ABSTRACT

Title of dissertation: ESSAYS ON MARKET MICROSTRUCTURE AND HIGH FREQUENCY TRADING

Wei Li, Doctor of Philosophy, 2014

Directed by: Charles E. Smith Chair Professor Albert S. Kyle
Department of Finance
Robert H. Smith School of Business

This dissertation includes two chapters on topics related to market microstructure and high frequency trading.

In the first chapter, I explore the effects of speed differences among front-running high frequency traders (HFTs) in a model of one round of trading. Traders differ in speed and their speed differences matter. I model strategic interactions induced when HFTs have different speeds in an extended Kyle (1985) framework. HFTs are assumed to anticipate incoming orders and trade rapidly to exploit normal-speed traders' latencies. Upon observing a common noisy signal about the incoming order flow, faster HFTs react more quickly than slower HFTs. I find that these front-running HFTs effectively levy a tax on normal-speed traders, making markets less liquid and prices ultimately less informative. Such negative effects on market quality are more severe when HFTs have more heterogeneous speeds. Even when infinitely many HFTs compete, their negative effects in general do not vanish. I analyze policy proposals concerning HFTs and find that (1) lowering the frequency of

trading reduces the negative impact of HFTs on market quality and (2) randomizing the sequence of order execution can degrade market quality when the randomizing interval is short. Consistent with empirical findings, a small number of HFTs can generate a large fraction of the trading volume and HFTs' profits depend on their speeds relative to other HFTs.

In the second chapter, I study the effects of higher trading frequency and front-running in a dynamic model. I find that a higher trading frequency improves the informativeness of prices and increases the trading losses of liquidity driven noise traders. When the trading frequency is finite, the existence of HFT front-runners hampers price efficiency and market liquidity. In the limit when trading frequency is infinitely high, however, information efficiency is unaffected by front-running HFTs and these HFTs make all profits from noise traders who do not smooth out their trades.

ESSAYS ON MARKET MICROSTRUCTURE AND HIGH FREQUENCY TRADING

by

Wei Li

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2014

Advisory Committee:
Professor Albert S. "Pete" Kyle, Chair
Professor Peter Cramton
Professor Mark Loewenstein
Professor Anna A. Obizhaeva
Professor Nagpurnanand R. Prabhala
Professor Yajun Wang

Dedication

To my father Lianluo Li and my mother Jieying Wan.

Acknowledgments

First I would like to thank my advisor, Professor Albert "Pete" Kyle, and my committee members Professor Mark Loewenstein, Professor Nagpurnanand R. Prabhala, Professor Anna A. Obizhaeva, Professor Yajun Wang, and Professor Peter Cramton for sparing precious time serving on my dissertation committee.

I express my utmost gratitude to Pete for showing me the beauty and excitement of economics research. Without him, I would never have the courage to pursue my passion. I am forever indebted to Pete. The only thing I can do is to pass along Pete's generosity and encouragement to my future students. I also thank Mark for continuous guidance and invaluable support. The long walks with Mark around the Van Munching Hall are some of my warmest memories at Maryland. I am extremely grateful for the tutelage of Professor Prabhala. I have benefited tremendously from his advice on numerous occasions. I would like to thank Anya for spending countless hours discussing various topics and offering insights. I also thank Yajun for constant encouragement and being always available to help. Special thanks are due to Professor Cramton for teaching me game theory and spending precious time with me.

I owe my thanks to my fellow PhD students, our PhD coordinator Justina Blanco, and my wife Ya Zhou. I could not have enjoyed the time at Maryland so much without you.

Finally, I would like to acknowledge the NASDAQ OMX Group Educational Foundation for the Dissertation Fellowship.

Contents

List of Figures

Chapter 1: High Frequency Trading with Speed Hierarchies

1.1 Introduction

High frequency trading (HFT) has emerged as a prominent feature of today's financial markets. Empirical studies find that high frequency traders (HFTs) have high trading volume, very short holding horizons, and invest astonishing amounts of capital to be slightly faster.[1] Consistent with these stylized facts, I model the front-running HFTs who anticipate incoming orders, trade rapidly, and have short holding horizons. In contrast to existing models of front-running, such as Brunnermeier and Pedersen (2005) and Carlin, Lobo and Viswanathan (2007), I allow HFTs to have different speeds to examine the impact of speed competition. Speed differences among HFTs affect all traders: (1) profits of a high frequency trader depend on her speed relative to other HFTs and (2) the aggregate profits of all HFTs, effectively a "speed tax" levied on other traders, depend on the distribution of HFTs' relative speeds. The presence of HFTs makes markets less liquid and prices ultimately less informative. Such negative effects are more severe when HFTs have more heterogeneous speeds.[2]

[1] See Kirilenko, Kyle, Samadi and Tuzun (2011) and Laughlin, Aguirre and Grundfest (2013).

[2] This paper focuses on the front-running HFTs who trade mainly with market orders. Other types of HFT strategies could have different impact on market quality.

I introduce fast traders into an extended Kyle (1985) framework with trading and quoting latencies. There are three types of "normal-speed" traders: (i) a strategic informed trader who privately observes the true value of a risky asset, (ii) noise traders who trade randomly for non-informational motives, and (iii) a continuum of competitive market makers who passively absorb order flow imbalance. Normal-speed traders' actions suffer from short latencies and fast traders exploit these latencies with their speed advantage.

Specifically, at the beginning of a trading round, competitive market makers post a linear pricing function that others can trade against. The slope of the pricing function is fixed during the trading round because market makers are not fast enough. After observing the pricing function, the informed and noise traders submit market orders. Before the orders arrive in the market, fast traders observe a common noisy signal about the orders and rapidly front-run by trading in the same direction at better prices ahead of the orders. When the informed and noise traders' orders arrive slightly later, fast traders reverse their early trades and exit their positions at profits. At the end of the trading round, competitive market makers update the final quoted price.

Fast traders effectively levy a speed tax and the tax makes the market less liquid for both the informed and noise traders. Because market makers cannot adjust the pricing function instantaneously, they suffer additional losses trading with fast traders. To make up for the additional losses, market makers increase the slope of the pricing schedule to charge more for absorbing order flow imbalance. Because trading is anonymous, market makers have to set a steeper pricing schedule for all other traders. Effectively, fast traders levy a speed tax on market makers and market makers shift the tax burden to the informed and noise traders by making liquidity more costly. Hence the informed trader profits less on given private information and noise traders lose more to trade the same amount.

Speed differences lead to short-term price momentum and reversals in a trading round. Prices exhibit short-term momentum when front-running fast traders sequentially "pick off" the stale pricing schedule to establish positions. The informed and noise traders' orders arrive slightly later and the orders are executed at worse prices due to the steeper pricing schedule. At the end of the trading round, knowing that prices have overshot, market makers partially reverse the final quoted price back to the informationally efficient level.

Prices are ultimately less informative when front-running fast traders are present. In the brief time before the informed and noise traders' orders arrive, fast traders' front-running trades bring information to the market and the intermediate prices are more informative. Information value of the intermediate prices, however, is quickly superseded by the more informative orders from the informed and noise traders. Ultimately, price informativeness is determined by the fraction of informed trader's orders in the aggregate order flow. Because the informed trader lowers trading intensity in response to higher liquidity costs, the aggregate order flow becomes less informative. This makes prices ultimately less informative after normal-speed traders have traded and fast traders have exited their positions.

A fast trader's effect on the informativeness of the order flow is similar to a prying messenger's effect on the information content of a letter. Suppose right before delivering a letter, a messenger glances at it and summarizes the letter to the receiver. Although the summary is informative, its information value is supplanted momentarily by the letter itself. Furthermore, expecting the letter to be pried into, the sender is less likely to write clearly and ultimately the receiver is less informed. The informed and noise traders share the direct cost of paying the speed tax; indirectly, decision makers who rely on price signals also suffer from less informative prices.

The negative impact of fast traders on market quality is more severe when they make more profits. Fast traders' aggregate profits in turn depend on the distribution of their relative speeds. I prove that for a given number of fast traders, their aggregate profits are minimized if every fast trader has the same speed and their aggregate profits are maximized if every fast trader has a different speed. Similarly, the entry of a fast trader could improve or degrade market quality depending on the entrant's speed relative to existing fast traders. If the entrant has the same speed as an existing fast trader, the aggregate fast trading profits decline. If the entrant has a different speed from all existing fast traders, the aggregate fast trading profits increase.

Intuitively, fast traders with the same speed compete against one another on quantity in a Cournot competition. In the limit when infinitely many fast traders compete on quantity, their aggregate profits converge to zero because they collectively push their entry price to equal their exit price. On the other hand, when more fast traders with different speeds are present, they collectively have more trading opportunities, march along the price schedule more gradually, and extract higher profits from market makers' stale pricing function. Intuitively, their aggregate profits are higher in this model because the "combined" trading frequency of all fast traders is higher than the trading frequency of any individual fast trader. In the limit when infinitely many fast traders with different speeds compete, the fastest front-runners still make positive profits and the slowest front-runners' profits converge to zero.

Market quality tends to settle on the suboptimal case in which fast traders have different speeds. First, fast traders have strong incentives to break away from a same-speed scenario. A higher speed leads to two-fold advantages for faster front-runners: they are able to establish *larger* positions at *better* prices than slower ones. In this model, a fast trader could quadruple the expected trading profits by moving up one spot in the ranking of relative speeds among fast traders. Second, the increasingly finer time granularity of modern markets opens up more space for fast traders to differ on speed. In a continuously operating market, no one could ever attain zero latency and traders can always beat competitors by any slight speed advantage. If fast traders engage in "arms race" in speed, they are unlikely to have the same speed and increased competition among fast traders may not improve market quality.

The model sheds light on several policy proposals concerning HFTs. (1) Lowering the entry cost to become HFTs may not improve market quality significantly if it does not reduce the heterogeneity of relative speeds among HFTs. (2) Converting a continuous market to a market with periodic uniform price auctions reduces the negative impact of front-running by enforcing Cournot competition among HFTs. (3) Randomizing the sequence of order execution is far less effective than the periodic uniform price auction. When the interval of randomizing is short, it can even degrade market quality. Intuitively, while randomizing reduces profits of the fastest HFT, it could help some slower HFTs; the net effect depends on the number of HFTs pooled in the same randomizing interval. (4) Requiring quotes to stay for a minimum duration or charging a fee for quote updating might aggravate the impact of front-running. Such rules could make the order flow easier to predict and extend the durations of profitable front-running.

My model also generate many empirical predictions. (1) In a continuous market, if front-runners trade on the same signal, only a few can survive; the markets for front-runners saturates quickly. (2) Entry of a fastest front-running HFT reduces the volume and profits of all existing front-running HFTs. It does not affect market quality significantly when the existing market for front-runners is almost saturated. (3) When front-running HFTs predict the order flow more accurately or when their relative speeds become more heterogeneous, their impact becomes more severe. Short-term price volatilities increase, short-term price momentum and reversal become stronger, normal-speed traders initiate less trading volume, and front-running HFTs' profits become a larger fraction of noise traders' implementation shortfall. (4) Faster front-running HFTs tend to trade more shares, have higher inventory levels, hold inventory for longer durations, and make more profit per share because they establish larger positions earlier at better entry prices.

1.1.1 Related literature

Angel, Harris and Spatt (2011) and Litzenberger, Castura and Gorelick (2012) recently survey literature related to HFT and modernization of the financial market.

This paper focuses on the impact of speed differences on the competition among front-running fast traders. Assuming predatory front-runners have the same speed, Brunnermeier and Pedersen (2005) find that "predation is most fierce if there are few predators". I show that when front-runners have different speeds, the impact of "predation" does not vanish even when infinitely many front-runners compete and their impact can be more severe when more front-runners are present. Hence, increasing the trading frequency tends to weaken the price or quantity competition and encourage socially wasteful speed competition among front-runners.

Many existing papers also assume that fast traders have homogeneous speeds and thus are not suitable to explore the implications of speed differences *among* fast traders. For instance, some studies, such as Pagano and Röell (1993), Brunnermeier (2005), Bernhardt and Taub (2008), Cohen and Szpruch (2012), and Foucault, Hombert and Rosu (2012), consider the situation with one monopolistic fast trader; others, such as Hirshleifer, Subrahmanyam and Titman (1994), Brunnermeier and Pedersen (2005), Carlin et al. (2007), Jarrow and Protter (2012), Hoffmann (2012), and Biais, Foucault and Moinas (2013), assume that fast traders have the same deterministic speed. Recently Budish, Cramton and Shim (2013) model high frequency market makers with the same speed in probability. In their model, when a high frequency market maker's quote of one share becomes stale, it is randomly "sniped" by another market maker. This leads to a positive bid-ask spread without information asymmetry or risk aversion. They also advocate using periodic batch auctions to eliminate this inefficiency. Penalva and Cartea (2012) discuss the situation when each HFT randomly intercepts a profitable trade according to her skill level. Essentially in their model multiple HFTs have the same speed but different market shares. Martinez and Rosu (2013) discuss one case in which two groups of HFTs have different speeds. HFTs in their model behave very differently because they are faster *and* more informed.

