beta)$) as the measure of traders with perfect private information about v. Also, parameter γ captures the arrival frequency of public news—fair disclosure—that tells all traders v. An intentional delay δ makes public disclosure more likely to happen before informed traders may take liquidity. Therefore, δ can be seen as the regulation that requests and promotes the swift dissemination of public news.

Following the discussion in Section 2.2, the regulation generates an additional opportunity for the informed trader to exploit her informational advantage. As a result, fair disclosure may promote traders' information acquisition activity, i.e., the crowding-in effect. Although the formal analyses are left as a future research topic, my model predicts that the crowding-in effect should exist as long as regulations try to mitigate asymmetric information, whereas traders can strategically decide on the quality (speed and precision) of their private information.

2.6 Conclusion

Order execution delays, such as speed bumps, have been widely adopted in many exchanges as one of the simplest and easiest ways to slow down fast informed trading. The purpose of a delay is to slow down HFT and mitigate adverse selection for liquidity providers by imposing an intentional delay on HFTs' order execution. In contrast to its intended purpose, however, my model shows that a longer delay can facilitate the equilibrium speed acquisition of HFTs, i.e., it has a *crowding-in* effect. The crowding-in effect competes against a delay's intended effect, and the equilibrium speed acquisition and adverse selection exhibit rich reactions to a longer delay. Thus, this chapter complements the literature by theoretically showing that order execution delays can have a positive, negative, and ambiguous impact on the market.

This chapter describes how fast HFTs obtain private information and act on it. Although this follows the literature and provides a tractable framework, it ignores another dimension of information acquisition, namely, the precision of a signal. Several studies, such as Huang and Yueshen (2018) and Dugast and Foucault (2018), argue that the precision of private information and the information processing speed may have a negative relationship. Hence, analyzing the impact of a delay on information acquisition, both in terms of speed and precision, would be a good extension of my model to capture a more holistic view on the impact of an exogenous speed restriction.

This chapter discusses the optimal length of delays but does not analyze the market design problem, i.e., the optimal *structure* of delays to improve market quality or trader welfare. The question regarding the optimal design is important, as we have seen a large variety of speed bumps in reality. For example, A-NEO in Canada classifies traders into HFTs and non-HFTs by using their IDs and latency-(in)sensitive trading behavior and imposes speed bumps only on HFTs. Some other exchange platforms have proposed randomizing the distribution of a speed bump by using AI to prevent HFTs from forecasting the pattern of delays. Also, the model's prediction regarding exchanges' incentive to adopt a speed bump is optimistic, and modeling the exchange competition with a speed bump is another topic of the future research.

Chapter 3

A Model of a Decentralized Exchange with Constant Product Market Makers

3.1 Introduction

The limit order book has been a core trading mechanism in the modern electronic financial market. Traders called market makers provide trading opportunities by placing *limit orders*—a price quote at which they are willing to buy or sell a certain amount of an asset. Limit orders are stored on a book (i.e., the limit order book, LOB) and publicly displayed. If other traders find a good deal on the book, they try to consume the trading opportunities by placing marketable limit orders or *market orders*. An incoming market order is matched with standing limit orders on the book, and a trade is executed at the proposed price.

The recent hype surrounding cryptocurrencies and blockchain, however, changes the landscape of market structure. In particular, many exchange platforms have been built on the smart contract on the Ethereum blockchain, and transactions are executed in a decentralized manner. Those platforms are called decentralized exchanges (DEXs). They are built on a trustless record-keeping system run by the network of innumerable computer nodes on the blockchain and robust to cyber attacks and a single point of failure. Moreover, from the market-microstructure perspective, DEXs have proposed and implemented trading via a novel mechanism for trade execution called *automated market makers* (AMMs).[1]

[1] The example of decentralized exchanges (DEXs) for cryptocurrency includes Uniswap, Curve, IDEX, 0x, Waves, EtherDelta, dYdX, Balancer, Sushiswap, and so on, where all transactions are settled P2P through users' wallets and smart contracts. In the traditional centralized exchanges (CEXs), such as Coinbase, Bittrex, and Binance, a centralized authority manages trader information and funds. Although CEXs have achieved large liquidity and trading volume, they are vulnerable to cyber-attacks and single point of failures. In contrast, information on DEXs is recorded on the blockchain with KYC (Know Your

The upper panel of Figure 3.1 illustrates the exponential growth in trading volume on DEXs. It hits the record high in February 2021 (about \$74B monthly volume), obtaining more than 10% of trading share to the traditional centralized exchanges for cryptocurrency. The figure also plots trading volumes on DEXs with the traditional order book system compared to those with AMMs, showing that AMMs have considerably contributed to the recent hype in decentralized exchanges.

The automated market maker is a single-function algorithm that determines asset prices or exchange rates. It does not require the physical presence of active market makers or dealers for order execution and pricing. Before the advent of AMMs, DEXs have operated with the traditional order-book mechanism and had a hard time to attract sufficient liquidity. This is because trade execution with an order book involves a complicated matching mechanism, and embedding a large amount of information in a smart contract *on* the blockchain is extremely costly. In contrast, AMMs require much smaller memory than the traditional order-book algorithm and allow a substantial part of trades to be on the blockchain. There are several types of AMMs but, as shown by the lower panel of Figure 3.1, the Constant Product Market Makers proposed by Uniswap (and also adopted by Sushiswap) have been a dominant market structure—more than 70% of transactions on the DEXs with AMMs are handled by the Constant Product Market Makers. As formally described below, the CPMM is a simple algorithm that derives an execution price by requiring the (squared) geometric mean of the pools size to stay constant.

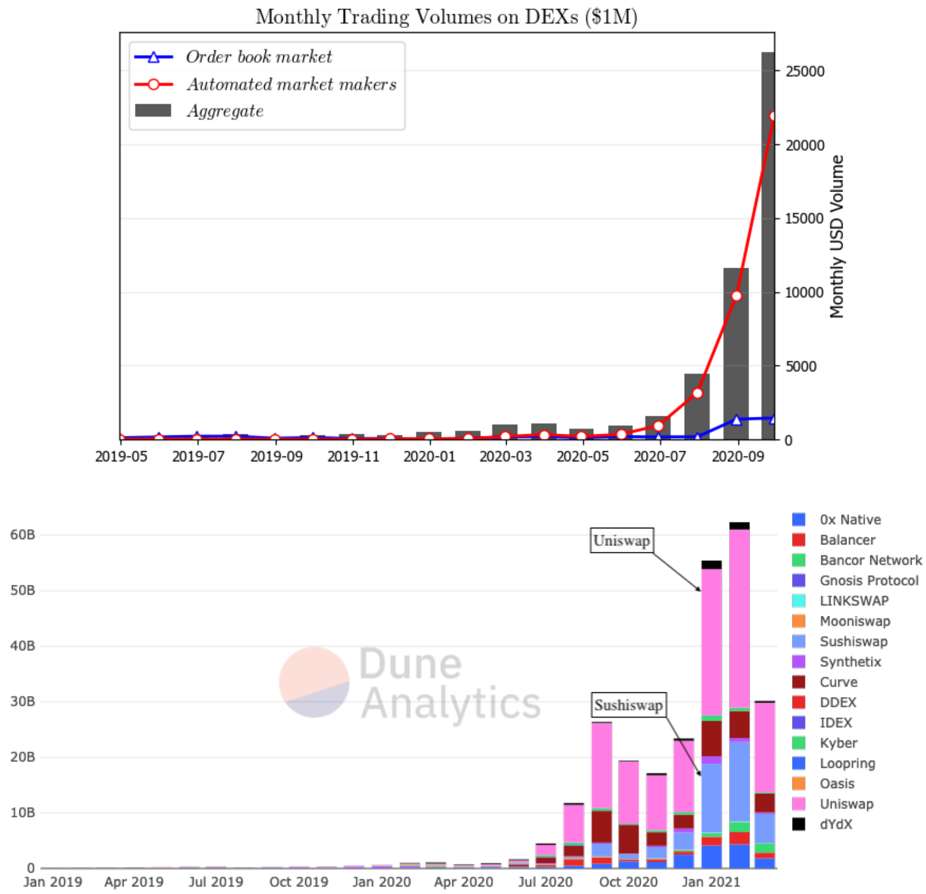
In this chapter I study the liquidity impact of the Constant Product Market Makers (CPMMs) on the traditional centralized exchange (CEX). Liquidity on the DEX is measured by the amount of assets "locked" in the platform, while CEX liquidity is measured by the bid-ask spread, as in the traditional microstructure theory. I find that DEX liquidity is positively associated with CEX liquidity, i.e., they complement each other. The result arises because the novel market-making algorithm of the CPMM triggers asymmetric behavior of information-driven order flow and noise trading across the CEX and the DEX.

In my model, a DEX with the CPMM operates in parallel with a CEX with the limit order book. There are informed for-profit traders, liquidity traders, and market makers, and they are endogenously differentiated between two trading venues: the CEX and the DEX. As the traditional theory of market microstructure suggests, the bid-ask spread on the CEX reflects the cost of asymmetric information for uninformed market makers. I first analyze the consequence of an exogenous variation in the DEX liquidity for traders' behavior—in particular, their venue choice—and its impact on the CEX liquidity. I then endogenize liquidity provision by market makers on the DEX and describe how market liquidity on the DEX and the CEX jointly reacts to a more severe informational friction.

To formalize traders' venue choice, this chapter focuses on two important features of a constant product market: its pricing algorithm and delay in transactions. On the constant product market, market makers create *liquidity pools* by depositing traded assets

Customer) data and stay immutable, providing users with an efficient and robust transaction method.

Figure 3.1: Monthly trading volume on DEXs



Source: Dune Analytics (duneanalytics.com) (say, cash and a token) into an exchange.[2] For instance, suppose that the liquidity pools reserve x and y unit of cash and token prior to a trade. If a trader buys δ unit of the token by paying $p\delta$ of cash (i.e., p is the token price or the exchange rate), she subtracts the token from the pool ($y \to y' = y - \delta$) and adds price-adjusted cash to the pool ($x \to x' = x + p\delta$), triggering a change in the liquidity pools from (x, y) to (x', y'). The CPMM algorithm requires the (squared) geometric mean of the liquidity pools to be constant $k = xy = x'y'$ with some pre-determined k. This single equation derives the execution price (or the exchange rate) p for this trading order as a function of x, y, and δ, as $p = \frac{x}{y-\delta}$.[3] When a market maker exits the market, she withdraws and liquidates her contribution to the pools. Thus, the accumulated gain or loss in the pools' value caused by trade δ (i.e., $x' - x$ and $y' - y$) is distributed to market makers on a pro rata basis.

As the second feature, trade execution on the DEX is relatively slow compared to the CEX because it involves validation by the blockchain miners with the limited block capacity and infrequent creation of blocks.[4] My model captures this delay in transactions by assuming that liquidity traders' activity is motivated by needs for immediacy (e.g., margin constraints and hedging requirements), and they incur a utility cost if their orders are not settled immediately on the DEX (e.g., Zhu, 2014; Huberman, Leshno, and Moallemi, 2019; Lehar and Parlour, 2019).

In this environment, my model first shows that ample liquidity on the DEX complements that on the CEX. The execution price for a market order with size δ on the DEX is $p = \frac{x}{y-\delta}$, and the additional liquidity to the DEX's liquidity pools, represented by an increase in x and y with $k = xy$, mitigates the price impact of δ. However, a change in the price impact affects traders differently depending on their trading motive. On the one hand, an informed trader is informed of the (future) asset fundamentals and anticipates the trading direction of other informed traders. Thus, informed traders tend to cluster on the same side of the DEX and to incur a large price impact. The additional DEX liquidity weakens the price impact of clustered informed trading and attracts more informed traders to the DEX. On the other hand, liquidity traders do not enjoy the reduction in the price impact. This is because their trading behavior is random, and each liquidity trader expects that random buy and sell orders by other liquidity traders (almost) cancel out each other on the DEX. It results in a small expected price impact, and a deeper liquidity on the DEX has only a limited impact on liquidity traders' behavior. As a result, when DEX liquidity exogenously improves, the CEX experiences a larger outflow of informed

[2] In the real financial markets, most traded assets on DEXs are digital tokens, and fiat currency is not supported. One interpretation of "cash" in this context is some ERC-20 tokens pegged to USD, such as Tether (USDA). Also, I intend to build a general model of constant product markets without restricting our attention to exchange of tokens.

[3] See Section 3.2.3 for more details.

[4] Due to the innovations in information technologies and communication methods the average throughput on the CEX is measured by the scale of microsecond or even nanosecond (see Budish, Cramton, and Shim, 2015; Aquilina, Budish, and O'Neill, 2020). Compared to the ultra-fast trade processing on the CEX, a transaction on the DEX must be validated by the blockchain miners, which takes, on average, several seconds to a couple of minutes (see dune analytics).

traders to the DEX than outflow of liquidity traders. This implies a less severe adverse selection problem, a narrower bid-ask spread, and a deeper liquidity on the CEX (Glosten and Milgrom, 1985).

Secondly, I endogenize liquidity provision by market makers on the DEX with the CPMM and relate it to the traditional market making by the order-book system. I show that asymmetric information between informed traders and market makers brings about the adverse selection cost for market makers on the DEX. Whenever the locked liquidity is taken by an informed trader, the expected value of the liquidity pools deteriorates. This is because an informed trader subtracts a more valuable asset from the liquidity pools by adding a less valuable asset to the pools based on her private information. In contrast, the expected value of the liquidity pools does not deteriorate when a trade is initiated by a liquidity trader. The above logic is reminiscent of classical models of market microstructure, such as Glosten and Milgrom (1985) and Kyle (1985). Even with the new market-making mechanism, market makers still suffer from the cost of adverse selection due to informed trading. I characterize the break-even condition for competitive market makers on the DEX and pin down the equilibrium size of DEX liquidity, which is naturally decreasing in the trading activity of informed traders relative to liquidity traders.

Finally, I characterize how equilibrium liquidity on the CEX and the DEX reacts to changes in market conditions, such as the volatility of the asset. My model demonstrates that informed buyers tend to cluster on the CEX, whereas informed sellers tend to cluster on the DEX, when the asset becomes more volatile. A higher asset volatility implies a stronger informational advantage of informed traders and exacerbates adverse selection. Thus, the bid-ask spread widens on the CEX, and the liquidity pools shrink on the DEX. The asymmetric reaction of buyers and sellers is hard-wired in the convexity of the CPMM's pricing algorithm. Namely, since the execution price is determined so that the liquidity pools (x, y) shift along the convex curve, $y = \frac{k}{x}$, a positive shift in x (a sell order) requires a smaller adjustment in y than the case of a negative shift in x (a buy order). This implies that selling the asset incurs a smaller price impact than buying the asset (see Figure 3.4). Due to the asymmetric price impact, informed buyers suffer more from the shallower liquidity on the DEX than informed sellers. Thus, DEX informed buyers tend to migrate away from the DEX to the CEX, while sellers tend to stick to (or migrate into) the DEX.

My results have several empirical implications. Due to the hard-wired asymmetric price impact for buy and sell orders on the DEX, bid and ask prices on the CEX are also asymmetrically distributed around the expected value of the asset. The asymmetric bid and ask prices are well documented in the literature (e.g., Ho and Stoll, 1981; Bossaerts and Hillion, 1991), and my model proposes the advent of the CPMM as a new source of asymmetric bid and ask prices. Also, my model suggests that "sell" order flow tends to be more informative than "buy" order flow on the DEX when the asset volatility increases, as informed sellers and buyers exhibit asymmetric reactions to the asset volatility. Moreover, in my model, adding a DEX with the CPMM algorithm tightens the bid-ask spread on the CEX. These implications can be tested by analyzing the listing of new cryptocurrency pairs on a DEX, such as Uniswap. There are several ERC-20 tokens pegged to some heavily traded cryptocurrencies. For example, Wrapped Bitcoin (WBTC) is one of the ERC-20 tokens pegged to Bitcoin and is listed on Uniswap. My model indicates that, for example, we observe a shrink in the bid-ask spread of BTC/ETH exchange rate on centralized exchanges (e.g., Coinbase) when Uniswap announced that it starts to trade WBTC/ETH.

This chapter is built on the large body of literature on market microstructure. In particular, Glosten and Milgrom (1985) and Kyle (1985) provide models of market liquidity with asymmetric information. Following the idea of Bagehot (1971), they show that competitive liquidity providers try to countervail the adverse selection cost of informed trading by making a market less liquid (i.e., posting a wider bid-ask spread; increasing the price impact of order flows). This chapter applies their canonical idea to the new context of decentralized exchanges. I show that adverse selection still plays a key role in explaining liquidity provision in a constant product market.

Also, the modern financial market has experienced substantial fragmentation of trading exchanges, and several papers have addressed implications of coexisting exchange platforms with different market microstructures. For example, Ye (2011), Zhu (2014), and Ye (2016) consider the addition of so-called dark pools and analyze the reaction of informed and uninformed traders, as well as market quality. Lee (2019) investigates traders' behavior when multiple exchanges have different degrees of latency and transparency. More recently, exchanges impose "speed bumps" to slow down high-frequency trading and protect market makers against latency arbitrage. Brolley and Cimon (2020) analyze the order flow segmentation with speed bumps to address the liquidity impact of speed bumps. My model also sheds light on the liquidity impact of heterogeneous market structures in the era of decentralization and blockchain by incorporating endogenous order flow segmentation.

My model also contributes to the research on the blockchain, cryptocurrency, and decentralized exchanges. The literature is expanding (see Harvey, 2016 and Chen, Cong, and Xiao, 2019 for comprehensive reviews), and many papers have analyzed the blockchain protocol as a new method or a platform for value transfer, e.g., Chiu and Koeppl (2017), Malinova and Park (2017), Cong, Li, and Wang (2018), Pagnotta and Buraschi (2018), Schilling and Uhlig (2018), Abadi and Brunnermeier (2018), Huberman, Leshno, and Moallemi (2019), and Lehar and Parlour (2019). However, those studies either consider order book markets or abstract away from the formal description of matching or pricing algorithm on the blockchain. My model complements the literature by characterizing the equilibrium liquidity provision on a blockchain-based exchange with the automated market making system when it coexists with the traditional exchange with the order-book mechanism.

Formal descriptions and implementational details of decentralized exchanges are provided by, for example, Warren and Bandeali (2017), Adams, Zinsmeister, and Robinson (2020), and Zhang, Chen, and Park (2018). Although the analyses on AMMs is in its infancy, Angeris, Kao, Chiang, Noyes, and Chitra (2019) provide a formal model of the optimal arbitrage problem with constant product market makers. Also, Angeris and Chitra (2020), Evans (2020), and Angeris, Evans, and Chitra (2020) analyze more general constant *function* market makers. My model focuses on the CPMM and generalize the above models by incorporating asymmetric information between traders and traders' endogenous venue choice with coexisting two different market-making algorithms (the AMM and the LOB).

3.2 Technology review

This section is devoted to preliminary discussions. I briefly describe the blockchain technology and trade execution/settlement on decentralized exchange platforms operated with the Constant Product Market Makers. Readers can refer to Antonopoulos (2014) and Antonopoulos and Wood (2018) for more details on the blockchain technology, and Chen, Cong, and Xiao (2019) for more wholistic reviews on blockchain economics.

3.2.1 Blockchain technology

The blockchain can be seen as a novel way of managing and tracking transactions information. In the traditional world, we typically maintain a ledger that records participants' state information in a centralized manner, e.g., a bank acts as an intermediary. Bilateral transactions with no intermediation by a credible third party incur asymmetric information and settlement risk.

In contrast, on the blockchain platform, a ledger is not held by a particular entity, but is distributed across all participants in the network, called record keepers or blockchain miners. The distributed ledger system requires information about blockchain users to be a consensus among all record keepers. This highlights its first difference from traditional transactions, in which only a centralized authority keeps track of information. Due to its distributed nature, the blockchain is robust to a single point of failure and does not incur costs of building credibility.

A transaction with a distributed record-keeping system by blockchain goes as follows. Suppose that Alice wants to buy a cup of coffee at Bob's cafe by paying Bitcoin. Information about this transaction must be validated by blockchain miners for settlement. More precisely, the transaction is added to a block by a miner. A sequence of blocks are encrypted and become a blockchain. In the Bitcoin blockchain, for example, each miner in the network maintains a temporary list of unconfirmed transactions, called a *mempool*. Transactions in the mempool are yet to be recorded on the blockchain, and information on the mempool is public to the network. A miner picks one of the transactions in the pool and tries to validate it by executing costly computation following a certain algorithm. The fastest miner who solves the problem adds transaction information to a block (i.e., she mines a block). The reward for mining a block is a fee: when Alice initiates a transaction, she attaches a fee to her transaction, and the validating miner obtains the attached fee.[5]

In general, it is extremely difficult for one miner in the network to overturn the consensus. In the case of Bitcoin or Ethereum, for example, they leverage their computing power to solve a time-consuming cryptographic problem. This process is called proof of work (PoW), and the miner who performs it fastest is entitled to add a new block a chain.[6] Of course there can be multiple chains of blocks, because each miner can choose to which blockchain she adds a newly mined block. Following Nakamoto (2008), however, the longest chain is regarded as a valid chain. Therefore, if a malicious agent attempts to add fraudulent information to the transaction history (e.g., a double-spending problem), she must outpace all miners in the network and secretly generate a longer chain than other chains, which requires prohibitively high computing power. That is, information on the blockchain is (almost) free from tampering.