Unlike in Goldman and Sosin (1979), Hirshleifer et al. (1994) and Martinez and Rosu (2013), HFTs in my paper are faster than the informed trader but the informed trader has better information about the fundamentals. This assumption reflects the underlying costs of information production. Information processing takes time. So traders face the trade-off between information accuracy and trading speed. This trade-off is not reflected in existing models in which some traders are both faster *and* (weakly) more informed than all others. By contrast, in my paper fast traders start trading earlier on a less accurate signal while the informed trader trades later on a more accurate signal.

In addition, fast traders in this paper focus on information about incoming order flow not the fundamental value. In the short-term, the resale value of a risky security is more likely driven by the order flow rather than the fundamental value. Hence, consistent with the insight of Froot, Scharfstein and Stein (1992), fast traders with limited holding horizons and limited resource should focus on producing information about the short-term order flow and tend to become less informed about long-term fundamentals. Consequently, in contrast to Martinez and Rosu (2013), in this paper the presence of fast traders reduces overall price informativeness ex post because it is an impediment to the slower but better informed trader. In Brunnermeier (2005), a short-run trader has noisy information about an incoming public news release. This trader trades less when his signal is less informative about the fundamentals. In my model, HFTs trade the same even when the fraction of informed trading volume is low because HFTs make the same profits no matter whom they front-run.

In this paper, fast traders are partially informed because of their information on the order flow. They differ from an informed trader in existing models in two respects: (1) fast traders have higher speeds and shorter holding horizons than the better informed trader, and (2) fast traders have no source of information that is independent from slow traders' order flow. Such features imply that the more informed yet slower trader cannot avoid being front-run and she does not need to speed up to avoid information decay: fast traders cannot learn her information if the informed trader does not trade. Therefore, unlike in existing models with multiple informed traders such as Holden and Subrahmanyam (1992), Foster and Viswanathan (1996), Vayanos (1999), Back, Cao and Willard (2000), Bernhardt and Miao (2004), and Li (2013), in this paper, the better informed trader always reduces her trading intensity in the presence of less informed fast traders and prices become less informative.

Some theoretical papers focus on other aspects of HFTs and algorithmic trading. Cvitanic and Kirilenko (2010) model HFTs as a machine which immediately "snipes" out a human order when its price deviates too far from a benchmark level. Gerig and Michayluk (2010) model HFTs as automatic market makers who use the relationships between multiple securities to price order flow in an extension of the Glosten and Milgrom (1985) model. Jovanovic and Menkveld (2012) model HFTs as competitive intermediaries who can process hard information. Pagnotta and Philippon (2012) use a search model to investigate the exchanges' incentives to lower latencies. Yueshen (2013) models the strategic interactions among limit order traders when they cannot condition orders on positions in the queue. Weller (2013) develops a model in which fast market makers specialize in immediacy provision and slow market makers specialize in risk bearing.

My model is consistent with many empirical characteristics of HFTs. My model predicts that only a few HFTs can survive in the market. Many empirical studies identify only a very small number HFTs. My model predicts that when there are already a handful of HFTs, the entry of new HFTs does not affect the aggregate HFT trading profits significantly. Budish et al. (2013) find that the profitability of HFTs' trading opportunities remains almost constant over the years 2005-2011 despite the fierce competition among HFTs. HFTs in this paper are front-runners. Hirschey (2013) finds that HFTs have the ability to anticipate non-HFTs' large trades. Clark-Joseph (2013) finds that aggressive HFTs use smallest orders to explore market conditions and choose the timing to front-run large incoming demands. In my model HFTs mainly use market orders. Studies using individual account data, such as Baron, Brogaard and Kirilenko (2012), Breckenfelder (2013), Brogaard, Hagströmer, Norden and Riordan (2013a), and Hagströmer and Nordén (2013) all find that some HFTs predominantly trade with market orders and they tend to make more profits than other HFTs.

My model predicts that HFTs' trades are informative and improve short-term intermediate price informativeness. Brogaard, Hendershott and Riordan (2013b) find that the marketable orders of HFT have high predictive power about future price changes in less than 5 seconds. Brogaard et al. (2013a) find that co-located traders have an informational advantage. Zhang (2012) finds that HFTs profit on “hard” information and their profits realize quickly. Carrion (2013) also finds HFTs have intra-day market timing capability. My model also predicts that front-running HFTs can reduce information efficiency in the long run. Differentiating the two information efficiencies poses new empirical challenges.

In this paper, front-running HFTs increase short-term volatility and reduce long-term volatility. Breckenfelder (2013) finds that HFTs increase intra-day volatility but not inter-daily volatility. Jiang, Lo and Valente (2013) and Boehmer, Fong and Wu (2012) find that HFT or algorithmic trading tend to increase short-term volatility.

In this model, HFTs follow similar strategies and their profits depend on their relative speeds. Chaboud, Chiquoine, Hjalmarsson and Vega (2013) find algorithmic traders tend to use correlated strategies. Gai, Yao and Ye (2012) and Egginton, Van Ness and Van Ness (2013) find evidence that some traders use the “quote-stuffing” strategy to slow down other traders.

My model predicts that although their trading volume is high, front-running HFTs reduce market liquidity. Hendershott and Moulton (2011) find that automation increases bid-ask spreads. Tong (2013) finds that HFTs increase the trading costs of institutional traders. Many empirical studies find that algorithmic trading or high frequency trading improve liquidity. This could be due to three main reasons. First, some studies, such as Hasbrouck and Saar (2013), Hendershott, Jones and Menkveld

(2011), Hendershott and Riordan (2012), Menkveld (2013), investigate either all algorithmic traders or market making HFTs. Isolating the effects of front-running HFTs and assessing their relative importance are open empirical questions. Second, traditional measures could underestimate illiquidity in the presence of front-running HFTs. For example, bid-ask spreads and depth of the limit order book may not capture market liquidity when HFTs can quickly cancel limit orders. In addition, a large order would move the price *before* it arrives in the market when front-runners are present. Third, liquidity might be improved because of other contemporaneous factors.

1.2 Benchmark model of a monopolistic fast trader

I introduce delays in trading and quoting into the static model of Kyle (1985) and add a new type of trader fast enough to exploit the short delays. The section presents the benchmark model with only one fast trader.

1.2.1 Model setup

Assets and Traders Traders trade two assets: a risk free numeraire asset with zero interest rate and a risky asset with normally distributed fundamental value $v \sim \mathcal{N}(v_0, \sigma_v^2)$. All traders are risk neutral.[3] As in Kyle (1985), three types of traders have normal speeds: (1) a strategic monopolistic informed trader privately observes the true value v of the risky asset; (2) noise traders randomly trade normally

[3] In Appendix A.1.4, I discuss the implication of fast traders' risk aversion.

distributed $z \sim \mathcal{N}(0, \sigma_z^2)$ shares for exogenous non-informational motives;[4] and (3) competitive fringe market makers set the pricing functions, absorb the residual order flow imbalances, and make zero expected profit.[5]

I introduce a new type of traders: the fast traders who anticipate the size of the incoming market orders and rapidly trade twice in one trading round. Unlike normal-speed traders, fast traders do not carry inventory when the trading round ends. In the benchmark model, only one fast trader is present. In the general model of Section 1.3, multiple strategic fast traders with possibly different speeds compete with one another.[6]

Timeline and Information Structure The paper presents models of *one* trading round. Figure 1.1 illustrates the timeline of the benchmark model. At

[4] Although noise traders on average lose in trading, they are not necessarily irrational. For example, they could have idiosyncratic liquidity demands unrelated to the valuation of the risky asset.

[5] Following the literature, I name a continuum of competitive fringe traders *market makers*. They do not, however, act like specialists or designated market makers in a dealer market. These market makers represent the large population of traders who have no information or speed advantage, and also no incentives to initiate trades.

[6] Empirical studies such as Baron et al. (2012) and Breckenfelder (2013) have documented a large heterogeneity among HFTs in terms of order aggressiveness, i.e., the ratio of market versus limit orders. In this paper fast traders could be interpreted as the front-running HFTs who predominantly trade with market orders and reap the largest trading profits among all HFTs. As pointed out by Hasbrouck and Saar (2009), however, the difference between limit orders and market orders is not necessarily crucial. In practice, front-running strategies could be implemented with a mixture of market orders and limit orders.

time 0, the trading round starts. Market makers set a publicly observable pricing function $P(\cdot,\cdot)$. An order of y_t shares arrives at time $t \in (0,1]$ and is filled by the market makers at the average price of $p_t = P(y_t, \mathcal{F}_t)$. $\mathcal{F}_t$ denotes market makers' information by time t. To model market makers' latency, the pricing function $P(\cdot,\cdot)$ is fixed in time interval $(0, 1^+)$.

At time 0^+ after observing the pricing function $P(\cdot,\cdot)$ the informed trader submits a market order of x shares and noise traders submit a market order of z shares. Their orders suffer from a short latency and will not arrive in the market until time 1. Since the market is continuously operating, trades and quote updates may take place between time 0^+ and time 1. The delay is so short that no one other than the fast trader could exploit it.

Right after time 0^+ the fast trader observes a private signal $I_y = y + e_y$ about the incoming order flow $y = x + z$ where $e_y \sim \mathcal{N}(0, \sigma_e^2)$ denotes the normally distributed observation error. The quality of the signal I_y is represented by ρ, the squared correlation between I_y and y, i.e.,

$$\rho \equiv \mathbf{C}\mathrm{orr}^2(y, I_y) \in (0,1] \tag{1.2.1}$$

I take information quality ρ as exogenously given.[7] A more informative signal I_y has a higher ρ. If I_y reveals y precisely, $\rho = 1$; if I_y is almost all noise, $\rho \to 0$. The projection theorem for normal random variables implies that $\hat{y} \equiv \mathbf{E}[y|I_y] = \rho \cdot I_y$.

At time 1^-, based on her signal I_y the fast trader trades u shares. Market

[7] A fixed ρ implies that variance of the observation error $\sigma_e^2 = \frac{1-\rho}{\rho}\left(\sigma_x^2 + \sigma_z^2\right)$. For example, if it is know that the informed trader almost does not trade ($\sigma_x^2 \to 0$), the observation error is almost entirely about noise trading size z.

makers fill the order at the price of $p_{1^-} = P(u, \mathcal{F}_{1^-})$.

At time 1, the informed trader's order x and the noise traders' order z arrive in the market. At the same time, the fast trader submits a order of $-u$ shares to liquidate her position because she is not allowed to carry inventory when the trading round ends.[8] Trades are all anonymous and market makers fill all orders at the same price of $p_1 = P(x + z - u, \mathcal{F}_1)$.

Finally, at time 1^+, right after time 1, market makers look back at the order flow history of the trading round and update the final quoted price to be p_{1^+}. The trading round ends.

At time 0, it is common knowledge that value of the risky asset v, noise traders' order size z, and the fast trader's observation error e_y are mutually independent.

1.2.2 Discussion on speed

In many models of strategic trading, such as Holden and Subrahmanyam (1992), Foster and Viswanathan (1996), and Vayanos (1999), all traders move at the same frequency: they act once per round. The defining feature of high frequency traders, however, is not that they are fast but that they are *faster* than others. In this paper, the fast trader has speed advantage over all other traders in various ways.

First, while the fast trader is able to trade *twice* without latencies in one trading round, the informed and noise traders can only trade *once*. Moreover, they

[8] The fast trader trade the $-u$ shares with market orders. In practice, the fast trader might try to trade the $-u$ shares with limit orders to reduce transaction costs.

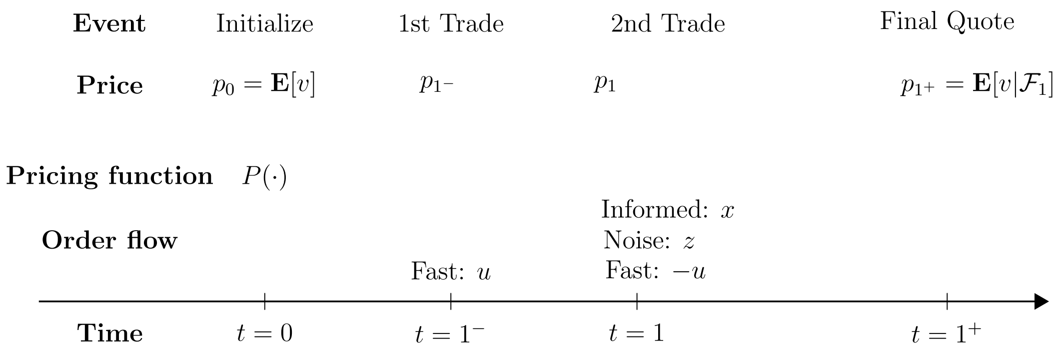

Figure 1.1: Timeline of the benchmark model of a monopoly fast trader.At time 0, market makers set the pricing function $P(\cdot)$. At time 1^-, the monopolistic fast trader trades u shares at price p_{1^-}. At time 1, the informed trader's order x, noise traders' order z, and the fast trader's second order $-u$ arrive simultaneously and are executed at price p_1. At time 1^+, market makers update the quote price to p_{1^+}.

place orders at time 0^+ and the orders are executed at time 1 after a short delay.

Second, the fast trader is also faster than market makers because market makers cannot update the pricing function when multiple trades take place within the short time window $[1^-, 1]$. Ideally market makers would like to adjust the pricing schedule $P(\cdot)$ whenever they observe new information. In practice many traders who submit limit order do not have the most powerful computers or the fastest connections. They can only update their limit orders with delays. During the delays, their limit orders become stale and faster traders could quickly pick off the stale orders. To capture the lagged adjustment of the limit orders, pricing function $P(\cdot)$ is set at time 0 and cannot be updated until time 1^+.

Although market makers are perfectly competitive, trade prices deviate from the informationally efficient levels. Effectively, lags in adjusting the pricing schedule

$P(\cdot)$ prevent market makers from using all information to price the orders in the time window $[1^-, 1]$.

The opening quote p_0 and the closing quote p_{1^+} are still informationally efficient by assumption. In practice limit order submitters eventually update their orders when trading activities wane and markets cool down.

1.2.3 Discussion on the holding horizon and information

The fast trader is averse to holding inventory and must liquidate her position by the end of a trading round. This assumption of short holding horizons, although restrictive, is consistent with most empirical studies of high frequency traders such as Kirilenko et al. (2011) and Baron et al. (2012). In fact, one characteristic used by SEC (2010) to define high frequency traders is that they have "very short time-frames for establishing and liquidating positions."

The short holding horizon has important implication on the fast trader's behavior and choice of information. Unlike the informed trader, the strategic fast trader incurs price impact twice in one trading round. When establishing a position, the fast trader must have a plan to exit within a short time window. The fast trader's profit is not determined by the difference between her entry price and the fundamental value, but by the difference between her entry and exit prices. Hence, the fast trader do not try to infer the long-term fundamental value of the risky asset but focus on predicting short-term price dynamics.

Consistent with the short holding horizon, I assume that the fast trader produces a signal I_y about the aggregate incoming order flow $y = x + z$, not about the fundamental value v. The fast trader has no incentive to differentiate orders from the informed trader and orders from the noise traders as long as the orders have the identical price impact per share. In fact, if noise traders' order size z dominates the informed trader's order size x as in the continuous time models of Kyle (1985) and Back (1992), the fast trader would try to correlate her trades mostly with the noise trading z.

The fast trader's advance information about incoming order flow could come from various sources. In practice, HFTs often have faster access to the exchanges' more detailed data feeds. Coupled with their computation power, HFTs could continuously track updates to the limit order books and quickly detect patterns in the order flow. They could also gain information advantage in more fragmented markets when orders are constantly routed between different trading venues.

Specifying the fast trader's information I_y as a noisy signal of the incoming order also encompasses several alternative information sources. For example, if a fast trader could parse a public news faster but less accurately, the fast trader effectively anticipates the informed trader's order x; if a fast trader is able to detect some retail noise traders' less sophisticated execution algorithm, the fast trader effectively predicts noise trading order flow z.[9]

[9] Budish et al. (2013) find that when observed at the millisecond precision, price correlation of highly related securities breaks down. Such short-lived price divergence can be another source of HFTs' information advantage.

1.2.4 Equilibrium

Definition 1.2.1 (Equilibrium conditions). The fast trader chooses her trade size using a strategy function $U(\cdot)$ and the informed trader chooses her trade size using a function $X(\cdot)$. Market makers commit to a pricing function $P(\cdot)$ and set the final quote using a function $Q(\cdot)$. The equilibrium is defined by four functions $U(\cdot), X(\cdot), P(\cdot)$ and $Q(\cdot)$ such that the following conditions hold:

1. *Informed trader profit maximization.* Given $P(\cdot)$, $U(\cdot)$, and the asset's true value v, the informed trader's profit $\pi^I = x(v - p_1)$ is maximized if she trades x^* shares, i.e.,

$$x^* = X(v; U(\cdot), P(\cdot)) = \underset{x}{\text{argmax}} \quad \mathbf{E}\left[\pi^I \,|v, P(\cdot), U(\cdot)\right] \tag{1.2.2}$$

   where p_1 is the execution price of her trade.

2. *Fast trader profit maximization.* Given $P(\cdot)$, $X(\cdot)$, and a signal about the incoming order flow $I_y = x + z + e_y$ the fast trader's profit $\pi^F = u(p_1 - p_{1^-})$ is maximized if she trades u^* shares at time 1^- and liquidates at time 1, i.e.,

$$u^* = U(I_y; P(\cdot), X(\cdot)) = \underset{u}{\text{argmax}} \quad \mathbf{E}\left[\pi^F \,|I_y, P(\cdot), X(\cdot)\right] \tag{1.2.3}$$

   where $p_1 - p_{1^-}$ is the difference between her entry and exit prices.

3. *Competitive pricing function.* Given $X(\cdot)$ and $U(\cdot)$, market makers choose a pricing function $P(\cdot)$ such that their expected profit $\mathbf{E}[\pi^M]$ at time 0 equals zero, i.e.,

$$0 = \mathbf{E}\left[\pi^M \,|P(\cdot), X(\cdot), U(\cdot)\right] \tag{1.2.4}$$

4. *Informationally efficient quotes.* Market makers set quotes p_0 and p_{1+} to be their expected value of v conditional on available information $\mathcal{F}_0$ and $\mathcal{F}_{1^+}$.

$$p_0 = \mathbf{E}[v] \tag{1.2.5}$$

$$p_{1^+} = \mathbf{E}[v|\mathcal{F}_{1^+}, X(\cdot), U(\cdot)] = Q(\mathcal{F}_{1^+}; X(\cdot), U(\cdot)) \tag{1.2.6}$$

Remark 1.2.1. Market makers' profits in Equation (1.2.4) is $\pi^M = up_{1^-} + (x + z - u)p_1 - (x + z)v$ because they trade $-u$ shares at price p_{1^-} and $-(x + z - u)$ shares at price p_1.

Remark 1.2.2. Setting p_{1+} has no impact on equilibrium because the game ends at time 1^+. When there are multiple rounds of trading, setting p_{1+} to be the posterior expectation makes it an appropriate initial reference quote for the next trading round.

The strategy functions $U(\cdot), X(\cdot), P(\cdot)$ and $Q(\cdot)$ can be very general. For tractability, I assume that market makers choose the pricing function $P(\cdot)$ from the following class of linear functions.

Assumption 1 (**Linear pricing function**)**.** Upon receiving the j-th market order of y_{t_j} shares at time $t_j \in (0, 1]$, market makers fill the order at the average price of

$$p_{t_j} = P(y_{t_j}, \mathcal{F}_{t_j}) = p_{t_{j-1}} + \lambda^T y_{t_j}, \quad j \geq 1 \tag{1.2.7}$$

If y_{t_j} is the first arriving order ($j = 1$), the reference price p_{t_0} is the initial quote p_0; if $j > 1$, $p_{t_{j-1}}$ is the average price of the previous traded market order. The price impact (market depth) factor λ^T is fixed in the trading round.

Assumption 1 reduces the choice of the pricing function $P(\cdot)$ to the choice of two parameters: the initial quote p_0 and the price impact factor λ^T. As discussed in the previous section, a fixed λ^T captures the latency in limit order book adjustment. It also seems reasonable that the price impact factor λ^T is fixed in short intervals when all market orders are anonymous. In practice traders break their large meta-orders into small trades and execute over time. The small trades are all stochastically sequenced together and there is no clear start or end of a trading round. It is unlikely that the limit order submitters would be able to infer the originator of orders and change pricing function accordingly, especially in very short intervals.

Lemma 1.2.1. *In the benchmark model,* $t_0 = 0$, $t_1 = 1^-$, *and* $t_2 = 1$. *Given Assumption 1, the traded prices are*

$$p_{1^-} = p_0 + \lambda^T u \tag{1.2.8}$$

$$p_1 = p_{1^-} + \lambda^T(x + z - u) = p_0 + \lambda^T(x + z) \tag{1.2.9}$$

It might seem that the execution price p_1 for the informed and noise traders are not affected by the fast trader's trading u because the fast trader completely liquidates her position at time 1. The observation, however, is not correct because in equilibrium the price impact factor λ^T is endogenously determined by the fast trader's trading intensity. We now examine the equilibrium.

Theorem 1.2.2 (Equilibrium of the benchmark model). *Given Assumption*

1, there is a unique equilibrium where

$$\text{Fast trading size: } u^* = U(I_y; p_0, \lambda^T) = \alpha \hat{y} = \alpha \rho I_y \tag{1.2.10}$$

$$\text{Informed trading size: } x^* = X(v; p_0, \lambda^T) = \beta\,(v - p_0) \tag{1.2.11}$$

$$\text{Market order pricing: } p_{t_j} = P(y_j, p_{t_{j-1}}) = p_{t_{j-1}} + \lambda^T y_j \tag{1.2.12}$$

$$\text{Initial quote: } p_0 = v_0 \tag{1.2.13}$$

$$\text{Final quote: } p_{1+} = Q(u, y - u) = p_0 + \lambda^P y \tag{1.2.14}$$

The endogenous parameters α, β, λ^T, and λ^P are:

$$\alpha = \frac{1}{2}, \quad \beta = \frac{\sigma_z}{\sigma_v}\theta, \quad \lambda^T = \frac{\sigma_v}{\sigma_z}\frac{1}{2\theta}, \quad \lambda^P = \frac{\sigma_v}{\sigma_z}\frac{\theta}{1+\theta^2} \tag{1.2.15}$$

where the market quality parameter

$$\theta \equiv \sqrt{\frac{1-\rho/4}{1+\rho/4}} \in [\sqrt{0.6}, 1] \tag{1.2.16}$$

Proof. See Appendix A.1. The proof covers the more general case where the fast trader is risk averse with an exponential utility functions. Theorem 1.2.2 here is a special case when risk aversion coefficient equals zero. □

1.2.5 Equilibrium analysis

The strategy functions $X(\cdot)$, $U(\cdot)$, and $Q(\cdot)$ are all linear if the pricing function $P(\cdot)$ is linear. The equilibrium is then fully characterized by the four endogenous parameters α, β, λ^T, and λ^P illustrated in Figure 1.2:

The parameter α is fast trader's trading intensity. The fast trader first observes a signal I_y of the incoming order flow y and estimate the order flow to be $\hat{y} = \rho I_y$.

Then, the fast trader choose to trade $u = \alpha\hat{y}$ shares at time 1^- and $-\alpha\hat{y}$ at time 1. A higher intensity α indicates that the fast trader trades more given an estimated order flow $\hat{y}$.

Trading intensity β characterizes the informed trader's strategy. The informed trader first calculates the pricing error $v - p_0$ using her private information v and the initial quote p_0. She then submits a market order of $x = \beta(v - p_0)$ shares. A higher β indicates that the informed trader trades more aggressively based on the same pricing error $v - p_0$.

The parameter λ^T represents the *temporary* price impact per share of an market order on the transaction price p_t.[10] Transaction price responds to the order flow according to the pricing function $\Delta p_t = \lambda^T y_t$. Effectively, market makers charge $\lambda^T y^2$ to execute a market order of y shares. Competitive market makers set λ^T just enough such that their revenues for executing trades exactly offset their loss in trading with the informed and the fast trader. A higher λ^T means that it costs more to execute a market order of any given size.