Moreover, Ethereum allows users to add complex scripts to the blockchain which describe the conditions under which transaction is verified and recorded. It implies that a transaction takes place only if the conditions in the code are fulfilled, and it is done automatically without any centralized third-party agencies. This type of automated contracts are called a *smart contract* following Szabo (1997).

3.2.2 Decentralized exchange

Building a trading platform on the blockchain—i.e., a decentralized exchange—looks a natural strategy to extricate financial trading from a centralized information management and to make it robust to cyber attacks or a single point of failures. However, maintaining a limit order book by a smart contract on the Ethereum blockchain is costly and tends to be slow, due to the time-consuming mining process and the limited capacity of the blockchain.

There are two major solutions: a hybrid system and the Automated Market Makers (AMM). Many DEXs have adopted some "hybrid" mechanisms that involve both on-chain and off-chain features for order execution and settlement. For example, `0x` is built on so-called the *relayer mechanism* (see Warren and Bandeali, 2017). It provides an off-chain order book, on which traders can broadcast their trading intention and find their counterparties, as in the traditional centralized limit order markets. Since the order book is maintained off-chain, it refreshes swiftly. Once traders agree on a trade (i.e., trade execution), the order is settled on the blockchain via smart contracts. Note that the hybrid system still involves centralized protocol to a certain extent, as the relayers reserve some centralized power.

The second type of DEXs operate with the AMM. As mentioned in the introduction,

[5] A miner also obtains a block reward, which is a constant amount of Bitcoin (or other cryptocurrency in other blockchains), when she mines a block. Although the block reward incentivizes miners to leverage their computing power, the amount of reward periodically shrinks and converges to zero in the future.

[6] There are several ways to reach a consensus, and different blockchains (including ETH 2.0) adopt different processes. For example, Saleh (2018) analyzes the viability of the proof of stake (PoS).

it is a single-function algorithm that determines a price for order execution. As it is more simple than a limit-order matching mechanism, it requires much smaller computational capacity, making trade execution and settlement on the blockchain easier and faster. In this chapter, I focus on the CPMM because it attains the largest share in cryptocurrency trading among automated markets. Also, the CPMM is relatively simple and can provide us with clear analytical insights into the new market structures.

3.2.3 Constant Product Market Making

This section briefly explains how constant product market makers (CPMMs) determine the execution price of a trade. It follows the specification by Zhang, Chen, and Park (2018) and more detailed discussions are provided by, for example, Adams, Zinsmeister, and Robinson (2020).

Consider token X and token Y. Since there are ERC-20 tokens pegged to USD, one of the tokens can be thought of as cash. Market makers inject tokens into an exchange following a certain rule described below. The exchange aggregates locked tokens and creates *a liquidity pool*. Suppose that the exchange reserves x unit of token X and y unit of token Y. The CPMM requires the geometric mean of the liquidity pools to be constant. That is, with some constant k, it must hold that $k = xy$.

If a trader wants to buy Δx of token X by selling $\Delta y = p\Delta x$ of token Y at price p, she adds Δy of token Y to the pool and withdraws Δx of token X from the pool. It triggers the following change in the pools:

$$\begin{aligned} x &\to x' = x - \Delta x, \\ y &\to y' = y + \Delta y. \end{aligned}$$

Note that the price of token X in terms of token Y is $p = \frac{\Delta y}{\Delta x}$. Since the geometric mean of the pool must be constant, the price must satisfy the following equation.

$$k = x'y' = (x - \Delta x)(y + p\Delta y).$$

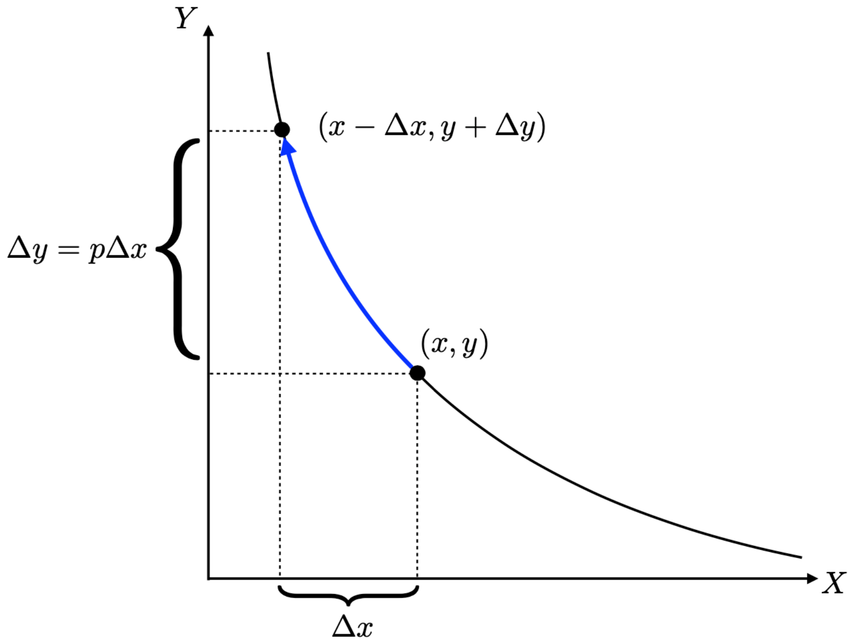
Thus, the above equation determines p as a function of the current state of the pool, (x, y), and the trading quantity Δx. In particular, I obtain

$$p = \frac{y}{x - \Delta x}.$$

Thus, the larger quantity the trader wants to buy ($\Delta x > 0$), the higher price she must pay, i.e., the price is an upward-sloping curve against the trading quantity. The price impact is mitigated when the exchange has a large amount of tokens in its liquidity pool.

Also by considering a small trading volume, $\Delta x \to 0$, the execution price for an infinitesimal trade is given by $p = y/x$, that is, the relative size of liquidity pools. Figure 3.2 shows a change in the pools' state caused by the above transaction: the exchange rate

Figure 3.2: Constant Product Market Makers



Note: This figure illustrates a change in the state of liquidity pools when an incoming market order is buying Δx unit of token X. The CPMM requires the liquidity pools to stay on the convex curve by adjusting a change in token Y or, equivalently, the execution price p.

for an infinitesimal trade is determined by the slope of the curve specified by $k = xy$. Importantly, this implies that the price is convex in the trading volume.

When a market maker (or a liquidity provider) supplies liquidity via the CPMM, she is required to lock both token X and token Y. The amount of supplied liquidity must be adjusted following the current asset price on the DEX. For instance, suppose that the liquidity pools have x and y. Then, the price of token X in terms of token Y for an infinitesimal trade is $\frac{y}{x}$. If market maker i wants to supply x_i unit of token X, she must also supply the corresponding amount of token Y so that the slope of the CPMM does not change, meaning that she also locks y_i such that $\frac{y+y_i}{x+x_i} = \frac{y}{x}$ or, equivalently, $y_i = \frac{y}{x}x_i$. According to this rule, market makers inject liquidity to the pools.

Although the geometric mean of the liquidity pools stays constant, a transaction causes a change in the liquidity pools, i.e., $(x, y) \rightarrow (x', y')$. If market maker i injects (x_i, y_i) before a trade, and the aggregate size of the pool is (x, y), she obtains the share of the aggregate liquidity pools, which is $\frac{x_i}{x}$ and $\frac{y_i}{y}$. Upon settlement of a trade, the market maker can withdraw her share from the liquidity pool and realizes her returns. Since the post-trade pool state is (x', y'), the market maker obtains the gross return of $\pi = x_i\frac{x'}{x} + y_i\frac{y'}{y}$.

3.3 Model

Consider a one-shot trading game in a two-period model. A single risky asset with common value $\tilde{v}$ is traded on two exchanges between three types of traders: for-profit traders, liquidity traders, and market makers.[7]

Events and traders. One of two possible event types may trigger a trade at $t = 1$: either an innovation in the fundamental value of the asset (a common-value shock) or a liquidity shock (a private-value shock).[8] With probability $\eta \in (0, 1)$, the common value of the asset experiences an innovation and becomes $\tilde{v} = v_0(1 + \tilde{\sigma})$, where $\tilde{\sigma}$ is stochastic growth rate of the asset's value and $\tilde{\sigma} = \pm\sigma$ with the same probability. v_0 represents the prior expected value of the asset and, without loss of generality, I normalize $v_0 = 1$.

There is a continuum of risk-neutral *for-profit traders* with a unit measure. They are sophisticated institutional investors: when a shock happens to $\tilde{v}$, they immediately observe the realized value of the shock and choose their trading venue (either the DEX or the CEX, as defined below). Each for-profit trader can decide on the trading venue contingent on private information, i.e., she is equipped with a smart order router (SOR)

[7]Although the term "asset" is used throughout the chapter, it does not restrict the types of the traded assets: following the fact that digital tokens are the major traded assets on DEXs, the asset in the model can be thought of as cryptocurrencies and ECR-20 tokens. Accordingly, the asset's value or the price of the asset can be seen as the exchange rate between a certain pair of ECR-20 tokens (or fiat currency).

[8]See, for example, Menkveld and Zoican (2017) and Brolley and Zoican (2020) for models with these shocks as a trigger of transactions.

that covers both the CEX and the DEX.[9] A for-profit trader sends a single-unit *market order* to take an arbitrage opportunity.[10] In what follows, for-profit traders are also referred to as *informed traders.*

With probability $1-\eta$, a shock hits the private value of *liquidity traders.* The liquidity traders are impatient traders, and the shock triggers their needs for immediacy, such as hedging motives, margin constraints, and other immediate borrowing and lending requirements. I assume that mass z_{buy} of liquidity traders are hit by a positive private-value shock and arrive at the market to buy one unit of the asset, whereas mass z_{sell} of liquidity traders are hit by a negative shock and try to sell one unit of the asset. The mass of liquidity traders is stochastic and uniformly distributed, $z_i \sim U[0,z]$ for $i \in \{buy, sell\}$. Upon arrival at the market, each liquidity trader immediately places a single-unit market order to fulfill her trading needs. Regarding their venue choice, liquidity traders need to decide on their trading venue at $t=0$ before they enter the market. This is because they are unsophisticated retail investors, and maintaining multiple accounts at both exchanges (or subscribing to an SOR) is costly for them.[11] In Section 3.5, I relax this assumption and allow liquidity traders to choose trading venues contingent on the sign of the private-value shock.

There also exists a continuum of *market makers* with measure L. Among them, measure L_C of market makers participate in the CEX, and the remaining $L_D = L - L_C$ measure of market makers provide liquidity on the DEX. In what follows, I take L_D as exogenous, while Section 3.4 endogenizes it.

Figure 3.3 illustrates the timeline of the game and possible outcomes of the trigger event.

Exchange platforms. There are two exchange platforms: the CEX and the DEX. The CEX is a traditional centralized exchange and operated with a continuous limit order book (LOB). Market makers on the CEX competitively provide quotes by submitting a single-unit limit orders with bid and ask prices, as in Glosten and Milgrom (1985). Limit orders are displayed on the limit order book, and liquidity takers (i.e., informed or liquidity traders) trade at the proposed prices by submitting market orders. Incoming market orders are processed with price-time priority. The CEX is based on the centralized matching algorithm with high-speed information processing. Thus, it provides ultra-fast trade execution, causing almost no delays.[12]

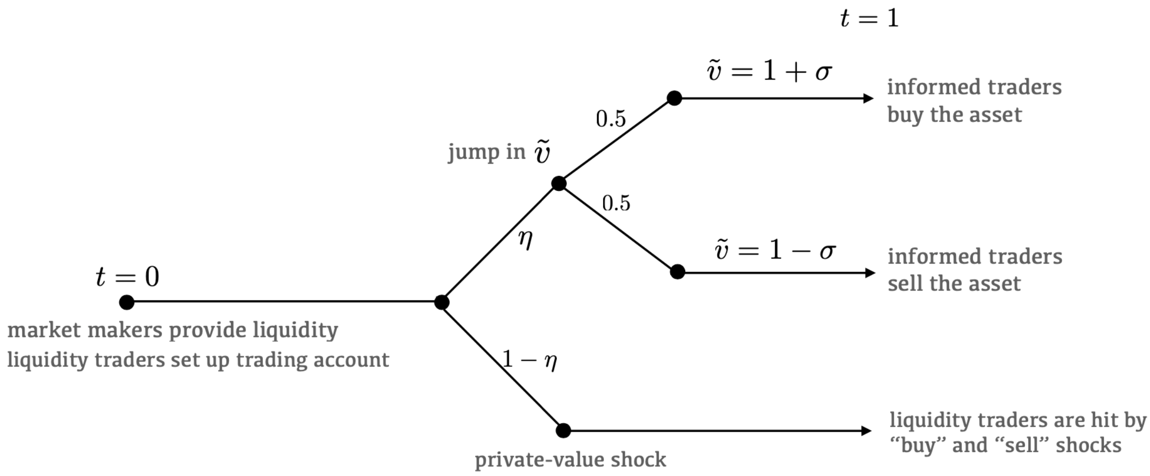
[9]Some third-party agencies provide routing services between CEXs and DEXs for trading firms, such as CoinRoutes (https://coinroutes.com/) and MainBloq (https://mainbloq.io/).

[10] Due to the risk neutrality, each trader is differentiated between two exchanges rather than splitting her order between two venues.

[11]As of March 2021, there exist only a limited number of cryptocurrency exchanges that provide order routine services across CEXs and DEXs. An investor may trade via institutional brokers, but a large portion of cryptocurrency trades are directly done by retail investors.

[12]From the high-frequency traders' perspective, even a centralized exchange with cutting-edge technologies causes a microsecond or nanosecond delay that may affect trading profits. However, the primary focus of this chapter is on the trading on the CEX compared to that on the DEX, and ignoring delays in

Figure 3.3: Timeline of the game



In contrast, the DEX is operated with the CPMM. As explained in Subsection 3.2.3, market makers competitively lock one unit of the asset (and cash) into the exchange, generating *liquidity pools*. A liquidity taker who wants to buy (resp. sell) the asset subtracts (resp. adds) the asset from the asset liquidity pool by adding (resp. subtracting) cash to the cash liquidity pool. The execution price is determined by the CPMM algorithm instead of quotes by market makers.

Trade execution with the CPMM involves smart contracts on Ethereum blockchain, and it takes much lower throughput than the CEX, causing a delay in completion of a transaction. I assume that all incoming orders placed at $t = 1$ are simultaneously executed (i.e., matched) at the end of the first period. However, executed transactions are stored in the mempool of the blockchain and wait to be validated by blockchain miners. Following Zhu (2014), I assume that a delay in order execution weighs negatively on the private utility of liquidity traders, as they are impatient and eager to fulfill their trading needs immediately. In the model, a liquidity trader on the DEX incurs $\gamma\sigma$ per unit of trade, where γ is a random parameter that measures the aversion toward a delay (or needs for immediacy) and drawn from $\gamma \sim U[0, 1]$.[13]

Differentiation. In what follows, I focus on the equilibrium in which the informed and the liquidity traders are endogenously differentiated between the CEX and the DEX. Note that each informed trader buys (resp. sells) the asset when the asset value experiences

order of microseconds or nanoseconds does not harm my discussions, as a delay on the DEX is the order of seconds or minutes, which is much longer than that on the CEX.

[13]The delay cost that is proportional to the asset volatility σ can be seen as margin constraint or unmodeled risk aversion (e.g., Brunnermeier and Pedersen, 2009).

a positive jump $\tilde{\sigma} = +\sigma$ (resp. a negative jump, $\tilde{\sigma} = -\sigma$). It becomes clear in the following discussion that the behavior of informed traders is asymmetric and depends on the trade direction due to the convex nature of the execution price on the DEX. Therefore, conditional on $\tilde{\sigma} = +\sigma$, I suppose $\beta_{buy} \in [0, 1]$ fraction of informed traders participate in the DEX to buy the asset. In contrast, with $\tilde{\sigma} = -\sigma$, measure $\beta_{sell} \in [0, 1]$ of informed traders sell on the DEX. I denote the measure of liquidity traders on the DEX as $\alpha \in [0, 1]$, which is not contingent on the sign of a private-value shock, as they decide on trading venues at $t = 0$ (see Section 3.5 for the case with asymmetric α).

3.3.1 Trade on the CEX

The partial equilibrium on the CEX mostly follows the conventional model by Glosten and Milgrom (1985). I denote the equilibrium bid and ask prices on the CEX as

$$\text{Ask} = 1 + a, \ \text{Bid} = 1 - b.$$

I sometimes call the deviation of Ask and Bid from the expected value of the asset $\mathbb{E}[\tilde{v}] = 1$ (i.e., a and $-b$) the ask and bid prices. Also, the (effective) bid-ask spread is defined as $S = a + b$.

Given the differentiation of traders, the expected profits for a market maker on each side of the market are given by

$$\pi^C_{M,ask} = \frac{1}{2}\left[\eta(1 - \beta_{buy})(a - \sigma) + (1 - \eta)(1 - \alpha)za\right],$$
$$\pi^C_{M,bid} = \frac{1}{2}\left[\eta(1 - \beta_{sell})(b - \sigma) + (1 - \eta)(1 - \alpha)zb\right].$$

In both equations, the first term represents a trade with an informed trader, which happens with probability $\frac{1}{2}\eta(1 - \beta_i)$, and the second term shows the case of liquidity trading, which happens with probability $\frac{1}{2}(1 - \eta)(1 - \alpha)z$. Following Zhu (2014), I focus on the equilibrium in which a market maker breaks even on the both sides of the market. Then, the break-even condition yields the following competitive bid and ask prices.

$$a = a(\beta_{buy}, \alpha) = \sigma \frac{(1 - \beta_{buy})\eta}{(1 - \beta_{buy})\eta + (1 - \eta)(1 - \alpha)z}, \tag{3.1}$$

$$b = b(\beta_{sell}, \alpha) = \sigma \frac{(1 - \beta_{sell})\eta}{(1 - \beta_{sell})\eta + (1 - \eta)(1 - \alpha)z}. \tag{3.2}$$

Note that the bid and the ask prices are potentially asymmetric when $\beta_{buy} \neq \beta_{sell}$. Otherwise, the comparative statics of the bid-ask spread are the same as the traditional models of market microstructure with asymmetric information. That is, the bid-ask spread is positively (resp. negatively) affected by the intensity of informed (resp. liquidity) trading, as it exacerbates (resp. mitigates) the adverse selection cost for a market maker.

Accordingly, the expected profits for an informed trader who trades on the CEX are given by

$$\pi_I^C(\tilde{\sigma}) = \begin{cases} \sigma - a & \text{if } \tilde{\sigma} = +\sigma \text{ and buys the asset,} \\ \sigma - b & \text{if } \tilde{\sigma} = -\sigma \text{ and sells the asset.} \end{cases} \tag{3.3}$$

Note that the informed trader's profits are computed conditional on the realized value of $\tilde{\sigma}$. Similarly, a liquidity trader's (ex-post) profits from trading on the CEX are given by[14]

$$\pi_{L,k}^C = \begin{cases} -a & \text{if } k = buy, \\ -b & \text{if } k = sell, \end{cases} \tag{3.4}$$

where subscript k indicates whether the private-value shock induces a trader to buy or sell the asset. Since z_{buy} and z_{sell} measures of liquidity traders are hit by "buy" and "sell" private-value shocks, the ex-ante expected trading cost for a liquidity trader at $t = 0$ is equivalent to the effective bid-ask spread:

$$\mathbb{E}\left[\sum_{k=buy,sell} z_k \pi_{L,k}^C\right] = -\frac{z}{2}S. \tag{3.5}$$

3.3.2 Trade on the DEX

Liquidity pools. I assume that the DEX has $\bar{c}$ and $\bar{x}$ of cash and the asset in its liquidity pools before the trading game starts.[15] The positive initial reserve is justifiable by the existence of unmodeled *passive* liquidity providers in the real financial markets, who inject liquidity into the pool and stay inactive. The non-arbitrage condition at the beginning of the game, together with the CPMM algorithm, implies that $\frac{\bar{c}}{\bar{x}} = v_0 = 1$, i.e., an infinitesimal trade on the DEX does not make any profits. This equation works as an exogenous initial condition for the DEX liquidity pools.

Each market maker at the DEX must follow the CPMM algorithm. For example, if market maker i wishes to supply c_i (cash) and x_i (the asset), it must satisfy $c_i = v_0 x_i$ to keep the price on the exchange at $t = 0$ constant at v_0, i.e., it does not change the execution price of an infinitesimal trade on the DEX because $\frac{\bar{c}+c_i}{\bar{x}+x_i} = v_0$. Since I have the unit-trading assumption, the above rule on liquidity provision implies that each market maker supplies one unit of the asset and cash simultaneously at the beginning of the game.

[14] I implicitly assume that a liquidity trader obtains private utility u if she fulfills her trading needs, with u being sufficiently large (e.g., $u > 1 + \sigma$). Therefore, all liquidity traders participate in the market upon hit by a shock. u does not affect the equilibrium conditions because a liquidity trader obtains u no matter where she trades.

[15] I assume that the size of the liquidity pools are larger than the potential size of liquidity taking orders, i.e., $\bar{c}, \bar{x} \geq \max\{1, z\}$. In other words, the liquidity pools do not dry up even if all traders participate in the DEX and take liquidity.

I denote the aggregate size of cash and the asset in the liquidity pools at $t = 0$ as C and X, where

$$C \equiv \bar{c} + \int c_i di = \bar{c} + L_D \tag{3.6}$$

and $X = v_0^{-1} C = C$. Also, the initial constant for the CPMM is given by $k \equiv CX = C^2$. The execution price of a trade is determined so that the geometric mean of the pools stays at the constant level, k.

Execution price. Consider an incoming market order with size x (e.g., $x > 0$ means a buy market order with mass x). By denoting the execution price for the order as p, the state of the liquidity pools changes as follows after the trade:

$$C \to C' = C + px \tag{3.7}$$

$$X \to X' = X - x. \tag{3.8}$$

For example, if a trader is buying the asset ($x > 0$), she subtracts x unit of the asset from the asset liquidity pool, X, by adding the corresponding amount of cash px to the cash pool, C. The CPMM sets the price p so that the post-trade state of the pools satisfies

$$k = C^2 = (C + px)(X - x),$$

leading to the following value.

$$p(x) = \frac{C}{C - x}. \tag{3.9}$$

Thus, the larger quantity the trader intends to buy (resp. sell), the higher (resp. lower) the execution price becomes. Moreover, the price impact of a trade can be defined as follows.

Lemma 3.1. *The price impact of a market order with measure x is given by $\frac{d \log p(x)}{dx} = \frac{1}{C-x}$. Also, the price impact is decreasing in C and increasing in x.*

The above lemma characterizes the market depth on the DEX. When the DEX has a larger quantity in its liquidity pools (C), the market becomes deeper, and a market order of a given size has a smaller price impact. Therefore, it is appropriate to use the pool size (C) as the measure of liquidity on the DEX.

Profits for informed traders on the DEX. When a jump in the asset's value occurs, each informed trader knows $\tilde{\sigma} = \pm\sigma$, which also implies that she is aware of the trading attempts of other informed traders. More precisely, she knows that a positive (resp. negative) jump triggers buy (resp. sell) market orders of aggregate size β_{buy} (resp. β_{sell}). Therefore, conditional on $\tilde{\sigma} = \pm\sigma$, the execution price on the DEX is given by

$$p = \begin{cases} p(\beta_{buy}) = \frac{C}{C-\beta_{buy}} & \text{when } \tilde{\sigma} = +\sigma, \\ p(-\beta_{sell}) = \frac{C}{C+\beta_{sell}} & \text{when } \tilde{\sigma} = -\sigma. \end{cases} \tag{3.10}$$

For simplicity, I assume that all incoming orders are aggregated and executed all at once, but Appendix C.4 shows that the expected execution price is the same as (3.10) when we consider sequential execution of orders. As a result of (3.10), an informed trader's expected profits from trading on the DEX are

$$\pi_I^D(\tilde{\sigma}) = \begin{cases} 1+\sigma - p(\beta_{buy}) & \text{if } \tilde{\sigma} = +\sigma \\ p(-\beta_{sell}) - (1-\sigma) & \text{if } \tilde{\sigma} = -\sigma. \end{cases} \tag{3.11}$$

Importantly, the execution prices for buy and sell orders are not symmetric around the expected value of the asset. We can think of $p(x)$ and $p(-x)$ as the ask and the bid prices on the DEX but $p(x) - 1 \neq 1 - p(-x)$ as long as $x \neq 0$. This is because of the convexity in the CPMM's pricing algorithm. As illustrated in Figure 3.4, suppose that a buy or a sell order with quantity Δx moves the state of the liquidity pools from LP_0 (the initial condition) to LP_1 in the figure. If it is a buy order ($\Delta x > 0$), LP_0 slides upward to $LP_{1,buy}$, while a sell order moves LP_0 downward to $LP_{1,sell}$, both along the curve $C = X^{-1}$. Since the curve is convex, even if the buy and the sell orders have the same size, the buy order requires a larger adjustment of the cash pool than the sell order, $\Delta c_{buy} > \Delta c_{sell}$. Hence, a trader must pay a higher price p to buy the asset than the price payment she obtains when she sells the asset.

Note that the above discussion is true even if I switch the role of cash and the asset. Economically, the convexity of the CPMM implies that the execution price or the exchange rate between assets is determined so that adding the liquidity to the pool bears a smaller price impact compared to consuming liquidity in the pool. This hard-wired asymmetry in the CPMM algorithm is absent in the limit order market because, without frictions or risk aversion, a buy and a sell market order bear the same trading cost, i.e., bid and ask prices are symmetric around the mid point.

Profits for liquidity traders on the DEX. With probability $1-\eta$, the trigger event is a private-value shock on the needs for immediacy of liquidity traders. Conditional on the direction of the private-value shock, the profits for a liquidity trader with the delay cost parameter at γ are given by

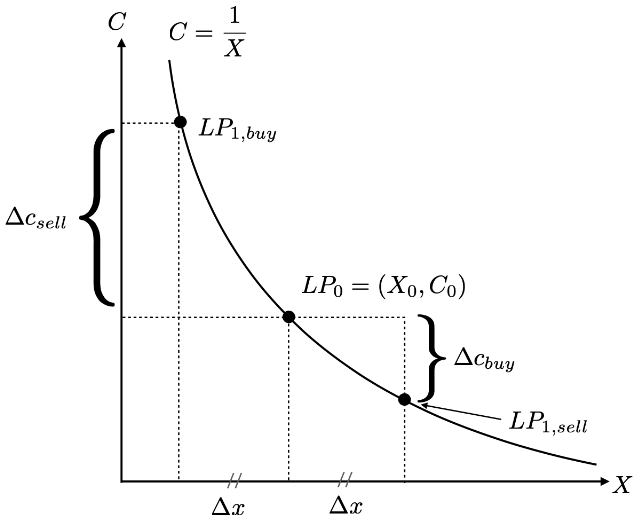
$$\pi_{L,k}^D(\gamma) = \begin{cases} 1 - \mathbb{E}_{\Delta z}[p(\alpha \Delta z)] - \gamma\sigma & \text{if } k = buy, \\ \mathbb{E}_{\Delta z}[p(\alpha \Delta z)] - 1 - \gamma\sigma & \text{if } k = sell. \end{cases} \tag{3.12}$$

Note that the aggregate order size, $\Delta z = z_{buy} - z_{sell}$, is uncertain for each liquidity trader, and $\mathbb{E}_{\Delta z}$ is the expectation over Δz. Compared to the informed trading, the price impact of liquidity trading tends to be weak because their trading direction is random, and buy and sell orders are netted out.

3.3.3 Equilibrium venue choice

I define the equilibrium with exogenous DEX liquidity (C, X) as follows:

Figure 3.4: Asymmetric price impact



Note: This figure discribes the price impact of buy and sell orders with the same size, Δx. Starting from (X_0, C_0), a buy market order with size Δx causes an upward shift of the liquidity pools, while a sell order triggers a downward shift. The execution price must keep (C, X) on the CPMM curve $(C = 1/X)$. Since the curve is convex, an addition of Δx to X requires a smaller adjustment of the value of C than the case of reduction of X, meaning that the price impact is larger for a buy order.

Definition 3.1. *The equilibrium of the model is defined by the measure of informed and liquidity traders who participate in the DEX, $\{(\beta_i)_{i=buy,sell}, \alpha\}$, the bid-ask prices, (a, b), and the execution prices on the DEX, p, such that (i) the informed traders are indifferent between trading on the CEX and the DEX given the occurrence of a jump in $\tilde{v}$ and the information about $\tilde{\sigma}$, (ii) the liquidity traders are differentiated by comparing the ex-ante expected profits on the CEX and the DEX, (iii) the market makers on the CEX break even at the bid and the ask prices, and (iv) the prices on the DEX follows the CPMM algorithm given the initial pools condition (C, X).*

With the profit functions for each trader on each venue given by (3.3), (3.4), (3.11), and (3.12), the indifference conditions for traders pin down the equilibrium mass of traders on the DEX.

Informed traders. The informed traders' indifference condition for the buy and the sell sides boils down to

$$\begin{aligned} 1 + a(\beta_{buy}, \alpha) &= p(\beta_{buy}), \\ 1 - b(\beta_{sell}, \alpha) &= p(-\beta_{sell}). \end{aligned}$$

By rearranging the above equations with (3.1), (3.2), and (3.10), β_{buy} and β_{sell} solve $B(\beta; +\sigma) = 0$ and $B(\beta; -\sigma) = 0$, respectively, where B is given by

$$B(\beta; \tilde{\sigma}) = \beta^2(1 + \tilde{\sigma})\eta - \beta[(1 + \tilde{\sigma})\eta + z(1 - \alpha)(1 - \eta) + \sigma\eta C] + \sigma\eta C. \tag{3.13}$$

Since we have $B(0; \tilde{\sigma}) > 0$ and $B(1; \tilde{\sigma}) < 0$ there is a unique solution of $B(\beta; \tilde{\sigma}) = 0$ in $\beta \in (0, 1)$.

Proposition 3.1. *(i) Given α, the equilibrium measure of informed traders on the DEX is given by $\beta^*_{buy} = \beta(+\sigma, \alpha)$ and $\beta^*_{sell} = \beta(-\sigma, \alpha)$ where*

$$\beta(\tilde{\sigma}, \alpha) = \frac{(1 + \tilde{\sigma})\eta + (1 - \eta)\alpha z + \sigma\eta C - \sqrt{D(\tilde{\sigma})}}{2(1 + \tilde{\sigma})\eta} \tag{3.14}$$

with $D(\tilde{\sigma}) = [(1 + \tilde{\sigma})\eta + (1 - \eta)\alpha z + \sigma\eta C]^2 - 4(1 + \tilde{\sigma})\sigma\eta^2 C$.
(ii) β^*_i is monotonically increasing in α, C, and σ.
(iii) $\beta^*_{buy} < \beta^*_{sell}$ for all $\alpha \in (0, 1)$.

Proof. Solving the quadratic equation in (3.14) yields the result. Point (ii) is obtained by the implicit function theorem. □

β^*_{buy} and β^*_{sell} are both a unique solution for each indifference condition in $\beta \in (0, 1)$, and they are stable, i.e., they do not diverge even if a small perturbation happens to parameter values. Intuitively, a small increase in the measure of DEX informed traders β^* causes a larger price impact on the DEX and a narrower bid-ask spread on the CEX due to mitigated adverse selection for CEX market makers. Thus, the CEX becomes more attractive than the DEX, meaning that a marginal informed trader on the CEX has no incentive to switch her trading venue.

Moreover, the trading intensity of the informed traders exhibits anticipated reactions to a change in parameter values. If a larger set of liquidity traders participate in the DEX (α increases), it exacerbates adverse selection for market makers on the CEX. It induces the market makers on the CEX to charge a wider bid-ask spread, and more informed traders are willing to participate in the DEX (β_i^* increases). Also, a larger C implies that the DEX provides deeper liquidity and attenuates the price impact of informed trading (see Lemma 3.1), inviting more informed traders to the DEX. Finally, a higher volatility of the asset (σ) also attracts informed traders to the DEX. On the one hand, the execution price at the DEX is not directly affected by the asset volatility. On the other hand, the bid-ask spread on the CEX becomes proportionally wider when the asset becomes more volatile. This is because σ captures a stronger informational advantage of informed traders and more severe adverse selection for market makers.

One of the novel implications of the CPMM algorithm emanates from asymmetry in execution prices due to its convexity explained above. Since buying the asset incurs a larger cost than the return of selling the asset, a negative innovation in the asset's value induces a disproportional reaction of informed sellers to informed buyers, leading to $\beta_{buy}^* < \beta_{sell}^*$. The asymmetric reaction survives in the general equilibrium with endogenous α and C and provides some empirical implications, as discussed later.

Liquidity traders. A liquidity trader with cost parameter γ participates in the DEX if and only if the profits from trading on the DEX dominate that from trading on the CEX:[16]

$$\frac{z}{2}\sum_{k=buy,sell}\pi_{L,k}^{D} \geq \frac{z}{2}\sum_{k=buy,sell}\pi_{L,k}^{C} = -\frac{z}{2}S,$$

where the last equality uses (3.5). The above inequality is reduced to

$$\gamma < \gamma^* \equiv \frac{S(\beta_{buy}, \beta_{sell}, \alpha)}{2\sigma}.$$

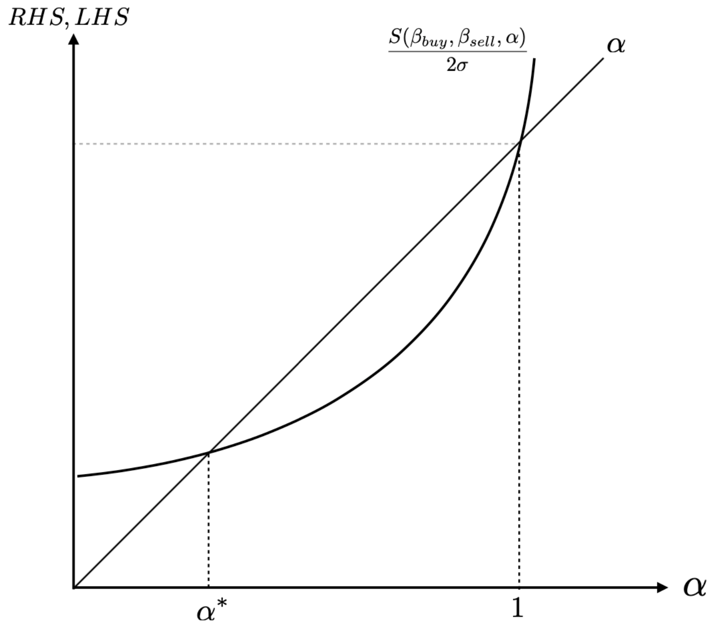
The LHS is the expected trading cost on the DEX. Since buying and selling happens with the same probability, the execution price does not affect the expected cost. However, the delay cost matters because a liquidity trader on the DEX bears it regardless of the trading direction. In contrast, the RHS represents the expected trading cost on the CEX, i.e., the bid-ask spread. Since $\gamma \sim U[0,1]$, I obtain the following result.

Proposition 3.2. *(i) Given $(\beta_{buy}, \beta_{sell})$, the equilibrium mass of liquidity traders on the DEX is determined by*

$$\alpha = \Pr(\gamma < \gamma^*) = \frac{S(\beta_{buy}, \beta_{sell}, \alpha)}{2\sigma}. \tag{3.15}$$

[16] I assume that a liquidity trader participates in the DEX when she is indifferent.

Figure 3.5: Expected trading costs for liquidity traders



If $\eta < \frac{z}{1+z}$*, equation (3.15) has two solutions:* $\hat{\alpha} = 1$ *and* $\alpha^* \in (0, 1)$*.* $\hat{\alpha}$ *is unstable and* α^* *is stable. If* $\eta > \frac{z}{1+z}$*,* $\hat{\alpha} = 1$ *is a unique (unstable) solution.*
(ii) α^* *is monotonically decreasing in* β_{buy} *and* β_{sell}*.*

Proof. See Appendix C.1. □

Figure 3.5 depicts the LHS and the RHS of equation (3.15) with $\eta < \frac{z}{1+z}$. Since the cost of the DEX trading crosses the cost of the CEX trading from above at α^*, it is stable, i.e., even if a small change happens to α^*, it converges to the original point. In what follows, I assume that the condition holds for multiple solutions and focus on the stable one $\alpha^* \in (0, 1)$.

Proposition 3.2 demonstrates that the measure of liquidity traders on the DEX is decreasing in the measure of informed traders on the DEX. When informed traders migrate away from the CEX to the DEX (β_i^* increases), CEX market makers face a less severe adverse selection problem. It induces the market makers to lower the bid-ask spread, and the trading cost on the CEX declines. Sine the execution price on the DEX for liquidity trading is not directly affected by the measure of informed traders, the CEX becomes more attractive for the liquidity traders, lowering α^*.

Overall, Propositions 3.1 and 3.2 indicate that informed traders "chase" liquidity traders between trading venues. When the liquidity traders migrate away from the CEX to the DEX (α declines), the informed traders are more willing to trade on the DEX as well. This is because of a more severe adverse selection problem on the CEX and a larger trading cost there. In contrast, when informed traders move to the DEX, the liquidity traders are inclined to migrate back to the CEX, because liquidity on the CEX improves due to a less severe adverse selection problem for CEX market makers. In the following subsection, I derive the equilibrium by putting together the above two channels and analyze the impact of the DEX liquidity (C) on the CEX liquidity (the bid-ask spread).

3.3.4 Liquidity impact of the CPMM

It is not trivial whether an additional liquidity on the DEX improves or deteriorates liquidity on the CEX. Since the bid-ask spread on the CEX is determined by the ratio of informed trading to liquidity trading, we need to investigate the reaction of $(\beta_{buy}, \beta_{sell})$ relative to that of α to a change in DEX liquidity C.

From Propositions 3.1 and 3.2, I obtain the following result.

Proposition 3.3. *(i) The equilibrium measures of liquidity traders and informed traders on the DEX,* $(\alpha^*, \beta^*_{buy}, \beta^*_{sell})$*, solve the following equations.*

$$\alpha = \frac{S(\beta_{buy}, \beta_{sell}, \alpha)}{2\sigma}, \tag{3.16}$$

$$\beta_i = \begin{cases} \beta^*(+\sigma, \alpha) & \text{for } i = buy, \\ \beta^*(-\sigma, \alpha) & \text{for } i = sell, \end{cases}$$

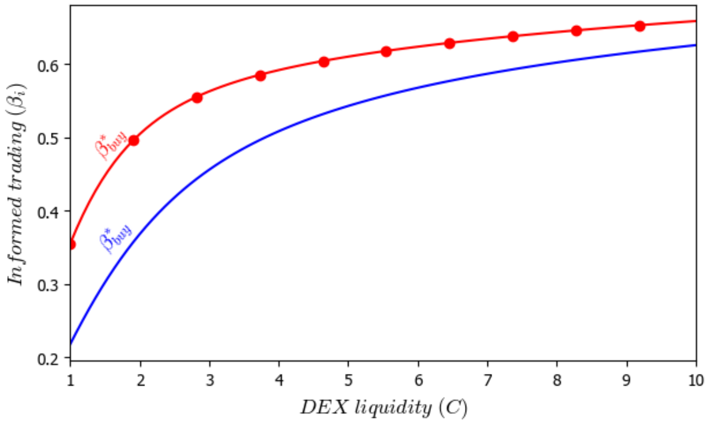
where β^* *is given by (3.14). There is a unique stable interior solution for the above equations.*
(ii) The equilibrium measure of liquidity traders on the DEX α^* *is a decreasing function of DEX liquidity* C*.*
(iii) The equilibrium bid-ask spread on the CEX is a decreasing function of C*.*

Proof. See Appendix C.2. □

Although the behavior of β^*_i is hard to obtain analytically, I can numerically check that informed traders are more inclined to participate in the DEX when it becomes more liquid, as shown by Figure 3.6. This is because the price impact for informed trading is decreasing in C. Thus, larger liquidity pools on the DEX render informed trading on the DEX less costly and attracts more informed traders to the DEX.

Moreover, Proposition 3.3 shows that deeper liquidity on the DEX causes migration of liquidity traders from the DEX to the CEX. Firstly, an increase in β^*_i mitigates the adverse selection cost for CEX market makers, and the bid-ask spread declines. Secondly, we know from equation (3.16) that a change in C does not have a direct impact on liquidity traders' behavior, as the execution price on the DEX does not matter in expectation. Hence, facing a decline in the trading cost on the CEX, more liquidity traders participate in the CEX. Note that this process involves a decline in the bid-ask spread or improved market liquidity on the CEX, as demonstrated by point (iv) in Proposition 3.3.

Figure 3.6: Informed trading on DEX



Note: This figure illustrates the values of β_i^* for different C. The values for other parameters are $\sigma = 0.1, z = 1.0, \eta = 0.2$. The upward sloping configuration is robust to other parameter values.

Therefore, my model suggests that liquidity on the DEX complements that on the CEX. Although I do not model underling blockchain mechanisms for the DEX in detail, some exogenous variations in blockchain parameters may affect CEX liquidity by altering liquidity provision on the DEX. For example, the amount of Ethereum locked in Uniswap has experienced a substantial drop (about 40%) in November 2020 after the announcement of a change in the fee structure on the platform.[17] The above result can be a simple building block to analyze the liquidity impact of such an event and helps discuss the broader implications of blockchain environment by incorporating its cross-market effects.

Moreover, compared to actively monitored and repriced limit orders on the traditional markets, liquidity provision with the CPMM (or AMMs, in general) is called "lazy liquidity" and regarded as inefficient in the real financial market. The claim is based on the fact that market makers on the DEX do not have control over the execution price of a trade and cannot incorporate information even if some news arrives. my model cannot address if it is inefficient, as it does not directly compare the welfare implication of liquidity on both venues. However, I show that an additional liquidity on the DEX improves that on the CEX, which may go counter to the claimed inefficiency of DEX liquidity.

3.4 Liquidity provision with the CPMM

Although the previous subsection establishes that DEX liquidity C improves liquidity on the CEX, C must be an endogenous variable in the equilibrium. Namely, the value of C in my model captures how many market makers provide liquidity on the DEX. Thus,

[17]See, for example, https://cointelegraph.com/.

this section endogenizes the liquidity provision by the market makers by allowing them to choose the venue to provide liquidity.

3.4.1 Market makers' profits on the DEX

Prior to the trading game (at $t = 0$), each market maker decides on the trading venue to supply liquidity. If she participate in the CEX, she decides on bid and ask prices for one unit of the asset. Since the market-making sector on the CEX is competitive, she proposes the competitive prices, a and b derived in (3.1) and (3.2), to obtain zero profits in expectation.