The slope λ^P represents the *permanent* price impact per share on the final quote p_{1^+} of the aggregated order size $x + z$. The difference between the closing quote p_{1^+} and the opening quote p_0 equals $\lambda^P(x + z)$. Because market makers are competitive, the quote update $p_{1^+} - p_0$ is determined by the information content of the order flow $x + z$. A higher λ^P indicates that the aggregate order flow $x + z$ contains more information about the fundamental value v and thus the quote update

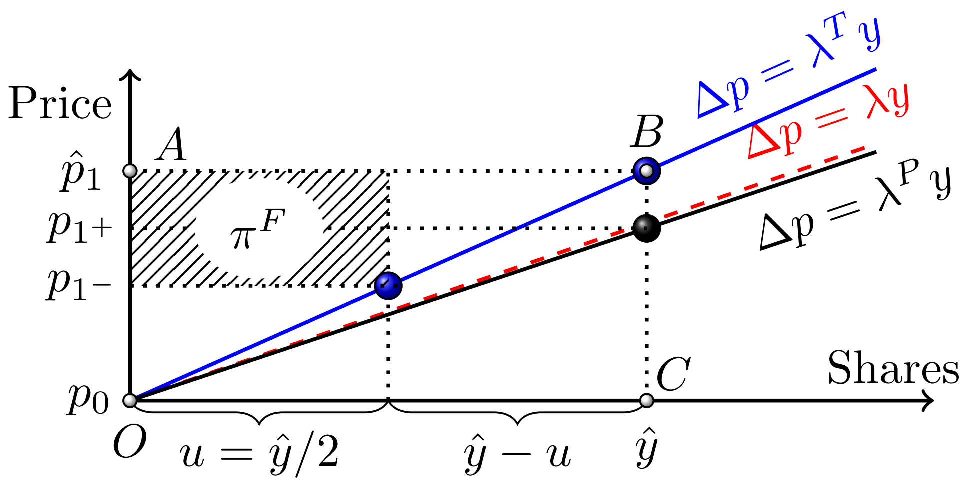

Figure 1.2: Equilibrium strategies. $\Delta p = \lambda y$ is the pricing function without fast trader. When a fast trader is present, market makers raise λ^T and lower λ^P. At time 0, market makers set p_0 and λ^T; at time 1^-, the fast trader trades $u = \hat{y}/2$ shares at price $p_{1^-} = p_0 + \lambda^T u$; at time 1, the informed trades x shares, the noise traders z shares, and the fast trader $-u$ shares at the price $p_1 = p_{1^-} + \lambda^T(x + z - u) = p_0 + \lambda^T y$; finally at time 1^+, market makers set the quote to $p_{1^+} = p_0 + \lambda^P y$. The shaded rectangle is the fast trader's profit π^F. Its area is 1/4 of rectangle $OABC$ which corresponds to market makers' price impact surplus $\lambda^T \hat{y}^2$.

$p_{1^+} - p_0$ is more sensitive to $x + z$.

The fast trading intensity α always equals 1/2. Both the fast trader and the informed trader are risk neutral, strategic, and monopolistic. They maximize expected profit by pushing the price half way toward the level at which they expect to exit their positions.[11]

The other three endogenous parameters β, λ^T, and λ^P are determined by three exogenous parameters: volatility of the fundamental value σ_v, volatility of noise trading σ_z, and information quality of the fast trader $\rho \in [0, 1]$. The equilibrium effects of σ_z and σ_v are similar to the Kyle (1985). In addition, we can set $\sigma_v = \sigma_z = 1$ by choosing certain units of currency and trade size.

Fast trader's information quality ρ, however, is invariant to change of units. When the fast trader has no information ($\rho = 0$), the equilibrium reduces to the equilibrium of the static model of Kyle (1985). When the fast trader has some information $\rho > 0$, the equilibrium differs qualitatively.

Corollary 1.2.3. *Other things equal, when the fast trader's information becomes more accurate ($\rho \uparrow$), temporary price impact increases ($\lambda^T \uparrow$), permanent price impact decreases ($\lambda^P \downarrow$), informed trading intensity declines ($\beta \downarrow$), and the fast trading intensity α is unchanged.*

[10] λ^T is not entirely temporary. It includes the permanent price impact λ^P.

[11] When the fast trader is risk averse, her trading intensity is indeed lower when $\sigma_v \uparrow$, $\sigma_z \uparrow$, or her information quality $\rho \downarrow$. See Appendix A.1 for a generalize Theorem where the fast trader has a negative exponential utility and Appendix A.1.4 for a brief discussion of the equilibrium impact of the fast trader's risk aversion.

Proof. Follows from Theorem 1.2.2. □

Figure 1.3 illustrates the equilibrium impact of fast trader's information quality ρ. Due to the existence of fast traders, market makers cannot break-even if they set the price p_1 to equal their posterior expectation. The fast trader profitably intercepts u shares of the order flow $x + z$: she acquires u shares from market makers at a discounted price p_{1^-} and supplies the shares back to the informed and noise traders at a profit. To make up for the loss to the fast trader, market makers have to charge more to absorb the same order imbalance. They raise the temporary price impact factor λ^T above the permanent price impact λ^P implied by the informativeness of the order flow. Market makers' break-even price at time 1 thus differs from their posterior conditional expectation.

When the signal I_y is more accurate signal ($\rho \uparrow$), the fast trader makes more profit. Market makers raise the temporary price impact ($\lambda^T \uparrow$) more to break even. The informed trader responds by reduce her trading intensity ($\beta \downarrow$). Because the informed trader's order x is the only informative component of the order flow $x + z$, the aggregate order flow $x + z$ becomes less informative ($\lambda^P \downarrow$).

This section develops the benchmark model with one fast trader and illustrates that the fast trader can profit on a less accurate signal I_y if she is faster than others. In addition to the "information rents", market makers have to pay the "speed rents". Because market makers are not the fastest, they protect themselves by setting a steeper pricing schedule. The informed traders reduce her trading intensity faced with a higher cost of transacting. These results are not surprising given the assumed behavior of the fast traders. Profits attract the entry of similar fast traders. How does the equilibrium change if more fast traders compete? In the next section, I develop the general model in which multiple strategic front-runners with different speeds compete.

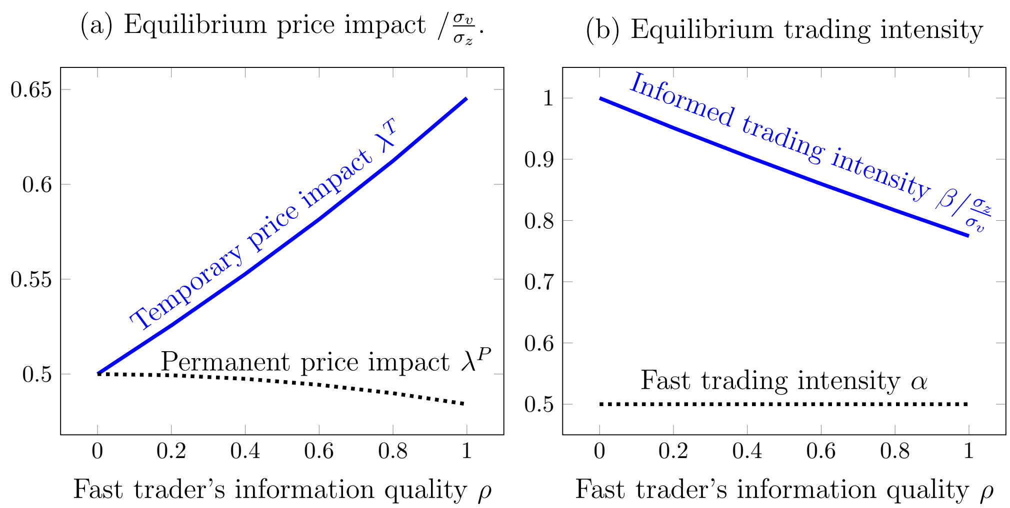

Figure 1.3: Equilibrium parameters of the benchmark model normalized by volatility of fundamental value σ_v and volatility of noise trading σ_z. Theorem 1.2.2 implies that when the fast trader's information I_y is more informative (higher ρ) about the incoming order, temporary price impact per share λ^T increases, permanent price impact per share λ^P decreases, informed trading intensity β decreases, and fast trading intensity α stays the same.

1.3 Model of multiple fast traders

In the benchmark model, after controlling for noise trading volatility σ_z and fundamental uncertainty σ_v, the monopolistic fast trader's information quality ρ

fully characterizes the equilibrium. I now consider the case in which multiple high frequency traders with different speeds compete in the same market. I characterize the equilibrium with multiple fast traders and examine parameters that affect the equilibrium apart from the fast traders' information quality ρ.

1.3.1 Generalized model setup

Normal-speed traders have the same action timings as in the benchmark model: the informed and noise traders trade once at time 1; market makers set the pricing function $P(\cdot)$ at time $t_0 = 0$ and update the quote to p_{1^+} when the trading round ends.

I modify the timeline to accommodate multiple fast traders. There are N strategic fast traders. As in the benchmark model, each fast trader is able to trade twice until time 1 and fast traders are not allowed to carry inventory after time 1. At time 0^+ a signal $I_y = y + e_y$ with exogenous quality $\rho = \mathbf{C}\text{orr}^2(I_y, y) \in (0, 1]$ is generated. To focus on the effect of speed differences, the *same* signal is distributed to all N fast traders.[12]

Each fast trader receives the signal I_y with different delays, analyzes it at different speeds, and submits an order with different latencies. Fast traders' orders also suffer from latencies but their latencies are much shorter than normal-speed traders'. Between time 0^+ and time 1, fast traders' orders sequentially arrive in

[12] Possible information cascades phenomena as in Bikhchandani, Hirshleifer and Welch (1998) are not modeled because the common signal assumption makes it unnecessary for slower fast traders to learn from earlier trades.

J instants $0^+ < t_1 < t_2 \cdots < t_J < 1$. At time t_j, n_j orders arrive simultaneously. Naturally $N = \sum_j n_j$. The speed of a fast trader is measured by the her order arrival time t_j. The difference between t_j and 0^+ includes the signal transmitting time, the signal processing time, and her order latency. A fast trader is indexed by $j \in \{1, 2, \cdots, J\}$ and $k \in \{1, 2, \cdots, n_j\}$.

Definition 1.3.1 (**Speed profile of fast traders**). The speed profile of fast traders is a vector of J numbers $\{n_1, n_2, \cdots, n_J\}$ where $n_j \geq 0$ is the number of fast traders arriving at time t_j. The speed profile is common knowledge among all traders.

Two special speed profiles are particularly important because their equilibrium properties are on the two extremes among all possible speed profiles.

Definition 1.3.2 (Stackelberg-N speed profile). Each of the N fast traders arrives at a different moment and the speed profile is $\{1, 1, \cdots, 1\}$.

Definition 1.3.3 (Cournot-N speed profile). All N fast traders arrive at the same time and the speed profile is $\{N\}$.

Upon arriving at time t_j, fast trader (j, k) uses all available information, including the signal I_y and the last traded price $p_{t_{j-1}}$, to chooses a trade size $u_{j,k}$. I introduce two notations of fast traders' order sizes: u_j denotes the total order size from fast traders arriving *at* time t_j and S_{j-1} denotes the total order size from fast traders arriving *before* time t_j.

Definition 1.3.4 (**Fast traders' order sizes**). For $j = 1, 2, \cdots, J$,

$$u_j \equiv \sum_{k=1}^{n_j} u_{j,k}, \quad S_{j-1} \equiv \sum_{i=0}^{j-1} u_i = \sum_{i=0}^{j-1} \sum_{k=1}^{n_i} u_{i,k} \tag{1.3.1}$$

$u_{0,k} = 0$ and $n_0 = 1$ for completeness.

Since fast traders could not carry inventory beyond time 1, each fast trader must completely exit her position $u_{j,k}$ using the second trade. When is the best time to exit? In the benchmark model the monopolistic trader exits at time 1 because exiting any earlier would make her profit zero. When there are multiple fast traders, it is still optimal for *all* fast traders to exit at time 1 simultaneously. All fast traders have the same information and same preference (risk neutral). A fast trader arriving earlier knows that a later fast trader would only enter a position if it is profitable. If it is profitable for a later fast trader to enter, then it is profitable for the earlier fast trader to wait. Hence all fast traders liquidate at time 1 when they cannot wait any longer. The time 1 net order flow is $y - \sum_{j,k} u_{j,k}$.

Market makers observe $J+1$ net market orders $\{u_1, u_2, \cdots, u_J, y - \sum_i u_i\}$ at $J+1$ moments $\{t_1, t_2, \cdots, t_J, 1\}$. They execute the orders at the prices of $\{p_{t_1}, p_{t_2}, \cdots, p_{t_J}, p_1\}$ according to the pricing function $P(\cdot)$ set at time 0.