In contrast, if she participates in the DEX, she injects one unit of cash and the asset to the liquidity pools. The expected cost of liquidity provision is $1 + \mathbb{E}[\tilde{v}] = 2$, i.e., one unit of cash and the asset, with the latter having the expected value $\mathbb{E}[\tilde{v}]$. By locking liquidity, she obtains $w = \frac{1}{C}$ share of the post-trade aggregate liquidity pools. With a trade of size x, the post-trade liquidity pools have C' and X' in equations (3.7) and (3.8).

Since the order type and the trigger event are uncertain for market makers at the liquidity provision stage, the expected liquidation value of the liquidity pools is

$$V_{LP} = \mathbb{E}\left[\frac{C^2}{C - \tilde{x}} + \tilde{v}(C - \tilde{x})\right]$$

where $\mathbb{E}$ is the expectation operator regarding the type of traders (i.e., informed or uninformed) and the type of the trigger event (i.e., an innovation in $\tilde{v}$ or a private-value shock). By expanding the expectation and applying the trading strategy of liquidity takers, it holds that

$$\begin{aligned} V_{LP} = \frac{\eta}{2}&\left[\frac{C^2}{C - \beta_{buy}} + (1+\sigma)(C - \beta_{buy}) + \frac{C^2}{C + \beta_{sell}} + (1-\sigma)(C + \beta_{sell})\right] \\ &+ (1-\eta)\int_{-z}^{z}\left(\frac{C^2}{C - \alpha\Delta z} + C - \alpha\Delta z\right) dG(\Delta z) \end{aligned} \tag{3.17}$$

where G is the cdf of $\Delta z \equiv z_{buy} - z_{sell}$.[18] The first line shows the value of the post-trade liquidity pools when the trigger event is an innovation in the fundamental value of the asset. In this case, the jump in $\tilde{v}$ is either positive or negative with the same probability and triggers informed trading with measure β_{buy} (if $\tilde{\sigma} = +\sigma$) or β_{sell} (if $\tilde{\sigma} = -\sigma$). With the complementary probability, a shock hits on the private value of liquidity traders. It

[18] Since $\Delta z = z_{buy} - z_{sell}$ with $z_i \sim U[0, z]$, the pdf of Δz is

$$g(\Delta z = q) = \begin{cases} 0 & \text{if } q \notin [-z, z], \\ \frac{1}{z^2}(z - |q|) & \text{if } q \in [-z, z]. \end{cases}$$

results in the market orders with stochastic size $\Delta z = z_{buy} - z_{sell}$, as captured by the second line.

Each market maker has share w of the post-trade liquidity pools, meaning that her expected profits from supplying liquidity, net of the initial cost, is given by $\pi_M^D \equiv wV_{LP} - 2$, that is,

$$\pi_M^D = \underbrace{\frac{\eta}{2}\left[\frac{\beta_{buy}}{C-\beta_{buy}} - (1+\sigma)\frac{\beta_{buy}}{C} - \frac{\beta_{sell}}{C+\beta_{sell}} + (1-\sigma)\frac{\beta_{sell}}{C}\right]}_{=\pi_{M,IT}^D}$$
$$+ (1-\eta)\underbrace{\left[\int_{-z}^{z}\left(\frac{C}{C-\alpha\Delta z}\right) dG(\Delta z) - 1\right]}_{=\pi_{M,LT}^D}.$$

I denote the net profits from informed and liquidity trading as $\pi_{M,IT}^D$ and $\pi_{M,LT}^D$, respectively.

Proposition 3.4. *(i) A market maker on the DEX obtains negative profits from informed trading and positive profits from liquidity trading, that is,* $\pi_{M,IT}^D < 0$ *and* $\pi_{M,LT}^D > 0$.
(ii) The (negative) profits from informed trading is decreasing in β_{buy}*,* β_{sell}*, and* σ*. It takes a U-shaped curve against* C*, and converges to* 0 *as* $C \to \infty$.
(iii) The (positive) profits from liquidity trading is increasing in α *and decreasing in* C.

Proof. See Appendix C.3. □

As in the case of limit order markets, market makers on the DEX gain from trading with liquidity traders but lose from trading with informed for-profit traders. In other words, informational advantage of informed traders causes the risk of negative profits for the market makers on the DEX. On the DEX, the fact that liquidity is taken by an informed trader implies that the value of liquidity pools inevitably declines. This is because an informed trader always subtracts one of the more valuable assets from the liquidity pools by adding a less valuable asset to the pools. As a result, a market maker ends up having a larger amount of a less valuable asset by giving up a more valuable asset. This result highlights the similarity of the CPMM to market making on the limit order book. Namely, informed trading involves adverse selection for market makers on the DEX, as in the case of limit order markets (e.g., Glosten and Milgrom, 1985; Kyle, 1985).

Following the above intuition, market makers lose more when informed trading on the DEX is more active, i.e., when β_i is high. Moreover, the adverse selection cost for DEX market makers ($\pi_{M,IT}^D$) is a decreasing function of σ. The more volatile the asset becomes, the more informational advantage the for-profit traders obtain by knowing $\tilde{\sigma} = \pm\sigma$. Thus, σ also captures the degree of adverse selection and the cost of liquidity provision on the DEX.

In contrast, the size of liquidity on the DEX has an ambiguous impact on the cost of informed trading. On the one hand, a larger pool size (C) induces more informed trading (Proposition 3.1), and the adverse selection problem deteriorates. On the other hand, a larger C implies that each market maker has a smaller stake in the aggregate liquidity, diminishing the negative impact of informed trading on her profits.

In contrast, uninformed liquidity trading improves the value of the liquidity pools. Firstly, since liquidity buy and sell orders are netted out and they are independent of $\tilde{\sigma}$, liquidity trading does not change the expected value of the asset pool ($\mathbb{E}[X'] = C - \alpha\mathbb{E}[\Delta z] = C$), leading to zero net profits from $X \rightarrow X'$. The strictly positive profits from liquidity trading emanate from the liquidity pool of cash. Since the execution price adjusts the post-trade liquidity pools along with the convex curve $C = X^{-1}$, Jensen's inequality implies that

$$\mathbb{E}_{\Delta z}\left[\frac{C}{C - \alpha\Delta z}\right] > \frac{C}{C - \alpha\mathbb{E}_{\Delta z}[\Delta z]} = 1.$$

Therefore, the positive impact of liquidity trading is hard-wired in the CPMM's convex pricing algorithm and works as an implicit reward for liquidity providers on the DEX.

The profit from liquidity trading is magnified when the volatility of liquidity trading is large. For example, both a larger set of liquidity traders on the DEX (α) and a wider variation in liquidity trading (i.e., $Var(z_i)$) improve $\pi^D_{M,LT}$. However, a larger liquidity size (C) diminishes a variation in the liquidity trading and reduces $\pi^D_{M,LT}$. These properties are easily derived from Jensen's inequality, as a change in $\alpha\Delta z$ or C alters the convexity of the pricing curve.

3.4.2 Equilibrium with endogenous DEX liquidity

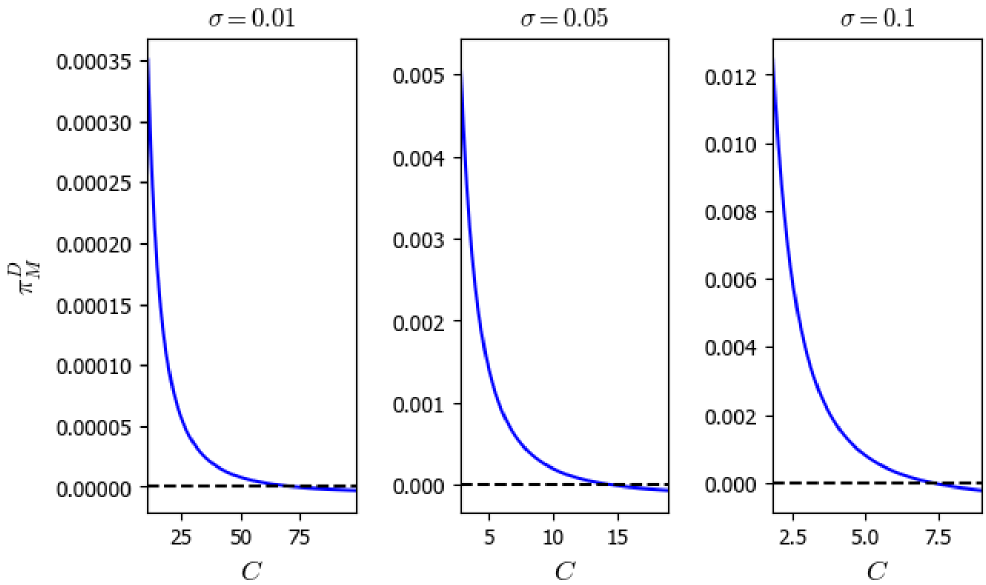
We are poised to characterize the full equilibrium with endogenous liquidity on the DEX. I focus on the equilibrium in which market makers are differentiated between two venues, meaning that each of them must be indifferent between them. Since market makers on the CEX break even due to competition, the measure of DEX market makers L_D is determined by the zero-profit condition on the DEX. That is, $C^* = \bar{c} + L^*_D$ solves

$$\pi^D_M = \frac{\eta}{2}\pi^D_{M,IT}(C) + (1-\eta)\pi^D_{M,LT}(C) = 0. \tag{3.18}$$

As shown by Figure 3.7, I numerically check that π^D_M is monotonically decreasing in C when $\pi^D_M < 0$, and the above equation has a unique stable solution, C^*, as long as $\sigma < \frac{C}{1+C}$ and some other parameter conditions are satisfied.

Moreover, Proposition 3.4 implies that C^* negatively reacts to an exogenous change in β^*_i relative to α. Intuition follows the traditional discussions on adverse selection: informed trading relative to liquidity trading makes it more costly for market makers on the DEX to supply liquidity. Thus, the size of the liquidity pools (or the measure of DEX market makers) must decline to guarantee the break-even condition, leading to a

Figure 3.7: Market makers' profit on DEX



Note: These figures are illustrated by using $z = 2.0$ and $\eta = 0.4$.

shallower market. This is also true for an increase in σ, as it exacerbates information asymmetry between informed traders and market makers.

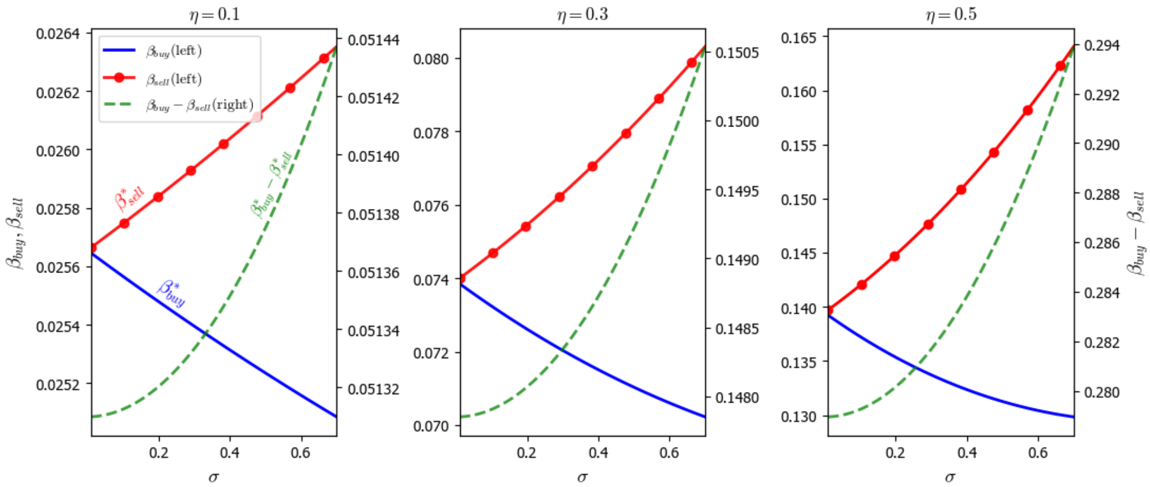
Comparative statics. In what follows, I take the volatility of the asset σ (or the measure of asymmetric information) to gauge a joint reaction of the traders' behavior and market liquidity. Parameter σ has a direct impact on liquidity on the DEX (C^*; via equation [3.18]) and informed trading ($\beta^*_{buy}, \beta^*_{sell}$; via equation [3.14]), whereas the measure of liquidity trading (α^*) is only indirectly affected by σ.

A higher volatility of the asset implies that informed traders possess a more informational advantage over market makers, and the adverse selection problem worsens both on the DEX and the CEX. Therefore, it confounds liquidity provision by DEX market makers, leading to a decline in C^*, as well as a wider effective bid-ask spread on the CEX.

As both the DEX and the CEX become more costly to trade, the reaction of β^*_{buy} and β^*_{sell} can be ambiguous. Remember that β^*_{buy} solves $B(\beta; +\sigma) = 0$, while β^*_{sell} is the solution of $B(\beta; -\sigma) = 0$, with B given by equation (3.13). Thus, the impact of σ on β_{buy} and β_{sell} *through market liquidity* can be analyzed by the following partial derivative:

$$\frac{\partial B(\beta; \tilde{\sigma})}{\partial \sigma} \sim \underbrace{C^* - \text{sign}(\tilde{\sigma})\sigma}_{>0} + \sigma \underbrace{\frac{\partial C^*}{\partial \sigma}}_{<0}. \tag{3.19}$$

Figure 3.8: Reaction of informed traders to σ



Note: These figures are illustrated by using $z = 2.0$ and $\eta < \frac{C}{1+C}$ to guarantee the existence of the stable equilibrium. The solid line shows β_{buy}, the dotted line shows β_{sell}, and the dashed line shows $\beta_{buy} + \beta_{sell}$.

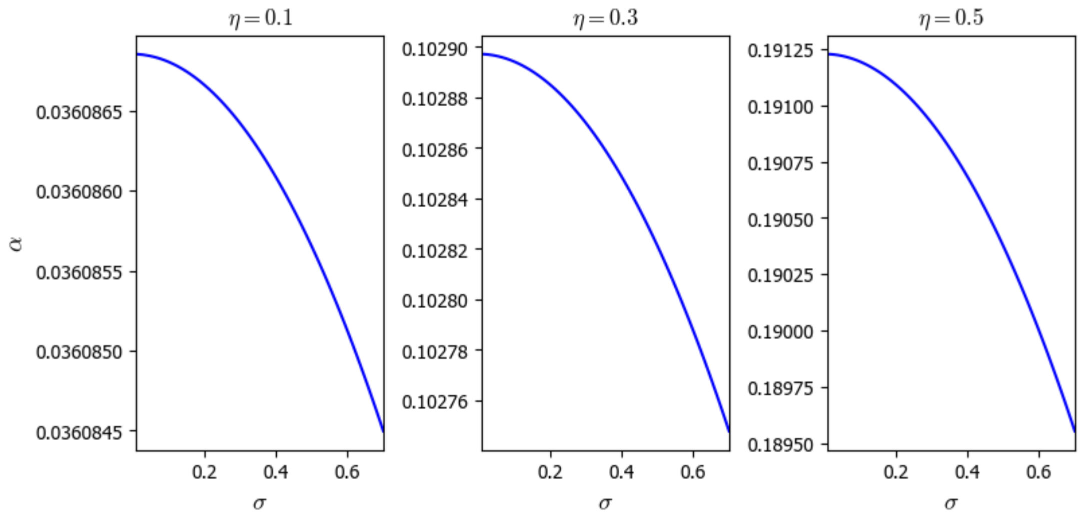
The first term shows the impact of σ on β_i via a change in the bid-ask spread (CEX liquidity), and the second term is the impact via a change in C^* (DEX liquidity). It shows that β_{buy} is more resilient to a wider bid-ask spread than β_{sell}, i.e., the first component in (3.19) is larger for β_{sell} because $\tilde{\sigma} = -\sigma$. This is intuitive because the convexity of the CPMM makes it less costly for a liquidity taker to add liquidity to the pool than consume liquidity in the pool, meaning that an incentive to migrate to (or stick to) the DEX is stronger for selling informed traders. As a result, informed traders on the DEX exhibit asymmetric reaction to a volatility shock: when selling (resp. buying) the asset, informed traders tend to cluster on the DEX (resp. the CEX).

The above logic is confirmed by Figure 3.8, which shows that the DEX-liquidity effect (the second term of [3.19]) dominates the CEX-liquidity effect (the first term of [3.19]) for buying informed traders, leading to decreasing β^*_{buy} against σ, while β^*_{sell} shows the opposite reaction. Moreover, since β_{sell} exhibits a stronger (positive) reaction to σ than a negative reaction of β_{buy}, the DEX involves more informed trading in expectation, i.e., $\beta^*_{buy} + \beta^*_{sell}$ increases.

Numerical result 1: *When the asset becomes more volatile, informed sellers tend to cluster on the DEX and informed buyers tend to cluster on the CEX. The net effect is positive in the sense that outflow of buyers is dominated by inflow of sellers to the DEX.*

Next, consider the behavior of liquidity traders. Figure 3.9 shows that liquidity traders tend to cluster on the CEX when the asset becomes more volatile. They compare the

Figure 3.9: Reaction of liquidity traders to σ



Note: These figures are illustrated by using $z = 2.0$ and $\eta < \frac{C}{1+C}$ to guarantee the existence of the stable equilibrium.

delay cost on the DEX ($\gamma\sigma$) to the expected trading cost on the CEX (S; the bid-ask spread). Since both of them are proportional to the asset volatility, σ has no direct impacts on liquidity traders' venue choice. Instead, what matters is the *normalized* bid-ask spread, $\frac{S}{2\sigma}$, which captures the adverse selection problem for the CEX market makers that emanates from informed traders' venue choice.

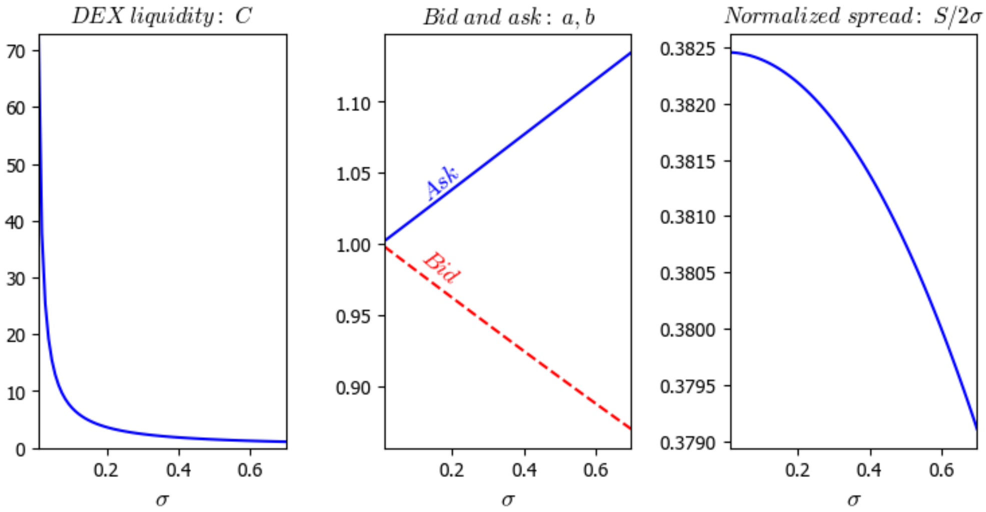
In the above discussion, we have established that informed traders tend to cluster on the DEX in expectation (i.e., $\frac{\beta_{buy}+\beta_{sell}}{2}$ increases), and it imposes more severe adverse selection on DEX market makers, while mitigating that for CEX market makers. It tightens the normalized bid-ask spread on the CEX and attracts liquidity traders to the CEX.

Numerical result 2: *When the asset becomes more volatile, liquidity traders tend to cluster on the CEX.*

Finally, Figure 3.10 summarizes the reaction of market liquidity to a change in the asset volatility incorporating the above behavior of traders. Through their venue choice, traders' reactions may have indirect effects and undermine the direct impact of σ on market liquidity, but they cannot offset or dominate the direct effect.

Numerical result 3: *When the asset becomes more volatile, liquidity on the DEX, measured by the amount of cash locked in the pool, and liquidity on the CEX, measured by the bid-ask spread, both deteriorate. The normalized bid-ask spread on the CEX, however, improves.*

Figure 3.10: Reaction of market liquidity to σ



Note: These figures are illustrated by using $z = 2.0$ and $\eta = 0.3$. They are rbust to other parameter values as long as $\eta < \frac{C}{1+C}$ is satisfied to guarantee the existence of the stable equilibrium.

The normalized bid-ask spread becomes narrower because informed traders, in expectation, tends to cluster on the DEX, while liquidity traders are more likely to trade on the CEX.