1.3.2 Equilibrium

Equilibrium conditions differ from Definition 1.2.1 only in the condition about fast traders. I select the symmetric equilibrium where fast traders arriving at the same time trade the same quantity.

Definition 1.3.5 (**Modified equilibrium condition for fast traders**)**.** Given the pricing function $P(\cdot)$, the informed trader's strategy $X(\cdot)$, a signal $I_y = y + e_y$ about the incoming order flow y, the price of the last trade $p_{t_{j-1}}$, and strategies

$U_{j,l}()$ of other fast traders ($l \neq k$) arriving simultaneously at time j, fast trader (j,k) maximizes her profit $\pi^F_{j,k}$ if she trades $u^*_{j,k}$ shares at time t_j and liquidate her position at time 1, i.e.,

$$
\begin{aligned}
u^*_{j,k} &= U_{j,k}(I_y, p_{t_{j-1}}; P(\cdot), X(\cdot), U_{j,l}(\cdot) \text{ for } l \neq k) \qquad (1.3.2)\\
&= \underset{u}{\text{argmax}} \quad \mathbf{E}\left[\pi^F_{j,k} = u(p_1 - p_{t_j}) \,\middle|\, I_y, P(\cdot), X(\cdot), p_{t_{j-1}}, U_{j,l}(\cdot) \text{ for } l \neq k\right]
\end{aligned}
$$

In the symmetric equilibrium $u^*_{j,k} = u^*_{j,l}$ for all $j \in \{1, 2, \cdots, J\}$ and $k, l \in \{1, 2, \cdots, n_j\}$.

Theorem 1.3.1 (Equilibrium with oligopolistic fast traders). *Given a linear pricing function as in Assumption 1, there is a unique symmetric equilibrium where*

$$(j,k)\text{-th fast trading size: } u^*_{j,k} = U_{j,k}(\hat{y}, S_{j-1}) = \frac{\alpha_j}{n_j}(\hat{y} - S_{j-1}) \qquad (1.3.3)$$

$$\text{Informed trading size: } x^* = X(v, p_0) = \beta(v - p_0) \qquad (1.3.4)$$

$$\text{Pricing function: } p_{t_j} = P(y_j, p_{t_{j-1}}) = p_{t_{j-1}} + \lambda^T y_j, \qquad (1.3.5)$$

$$\text{Initial quote: } p_0 = v_0 \qquad (1.3.6)$$

$$\text{Final quote: } p_{1+} = Q(u, y - u) = v_0 + \lambda^P y \qquad (1.3.7)$$

for all $1 \leq j \leq J$ *and* $1 \leq k \leq n_j$. *The endogenous parameters* α_j, β, λ^T, *and* λ^P *are:*

$$\alpha_j = \frac{n_j}{1 + n_j}, \quad \beta = \frac{\sigma_z}{\sigma_v}\theta, \quad \lambda^T = \frac{\sigma_v}{\sigma_z}\frac{1}{2\theta}, \quad \lambda^P = \frac{\sigma_v}{\sigma_z}\frac{\theta}{1 + \theta^2} \qquad (1.3.8)$$

where

$$\theta = \sqrt{\frac{1 - \rho\gamma}{1 + \rho\gamma}}, \quad \gamma = \sum_{j=1}^{J}\left(\alpha_j(1 - \alpha_j)\prod_{i=0}^{j-1}(1 - \alpha_i)^2\right) \qquad (1.3.9)$$

Let $\alpha_0 = 0$ for completeness.[13]

Proof. See Appendix A.2.2. $\square$

Remark 1.3.1. When there is only one fast trader, $J = 1$, $n_1 = 1$, and $\gamma = 1/4$. Equilibrium reduces to the special case given in Theorem 1.2.2.

Remark 1.3.2. Observing the last traded price $p_{t_{j-1}}$ is equivalent to observing the shares trade by earlier fast traders S_{j-1} because $p_{t_{j-1}} = p_0 + \lambda^T S_{j-1}$. In addition, $\hat{y} = \mathbf{E}[y|I_y] = \rho I_y$. So the fast trader's strategy can be equivalently expressed as $u^*_{j,k} = \frac{\alpha_j}{n_j}\left(\rho I_y - \frac{p_{t_{j-1}} - p_0}{\lambda^T}\right)$. In equilibrium, since all fast traders observe the same signal I_y, each fast trader can calculate the earlier fast traders' order sizes. Hence, they do not have to observe the last traded price.

Remark 1.3.3. As discussed in the benchmark model, one can always set $\sigma_v = \sigma_z = 1$ by changing the units of currency and order size. In the following discussion, the effects of σ_v and σ_z are normalized.

Fast trader (j,k) is k-th of the n_j fast traders arriving at time t_j. Each of the n_j traders acts like a Cournot competitor and trades $\frac{1}{n_j+1}$ of the residual incoming order size $\hat{y} - S_{j-1}$. After time-t_j fast traders trade u_j shares, the residual order flow is reduced to $\hat{y} - S_j = \hat{y} - S_{j-1} - u_j$. Traders arriving next at time t_{j+1} follow a similar Cournot strategy and trade $\frac{n_{j+1}}{1+n_{j+1}}$ of the residual order flow $\hat{y} - S_j$. Figure 1.4 illustrates the equilibrium strategies when the speed profile is $\{2, 1\}$.

[13] $\rho \in [0, 1]$ by definition. Proposition A.3.1 on page 126 proves that $0 \leq \gamma < 1$ for any $J \geq 0$. Hence, θ is well defined.

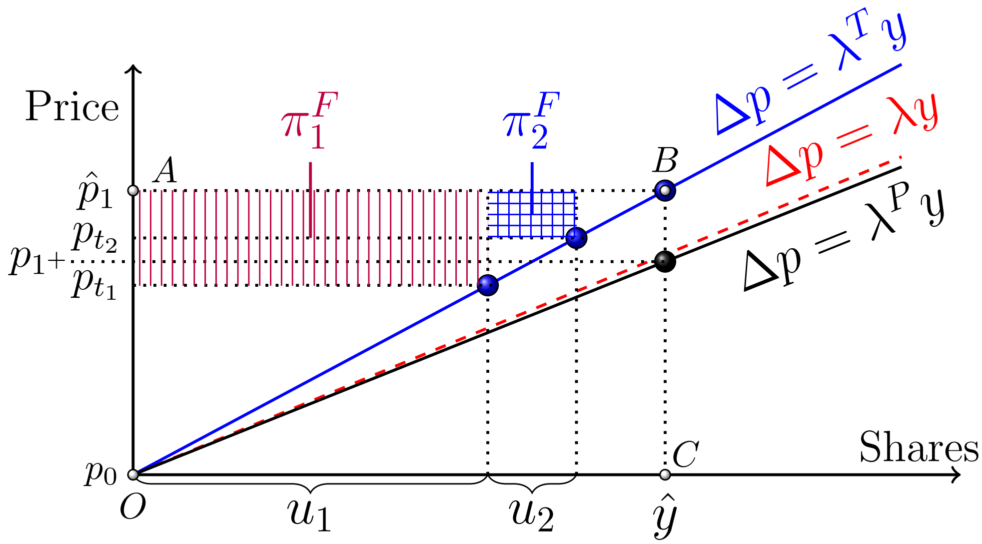

Figure 1.4: Equilibrium strategies when speed profile is $\{2,1\}$. $\Delta p = \lambda y$ is the pricing function without fast trader. When fast traders are present, market makers raise λ^T and lower λ^P. At time 0, market makers set p_0 and λ^T; at time t_1, two fast traders arrive and together trade $u_1 = 2\hat{y}/3$ shares at price $p_{t_1} = p_0 + \lambda^T u_1$; at time t_2, one fast trader trades $u_2 = \hat{y}/6$ shares at price $p_{t_2} = p_0 + \lambda^T(u_1 + u_2)$; at time 1, the informed trades x shares, the noise traders trade z shares, and the fast traders $-u_1 - u_2$ shares at the price $p_1 = p_0 + \lambda^T y$; at time 1^+, market makers set the quote to $p_{1^+} = p_0 + \lambda^P y$. The shaded rectangles are the profits of the two groups of fast traders π_1^F and π_2^F. The sum of π_1^F and π_2^F equals 1/4 of rectangle $OABC$ which corresponds to market makers' price impact surplus $\lambda^T y^2$.

Normal-speed traders' strategies $X(\cdot)$, $P(\cdot)$, and $Q(\cdot)$ and the choice parameters β, λ^T, and λ^P in Eq.(1.3.8) have exactly the same functional forms as in the benchmark model of Theorem 1.2.2. The key difference between the benchmark model and the general model lies in the new parameter γ.

Definition 1.3.6 (**Speed friction** γ)**.** Given a speed profile $\{n_1, n_2, \cdots, n_J\}$ of fast traders, the speed friction is defined as follows

$$\gamma = \sum_{j=1}^{J} \left(\alpha_j (1-\alpha_j) \prod_{i=0}^{j-1} (1-\alpha_i)^2 \right) \tag{1.3.10}$$

where $\alpha_j = \frac{n_j}{1+n_j}$ for all $1 \leq j \leq J$. Proposition A.3.1 shows that $\gamma \in [0, 1)$.

Speed friction γ is the crucial parameter that summarizes all equilibrium-relevant information in the profile of fast traders' relative speeds. Roughly speaking, speed friction γ is increasing in the heterogeneity of fast traders' relative speeds. For example, we calculate the speed friction for several special types of speed profile as follows.

Proposition 1.3.2. *Speed friction γ is easily calculated from Definition 1.3.6.*

1. *Monopoly speed profile $\{1\}$. $\gamma = \frac{1}{4}$.*

2. *Stackelberg-N speed profile $\{1, 1, \cdots, 1\}$. $\gamma = \frac{1}{3}\left(1 - \frac{1}{4^N}\right)$.*

3. *Cournot-N speed profile $\{N\}$. $\gamma = \frac{N}{(1+N)^2}$.*

Fast traders have more heterogeneous speeds in a Stackelberg speed profiles than in a Cournot speed profile. And we can see that given the number of fast traders, a Stackelberg speed profile has higher speed friction than the Cournot speed profiles.

Before investigate the properties of speed friction γ in detail, let's first examine its impact on equilibrium parameters. Speed friction γ and fast traders' information quality affect equilibrium through their impact on θ. Later, we will see that that market quality is better when θ is higher.

Definition 1.3.7 (**Market quality parameter** θ)**.**

$$\theta \equiv \sqrt{\frac{1-\rho\gamma}{1+\rho\gamma}} \in [0,1] \tag{1.3.11}$$

where ρ is fast traders' information quality and γ is the speed friction.

Proposition 1.3.3. *Keeping σ_v and σ_z constant, when speed friction decreases ($\gamma \downarrow$) or fast traders' information becomes less accurate ($\rho \downarrow$), market quality parameter goes up ($\theta \uparrow$), informed trading intensity goes up ($\beta \uparrow$), temporary price impact declines ($\lambda^T \downarrow$), and permanent price impact increases ($\lambda^P \uparrow$). Fast trading intensity α_j is not affected.*

Proof. Follows from Theorem 1.3.1. □

For given levels of uncertainty σ_v and σ_z, market quality parameter θ fully characterizes the equilibrium for normal-speed traders. By definition, market quality θ is solely determined by $\rho\gamma$, the product of fast traders' information quality and speed friction γ. Therefore, in equilibrium normal-speed traders only care about fast traders' information precision ρ and speed friction γ after controlling for σ_v and σ_z.