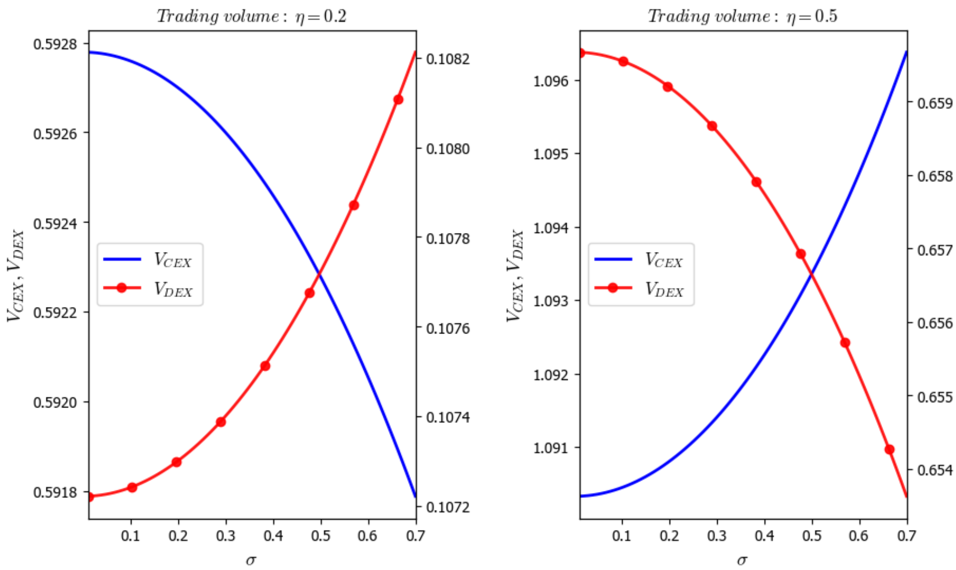
DEX trading share The share of the DEX in terms of trading volume can be used as a measure of traders' activity on the DEX. The expected trading volumes on the CEX and the DEX, as well as the aggregate volume, are defined by the following.

$$\begin{aligned} V_{CEX} &= \eta \left(1 - \frac{1}{2} \sum_{i=buy,sell} \beta_i \right) + (1-\eta)(1-\alpha)z, \\ V_{DEX} &= \eta \frac{1}{2} \sum_{i=buy,sell} \beta_i + (1-\eta)\alpha z, \\ V &= V_{CEX} + V_{DEX} = \eta + (1-\eta)z. \end{aligned}$$

Note that the aggregate trading volume is constant and perfectly determined by the probability of the trigger event (η) and the expected size of liquidity trading $z = \frac{1}{2}\mathbb{E}[z_{buy} + z_{sell}]$.

Figure 3.11 plots the trading volumes on the DEX and the CEX (V_{DEX}, V_{DEX}) against the asset volatility. The behavior of trading volumes on exchanges are not robust and dependent on the probability of the trigger event η. When the asset becomes more volatile, the measure of DEX liquidity traders increases, while that of informed traders declines

Figure 3.11: Trading volumes



Note: These figures are illustrated by using $z = 2.0, \eta = 0.3$ and $\sigma = 0.05$. They are rbust to other parameter values as long as $\eta < \frac{C}{1+C}$ is satisfied to guarantee the existence of the stable equilibrium.

in expectation. From the above equations for V_i, these two effects compete against each other with η being the weight on the informed traders' behavior. Thus, the DEX trading volume (as well as its share) tends to decrease when a common value shock is more likely to be the trigger of transactions (the right panel), while it tends to increase when transactions are motivated by private values.

3.5 Discussion

In this section, I first check the robustness of the above results by relaxing the assumption regarding liquidity traders' venue choice. I also discuss some testable implications provided by my model.

3.5.1 Robustness

For tractability, the above analyses assume that liquidity traders must decide on their trading venue prior to the trigger event. In this subsection, I allow liquidity traders to choose their trading venue contingent on the realized value of a private-value shock. Due to the convexity of the CPMM pricing, I focus on the equilibrium in which the fractions of buying and selling liquidity traders on the DEX are asymmetric and given by $\alpha_{buy} \in (0, 1)$ and $\alpha_{sell} \in (0, 1)$, respectively.

By applying the same logic as the previous sections, informed traders' indifference conditions are given by

$$\begin{aligned}1 + a(\beta_{buy}, \alpha_{buy}) &= p(\beta_{buy}),\\ 1 - b(\beta_{sell}, \alpha_{sell}) &= p(-\beta_{sell}),\end{aligned}$$

where $p(x)$ is given by equation (3.9), and the ask and the bid prices are given by (3.1) and (3.2) with asymmetric α. As a result, the equilibrium measure of informed buyers and sellers can be expressed by reusing the previous equations.

Corollary 3.1. *Given $\alpha \equiv (\alpha_{buy}, \alpha_{sell})$, the equilibrium measure of informed traders on the DEX is $\beta^*_{buy} = \beta(+\sigma, \alpha_{buy})$ and $\beta^*_{sell} = \beta(-\sigma, \alpha_{sell})$, where $\beta(\tilde{\sigma}, \alpha_j)$ is the solution of $B(\tilde{\sigma}; \alpha_j)$ with B given by equation (3.13). β^*_i is increasing in C, σ, and α_i for $i \in \{buy, sell\}$.*

Thus, the reaction of informed traders in the partial equilibrium stays the same as the previous case with symmetric α in Proposition 3.1.

Now, consider the venue choice for liquidity traders. When a liquidity trader buys (resp. sells) the asset on the CEX, her trading cost (resp. reward) is the ask (resp. bid) price. In contrast, she pays or obtains the following symmetric price on the DEX:

$$\begin{aligned}p_n(\alpha_{buy}, \alpha_{sell}) &= \mathbb{E}_{(z_{buy}, z_{sell})}\left[\frac{C}{C - (\alpha_{buy} z_{buy} - \alpha_{sell} z_{sell})}\right]\\ &= \frac{C}{z^2 \alpha_{buy} \alpha_{sell}} \log \frac{(C + \alpha_{sell} z)^{C + \alpha_{sell} z}(C - \alpha_{buy} z)^{C - \alpha_{buy} z}}{C^C (C - z\Delta\alpha)^{C - z\Delta\alpha}}\end{aligned} \tag{3.20}$$

where $\mathbb{E}_{(z_{buy}, z_{sell})}$ is the expectation regarding $z_i \sim U[0, z]$, and the second line expands the expectation. When α is symmetric, the net expected amount of liquidity trading is zero, as z_{buy} and $-z_{sell}$ are symmetrically distributed, i.e., buy and sell orders are netted out. In contrast, the asymmetric behavior of buy and sell liquidity traders prevents the orders from completely offsetting each other.

A liquidity traders with delay cost γ compares the trading cost on the DEX (the LHS) and the CEX (the RHS):

$$\gamma\sigma \gtreqless \begin{cases} 1 + a(\beta_{buy}, \alpha_{buy}) - p_n(\alpha_{buy}, \alpha_{sell}) & \text{if a "buy" liquidity shock hits,}\\ p_n(\alpha_{buy}, \alpha_{sell}) - (1 - b(\beta_{sell}, \alpha_{sell})) & \text{if a "sell" liquidity shock hits.}\end{cases}$$

Since γ uniformly distributes in $[0, 1]$, I obtain the following:

Corollary 3.2. *Given $(\beta_{buy}, \beta_{sell})$, the equilibrium measures of liquidity buyers and sellers on the DEX are given by the solution of the following equations.*

$$\alpha_{buy} = \frac{1 + a(\beta_{buy}, \alpha_{buy}) - p_n(\alpha_{buy}, \alpha_{sell})}{\sigma},$$

$$\alpha_{sell} = \frac{p_n(\alpha_{buy}, \alpha_{sell}) - (1 - b(\beta_{sell}, \alpha_{sell}))}{\sigma}.$$

For $i \in \{buy, sell\}$, α_i is decreasing in β_i and increasing in α_j for $j \neq i$.

The above result shows that the reaction of α_i in the partial equilibrium is the same as the previous analyses. The additional result brought by the asymmetric α is the strategic complementarity between liquidity buyers and sellers. Namely, liquidity buyers are more willing to trade on the DEX when more liquidity sellers participate in the DEX, and vice versa. This is because a larger trading volume on the opposite side of the market offsets the buy liquidity orders, leading to a smaller shift in the liquidity pools and a weaker price impact. Therefore, liquidity begets liquidity on the DEX with the CPMM, as in the traditional limit order markets (e.g., Pagano, 1989).

Finally, the expected profits for a market maker on the DEX is given by

$$\pi_M^D = \underbrace{\frac{\eta}{2}\left[\frac{\beta_{buy}}{C - \beta_{buy}} - (1+\sigma)\frac{\beta_{buy}}{C} - \frac{\beta_{sell}}{C + \beta_{sell}} + (1-\sigma)\frac{\beta_{sell}}{C}\right]}_{=\pi^D_{M,IT}} + (1-\eta)\underbrace{[p_n(\alpha_{buy}, \alpha_{sell}) - 1]}_{=\pi^D_{M,LT}}.$$

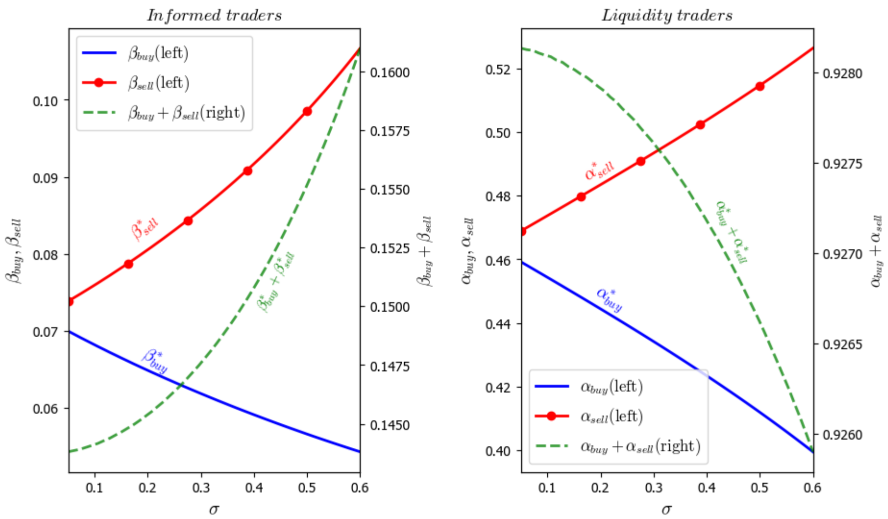
Once again, it is easy to check that $\pi^D_{M,IT} < 0$ and $\pi^D_{M,LT} > 0$, meaning that a market maker loses from informed trading and gains from liquidity trading. A larger mass of informed trading on the DEX, as well as a higher volatility of the asset, reduces DEX market makers' profits by worsening adverse selection.

Result. Figure 3.12 plots the reaction of informed traders (the left panel) and liquidity traders (the right panel) on the DEX to an increase in the volatility of the asset. The asymmetric reaction of informed buyers and sellers on the left panel shows that the result in the previous analyses is robust to a change in the assumption on liquidity traders' venue choice. The right panel, however, shows that allowing a contingent venue choice adds a new implication regarding liquidity traders' behavior on the DEX.

Numerical result 4: *When the asset becomes more volatile, liquidity buyers tend to cluster on the CEX, while liquidity sellers tend to cluster on the DEX. The net effect is negative, i.e., outflow of liquidity traders from the DEX dominates inflow to the DEX.*

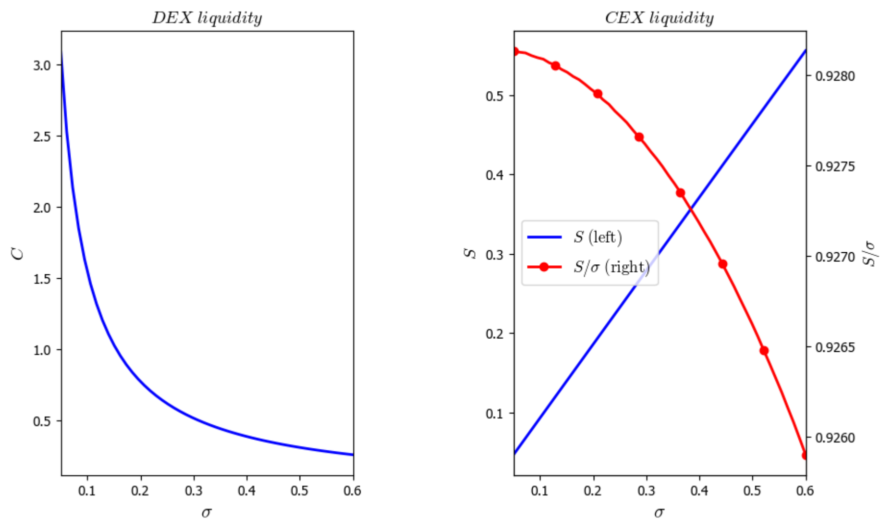
The net behavior of liquidity traders $\alpha_{buy} + \alpha_{sell}$ is different from that of informed traders. Intuitively, a liquidity trader on the DEX is not directly affected by the convexity of the CPMM algorithm *per se*, as she is uncertain about the aggregate trading volume (given by [3.20]). Thus, the asymmetric reaction of liquidity traders is driven by the asymmetric reaction of informed traders, that is, β_{buy} and β_{sell}. Since the expected mass of informed traders increases on the DEX, the bid-ask spread on the CEX shrinks which, in turn, induces liquidity traders to participate more on the CEX in expectation. Therefore, $\alpha_{buy} + \alpha_{sell}$ declines with σ.

Figure 3.12: Reaction of informed and liquidity traders



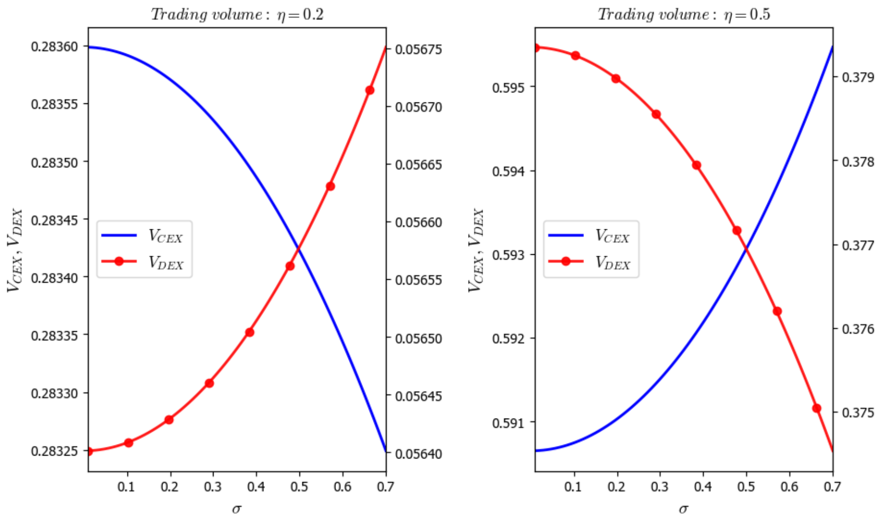
Note: This figures are illustrated by using $z = 2.0$ and $\eta = 0.3$. They are robust to other parameter values, as long as it holds that $z < \frac{C}{1+C}$.

Figure 3.13: Reaction of market liquidity



Note: This figures are illustrated by using $z = 2.0$ and $\eta = 0.3$. They are robust to other parameter values, as long as it holds that $z < \frac{C}{1+C}$.

Figure 3.14: Reaction of trading volumes



Note: This figures are illustrated by using $z = 2.0$ and $\eta = 0.3$. They are robust to other parameter values, as long as it holds that $z < \frac{C}{1+C}$.

Given the venue choice by traders, Figure 3.13 shows the reactions of market liquidity. The left panel shows the comparative static of DEX liquidity, measured by C^*, while the right panel illustrates the bid-ask spread (S) and the normalized bid-ask spread (S/σ). Since the *net* behavior of liquidity traders stays the same as the previous sections, so does the impact of the asset volatility on market liquidity.

Finally, the reaction of trading volumes is similar to the case of symmetric α. When transactions are more likely to be driven by the common-value shock (resp. the private-value shock), the share of the DEX declines (resp. increases), as shown by Figure 3.14. The same logic applies as the previous sections, as the expected measure of liquidity traders ($\alpha_{buy} + \alpha_{sell}$) still declines even if we consider a state-contingent venue choice of liquidity traders. Thus, we can check that the ambiguous behavior of trading volumes provided in the previous section is robust.

3.5.2 Empirical implications

I can derive novel empirical implications from my model. In the following discussion I consider a change in the asset volatility σ (or the degree of adverse selection) and the addition of a DEX with the CPMM to the traditional financial market. These events serve as an exogenous variation in financial market to propose testable implications.

More concretely, the addition of a DEX can be seen as a change in the status quo, in which all transactions are conducted via centralized exchanges. It is a relevant measure of the impact of DEXs because, in reality, many ERC-20 tokens are cross-listed between some CEXs and DEXs. Also, we have several tokens that replicate some major cryptocurrencies. For example, Uniswap has listed the ETH/WBTC pair on December 2020. WBTC is an ERC-20 token that is pegged to Bitcoin. Thus, the listing of WBTC on Uniswap can be seen as the advent of a DEX for the Ethereum and Bitcoin pair, which is previously traded mostly on the centralized exchanges, such as Coinbase and Binance.

Firstly, the equilibrium prices of the asset on the CEX is affected by the CPMM.

Conjecture 3.1. *The addition of the DEX with the CPMM induces or strengthens asymmetry in the bid and the ask prices.*

The first conjecture is a natural consequence of the convex pricing of the CPMM. It generates asymmetry in the price impact on the DEX for buy and sell orders. The asymmetry, in turn, must affect bid and ask prices on the CEX with different magnitudes via traders' venue choice. There is a large body of literature for the asymmetric bid and ask prices, such as Ho and Stoll (1981) and Stoll (1989). In terms of the bid-ask spread that stems from adverse selection, studies have highlighted the asymmetry due to some microstructure constraints, such as a discrete tick size (Anshuman and Kalay, 1998), and the asymmetric distribution for the value of assets (Bossaerts and Hillion, 1991). My model proposes a new and exponentially growing market structure that brings about the asymmetric prices on the traditional limit-order markets.

Conjecture 3.2. *All else being equal, an increase in the asset volatility (or the degree of adverse selection) is associated with a higher order informativeness on the DEX and a lower order informativeness on the CEX.*

This conjecture is the result of Subsection 3.4.2. A higher degree of adverse selection makes it costly to trade on both venues. Informed traders tend to cluster on the same side of the market on the DEX, bearing a larger price impact, compared to liquidity traders with random trading behavior. Thus, order flow on the DEX tends to be information driven, while that on the CEX tends to be private-value driven.

As the literature on the CPMM is still in its infancy, an empirical measure of informativeness on the DEX is yet to be constructed. In contrast, we have some metrics of informed trading on the traditional markets, such as PIN by Easley and O'hara (1987). My model suggests that the addition of a DEX with the CPMM strengthens a positive reaction of informativeness of order flow to a change in the volatility of the asset.

Related to the informativeness of order flow, my model suggests that buy and sell orders may react in the different ways even if the magnitude of a trigger event is the same. This, in turn, implies that return predictability of order flow is asymmetric between sell and buy orders.

Conjecture 3.3. *Buy orders on the DEX are more likely to be followed by a positive innovation in returns than sell orders followed by a negative innovation. The opposite is true on the CEX.*

Moreover, the above prediction regarding the informativeness of order flow has a direct implication for market liquidity.

Conjecture 3.4. *All else being equal, an increase in the asset volatility (or the degree of adverse selection) is associated with a decline in the amount of assets locked in the DEX (e.g., Uniswap), a wider effective bid-ask spread, and a narrower normalized bid-ask spread on the CEX.*

In terms of the above-mentioned example, my model predicts that the correlation between the effective bid-ask spread for the ETH/BTC pair on centralized exchanges (e.g., Coinbase) and its return volatility tends to be stronger after Uniswap starts covering the ETH/WBTC pair compared to that prior to Uniswap.

3.5.3 Limitations of the model

Information horizon. One of the limitations of my model is that it does not accommodate informed traders with a longer information horizon. In my model, I consider a one-shot trading game, in which informed traders try to exploit their informational advantage knowing that the information is perfectly revealed (or the asset is liquidated) after a trade.

When traders act on some long-lived private information, I need to incorporate some other important features of the information management on the DEX, namely, public nature of blockchain information. As mentioned in Subsection 3.2.1, trading intentions on the DEX are stored in the mempool and wait for validation by blockchain miners. In most cases, the state of the mempool is publicly disseminated and observable for miners and traders. As suggested by Malinova and Park (2017), a trader may extract private information of other traders by observing publicly available information on the mempool, generating the front-running risk. My current model with competitive traders cannot fully address the long-run issues, and a strategic aspect of informed trading must be embedded in the future research.

Endogenous delay. In my model, I cut corners in introducing a trading delay on the DEX by assuming that a liquidity trader incurs a linear delay cost per transaction. It can be thought of as a situation where the mass of liquidity trading is sufficiently small compared to the block capacity so that all trades are settled with a constant (and deterministic) delay, i.e., the block time.

In general, however, a trader can shorten the expected waiting time by paying a higher transaction fee to blockchain miners. Since a miner tends to process transactions with higher fees, proposing a higher fee can put trader's transaction forward in a queue. For example, Huberman, Leshno, and Moallemi (2019) formulate the expected delay cost (the sum of the waiting time and fee payment) as an increasing function of the measure of traders waiting for verification.

Endogenizing the delay cost will certainly add new implications to my model. At the same time, however, I also believe that the endogenous delay cost strengthens my results. If the delay cost is an increasing function of the measure of liquidity traders on the DEX (α), as suggested by Huberman, Leshno, and Moallemi (2019), liquidity traders are discouraged to participate in the DEX, leading to even larger outflow of liquidity traders from the DEX to the CEX. Thus, the endogenous delay cost may work as an additional driving force to mitigate adverse selection for CEX market makers and improve CEX market liquidity.