Figure 1.5 illustrates the impact of fast traders' information quality ρ and speed friction γ on equilibrium parameters.[14]

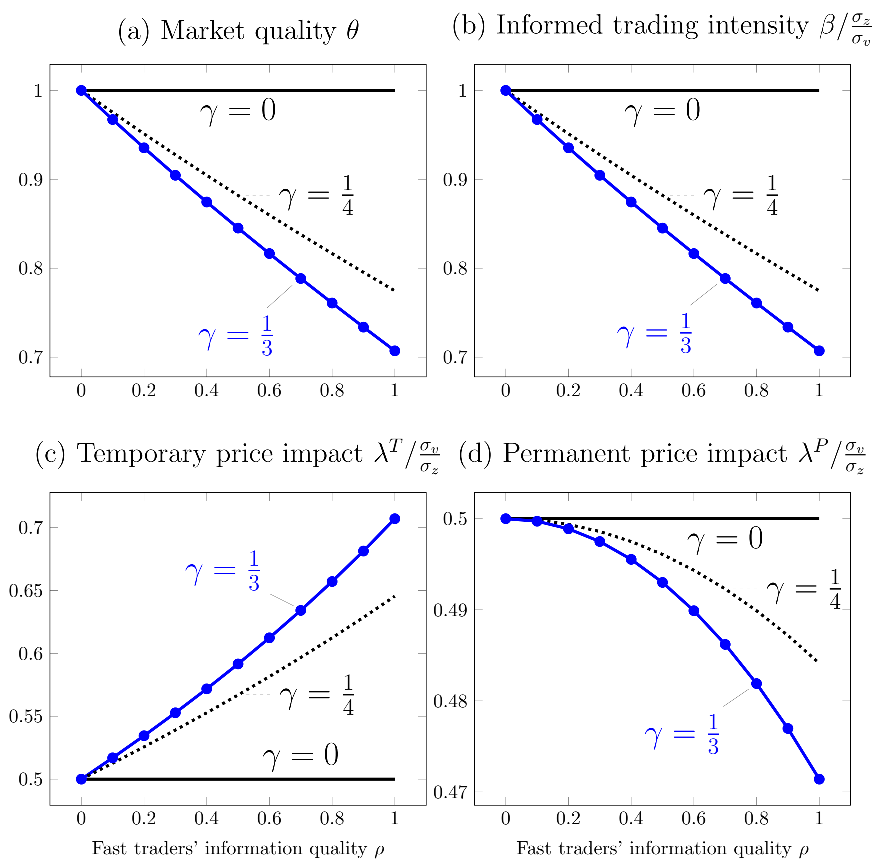

Figure 1.5: Equilibrium parameters of the general model normalized by volatility of fundamental value σ_v and volatility of noise trading σ_z. Theorem 1.3.1 shows that equilibrium is characterized by $\rho\gamma$ where $\rho \in [0, 1]$ is fast traders' information quality and $\gamma \in [0, 1]$ is level of speed friction. Market quality parameter θ, informed trading intensity β, and permanent price impact λ^P are decreasing in $\rho\gamma$; whereas temporary price impact λ^T is increasing in $\rho\gamma$.

[14] Because speed friction $\gamma = 1/4$ when only one fast trader is present, in the benchmark model, equilibrium market quality $\theta = \sqrt{\frac{1-\rho/4}{1+\rho/4}}$ is determined only by ρ.

1.3.3 Discussion on $\rho\gamma$ as a speed tax rate

Intuitively, fast traders levy a "speed tax" on the market makers with $\rho\gamma$ being the effective expected tax rate. When market makers receive an order of y shares, they mark the price up by $\lambda^T y$ and fill the order. The price impact surplus for executing the trade is $\lambda^T y^2$. Market makers use the surplus to offset the loss to the informed trader and to pay the speed tax to fast traders. Later we will see that fast traders' aggregate profit is $\gamma\,\mathbf{E}[\lambda^T \hat{y}^2] = \rho\gamma\,\mathbf{E}\left[\lambda^T y^2\right]$. Conditional on her signal I_y, fast traders take away a fraction γ of the expected price impact revenue $\mathbf{E}[\lambda^T \hat{y}^2]$. As shown in Figure 1.2, the monopolistic fast trader takes away 1/4 of the total price impact revenue and thus $\gamma = 1/4$ for the benchmark model. Unconditioned on I_y, the effective tax rate on market makers' price impact surplus is $\rho\gamma$.[15]

We can write an alternative model of financial transaction tax with this intuition. Suppose after a trading round, with probability ρ market makers are taxed at the rate of γ on their price impact surplus $\lambda^T y^2$. This setup would generate the same equilibrium strategies for the informed trader and market makers.

The effective tax rate $\rho\gamma$ goes down either because fast traders have less accurate signal ($\rho \downarrow$) or fast traders take away a smaller fraction ($\gamma \downarrow$) of market makers' price impact revenue. When the speed tax rate drops ($\rho\gamma \downarrow$), market makers are able to use a larger fraction of the surplus $\mathbf{E}[\lambda^T y^2]$ to cover loss to the informed trader. Competition among market makers then drives down the temporary price

[15] Subrahmanyam (1998) models the impact of a quadratic financial transaction tax levied on the informed and noise traders' orders.

impact λ^T. Seeing a lower cost to trade ($\lambda^T \downarrow$), the informed trader increases trading intensity ($\beta \uparrow$). As a result, the aggregate order flow contains more orders from the informed trader. The permanent price impact λ^P increases since the order flow is more informative.

1.4 Speed competition among fast traders

While they are already much faster than most other traders, HFTs keep investing in the latest speed technology to shave a few millisecond or even microseconds off the latency. They co-locate computers with the exchange's central matching engine; they build fiber optic cables under the Arctic ocean; they build algorithms directly into the hardware. Such investment is driven by the competition among peer HFTs.

The exchanges seem to be catering to the incessant demand for speed. They continue to reduce latencies of order processing; they build large data centers and lease co-location spots to traders; they provide sophisticated order types to facilitate faster and conditional order execution; and they provide faster data feeds with increasing granularity.

Regulators or market designers can introduce new rules to mitigate the impact of front-running. The impact of such rules on normal-speed traders boils down to their effects on fast trader's information quality ρ and the speed friction γ. It is relatively easy to conjecture a policy measure's impact on ρ. In this section, I investigate the properties of speed friction and examine the effect of speed competition on fast traders and on normal-speed traders. Specifically, I address the following questions:

(1) Why is speed vital to HFTs? (2) How does increased speed competition among HFTs affect speed friction γ?

1.4.1 Fast traders' profits and relative speeds

We can calculate the expected profit of all fast traders from Theorem 1.3.1.

Proposition 1.4.1 (Fast traders' profit). *In equilibrium, the expected profit of fast trader (j,k) is*

$$\mathbf{E}\left[\pi^F_{j,k}\right] = \frac{\sigma_v \sigma_z}{2} \times \rho \left(\frac{1}{\theta} + \theta\right) \frac{\alpha_j(1-\alpha_j)}{n_j} \prod_{i=0}^{j-1}(1-\alpha_i)^2 \tag{1.4.1}$$

The profit of all fast traders arriving at time t_j is

$$\sum_{k=1}^{n_j} \mathbf{E}\left[\pi^F_{j,k}\right] = \frac{\sigma_v \sigma_z}{2} \times \rho \left(\frac{1}{\theta} + \theta\right) \times \alpha_j(1-\alpha_j) \prod_{i=0}^{j-1}(1-\alpha_i)^2 \tag{1.4.2}$$

and the profit of all fast traders is

$$\mathbf{E}\left[\pi^F\right] = \sum_{j,k} \mathbf{E}\left[\pi^F_{j,k}\right] = \frac{\sigma_v \sigma_z}{2} \times \frac{1-\theta^2}{\theta} \tag{1.4.3}$$

where $\theta = \sqrt{\frac{1-\rho\gamma}{1+\rho\gamma}}$, $\alpha_j = \frac{n_j}{1+n_j}$, and n_j is the number of fast traders of the j-th fastest speed as in Theorem 1.3.1.

Proof. See Appendix A.2.3. □

Not surprisingly, aggregate fast trading profit $\mathbf{E}[\pi^F]$ is increasing in the effective speed tax rate $\rho\gamma$. Fast traders also make more profits when there is more uncertainty about the fundamental value ($\sigma_v \uparrow$) or there is more noise trading ($\sigma_z \uparrow$). Fast traders' profits come from the price impact of others' trades. When there is more fundamental uncertainty ($\sigma_v \uparrow$), the price impact of trades is higher and front-running an order of given size is more profitable. When there is more noise trading ($\sigma_z \uparrow$), trading volume is higher and there are more orders to front-run.[16]

For an individual fast trader (j, k), the expected profit $\mathbf{E}[\pi^F_{j,k}]$ depends not on her absolute speed but on her relative speed. As long as n_j fast traders arrive at a time in (t_{j-1}, t_{j+1}), their speed ranking stays the same and their expected profits do not change.[17]

Proposition 1.4.2 (Fast traders' profits and relative speed). *Suppose $n_j > 0$ fast traders arrive at time t_j and $n_{j+1} > 0$ arrive at time t_{j+1}. Then,*

$$\frac{\text{Expected profit of } \textbf{all} \text{ fast traders at time } t_j}{\text{Expected profit of } \textbf{all} \text{ fast traders at time } t_{j+1}} = \frac{n_j(1+n_{j+1})^2}{n_{j+1}} \geq 4$$

$$\frac{\text{Expected profit of } \textbf{one} \text{ fast trader at time } t_j}{\text{Expected profit of } \textbf{one} \text{ fast trader at time } t_{j+1}} = (1+n_{j+1})^2 \geq 4$$

The ratios are minimized when $n_j = n_{j+1} = 1$.

Proof. Follows from Eq. (1.4.2) of Proposition 1.4.1. □

Speed establishes a pecking order. Fast traders have very strong incentives to be relatively fast because higher speed translates to a two-fold advantage: faster traders are able to acquire *larger* positions at *better* prices. For example, if each fast

[16] This is consistent with the popular belief: "... two things [HFT] needs the most: trading volume and price volatility", Bloomberg Businessweek, http://goo.gl/NEQOV

[17] In this model fast traders' arrival timings are deterministic. Absolute speed could also matter when arrive times are random because a larger advantage in absolute speed translates to a higher probability of arriving earlier.

trader has a different speed, then the j-th fast trader could pick up twice as many shares as the $(j+1)$-th fast trader and the price discount is twice as large. As a result, a fast trader could quadruple her expected trading profits by moving up one spot in the speed ranking among fast traders. Consistent with the model prediction, Baron et al. (2012) find aggressive HFTs' profits increase with their relative speeds.

The increasing return to being relatively fast offers one explanation of high frequency traders' obsession with speed. It might also explain the exchanges' motive to increase the trading frequency. Higher trading frequencies tend to create more space for speed competition. For example, suppose the informed and noise traders' latency is 1 second. If the trading platform allows 10 trades per second, fast traders would have the highest profit if their orders arrive within the first 0.1 second. Suppose the exchange upgrade the trading platform. Normal-speed traders' latency is 0.1 second and the trading frequency is 1000 times per second. Then fast traders need to arrive within the first 0.001 second to reap the highest profit. Exchange could extract large rents by offering tiered access speeds to the markets.

1.4.2 Speed competition and speed friction γ

I have shown that the fast traders have strong incentives to engage in speed competition because their profits decay rapidly as they go down in the speed ranking. The impact of speed competition is not limited to fast traders. In this section, I show that speed competition among fast traders affects speed friction γ and in turn affects normal-speed traders by changing the effective speed tax rate $\rho\gamma$.

Roughly speaking, speed friction γ increases with the heterogeneity of fast traders' *relative* speeds. Recall that the speed profile $\{n_1, n_2, \cdots, n_J\}$ describes the number of fast traders arrive at each of the J moments and speed friction γ is defined as

$$\gamma = \sum_{j=1}^{J} \left(\alpha_j (1 - \alpha_j) \prod_{i=0}^{j-1} (1 - \alpha_i)^2 \right) \tag{1.4.4}$$

where the equilibrium fast trading intensity is $\alpha_j = \frac{n_j}{n_j+1}$ for all j. Speed friction γ is not affected by other exogenous parameters σ_v, σ_z, v_0, or ρ. It is determined solely by fast traders' speed profile $\{n_1, n_2, \cdots, n_J\}$.[18]

Definition 1.4.1 (Equivalent speed profiles). Two speed profiles are equivalent if they have the same speed friction γ.

Proposition 1.4.3. *Adding* 0 *to (or removing* 0 *from) a speed profile does not change speed friction.*

Proof. $\{n_1, n_2, \cdots, n_{j-1}, 0, \cdots n_J\}$ and $\{n_1, n_2, \cdots, n_{j-1}, n_{j+1}, \cdots n_J\}$ are equivalent speed profiles because setting $n_j = 0$ is equivalent to remove all n_j related terms from γ. □

Proposition 1.4.4. *Only the relative speeds of fast traders affect the speed friction* γ.

Proof. Follows from Proposition 1.4.3 because changing absolute speed without changing relative speed is equivalent to changing the ranking of some 0s in the speed profile. □

[18] If fast traders are risk averse, ρ and γ become codependent through α_j (Appendix A.1).