3.6 Conclusion

This chapter studies the equilibrium impact of the adoption of a decentralized exchange (DEX) with a novel market-making algorithm called the Constant Product Market Makers (CPMMs). In the real financial market, DEXs with the CPMM and the traditional centralized exchanges (CEXs) with the limit order mechanism interact with each other. I built a model of coexisting exchanges, in which traders are endogenously differentiated between the DEX and the CEX. It characterizes joint behavior of market liquidity on both venues.

As in the traditional market microstructure models, liquidity on both trading venues is determined by information asymmetry between informed for-profit traders and uninformed market makers. My model shows that DEX liquidity complements CEX liquidity. I also find that the convexity of the CPMM pricing equation leads to asymmetric reaction of informed traders. In particular, when the asset becomes more volatile, informed buyers tend to cluster on the CEX, while informed sellers tend to cluster on the DEX. Thus, my model predicts that sell orders on the DEX has higher informativeness compared to buy orders on the DEX.

In my model, I focus on a one-shot trading environment and abstract away from long-lived private information. When information horizon becomes longer, informed traders must incorporate the speed of information revelation via their trading orders (as in Kyle, 1985). Moreover, price discovery in the long-run is one of the two pillars that determines trader welfare. Thus, constructing a long-run model based on the current analyses is the topic for future research.

Bibliography

Abadi, J. and M. Brunnermeier (2018). *Blockchain economics*. Tech. rep. mimeo Princeton University.

Adams, H., N. Zinsmeister, and D. Robinson (2020). “Uniswap v2 core”. In: *URl: https://uniswap.org/whitepaper. pdf*.

Aït-Sahalia, Y. and M. Saglam (2013). *High frequency traders: Taking advantage of speed*. Tech. rep. National Bureau of Economic Research.

Aldrich, E. M. and D. Friedman (2018). *Order protection through delayed messaging*. Tech. rep. WZB Discussion Paper.

American, N. (2019). *NYSE American to Eliminate Speed Bump, Restore Floor-Based Trading*. (Visited on 2019).

Anderson, L., E. Andrews, B. Devani, M. Mueller, and A. Walton (2018). *Speed segmentation on exchanges: Competition for slow flow*. Tech. rep. Bank of Canada Staff Working Paper.

Angeris, G. and T. Chitra (2020). “Improved Price Oracles: Constant Function Market Makers”. In: *arXiv preprint arXiv:2003.10001*.

Angeris, G., A. Evans, and T. Chitra (2020). “When does the tail wag the dog? Curvature and market making”. In: *arXiv preprint arXiv:2012.08040*.

Angeris, G., H.-T. Kao, R. Chiang, C. Noyes, and T. Chitra (2019). “An analysis of Uniswap markets”. In: *Cryptoeconomic Systems Journal*.

Anshuman, V. R. and A. Kalay (1998). “Market making with discrete prices”. In: *The Review of Financial Studies* 11.1, pp. 81–109.

Antonopoulos, A. M. (2014). *Mastering Bitcoin: unlocking digital cryptocurrencies*. O’Reilly Media, Inc.

Antonopoulos, A. M. and G. Wood (2018). *Mastering ethereum: building smart contracts and dapps*. O’reilly Media.

Aoyagi, J. (2020). “The Dark Side of Regulating Fast Informed Trading”. In: *SSRN 3221990*.

Aquilina, M., E. B. Budish, and P. O’Neill (2020). “Quantifying the high-frequency trading “arms race”: A simple new methodology and estimates”. In: *Chicago Booth Research Paper* 20-16.

Bagehot, W. (1971). “The only game in town”. In: *Financial Analysts Journal* 27.2, pp. 12–14.

Baldauf, M. and J. Mollner (2017). “High-frequency trading and market performance”. In: *SSRN Electronic Journal*.

Banerjee, S. and B. M. Breon-Drish (2018). “Dynamic information acquisition and entry into new markets”. In: *SSRN 2846059*.

Banerjee, S., J. Davis, and N. Gondhi (2018). “When transparency improves, must prices reflect fundamentals better?” In: *The Review of Financial Studies* 31.6, pp. 2377–2414.

Baruch, S. and L. R. Glosten (2013). “Flickering quotes”. In: *Columbia University*.

— (2019). “Tail expectation and imperfect competition in limit order book markets”. In: *Journal of Economic Theory* 183, pp. 661–697.

Bertomeu, J., A. Beyer, and R. A. Dye (2011). “Capital structure, cost of capital, and voluntary disclosures”. In: *The Accounting Review* 86.3, pp. 857–886.

Beyer, A., D. A. Cohen, T. Z. Lys, and B. R. Walther (2010). “The financial reporting environment: Review of the recent literature”. In: *Journal of Accounting and Economics* 50.2-3, pp. 296–343.

Biais, B., T. Foucault, and S. Moinas (2015). “Equilibrium fast trading”. In: *Journal of Financial Economics* 116.2, pp. 292–313.

Boehmer, E., K. Fong, and J. Wu (2015). “International evidence on algorithmic trading”. In: *SSRN Electronic Journal*.

Bongaerts, D. and M. Van Achter (2016). “High-frequency trading and market stability”. In: *SSRN Electronic Journal*.

Bossaerts, P. and P. Hillion (1991). “Market Microstructure Effects of Government Intervention in the Foreign Exchange Market”. In: *The Review of Financial Studies* 4.3, pp. 513–541.

Brogaard, J., B. Hagströmer, L. Nordén, and R. Riordan (2015). “Trading fast and slow: Colocation and liquidity”. In: *The Review of Financial Studies* 28.12, pp. 3407–3443.

Brolley, M. and D. A. Cimon (2017). “Order Flow Segmentation, Liquidity and Price Discovery: The Role of Latency Delays”. In: *SSRN Electronic Journal*.

— (2020). “Order-flow segmentation, liquidity, and price discovery: the role of latency delays”. In: *Journal of Financial and Quantitative Analysis* 55.8, pp. 2555–2587.

Brolley, M. and M. Zoican (2019). “Liquid speed: On-demand fast trading at distributed exchanges”. In: *arXiv preprint arXiv:1907.10720*.

— (2020). “Liquid speed: A congestion fee for low-latency exchanges”. In: *Rotman School of Management Working Paper* 3377346.

Brunnermeier, M. K. and L. H. Pedersen (2009). “Market liquidity and funding liquidity”. In: *The review of financial studies* 22.6, pp. 2201–2238.

Budish, E., P. Cramton, and J. 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.

Budish, E., R. Lee, and J. Shim (2018). “Will the Market Fix the Market? A Theory of Stock Exchange Competition and Innovation”. In: *Manuscript in Preparation*.

Bushee, B. J., D. A. Matsumoto, and G. S. Miller (2004). "Managerial and investor responses to disclosure regulation: The case of Reg FD and conference calls". In: *The Accounting Review* 79.3, pp. 617–643.

Cartea, A., S. Jaimungal, and J. Walton (2019). "Foreign exchange markets with last look". In: *Mathematics and Financial Economics* 13.1, pp. 1–30.

Chakrabarty, B., J. Huang, and P. K. Jain (2019). "Effects of a Speed Bump on Market Quality and Exchange Competition". In: *Available at SSRN 3280645*.

Chen, H., S. Foley, M. Goldstein, and T. Ruf (2017). "The Value of a Millisecond: Harnessing Information in Fast, Fragmented Markets". In: *SSRN Electronic Journal*.

Chen, J. and R. Lu (2019). "Does Public Disclosure Crowd Out Private Information Production?" In: *SSRN 3294142*.

Chen, L., L. W. Cong, and Y. Xiao (2019). "A Brief Introduction to Blockchain Economics". In: *Available at SSRN*.

Cheynel, E. and C. B. Levine (2020). "Public disclosures and information asymmetry: A theory of the mosaic". In: *The Accounting Review* 95.1, pp. 79–99.

Chiu, J. and T. V. Koeppl (2017). "The economics of cryptocurrencies–bitcoin and beyond". In: *Available at SSRN 3048124*.

Chiyachantana, C. N., C. X. Jiang, N. Taechapiroontong, and R. A. Wood (2004). "The impact of Regulation Fair Disclosure on information asymmetry and trading: An intraday analysis". In: *Financial Review* 39.4, pp. 549–577.

Coller, M. and T. L. Yohn (1997). "Management forecasts and information asymmetry: An examination of bid-ask spreads". In: *Journal of accounting research* 35.2, pp. 181–191.

Colombo, L., G. Femminis, and A. Pavan (2014). "Information acquisition and welfare". In: *The Review of Economic Studies* 81.4, pp. 1438–1483.

Cong, L. W., Y. Li, and N. Wang (2018). "Tokenomics: Dynamic adoption and valuation". In:

Conrad, J., S. Wahal, and J. Xiang (2015). "High-frequency quoting, trading, and the efficiency of prices". In: *Journal of Financial Economics* 116.2, pp. 271–291.

Dennert, J. (1993). "Price competition between market makers". In: *The Review of Economic Studies* 60.3, pp. 735–751.

Diamond, D. W. (1985). "Optimal release of information by firms". In: *The Journal of Finance* 40.4, pp. 1071–1094.

Du, S. and H. Zhu (2017). "What is the optimal trading frequency in financial markets?" In: *The Review of Economic Studies* 84.4, pp. 1606–1651.

Dugast, J. and T. Foucault (2018). "Data abundance and asset price informativeness". In: *Journal of Financial Economics* 130.2, pp. 367–391.

Dye, R. A. (2001). "An evaluation of "essays on disclosure" and the disclosure literature in accounting". In: *Journal of Accounting and Economics* 32.1-3, pp. 181–235.

Easley, D., N. M. Kiefer, and M. O'Hara (1997). "One day in the life of a very common stock". In: *The Review of Financial Studies* 10.3, pp. 805–835.

Easley, D. and M. O'hara (1987). "Price, trade size, and information in securities markets". In: *Journal of Financial economics* 19.1, pp. 69–90.

Eleswarapu, V. R., R. Thompson, and K. Venkataraman (2004). "The impact of Regulation Fair Disclosure: Trading costs and information asymmetry". In: *Journal of financial and quantitative analysis* 39.2, pp. 209–225.

Evans, A. (2020). "Liquidity Provider Returns in Geometric Mean Markets". In: *arXiv preprint arXiv:2006.08806*.

Foucault, T., O. Kadan, and E. Kandel (2013). "Liquidity cycles and make/take fees in electronic markets". In: *The Journal of Finance* 68.1, pp. 299–341.

Foucault, T., R. Kozhan, and W. W. Tham (2016). "Toxic arbitrage". In: *The Review of Financial Studies* 30.4, pp. 1053–1094.

Foucault, T., A. Roell, and P. Sandas (2003). "Market making with costly monitoring: An analysis of the SOES controversy". In: *The Review of Financial Studies* 16.2, pp. 345–384.

Frino, A., V. Mollica, and R. I. Webb (2014). "The impact of co-location of securities exchanges' and traders' computer servers on market liquidity". In: *Journal of Futures Markets* 34.1, pp. 20–33.

Gao, P. and P. J. Liang (2013). "Informational feedback, adverse selection, and optimal disclosure policy". In: *Journal of Accounting Research* 51.5, pp. 1133–1158.

Gintschel, A. and S. Markov (2004). "The effectiveness of Regulation FD". In: *Journal of Accounting and Economics* 37.3, pp. 293–314.

Glosten, L. R. and P. R. Milgrom (1985). "Bid, ask and transaction prices in a specialist market with heterogeneously informed traders". In: *Journal of Financial Economics* 14.1, pp. 71–100.

Glosten, L. R. and T. J. Putnins (2016). "Welfare costs of informed trade". In: *Columbia Business School Research Paper*, pp. 16–58.

Goldstein, I. and L. Yang (2017). "Information disclosure in financial markets". In: *Annual Review of Financial Economics* 9, pp. 101–125.

Gong, F. and H. Liu (2012). "Inside trading, public disclosure and imperfect competition". In: *International Review of Economics & Finance* 24, pp. 200–223.

Grossman, S. J. and J. E. Stiglitz (1980). "On the Impossibility of Informationally Efficient Markets". In: *The American Economic Review* 70.3, pp. 393–408. ISSN: 0003-0651. DOI: 10.1177/0003065108330523. URL: http://sci-hub.cc/http://www.jstor.org/stable/1805228.

Haas, M. and M. Zoican (2016). "Beyond the frequency wall: Speed and liquidity on batch auction markets". In: *SSRN Electronic Journal*.

Han, B., Y. Tang, and L. Yang (2016). "Public information and uninformed trading: Implications for market liquidity and price efficiency". In: *Journal of Economic Theory* 163, pp. 604–643. ISSN: 10957235. DOI: 10.1016/j.jet.2016.02.012. URL: http://dx.doi.org/10.1016/j.jet.2016.02.012.

Han, J., M. Khapko, and A. S. Kyle (2014). "Liquidity with high-frequency market making". In:

Harvey, C. (2016). “Cryptofinance”. In: *Available at SSRN 2438299*.
Hasbrouck, J. and G. Saar (2013). “Low-latency trading”. In: *Journal of Financial Markets* 16.4, pp. 646–679. ISSN: 1386-4181. DOI: https://doi.org/10.1016/j.finmar.2013.05.003. URL: http://www.sciencedirect.com/science/article/pii/S1386418113000165.
Hendershott, T. and P. C. Moulton (2011). “Automation, speed, and stock market quality: The NYSE’s hybrid”. In: *Journal of Financial Markets* 14.4, pp. 568–604.
Hirshleifer, J. (1971). “The Private and Social Value of Information and the Reward to Inventiv The Private and Social Value of Information and the Reward to Inventive Activity”. In: *The American Economic Review* 61, pp. 561–574.
Ho, T. and H. R. Stoll (1981). “Optimal dealer pricing under transactions and return uncertainty”. In: *Journal of Financial economics* 9.1, pp. 47–73.
Hoffmann, P. (2014). “A dynamic limit order market with fast and slow traders”. In: *Journal of Financial Economics* 113.1, pp. 156–169.
Hu, E. (2018). “Intentional access delays, market quality, and price discovery: Evidence from IEX becoming an exchange”. In: *SEC Working Paper*.
Huang, S. and B. Z. Yueshen (2018). “Speed Acquisition”. In: *Available at SSRN 2845864*.
Huberman, G., J. Leshno, and C. C. Moallemi (2019). “An economic analysis of the bitcoin payment system”. In: *Columbia Business School Research Paper* 17-92.
Huddart, S., J. S. Hughes, and C. B. Levine (2001). “Public disclosure and dissimulation of insider trades”. In: *Econometrica* 69.3, pp. 665–681.
Jones, C. M. (2013). “What do we know about high-frequency trading?” In: *Columbia University Working Paper*.
Kanodia, C. and H. Sapra (2016). “A real effects perspective to accounting measurement and disclosure: Implications and insights for future research”. In: *Journal of Accounting Research* 54.2, pp. 623–676.
Khapko, M. and M. Zoican (2019). “Do Speed Bumps Curb Speed Investment? Evidence from a Pilot Experiment”. In: *Evidence from a Pilot Experiment (March 12, 2019)*.
Kim, O. and R. E. Verrecchia (1994). “Market liquidity and volume around earnings announcements”. In: *Journal of accounting and economics* 17.1-2, pp. 41–67.
Krinsky, I. and J. Lee (1996). “Earnings announcements and the components of the bid-ask spread”. In: *The Journal of Finance* 51.4, pp. 1523–1535.
Kyle, A. S. (1985). “Continuous auctions and insider trading”. In: *Econometrica*, pp. 1315–1335.
Kyle, A. S. and J. Lee (2017). “Toward a fully continuous exchange”. In: *Oxford Review of Economic Policy* 33.4, pp. 650–675.
Laughlin, G. (2014). *Insights Into High Frequency Trading from the Virtu Initial Public Offering.*
Lee, T. (2019). “Latency in fragmented markets”. In: *Review of Economic Dynamics* 33, pp. 128–153.
Lehar, A. and C. A. Parlour (2019). *Liquidity Demand and Bitcoin Transaction Fees.* Tech. rep. Working paper.

Liu, W.-M. (2009). “Monitoring and limit order submission risks”. In: *Journal of Financial Markets* 12.1, pp. 107–141.

Malinova, K. and A. Park (2017). “Market Design with Blockchain Technology”. In: *Available at SSRN 2785626*.

McNichols, M. and B. Trueman (1994). “Public disclosure, private information collection, and short-term trading”. In: *Journal of Accounting and Economics* 17.1-2, pp. 69–94.

Menkveld, A. J. (2016). “The economics of high-frequency trading: Taking stock”. In: *Annual Review of Financial Economics* 8, pp. 1–24.

Menkveld, A. J. and M. A. Zoican (2017). “Need for speed? Exchange latency and liquidity”. In: *The Review of Financial Studies* 30.4, pp. 1188–1228.

Morris, S. and H. S. Shin (2002). “Social Value of Public Information”. In: *The American Economic Review* 92.5, pp. 1521–1534.

Nakamoto, S. (2008). “Bitcoin: A peer-to-peer electronic cash system”. In:

O’Hara, M. (2015). “High frequency market microstructure”. In: *Journal of Financial Economics* 116.2, pp. 257–270.

Oomen, R. (2017). “Last look”. In: *Quantitative Finance* 17.7, pp. 1057–1070.

Pagano, M. (1989). “Trading volume and asset liquidity”. In: *The Quarterly Journal of Economics* 104.2, pp. 255–274.

Pagnotta, E. S. and T. Philippon (2018). “Competing on speed”. In: *Econometrica* 86.3, pp. 1067–1115.

Pagnotta, E. and A. Buraschi (2018). “An equilibrium valuation of bitcoin and decentralized network assets”. In: *Available at SSRN 3142022*.

Riordan, R. and A. Storkenmaier (2012). “Latency, liquidity and price discovery”. In: *Journal of Financial Markets* 15.4, pp. 416–437.

Rosu, I. (2018). “Dynamic Adverse Selection and Liquidity”. In: *Paris December 2018 Finance Meeting EUROFIDAI-AFFI*.

Saleh, F. (2018). “Blockchain without waste: Proof-of-stake”. In:

Schilling, L. and H. Uhlig (2018). “Some Simple Bitcoin Economics”. In: *NBER Working Paper* No. 24483.

Shkilko, A. and K. Sokolov (2016). “Every cloud has a silver lining: Fast trading, microwave connectivity and trading costs”. In: *SSRN Electronic Journal*.

Sidhu, B., T. Smith, R. E. Whaley, and R. H. Willis (2008). “Regulation fair disclosure and the cost of adverse selection”. In: *Journal of Accounting Research* 46.3, pp. 697–728.

Stoll, H. R. (1989). “Inferring the components of the bid-ask spread: Theory and empirical tests”. In: *the Journal of Finance* 44.1, pp. 115–134.

Straser, V. (2002). “Regulation Fair Disclosure and information asymmetry”. In: *SSRN 311303*.

Szabo, N. (1997). “Formalizing and securing relationships on public networks”. In: *First Monday* 2.9.

Verrecchia, R. E. (1982). “The use of mathematical models in financial accounting”. In: *Journal of Accounting Research*, pp. 1–42.

Verrecchia, R. E. (2001). “Essays on disclosure”. In: *Journal of Accounting and Economics* 32.1-3, pp. 97–180.

Warren, W. and A. Bandeali (2017). “0x: An open protocol for decentralized exchange on the Ethereum blockchain”. In: *URl: https://github. com/0xProject/whitepaper*.

Ye, L. (2016). “Understanding the impacts of dark pools on price discovery”. In: *Available at SSRN 2874957*.

Ye, M. (2011). “A glimpse into the dark: Price formation, transaction cost and market share of the crossing network”. In: *Transaction Cost and Market Share of the Crossing Network (June 9, 2011)*.

Ye, M., C. Yao, and J. Gai (2013). “The externalities of high frequency trading”. In: *SSRN Electronic Journal*.

Zhang, Y., X. Chen, and D. Park (2018). “Formal specification of constant product (xy=k) market maker model and implementation”. In: *manuscript*.

Zhu, H. (2014). “Do dark pools harm price discovery?” In: *The Review of Financial Studies* 27.3, pp. 747–789.

Appendix A

Institutional details of intentional delays

This section briefly describes the institutional details of speed bumps. Although the model in the main text is built to analyze asymmetric speed bumps rather than a symmetric speed bump in IEX, I start this section by explaining IEX speed bump below, as it is a precursor of all other speed bumps.

A.1 Symmetric and deterministic delays

A.1.1 Who is protected?

A symmetric and deterministic speed bump is first adopted by IEX in 2013 and followed by NYSE American (while NYSE American has decided to remove it based on NYSE, 2019). It delays all incoming and outgoing orders by 350 microseconds. The type of orders protected by the speed bump is "Pegged Order." Pegged order is the type of non-displayed limit orders whose price is dynamically adjusted by reference to the national best bid and offer (NBBO).[1] Although IEX imposes a delay on incoming orders and outgoing information, the messaging of SIP-related information is not delayed. Thus, the price of pegged orders is dynamically adjusted by IEX with no delays. If the NBBO changes, a speed bump allows IEX to adjust the pegged orders, and HFTs cannot snipe them at the stale price.

[1]There are three types of pegged-order: Primary Peg (P-Peg), Discretionary Peg (D-Peg), and Midpoint Peg (M-Peg). P-Peg and D-Peg orders are resting at one tick below or above the NBBO. P-Peg orders have discretion to trade at the NBBO, while D-Peg orders have discretion to trade up to the midpoint. M-peg orders stay and are traded at the midpoint of the NBBO and has a higher priority than D-Peg orders at the mid-point. Whether the discretion of each order is exercised is determined by the "IEX signal" that determines if the NBBO is volatile (i.e., "scrambling") by using a specific measure. The discretion is not exercised if the signal is "on," meaning that the bid-ask in the market is volatile.

A.1.2 Speed bump infrastructure

All traders sending messages to IEX must enter the IEX's system from the Point of Presence (POP) in Secaucus, NJ. After entering via the POP, a message sent to IEX travels through a "coiled" fiber optic cable, which has a distance of 38 miles. After exiting the coil, the message travels an additional physical distance to the IEX trading system, located in Weehawken, NJ. Due to this travel distance, a message sent to IEX must incur 350 microseconds of additional travel time.

A.1.3 Asymmetric delays

An asymmetric speed bump has been adopted by a growing number of exchanges. It delays all orders except liquidity-providing orders. Details in the implementation varies depending on exchanges. For example, a speed bump by Chicago Stock Exchange delays all orders except for visible limit orders from approved liquidity providers, while an exception in TSX Alpha is provided to visible Post-Only orders. Post-Only is the type of limit orders that is automatically rejected if it has a potential to cross a market and remove liquidity from the limit order book.

Empirically identifying the impact of speed bumps is not straightforward, as a speed bump typically comes with changes in other market structures and trading rules. For example, along with a speed bump, TSX Alpha sets a minimum size requirement for liquidity providing orders and adopts an inverted maker-taker fee structure. Due to these structural changes, market makers must pay some additional cost (in terms of monetary or risk exposure) in return for protection by a speed bump.

A.1.4 Randomness in delays

Randomizing the length of delay is expected to generate an additional benefit: it mitigates asymmetric information more effectively compared to a deterministic speed bump. The advantage of random speed bumps stems from a situation where an informed trader splits a large order and sends them to multiple exchanges, i.e., "sweep" or "sprayed" orders.

If a speed bump is deterministic, a trader can send split orders to multiple exchanges by adding or subtracting some time lags to synchronize the execution timing of her orders on all exchanges. The simultaneous execution of sweep orders is made possible by the smart order router (SOR) that calculates and predicts the execution timing of each order by incorporating the deterministic delay imposed by a speed bump.

For example, consider an informed trader who wants to fill a large order (say $1,000$ shares). Exchanges A has 300 shares available, and Exchange B has 500 shares. Thus, the informed trader may spray orders to both exchange to fill 1,000 shares. Suppose that it takes t_A and t_B to send and execute orders on Exchanges A and B, respectively. Now, a speed bump is applied in Exchange B. If the length of delay δ is deterministic, the informed trader can stagger the timing or order entry to make $t_A = t_B + \delta$.[2] By the synchronized execution, there is no information leakage, and the trader can fulfill all orders.

Randomness in a speed bump makes it harder for the SOR to predict the execution timing, so that synchronizing executions of split orders does not really work. Due to the failure in the synchronized execution, market makers in Exchange B observe the execution of informed order on Exchange A before a part of split orders arrive at Exchange B. The time lag generated by a random speed bump allows the market makers to cancel or reprice their limit orders to avoid being picked off. Hence, liquidity providers bear less severe adverse selection than in the case with a deterministic speed bump.

[2] In reality, there must involve some unexpected delays due to random factors, such as precipitation and temperature, and even a SOR cannot perfectly synchronize the order arrival timing.

Appendix B

Proofs for Chapter 2

B.1 Derivation of the equilibrium spread

In what follows, I denote $\psi = \phi + \beta + \gamma$ in the single-HFT model and $\psi = \phi_S + \beta + \phi_M$ in the multiple-HFTs model to save the space. Equation (2.1) is rewritten as

$$\begin{aligned} V_M(s) =& \mathbb{E}_{\tilde{\delta},t_0}\left[s\int_{\tilde{\delta}}^{\tilde{\delta}+t_0} \beta e^{-\beta(t-\tilde{\delta})} e^{-\gamma(t-t_0)\mathbb{I}_{\{t\geq t_0\}}} dt \right. \\ & \left. +[\phi(s-\sigma)+\beta s]\int_{t_0+\tilde{\delta}}^{\infty} e^{-\beta(t-\tilde{\delta})} e^{-\gamma(t-t_0)} e^{-\phi(t-t_0-\tilde{\delta})} dt \right]. \end{aligned} \tag{B.1}$$

By letting $b = \delta^{-1}$,

$$\begin{aligned} V_M(s) =& \mathbb{E}_{\tilde{\delta},t_0}\left[\mathbb{I}_{t_0>\tilde{\delta}} s \int_{\tilde{\delta}}^{t_0} \beta e^{-\beta(\tau-\tilde{\delta})} d\tau + s\int_{\max(t_0,\tilde{\delta})}^{\tilde{\delta}+t_0} \beta e^{-\beta(\tau-\tilde{\delta})} e^{-\gamma(\tau-t_0)} d\tau \right. \\ & \left. +[\phi(s-\sigma)+\beta s]\int_{t_0+\tilde{\delta}}^{\infty} e^{-\beta(\tau-\tilde{\delta})} e^{-\gamma(\tau-t_0)} e^{-\phi(\tau-t_0-\tilde{\delta})} d\tau \right] \\ =& \mathbb{E}_{t_0}\left[\int_0^{t_0} be^{-b\tilde{\delta}} s(1-e^{-\beta(t_0-\tilde{\delta})}) d\tilde{\delta} + \int_0^{\infty} be^{-b\tilde{\delta}} s e^{\beta\tilde{\delta}} e \frac{\beta}{\beta+\gamma}(e^{-(\beta+\gamma)\max(t_0,\tilde{\delta})} - e^{-(\beta+\gamma)(\tilde{\delta}+t_0)}) d\tilde{\delta} \right] \\ & + \frac{1}{\psi}\frac{z}{z+\beta}\frac{1}{1+\delta\gamma}[\phi(s-\sigma)+\beta s]. \end{aligned}$$

The first line is

$$\begin{aligned} \text{First line} =& s\mathbb{E}_{t_0}\left[1 - e^{-bt_0} - e^{-\beta t_0}\frac{b}{b-\beta}(1-e^{-(b-\beta)t_0}) + b\frac{\beta}{b-\beta}\frac{1}{b+\gamma}(e^{-\beta t_0} - e^{-bt_0}) \right] \\ =& \beta s \frac{b}{z+b}\frac{(b+1)+z}{b+\gamma}\frac{1}{\beta+z}. \end{aligned}$$

Thus, in aggregate,

$$V_M(s) = \frac{1}{\psi}\frac{z}{z+\beta}\frac{1}{1+\delta\gamma}\left[\phi(s-\sigma) + \beta s + \psi\beta s\frac{z^{-1}[(1+\delta)+\delta z]}{\delta z+1} \right],$$

which leads to the benchmark spread with $\lambda = \frac{z^{-1}[(1+\delta)+\delta z]}{\delta z+1}$.

B.2 Proof of Proposition 2.2

The cross-derivative of the FOC, evaluated at the optimal speed ϕ^*, is reduced to

$$\begin{aligned}\frac{\partial}{\partial\lambda}\left(\frac{dW_{HFT}}{d\phi}\right)|_{\phi=\phi^*} &= 2\beta^2\phi\frac{1+\lambda\eta}{K^3} - \gamma\frac{d\delta}{d\lambda}C'(\phi)\\ &= 2\phi\left[\beta^2\frac{1+\lambda\eta}{(\phi+\beta(1+\lambda\psi))^3} - c\gamma\frac{d\delta}{d\lambda}\right]\end{aligned} \tag{B.2}$$

where $K \equiv \phi + \beta(1+\lambda\psi)$, and the second line uses the quadratic cost function.

The first term of (B.2) is bounded from above, and there exists $L < \infty$ such that

$$\forall c,\ L > \beta^2\frac{1+\lambda\eta}{(\phi+\beta(1+\lambda\psi))^3}.$$

In contrast, $c\gamma\frac{d\delta}{d\lambda}$ is monotonically increasing in c and not bounded from above. Therefore, there exists $\bar{c}$, such that

$$c > \bar{c} \Rightarrow c\gamma\frac{d\varepsilon}{d\lambda} > L \Leftrightarrow \frac{d\phi^*}{d\lambda} < 0.$$

Next, we know from the FOC that $\phi^*|_{c=0} < \infty$, as long as $\lambda < \sqrt{\beta(\beta+\gamma)}$. Therefore, the first term of (B.2) is bounded from below, meaning that there exists $\underline{c}$ such that

$$c < \underline{c} \Rightarrow \frac{d\phi^*}{d\lambda} > 0.$$

B.3 General proof of Proposition 2.3

Denote that $s^* = s(\phi^*(\delta),\delta)$. By the chain rule,

$$\frac{ds^*}{d\delta} = \frac{\partial s}{\partial\delta} + \frac{d\phi^*}{d\delta}\frac{\partial s}{\partial\phi}.$$

The optimal speed ϕ^* satisfies the FOC, denoted as $H^*(\delta) = H(\phi^*(\delta),\delta) = 0$ with

$$H(\phi,\delta) = (\sigma - s(\phi,\delta))\frac{d\pi_{ND}(\phi)}{d\phi} - \pi_{ND}(\phi)\frac{\partial s(\phi,\delta)}{\partial\phi}.$$

By taking the derivative of H^* with respect to δ,

$$
\begin{aligned}
0 &= \frac{d\phi^*}{d\delta}\frac{\partial H^*}{\partial \phi} + \frac{\partial H^*}{\partial \delta} \\
&= \frac{d\phi^*}{d\delta}\left(-\frac{\partial s}{\partial \phi}(\phi^*, \delta)\frac{d\pi_{ND}(\phi)}{d\phi} + (\sigma - s(\phi^*, \delta))\frac{d^2\pi_{ND}(\phi)}{d\phi^2} - \frac{d\pi_{ND}(\phi)}{d\phi}\frac{\partial s(\phi^*, \delta)}{\partial \phi} - \pi_{ND}(\phi)\frac{\partial^2 s(\phi^*, \delta)}{\partial \phi^2}\right) \\
&- \frac{\partial s(\phi^*, \delta)}{\partial \delta}\frac{d\pi_{ND}(\phi)}{d\phi} - \pi_{ND}(\phi)\frac{\partial}{\partial \delta}\frac{\partial s(\phi, \delta)}{\partial \phi} \\
&= -\left[\pi_{ND}(\phi)\frac{\partial}{\partial \delta}\frac{\partial s(\phi, \delta)}{\partial \phi} + \frac{d\phi^*}{d\delta}\left(\frac{d\pi_{ND}(\phi)}{d\phi}\frac{\partial s(\phi^*, \delta)}{\partial \phi} + \pi_{ND}(\phi)\frac{\partial^2 s(\phi^*, \delta)}{\partial \phi^2}\right)\right] \\
&+ \frac{d\phi^*}{d\delta}(\sigma - s(\phi^*, \delta))\frac{d^2\pi_{ND}(\phi)}{d\phi^2} - \frac{ds^*}{d\delta}\frac{d\pi_{ND}(\phi)}{d\phi} \\
&= \frac{d}{d\delta}\left(-\pi_{ND}\frac{\partial s(\phi^*, \delta)}{\partial \phi} + (\sigma - s(\phi^*, \delta))\frac{d\pi_{ND}(\phi^*)}{d\phi}\right) - \frac{ds^*}{d\delta}\frac{d\pi_{ND}(\phi)}{d\phi} \\
&= -\frac{ds^*}{d\delta}\frac{d\pi_{ND}(\phi)}{d\phi}.
\end{aligned}
$$

The last equality comes from the FOC. Thus, the result in Proposition 2.3 is robust as long as the sniping probability $\pi(\delta, \phi)$ is separable, i.e., it is decomposed into the no-delay sniping probability π_{ND} and the delay effect $h(\delta)$. The separability holds if a delay is independent of the learning process of the HFT.

B.4 Proof of Proposition 2.4

Uniqueness of the mixed strategy

Firstly, suppose that HFM i puts a positive weight on $s_i = \sigma$. For $s_i = \sigma$ to obtain a positive weight, HFT j must charge prices above σ, which is not an equilibrium. Therefore, $s_i = \sigma$ cannot be an atom.

Secondly, suppose that HFT i puts positive wight w on $p \in (s_0, \sigma)$. For this to be an equilibrium, there must exist positive ϵ such that HFT j does not charge prices in $[s_0, p+\epsilon]$. If not, HFT j can exploit the profit discontinuity at p, and she undercuts HFT i to obtain positive profits. Also, if HFT j charges prices below p, it is not optimal for HFT i to put a positive weight on p. Thus, HFT j must charge prices above $p + \varepsilon$. In this case, however, it is optimal for HFT i to raise p.

Finally, suppose that $p = s_0$ has a positive weight. For $s_j \geq p = s_0$, the profits for

HFM j satisfy

$$
\begin{aligned}
V_{M,j}(s_j) &\sim \frac{\phi_S + \beta(1+\lambda\psi)}{\psi}(s_j - s_0)\Pr(s_j < s_i) \\
&+ \left[\frac{\phi_S + \beta(1+\lambda\psi)}{\psi}(s_j - s_0) - (1-g)\frac{\beta(1+\lambda\psi)}{\psi}s_j\right]\Pr(s_j = s_i) \\
&+ \left[\frac{\phi_S + \beta(1+\lambda\psi)}{\psi}(s_j - s_0) - \frac{\beta(1+\lambda\psi)}{\psi}s_j\right]\Pr(s_j > s_i) \\
&= \frac{\phi_S + \beta(1+\lambda\psi)}{\psi}(s_j - s_0) - \frac{\beta(1+\lambda\psi)}{\psi}s_j\left[\Pr(s_j > s_i) + (1-g)\Pr(s_j = s_i)\right] \\
&< \frac{\phi_S + \beta(1+\lambda\psi)}{\psi}(s_j - s_0) - \frac{\beta(1+\lambda\psi)}{\psi}s_j\Pr(s_j > s_i).
\end{aligned}
$$

Then, we can define

$$
\epsilon \equiv \frac{\beta(1+\lambda\psi)s_0\Pr(s_j > s_i)}{\phi_S + \beta(1+\lambda\psi)(1-\Pr(s_j > s_i))} > 0
$$

so that posting $s_j \in [s_0, s_0 + \epsilon]$ makes $V_{M,j}(s_j) < 0$. Thus, HFT j does not post prices in $[s_0, s_0+\epsilon]$. However, this implies that HFT i has an incentive to raise p from $p = s_0$ to $p = s_0 + \epsilon$, leading to the same discussion as the case with $p \in (s_0, 1)$. From above argument, if HFT i randomizes quote over $[s_0, \sigma]$ with an atom p, then p converges to σ. However, this contradict to the first case that shows $p = \sigma$ cannot be an atom.

Comparative statics

In what follows, the comparative statics are analyzed by variating λ instead of δ because it affects the equilibrium spread only via $\lambda(\delta)$ and $\lambda'(\delta) > 0$. Also, I use $\phi_i = \phi_S$ and $\phi_{-i} = \phi_M$ to make the correspondence of HFTs' role and their speed clear.

Firstly, note that $\log x$ has the following properties for general $x \in (0,1)$.

$$
\log x + 1 - x < 0, \tag{B.3}
$$

$$
2(\log x + 1 - x) + (1-x)^2 < 0. \tag{B.4}
$$

Also, I denote

$$
X = \log s_0 + 1 - s_0.
$$

Next, for later use, compute the following derivatives and cross-derivatives: with $\eta = \phi_M + \beta$,

$$
\frac{d\bar{s}(\phi_S, \phi_M)}{d\phi_S} = -\frac{1+\lambda\eta}{\beta}\frac{X}{(1+\lambda\psi)^2} > 0, \tag{B.5}
$$

$$
\frac{d^2\bar{s}(\phi_S, \phi_M)}{d\phi_S^2} = -\frac{1+\lambda\eta}{\beta\phi_S}\left[\frac{(1-s_0)^2(1+\lambda\eta) - 2\lambda\phi_S X}{(1+\lambda\psi)^3}\right] < 0,
$$

$$\frac{\partial \bar{s}(\phi_S,\phi_M)}{\partial \phi_M} = \lambda \frac{\phi_S}{\beta}\frac{X}{(1+\lambda\psi)^2} < 0,$$

$$\frac{\partial}{\partial \phi_M}\left(\frac{d\bar{s}(\phi_S,\phi_M)}{d\phi_S}\right) = \frac{\lambda}{\beta}\frac{1}{(1+\lambda\psi)^3}\left[X(1+\lambda\beta)+(1+\lambda\eta)(1-s_0)^2\right],$$

$$\frac{d\bar{s}(\phi_S,\phi_M)}{d\lambda} = \psi\frac{\phi_S}{\beta}\frac{X}{(1+\lambda\psi)^2} < 0.$$

$$\frac{\partial}{\partial \lambda}\left(\frac{d\bar{s}(\phi_S,\phi_M)}{d\phi_S}\right) = \frac{1}{\beta(1+\lambda\psi)^3}\left[[\phi_S+\psi(1+\lambda\eta)]X+(1+\lambda\eta)\psi(1-s_0)^2\right].$$

Also, $\phi_S^* = BR(\phi_M)$ satisfies the FOC:

$$0 = H(\phi_S,\phi_M,\lambda) = \frac{d\pi_{ND}}{d\phi_S}(1-\bar{s}(\phi_S,\phi_M)) - \pi_{ND}\frac{d\bar{s}(\phi_S,\phi_M)}{d\phi_S}. \tag{B.6}$$

The SOC is computed as

$$\begin{aligned}\partial H/\partial \phi_S &= (1-\bar{s}(\phi_S,\phi_M))\frac{d^2\pi_{ND}}{d\phi_S^2} - 2\frac{d\pi_{ND}}{d\phi_S}\frac{d\bar{s}(\phi_S,\phi_M)}{d\phi_S} - \pi_{ND}\frac{d^2\bar{s}(\phi_S,\phi_M)}{d\phi_S^2}\\ &= -2\frac{1}{\psi}\frac{d\bar{s}(\phi_S,\phi_M)}{d\phi_S} - \frac{\phi_S}{\psi}\frac{d^2\bar{s}(\phi_S,\phi_M)}{d\phi_S^2}\\ &= \frac{1}{\psi}\frac{(1+\lambda\eta)^2}{\beta(1+\lambda\psi)^3}\left[2X+(1-s_0)^2\right] < 0, \end{aligned} \tag{B.7}$$

where the last inequality uses (B.4).

Impact of λ on $\phi_S^* = BR(\phi_M)$ Since (B.7) holds, it suffices to show that $\partial H/\partial\lambda > 0$.

$$\begin{aligned}\partial H/\partial \lambda &= -\left(\frac{d\pi_{ND}}{d\phi_S}\frac{\partial \bar{s}(\phi_S,\phi_M)}{\partial \lambda} + \pi_{ND}\frac{\partial}{\partial\lambda}\frac{d\bar{s}(\phi_S,\phi_M)}{d\phi_S}\right)\\ &= -\frac{\phi_S}{\beta}\frac{1+\lambda\eta}{(1+\lambda\psi)^3}\left[2X+(1-s_0)^2\right]\\ &> 0, \end{aligned} \tag{B.8}$$

where the last inequality comes from (B.4).

Impact of ϕ_j on $\phi_S^* = BR(\phi_M)$ Since (B.7) holds, it suffices to show that $\partial H/\partial\phi_M > 0$. The cross derivative of FOC is

$$\begin{aligned}\partial H/\partial \phi_M = &-\frac{d\pi_{ND}}{d\phi_S}\frac{\partial \bar{s}(\phi_S,\phi_M)}{\partial \phi_M} - \pi_{ND}\frac{\partial}{\partial\phi_M}\frac{d\bar{s}(\phi_S,\phi_M)}{d\phi_S} + \frac{\partial}{\partial\phi_M}\frac{d\pi_{ND}}{d\phi_S}(1-\bar{s}(\phi_S,\phi_M))\\ &- \frac{\partial\pi_{ND}}{\partial\phi_M}\frac{d\bar{s}(\phi_S,\phi_M)}{d\phi_S}.\end{aligned}$$

By using the FOC, the first and the last terms sum up to

$$A_0 = -\frac{d\pi_{ND}}{d\phi_S}\frac{\partial \bar{s}(\phi_S,\phi_M)}{\partial \phi_M} - \frac{\partial \pi_{ND}}{\partial \phi_M}\frac{d\bar{s}(\phi_S,\phi_M)}{d\phi_S} = -\frac{\phi_S}{\beta\psi}\frac{X}{\psi(1+\lambda\psi)^2}(1+2\lambda\eta).$$

Also, the middle two terms amount to

$$\begin{aligned} A_1 &= -\pi_{ND}\frac{\partial}{\partial \phi_M}\frac{d\bar{s}(\phi_S,\phi_M)}{d\phi_S} + (1-\bar{s})\frac{\partial}{\partial \phi_M}\frac{d\pi_{ND}}{d\phi_S} \\ &= \pi_{ND}\left[-\frac{\partial}{\partial \phi_M}\frac{d\bar{s}(\phi_S,\phi_M)}{d\phi_S} + \left(\frac{d\pi_{ND}}{d\phi_S}\right)^{-1}\frac{d\bar{s}(\phi_S,\phi_M)}{d\phi_S}\frac{\partial}{\partial \phi_M}\frac{d\pi_{ND}}{d\phi_S}\right] \\ &= \frac{\pi_{ND}}{(1+\lambda\psi)^3}\left[-\frac{\lambda}{\beta}X\,(1+\lambda\beta) - \frac{1}{\beta}\lambda(1+\lambda\eta)(1-s_0)^2 + \frac{(1+\lambda\eta)(1+\lambda\psi)}{\psi\eta}X\right]. \end{aligned}$$

Therefore, in aggregate, $\partial H/\partial \phi_M = \frac{\tilde{\pi}_S}{(1+\lambda\psi)^3}A$ with

$$\begin{aligned} A &= -\frac{1}{\beta\psi\eta}\left\{X\left[\lambda\psi\eta\,(1+\lambda\beta) + \left(\phi(1+\lambda\eta)+\lambda\eta^2\right)(1+\lambda\psi)\right] + \psi\eta\lambda(1+\lambda\eta)(1-s_0)^2\right\} \\ &> \frac{1}{2}\frac{1}{\beta\psi\eta}\left\{\left[\lambda\psi\eta\,(1+\lambda\beta) + \left(\phi(1+\lambda\eta)+\lambda\eta^2\right)(1+\lambda\psi)\right] - 2\psi\eta\lambda(1+\lambda\eta)\right\}(1-s_0)^2 \\ &> 0 \end{aligned}$$

where the second inequality comes from (B.4).