The two results simplify analysis on speed friction γ. For example, consider three speed profiles $\{1, 0, 0, 0, 2\}$, $\{1, 0, 2\}$, and $\{0, 0, 0, 1, 2\}$. At the first glance they look quite different. In the three profiles, fast traders arrive at different times and the speed differences between the two groups of fast traders are not the same. The speed profiles, however, are all equivalent because after removing the 0s they all reduce to the speed profile $\{1, 2\}$. The equilibrium is exactly the same as long as 1 trader is the fastest and 2 traders are the second fastest.

Example 1: We illustrate the effect of speed competition on fast traders' profit and on the overall speed friction with an example. The example is illustrated in Figure 1.6. Suppose four fast traders, A,B,C, and D, are trading.

1. Initially four traders have the same speed and the starting speed profile is Cournot $\{4\}$. The speed friction $\gamma = \frac{4}{25} = 0.16$. All traders make the same expected profit.

2. Suppose the exchange increases the trading frequency. Trader A, B, and C subscribe to the upgraded co-location service. The speed profile becomes $\{3, 1\}$ and the speed friction γ increases to $\frac{13}{64} \approx 0.20$. According to Proposition 1.4.2, trader A,B, and C each makes 4 times as much profit as trader D. The ratio of the four traders' profits is $4 : 4 : 4 : 1$.

3. Suppose trader A purchases the most advanced computer and becomes even faster than B and C. The speed profile becomes $\{1, 2, 1\}$ and the speed friction γ increases to 0.31. The ratio of the four traders' profits is $36 : 4 : 4 : 1$.

4. Suppose trader B improves the algorithm and becomes faster than C but is slower than A. The speed profile becomes Stackelberg $\{1, 1, 1, 1\}$ and the speed friction γ further increases to 0.33. The ratio of the four traders' profits is $64 : 16 : 4 : 1$.

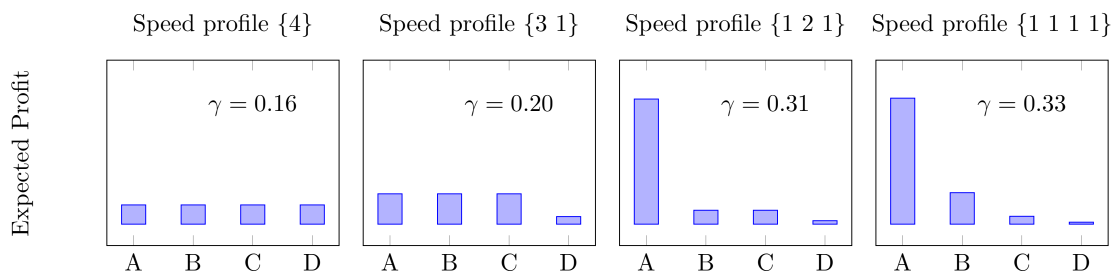

Figure 1.6: Effect of speed competition on fast traders' profits.

We see from the example that fast traders have strong incentive to become relatively faster. As the trading frequency increases, fast traders race to step ahead. Normal-speed traders could experience higher or lower speed frictions depending on how the speed profile evolves.

Example 2: Equation (1.4.4) is all one needs to compute the speed friction γ of any given speed profile. For example, suppose 3 fast traders are in the market. We know from Proposition 1.4.3 and 1.4.4 that all speed profiles can be reduced to one of $\{1,1,1\}$, $\{1,2\}$, $\{2,1\}$, and $\{3\}$. Their speed frictions are:

$$\begin{aligned}
\text{Stackelberg-3 profile } \{1,1,1\}: \quad \gamma_{\{1,1,1\}} &= \frac{1}{2}\left(1-\frac{1}{2}\right)+\frac{1}{2}\left(1-\frac{1}{2}\right)\left(1-\frac{1}{2}\right)^2 \\
&\quad +\frac{1}{2}\left(1-\frac{1}{2}\right)\left(1-\frac{1}{2}\right)^4 = \mathbf{\frac{21}{64}} \\
\text{Profile } \{1,2\}: \quad \gamma_{\{1,2\}} &= \frac{1}{2}\left(1-\frac{1}{2}\right)+\frac{2}{3}\left(1-\frac{2}{3}\right)\left(1-\frac{1}{2}\right)^2 = \mathbf{\frac{11}{36}} \\
\text{Profile } \{2,1\}: \quad \gamma_{\{2,1\}} &= \frac{2}{3}\left(1-\frac{2}{3}\right)+\frac{1}{2}\left(1-\frac{1}{2}\right)\left(1-\frac{2}{3}\right)^2 = \mathbf{\frac{1}{4}} \\
\text{Cournot-3 profile } \{3\}: \quad \gamma_{\{3\}} &= \frac{2}{3}\left(1-\frac{2}{3}\right) = \mathbf{\frac{2}{9}}
\end{aligned}$$

We have two important observations from the example and it turns out they are true in general.

1. Because $\gamma_{\{1,1,1\}} > \gamma_{\{1,2\}} > \gamma_{\{2,1\}} > \gamma_{\{3\}}$, the Cournot-3 speed profile $\{3\}$ has the lowest speed friction γ and the Stackelberg-3 speed profile $\{1,1,1\}$ has the highest among all possible speed profiles with 3 fast traders.

2. Recall that in the monopoly fast trader model, speed friction $\gamma = \frac{1}{4}$. Here, $\gamma_{\{1,1,1\}} > \gamma_{\{1,2\}} > \gamma_{\{2,1\}} = \frac{1}{4} > \gamma_{\{3\}}$. When 3 fast traders compete, the speed friction γ can be higher, equal, or lower than the speed friction when there is only one monopolistic fast trader.

Remark 1.4.1. Speed friction γ is increasing in the heterogeneity of fast traders' relative speeds. So $\gamma_{\{1,1,1\}} > \gamma_{\{2,1\}} > \gamma_{\{3\}}$. Speed friction γ includes additional information about the competition among fast traders. For example, speed profile $\{2,1\}$ and $\{1,2\}$ have the same speed heterogeneity but $\gamma_{\{2,1\}} < \gamma_{\{1,2\}}$. Intuitively, increased competition at time t_1 reduces aggregate fast trading profits more because fast traders at time 1 are more profitable.

Proposition 1.4.5. *Suppose N fast traders are in the market.*

1. *Speed friction γ is maximized if every fast trader has a different speed. $\gamma \leq \gamma_{\{1,1,\cdots,1\}} = \frac{1}{3}\left(1 - \frac{1}{4^N}\right)$.*

2. *Speed friction γ is minimized if every fast trader has the same speed. $\gamma \geq \gamma_{\{N\}} = \frac{N}{(1+N)^2}$.*

Proof. See Appendix A.3.3. □

Proposition 1.4.6. *Suppose* one *new fast trader enters the market.*

1. *If the new fast trader has the same speed as some existing fast traders, after the entry, speed friction γ declines;* [19]

2. *If the new fast trader has a different speed from all existing fast traders, after the entry, speed friction γ increases.*

Proof. See Appendix A.3.2. □

The entry of fast traders can increase or decrease the level of speed friction γ depending on the entrant's speed relative to existing fast traders.

[19] When fast traders are risk averse, same speed entries could increase speed friction. Intuitively, two risk averse Cournot fast traders in aggregate may trade like a monopolistic risk neutral fast trader. A risk averse fast trader chooses a trading intensity α lower than the risk neutral monopoly level of 1/2. As more of such risk averse fast traders enter, their aggregate trading intensity increases and could be closer to 1/2. Eventually, however, α keeps increasing beyond 1/2 when more such fast traders enter. Speed friction γ keeps dropping and converges to the same limit as the risk neutral case.

On one hand, suppose all existing fast traders and entrants have the same speed. Then, as new fast traders enter, they form a series of Cournot-N speed profiles. As N increases, the speed friction $\gamma = \frac{N}{(N+1)^2}$ keeps dropping and converges to zero when $N \to \infty$. Intuitively, when fast traders have the same speed, they compete on quantity in a Cournot competition. When more fast traders are present, they bid up their entry price (when they buy). The profit margin becomes smaller and their aggregate profits keep declining. Aggregate fast trading profits become a smaller fraction γ of market makers' expected price impact revenue. Since the effective speed tax rate $\rho\gamma$ decreases, market makers pay less speed tax to fast traders and they can lower the temporary price impact amplifier λ^T/λ^P. In the limit when $N \to \infty$, speed tax goes to zero and $\lambda^T/\lambda^P \to 1$. The impact of fast traders vanishes. [20]

On the other hand, suppose every entrant has a different speed. Then, as new fast traders enter, they form a series of Stackelberg-N speed profiles. The speed friction $\gamma = \frac{1}{3}\left(1 - \frac{1}{4^N}\right)$ keeps increasing with N and converges to $\frac{1}{3}$ as $N \to \infty$. Intuitively, each new fast trader is a local monopoly before the next trader arrives. Each of them trades away half of the residual order $\hat{y} - S_{j-1}$ and makes $\frac{1}{4}\rho$ of the residual profits. Multiple fast traders effectively split a large trade into small pieces and march along the supply curve. A series of monopolies are worse than one monopoly fast trader because their combined speed is higher than the speed of any individual fast trader. In aggregate, fast traders have more trading opportunities in a trading round. Although fast traders do not collude, in aggregate they are more

[20] Brunnermeier and Pedersen (2005) find that when infinitely many predators compete, their effect vanishes because predators effectively have the same speed in their model.

profitable. The market does not break down as $N \to \infty$ because the effective speed tax rate $\rho\gamma < \frac{\rho}{3} < 1$. Market makers can still afford to pay the speed tax and cover their loss to the informed trader by raising the temporary price impact λ^T. Entry of more fast traders does not drive the total speed tax to zero. It only drives the slowest fast trader's profit to zero.[21]

Figure 1.7 illustrates an example. Starting from the profile $\{2, 1\}$, speed friction γ keeps declining if new entrants all have the same speed as the fastest trader. The speed profile evolves from $\{2, 1\}$ to $\{3, 1\}$, $\{4, 1\}$, $\{5, 1\}$, and so on. Speed friction γ becomes closer to the Cournot-N lower bound.

Alternatively, if every new entrant is faster than all existing traders, speed friction γ keeps increasing. The speed profiles evolves from $\{2, 1\}$ to $\{1, 2, 1\}$, $\{1, 1, 2, 1\}$, $\{1, 1, 1, 2, 1\}$ and so on. Speed friction γ approaches the Stackelberg-N upper bound very quickly.

Remark 1.4.2. Entry of *multiple* fast traders at one new speed could reduce speed friction. For example, if an infinite number of traders with the highest speed enter, then speed friction γ goes to zero. A change from speed profile $\{n_1, n_2 = 0, n_3\}$ to speed profile $\{n_1, n_2 = n, n_3\}$ can be achieved in multiple steps by adding 1 to n_2 in each step. The first entry changes n_2 from 0 to 1 and it increases speed friction γ; the subsequent entries at t_2 reduce speed friction γ. The net effect depends on the number of fast traders entering at t_2.

[21] If N fast traders collude, the key result still holds: allowing them to have different speeds weakens the price competition among fast traders. When fast traders have different speeds, they

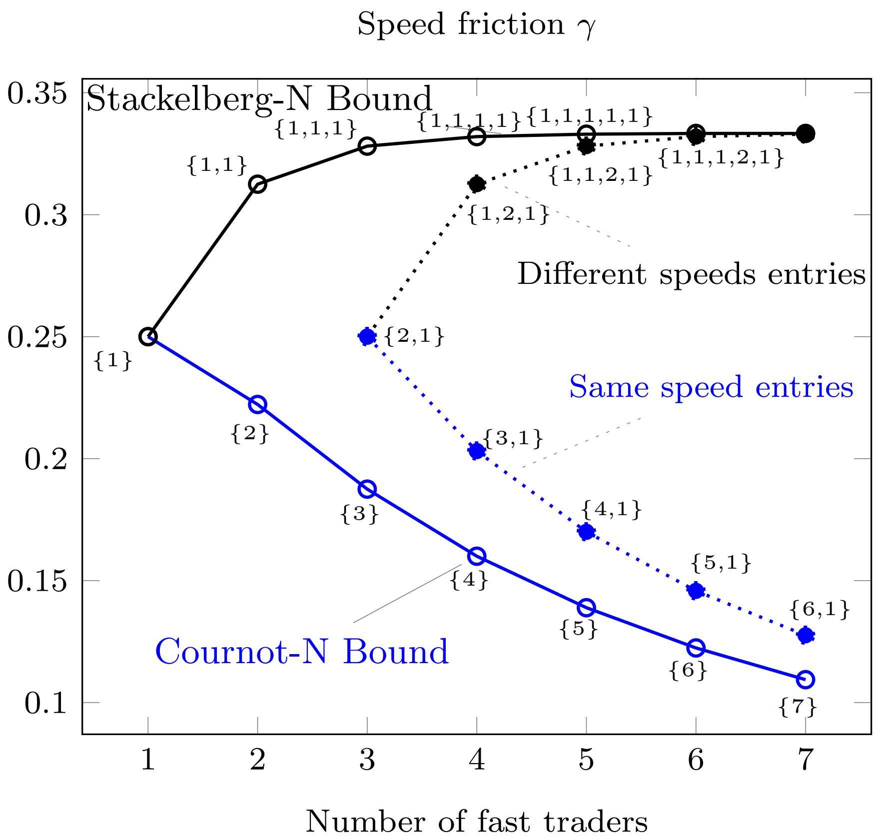

Figure 1.7: Effect of entry fast traders on speed friction γ. Given the number of fast traders N, a speed profile's level of speed friction γ is bounded by Stackelberg-N from above and by Cournot-N from below. Starting from speed profile $\{2, 1\}$, speed friction γ keeps going down if new fast traders keep entering at t_1 and speed profiles become $\{3, 1\} \to \{4, 1\} \to \{5, 1\} \to \{6, 1\}$; whereas speed friction γ keeps increasing if new fast traders keep entering with higher speeds than existing fast traders and speed profile becomes $\{1, 2, 1\} \to \{1, 1, 2, 1\} \to \{1, 1, 1, 2, 1\}$.