B.5 Analysis on strategic complementarity in Proposition 2.5

Consider the cross-derivative of the FOC (2.18) with respect to the rivals' speed ϕ_{-i}.

$$\begin{aligned} \frac{\partial^2 W_i}{\partial \phi_{-i}\partial \phi_i} = &\left[\overbrace{\frac{\partial(1-\frac{\bar{s}}{\sigma})}{\partial \phi_{-i}}\frac{\partial \pi_{ND}}{\partial \phi_i} + \pi_{ND}\frac{\partial^2(1-\frac{\bar{s}}{\sigma})}{\partial \phi_{-i}\partial \phi_i}}^{\text{effects (i) + (ii)>0}}\right] h(\delta,\phi_{-i}) + C'(\phi_i)\frac{1}{h}\overbrace{\frac{dh(\delta,\phi_{-i})}{d\phi_{-i}}}^{\text{effect (iii)<0}} \\ &+ \left[\overbrace{\frac{\partial \pi_{ND}}{\partial \phi_{-i}}\frac{\partial(1-\frac{\bar{s}}{\sigma})}{\partial \phi_i}}^{\text{effect (iv)}} + \overbrace{\frac{\partial^2 \pi_{ND}}{\partial \phi_{-i}\partial \phi_i}(1-\frac{\bar{s}}{\sigma})}^{\text{effect (v)}}\right] h(\delta,\phi_{-i}). \end{aligned} \tag{B.9}$$

As ϕ_{-i} increases, HFMs become less likely to be picked off. It affects HFT i's expected sniping profit through three channels: (a) by changing the behavior of the spread, (b) by strengthening the delay effect, $h(\delta,\phi_{-i})$, on the sniping probability, (c) and by reducing the no-delay sniping probability $\pi_{ND}(\phi_i, \phi_{-i})$. The third channel is new in the general model: even without a delay (i.e., $h = 1$), faster HFTs $-i$ makes it more difficult for HFT i to snipe.

Note that the first two channels, (a) and (b), bring about the same effect as an extension of a delay (δ). Hence, equation (B.9) has effects (i)-(iii) that replicate equation (2.6) in the benchmark model. The envelope condition with $C' = 0$ eliminates effect (iii).

Furthermore, channel (c) generates effects (iv) and (v) in equation (B.9). Effect (iv) implies that HFT i does not need to care about the adverse price movement caused by her speed-up. Even if she becomes faster, and the trading profit worsens, her *expected* profit is slightly affected because it is less likely to materialize due to her rivals' speed-up (ϕ_{-i}). Effect (v) captures the ambiguous effect of rivals' speed on the slope of the no-delay sniping probability, π_{ND}. When rivals are equipped with very high speed ϕ_{-i}, HFT i's speed-up does not really improve her sniping probability because an increase in ϕ_i is dwarfed by a large ϕ_{-i}. Hence, it holds that $\frac{\partial^2 \pi_{ND}}{\partial \phi_{-i} \partial \phi_i} < 0$ if ϕ_{-i} is large, and vice versa.

Although the sign of effect (v) is ambiguous by itself, effects (iv) and (v) are positive in total.

Lemma B.1. *In the equilibrium, the sum of effects (iv) and (v) is positive.*

Proof. The sum of effects (iv) and (v) is rewritten as

$$\begin{aligned}\frac{\partial \pi_{ND}}{\partial \phi_j}\frac{\partial(\sigma - \mathbb{E}[s_j])}{\partial \phi_i} + \frac{\partial^2 \pi_{ND}}{\partial \phi_j \partial \phi_i}(\sigma - \mathbb{E}[s_j]) &= \frac{\sigma - \mathbb{E}[s_j]}{\pi_{ND}}\left[-\left(\frac{\partial \pi_{ND}}{\partial \phi_i}\right)\left(\frac{\partial \pi_{ND}}{\partial \phi_j}\right) + \pi_i \frac{\partial^2 \pi_{ND}}{\partial \phi_j \partial \phi_i}\right] \\ &= \frac{\sigma - \mathbb{E}[s_j]}{\pi_{ND}}\frac{\phi_i^2}{\psi^4} > 0.\end{aligned}$$

□

Intuitively, effect (v) cannot be a large negative to dominate effect (iv) in the symmetric equilibrium, where all HFTs have the same speed. Overall, effects (iv) and (v) help the first two effects with generating strategic complementarity.

B.6 Proof of Propositions 2.6 and 2.7

Denote the speed level in the symmetric equilibrium as ϕ. Firstly, the FOC $H(\phi, \phi, \lambda) = 0$, defined by (B.6), is reduced to

$$0 = \hat{H}(\phi, \lambda) = \beta\eta(1 + \lambda\psi)^2 + \phi\psi(1 + \lambda\eta)(1 - s_0) + \phi(\phi + 2\eta(1 + \lambda\psi)) \log s_0.$$

It is easy to check $\hat{H}(0, \lambda) = \beta^2(1 + \lambda\beta)^2 > 0$. Also, at $\phi \to \infty$, $1 - s_0$ and $\log s_0$ converge to some constant values, and $\hat{H}(\phi, \lambda)$ can be written as a cubic function of ϕ with a negative coefficient on ϕ^3, meaning that $\lim_{\phi \to \infty} \hat{H}(\phi, \lambda) < 0$. As shown in the following,

$\hat{H}(\phi, \lambda)$ is a monotonically decreasing function of ϕ. Thus, the above argument implies that $\hat{H} = 0$ and $H = 0$ have a unique solution.

Next, by using the derivatives in Appendix B.4,

$$\frac{\partial H}{\partial \phi} = \frac{\psi - 4\eta}{\psi^3}(1 + \frac{\phi}{\beta(1+\lambda\psi)} \log s_0) + \frac{\beta + \eta + 2\lambda\eta\beta + \lambda\phi\psi}{\beta\psi^2}\frac{X}{(1+\lambda\psi)^2} + \frac{1}{\psi}\frac{1+\lambda\eta}{\beta}\frac{(1+\lambda\beta)(1-s_0)^2 - 4\phi\lambda X}{(1+\lambda\psi)^3}.$$

Inequalities (B.3) and (B.4) imply that $\frac{\partial H}{\partial \phi} < 0$. Since Appendix B.4 shows that $\frac{\partial H}{\partial \lambda} > 0$, the implicit function theorem implies $\frac{d\phi}{d\lambda} > 0$, where the explicit formula is

$$\frac{d\phi}{d\lambda} = \frac{\psi^3\phi(1+\lambda\eta)\left[2X + (1-s)^2\right]}{\beta(\psi - 4\eta)(1+\lambda\psi)^3(1 + \frac{\phi}{\beta(1+\lambda\psi)}\log s_0) + a_X X + \psi^2(1+\lambda\eta)(1+\lambda\beta)(1-s)^2]}, \tag{B.10}$$

with

$$a_X = \psi(1+\lambda\psi)[\beta + \eta + 2\lambda\eta\beta + \lambda\phi\psi] - 4\psi^2\phi\lambda(1+\lambda\eta).$$

Finally, we want to show that the expected spread is increasing in λ. At the symmetric equilibrium,

$$\begin{aligned} \frac{d\bar{s}}{d\lambda} &= \overbrace{\psi\frac{\phi}{\beta}\frac{1 - s_0 + \log s_0}{(1+\lambda\psi)^2}}^{=\frac{\partial \bar{s}}{\partial \lambda}} + \frac{d\phi}{d\lambda}\overbrace{\left[-\frac{1+\lambda\beta}{\beta}\frac{\log s_0 + 1 - s_0}{(1+\lambda\psi)^2}\right]}^{=\frac{\partial \bar{s}}{\partial \phi}} \\ &= \frac{X}{\beta(1+\lambda\psi)^2}\left[\psi\phi - (1+\lambda\beta)\frac{d\phi}{d\lambda}\right]. \end{aligned}$$

Since $X < 0$, inequality $\frac{d\bar{s}}{d\lambda} > 0$ is reduced to

$$\frac{d\phi}{d\lambda} > \frac{\psi\phi}{1+\lambda\beta}.$$

By using the explicit formula (B.10), the above condition is reduced to

$$\begin{aligned} \left[2\psi^2(1+\lambda\eta)(1+\lambda\beta) - a_X\right] X &< \beta(\phi - 3\eta)(1+\lambda\psi)^3(1 - \mathbb{E}[s]) \\ &= -\beta(\phi - 3\eta)(1+\lambda\psi)^3\frac{\phi\psi}{\beta\eta}(1+\lambda\eta)\frac{X}{(1+\lambda\psi)^2} \end{aligned}$$

where the last equality comes from the FOC. As a result, the condition is

$$\begin{aligned} \left[2\psi^2(1+\lambda\eta)(1+\lambda\beta) - a_X\right] X &< (3\eta - \phi)(1+\lambda\psi)\frac{\phi\psi}{\eta}(1+\lambda\eta)X \\ \therefore 2\psi^2(1+\lambda\eta)(1+\lambda\beta) - a_X &> (3\eta - \phi)(1+\lambda\psi)\frac{\phi\psi}{\eta}(1+\lambda\eta). \end{aligned}$$

We can compute that

$$
\begin{aligned}
\text{LHS} - \text{RHS} &= (1+\lambda\psi)\frac{\phi^2}{\eta} + 4\phi\lambda\psi(1+\lambda\eta) \\
&> 0.
\end{aligned}
$$

This inequality completes the proof.

B.7 Case of other distributions for delays

Suppose that $\delta \in [0, \bar{\delta}]$ has cdf $Q(\delta)$. Note that we allow $\bar{\delta} \to \infty$. From the same argument as the baseline model, the break-even spread is given by

$$
s = \frac{\phi}{\beta(1+\lambda\psi)+\phi}\sigma
$$

where

$$
\lambda(Q) = \frac{1}{\kappa^{-1}-\gamma}\left(1 - \frac{\mathbb{E}_Q[e^{-\kappa^{-1}\delta}]}{\mathbb{E}_Q[e^{-\gamma\delta}]}\right).
$$

Without loss of generality, assume that $\kappa\gamma < 1$.

Now, we compare two random variables with different distributions in a same class: $\delta \sim Q$ and $\hat{\delta} \sim \hat{Q}$ with $\mathbb{E}_Q[\delta] > \mathbb{E}_{\hat{Q}}[\hat{\delta}]$. Then, the restriction for the distribution Q that makes the generalized model with Q identical to the baseline model is $\lambda(Q) > \lambda(\hat{Q})$, which is identical to

$$
\frac{\mathbb{E}_{\hat{Q}}[e^{-\kappa^{-1}\delta}]}{\mathbb{E}_{\hat{Q}}[e^{-\gamma\delta}]} > \frac{\mathbb{E}_Q[e^{-\kappa^{-1}\delta}]}{\mathbb{E}_Q[e^{-\gamma\delta}]}.
$$

Therefore, as long as two distribution of speed bumps, Q and $\hat{Q}$, such that $\mathbb{E}_Q[\delta] > \mathbb{E}_{\hat{Q}}[\hat{\delta}]$, satisfies the above inequality, the results in the main model are robust.

For example, if Q is the uniform distribution, i.e., $Q = U[0, \bar{\delta}]$, then,

$$
\lambda = \frac{1}{\kappa^{-1}-\gamma}\left(1 - \gamma\kappa\frac{1-e^{-\kappa^{-1}\bar{\delta}}}{1-e^{-\gamma\bar{\delta}}}\right).
$$

Note that increasing $\bar{\delta}$ corresponds to a longer speed bump. It is straightforward to check that λ is monotonically increasing in $\bar{\delta}$, leading to the same results in the benchmark model.

Appendix C

Proofs for Chapter 3

C.1 Proof of Proposition 3.2

Equation (3.15) is

$$\alpha = \frac{S(\beta_{buy}, \beta_{sell}, \alpha)/\sigma}{2} = \frac{1}{2}\left[\frac{(1-\beta_{buy})\eta}{(1-\beta_{buy})\eta + z(1-\alpha)(1-\eta)} + \frac{(1-\beta_{sell})\eta}{(1-\beta_{sell})\eta + z(1-\alpha)(1-\eta)}\right]. \tag{C.1}$$

It holds that $S(\beta_{buy}, \beta_{sell}, 1) = 2\sigma$. Thus, the above equation has $\alpha = 1$ as a solution. Also, from the indifference conditions for informed traders, $0 < \beta_i < 1$ for all $\alpha \in [0,1]$.

Thus, the above equation has the second solution in $\alpha \in (0,1)$ if and only if

$$\frac{\partial S}{\partial \alpha}|_{\alpha=1} > 2\sigma.$$

It holds that

$$\frac{1}{\sigma}\frac{\partial S}{\partial \alpha}|_{\alpha=1} = z\frac{1-\eta}{\eta}\left(\frac{1}{1-\beta_{buy}} + \frac{1}{1-\beta_{sell}}\right) > 2z\frac{1-\eta}{\eta}.$$

Therefore, a sufficient condition for the existence of the second (interior) solution is $z\frac{1-\eta}{\eta} > 1$. The negative impact of β_i on α in the partial equilibrium is straightforward.

C.2 Proof of Proposition 3.3

Define $\gamma_{buy} \equiv \beta_{buy}/C$. Then, γ_{buy} is the (smaller) solution of $\Gamma(\gamma;\sigma) = 0$ where

$$\Gamma(\gamma;\tilde{\sigma}) = \gamma^2(1+\tilde{\sigma})\eta - \gamma\left(\frac{(1+\tilde{\sigma})\eta + z(1-\alpha)(1-\eta)}{C} + \sigma\eta\right) + \frac{\sigma\eta}{C}. \tag{C.2}$$

Given α, it holds that

$$\frac{\partial \Gamma}{\partial C} = \frac{\gamma\left[(1+\sigma)\eta + z(1-\alpha)(1-\eta)\right] - \sigma\eta}{C^2}.$$

Thus, the above partial derivative is negative if and only if

$$\gamma < \gamma_0 \equiv \frac{\sigma\eta}{(1+\sigma)\eta + z(1-\alpha)(1-\eta)}.$$

It is easy to check that

$$\Gamma(\gamma_0;\sigma) = -z\left(\frac{\sigma\eta}{(1+\sigma)\eta + z(1-\alpha)(1-\eta)}\right)^2 (1-\alpha)(1-\eta) < 0$$

meaning that $\gamma_{buy} < \gamma_0$. Therefore, for all α, $\gamma_{buy}(\alpha)$ is decreasing in C. The symmetric argument can be applied to show that $\gamma_{sell}(\alpha) = \beta_{sell}/C$ is also a decreasing function of C.

Proof for the existence of a unique solution in $\alpha \in (0,1)$ follows the proof of Proposition 3.2. We denote the solution as α^*.

Next, the equilibrium condition (C.1) is rewritten as

$$2\sigma\alpha = \frac{\gamma_{buy}(\alpha)}{1-\gamma_{buy}(\alpha)} + \frac{\gamma_{sell}(\alpha)}{1+\gamma_{sell}(\alpha)}.$$

The RHS is an increasing function of α and crosses the LHS from above at α^*. Since γ_i is decreasing in C for all α, the RHS shifts downward when α increases. Thus, α^* is a decreasing function of C.

Finally, note that the equilibrium bid-ask spread satisfies $2\sigma\alpha^* = S(\beta^*_{buy}, \beta^*_{sell}, \alpha^*)$, meaning that the bid-ask spread becomes narrower when C increases, as α^* declines.

C.3 Proof of Proposition 3.4

We can rewrite the market maker's net profits from informed trading on the DEX as follows.

$$\pi^D_{M,IT} = \frac{\beta_{buy}}{C}\left[\frac{C}{C-\beta_{buy}} - (1+\sigma)\right] + \frac{\beta_{sell}}{C}\left[(1-\sigma) - \frac{C}{C+\beta_{sell}}\right].$$

In the equilibrium, the indifference condition for an informed trader on the ask side imply that $\frac{C}{C-\beta_{buy}} = 1+\sigma\frac{\eta(1-\beta_{buy})}{\eta(1-\beta_{buy})+(1-\eta)(1-\alpha)z} < 1+\sigma$. Similarly, on the bid side, $\frac{C}{C+\beta_{sell}} > 1-\sigma$. Therefore, $\pi^D_{M,IT} < 0$.

From equation (C.2), $\gamma_i = \beta_i/C$ is decreasing in C and

$$\gamma_{buy,\infty} \equiv \lim_{C\to\infty} \gamma_{buy} = \frac{\sigma}{1+\sigma}$$
$$\gamma_{sell,\infty} \equiv \lim_{C\to\infty} \gamma_{sell} = \frac{\sigma}{1-\sigma}.$$

Therefore,

$$\lim_{C\to\infty} \pi^D_{M,IT} = 0.$$

The explicit formula for the net expected profit from liquidity trading is

$$N(\alpha, C) = \mathbb{E}\left[\frac{C}{C - \alpha\Delta z} - 1\right] = \frac{C}{\alpha^2 z^2}\left[\log\left(\frac{C+\alpha z}{C-\alpha z}\right)^{\alpha z}\left(\frac{C^2 - \alpha^2 z^2}{C^2}\right)^{C}\right] - 1.$$

The first order derivative with respect to C/α is negative. The impact of C on N in the sense of total derivative is negative because the equilibrium α is decreasing in C.

C.4 Sequential execution of orders at DEX

In the model, we assume that all market orders arriving at the DEX are simultaneously executed all at once. In this Appendix, we show that executing all at once (AAO) is the same as the sequential order execution.

Equivalence of post-trade liquidity pools Suppose that there are n informed traders, and each of them has measure $w = \frac{1}{n}$ and places δ unit of market buy order to the DEX (in the model, we assume $\delta = 1$). Note that the aggregate trading is of size δ. The initial state of the liquidity pool is denoted as (C_0, X_0) with $k \equiv C_0 X_0$. Note that the following discussion can be easily extended to the case with liquidity traders.

The first transaction is executed at price

$$p_1 = \frac{C_0}{X_0 - \delta w}$$

and the liquidity pool becomes

$$\begin{aligned} C_1 &= C_0 + p_1 \delta w = C_0 \frac{X_0}{X_0 - \delta w}, \\ X_1 &= X_0 - \delta w. \end{aligned}$$

By iterating, we obtain the following transition equations for the liquidity pools: for general $i = 1, 2, \cdots, n$,

$$\begin{aligned} C_i &= C_{i-1} \frac{X_{i-1}}{X_{i-1} - \delta w}, \\ X_i &= X_{i-1} - \delta w. \end{aligned}$$

The above equations imply that, after all (n) transactions are completed, the liquidity pools have

$$\begin{aligned} X_n &= X_0 - n\delta w = X_0 - \delta, \\ C_n &= C_0 \frac{X_0}{X_0 - n\delta w} = C_0 \frac{X_0}{X_0 - \delta}. \end{aligned}$$

Thus, the post-trade state of the pools with sequential execution is the same as that of AAO execution. The above result also implies that the profits for the market makers on the DEX stay the same even if we consider sequential execution of orders.

Equivalence of the execution price Next, consider the expected trading cost (i.e., the execution price) for an informed trader. We consider a continuum of traders with measure β (by setting $n \to \infty$ with $\delta = 1$ and $w = \beta/n$ in the above example) and assume that traders' orders are independently executed following a Poisson process. Suppose that $y \in [0, \beta)$ orders have been executed before an informed trader gets to execute her order. From the above discussion, her order faces the following liquidity pools.

$$C_y = C_0 \frac{X_0}{X_0 - y}, \ X_y = X_0 - y.$$

Since her order is infinitesimal, it is executed at price

$$p(y) = \frac{C_y}{X_y} = \frac{C_0 X_0}{(X_0 - y)^2}.$$

Due to the independent Poisson process, $y \sim U[0, \beta]$. Thus, the expected execution price is given by

$$p = \frac{1}{\beta} \int_0^\beta p(y) dy = \frac{C_0}{X_0 - \beta},$$

which is identical to the execution price of each order in the case with AAO trade execution.