Without knowing the cost of speed, we cannot pin down the long-term equilibrium level of speed friction γ. The following result suggests that under quite general conditions, the impact of fast traders does not vanish even when infinitely many fast traders compete.

Proposition 1.4.7 (Speed friction in the limit). *Suppose N fast traders are present and $n_1 > 0$ of the N traders are the* fastest. *In the limit when $N \to \infty$, speed friction vanishes if and only if $n_1 \to \infty$, i.e.,*

$$\lim_{N\to\infty} \gamma = 0 \iff \lim_{N\to\infty} n_1 = \infty \tag{1.4.5}$$

Proof. See Appendix A.3.4. □

Due to physical limits, e.g. the number of co-location spots, only a very limited number of traders can be the fastest in a continuous market. Even adding infinite number of traders at the second highest speed would not eliminate speed friction γ.

1.5 Market quality

I have shown that speed competition affects the relative profit of each fast trader and affects the speed friction γ. In this section, I study how fast traders' information quality ρ and speed friction γ affect the information aggregation and liquidity provision functions of a financial market.

---

extract more aggregate profits than when they have the same speed. See Appendix A.7.

1.5.1 Information efficiency

In the context of this paper, fundamental value of the risky asset v is the informed trader's private information. If people learn more about v by observing the trading process, the market is more informationally efficient.

Definition 1.5.1 (Information efficiency). Information efficiency at time t is defined as $\phi_t = 1 - \frac{\mathbf{V}\text{ar}[v|\mathcal{F}_t]}{\mathbf{V}\text{ar}[v]} \in [0,1]$ where $\mathcal{F}_t$ represents the public information at time t. Specifically,

$$\text{Intermediate information efficiency: } \phi_{t_1} = 1 - \frac{\mathbf{V}\text{ar}[v|\mathcal{F}_{t_1}]}{\mathbf{V}\text{ar}[v]} \tag{1.5.1}$$

$$\text{Ex post information efficiency: } \phi_{1^+} = 1 - \frac{\mathbf{V}\text{ar}[v|\mathcal{F}_{1^+}]}{\mathbf{V}\text{ar}[v]} \tag{1.5.2}$$

where t_1 is the time when the first fast trader arrives.[22]

Information efficiency ϕ_t measures how much uncertainty about the fundamental value v is resolved by time t. If people have learned the true value v precisely by time t, then $\mathbf{V}\text{ar}[v|\mathcal{F}_t] = 0$ and information efficiency $\phi_t = 1$; if people have not learned any information about v by time t, $\mathbf{V}\text{ar}[v|\mathcal{F}_t] = \mathbf{V}\text{ar}[v]$ and $\phi_t = 0$.

We need two information efficiency measures to describe the information revealed at different moments of the trading round. The intermediate information efficiency ϕ_{t_1} measures how much information about v is revealed by the first fast trader. Because fast traders observe the same signal I_y, no additional information about v is revealed in the time window $(t_1, 1)$. The ex post information efficiency ϕ_{1^+}

[22] Variance captures the level of uncertainty about the normally distributed v. For general distributions, we could use entropy in lieu of variance.

measures how much information about v is revealed by the end of the trading round.

Proposition 1.5.1 (Equilibrium information efficiency). *In equilibrium,*

1. *intermediate information efficiency* $\phi_{t_1} = \frac{\rho\theta^2}{1+\theta^2} = \frac{\rho(1-\rho\gamma)}{2}$*,*

2. *and ex post information efficiency* $\phi_{1^+} = \frac{\theta^2}{1+\theta^2} = \frac{1-\rho\gamma}{2}$*,*

Proof. See Appendix A.4.1. □

Corollary 1.5.2. *Speed friction γ and fast traders' information quality ρ determine the information efficiency of the market.*

1. *When speed friction is higher ($\gamma \uparrow$), intermediate and ex post information efficiencies are both lower ($\phi_{1^-} \downarrow$ and $\phi_{1^+} \downarrow$).*

2. *When fast traders' information quality is more accurate ($\rho \uparrow$), intermediate information efficiency goes up ($\phi_{1^-} \uparrow$) while ex post information efficiency goes down ($\phi_{1^+} \downarrow$).*

Proof. Follows from Proposition 1.5.1. □

Figure 1.8 illustrates the impact of fast traders' information quality ρ and speed friction γ on information efficiency. The intermediate information efficiency equals a fraction ρ of the ex post efficiency, i.e., $\phi_{t_1} = \rho\phi_{1^+}$. Fast traders effectively bring part of the information to the market earlier at time t_1.

A high speed friction γ impedes informed trading because it raises the temporary price impact λ^T. Ex post information efficiency ϕ_{1^+} is lower because the order flow contains less orders from the informed trader. The intermediate efficiency ϕ_{t_1} is also

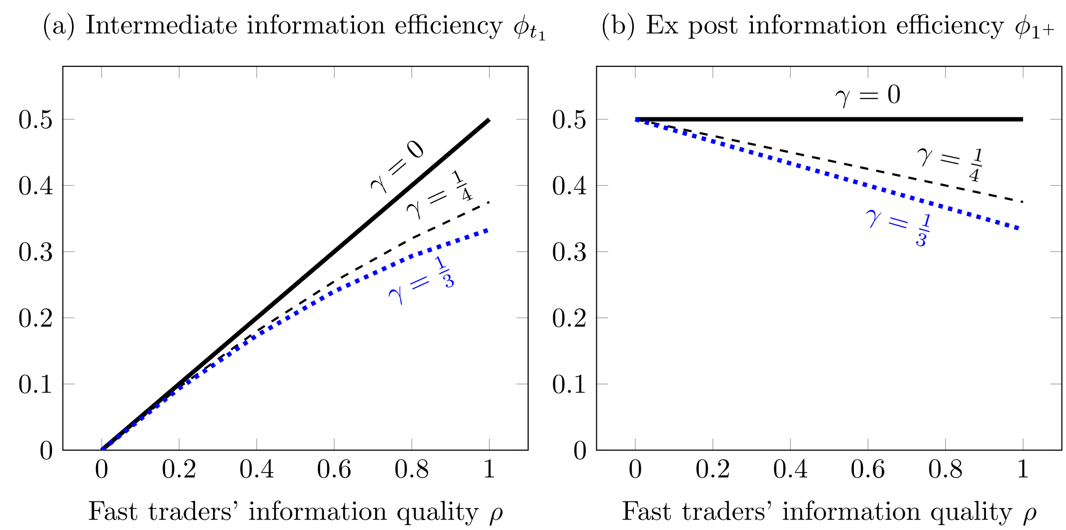

Figure 1.8: Equilibrium information efficiency. Intermediate information efficiency ϕ_{t_1} increases with fast traders' information quality ρ and decreases with speed friction γ; ex post information efficiency ϕ_{1+} decreases with ρ and γ. Intermediate information efficiency ϕ_{t_1} equals a fraction ρ of the ex post information efficiency ϕ_{1+} for all cases.

lower because fast traders bring a *fixed* fraction ρ of the ex post information to t_1. To improve information efficiency we need to lower speed friction γ.

In models of informed trading such as Holden and Subrahmanyam (1992), Foster and Viswanathan (1996), Back et al. (2000), Bernhardt and Miao (2004), and Li (2013), an informed trader often ramps up her trading intensity when faced with the competition from other informed traders with highly positively correlated information.[23] The elevated informed trading intensity then leads to more informative prices.

In this study, fast traders infer similar but less accurate information about v from the order flow signal I_y. Yet, faced with the pressure from fast traders, the informed trader reduces rather than increases her trading intensity. As a result ex post information efficiency ϕ_{1+} is lower.

The opposite impact on ex post information efficiency highlights the difference between fast traders and competing informed traders. Typically when a informed trader chooses a trading intensity, she faces the trade-off of price impact and information decay. If an informed trader slows down, her private information is traded away by competing informed traders; if an informed trader speeds up, her price impact is higher.[24]

[23] Foster and Viswanathan (1996) and Back et al. (2000) show that the trading game eventually turns into the "waiting game" stage when multiple informed traders refrain from trading. By the time the game reaches the waiting game stage, however, the informed traders' residual private information has become negatively correlated.

[24] Kyle, Obizhaeva and Wang (2013) recently develop a continuous time model where every trader is overconfident, has a flow of private information, and endogenously chooses an optimal

In this model, the informed trader does not have the incentive to speed up for two reasons: (1) she cannot reduce information decay by trading more intensively because fast traders always anticipate the order flow; and (2) if the informed does not trade, her information is not traded away by fast traders because fast traders do not have an independent source of information about fundamentals. Hence the informed trader always slow down the trading in response to a higher temporary price impact λ^T. As a result, the more information fast traders know, the slower the informed trader trades and the less information is revealed by the end of each trading round.

A higher ρ, however, improves intermediate information efficiency ϕ_{t_1}. Proposition 1.5.1 shows that $\phi_{t_1} = \rho\phi_{1^+}$. Despite the dampening effect on ϕ_{1^+}, in equilibrium intermediate information efficiency ϕ_{t_1} increases with ρ because fast traders bring a larger fraction ρ of the ex post information to intermediate moments.

Nevertheless, the social value of intermediate information efficiency is questionable. Intermediate information efficiency could be socially valuable if (1) people can use the intermediate information to make a welfare enhancing economic decision and (2) the cost of delaying the decision from time t_1 to time 1^+ is very high. Both conditions are unlikely to be true when holding horizons are at the minute or second level. It is hard to imagine a case when normal-speed traders and outside agents must use p_{1^-} to make economic meaningful decisions. Even if the information is crucial, they could wait until time 1^+ and use a more informative price p_{1^+}. After all, fast trading is profitable because other traders cannot react fast enough in the

trading speed.

time window $[1^-, 1^+]$.

Therefore, although we face a trade-off between intermediate and ex post information efficiency, from the social welfare perspective, one probably should not raise ρ to improve the intermediate information efficiency at the expense of ex post information efficiency. Fast traders' production of order flow information I_y is a classic example of socially wasteful production of "foreknowledge" (Hirshleifer (1971)).

Fast traders make intermediate prices more informative by trading earlier during time $(t_1, 1)$ on a noisy signal I_y. The closing quote p_{1^+}, however, is less informative because the informed trader reduces trading intensity. Fast traders' effect on the order flow is similar to a prying messenger's effect on a letter. Suppose a messenger (HFT) glances at a letter (order flow) and summarizes it to the receiver (market) right before delivering it. The summary (a HFT trade) is informative but its information value is fleeting: the letter itself is much more informative than the summary. Furthermore, the sender (the informed) is less likely to write clearly ex ante worrying about privacy issues and ultimately the receiver (market) is less informed.

In sum, to improve the more economically meaningful ex post information efficiency, one should try to reduce speed friction γ and fast traders' information quality ρ.

1.5.2 Market liquidity

In this paper, noise traders trade for non-informational motives; a market is less liquid if noise traders expect to lose more to trade the same number of shares. Market liquidity also affects the informed trader because trading is anonymous.

Vayanos and Wang (2012) point out that different measures of market liquidity are designed to capture different market frictions. In the end, however, all the measures attempt to capture the impact of market friction on traders' economic profits. Thus, in addition to the traditional measure λ^T, I also use normal-speed traders' expected profit to measure market liquidity.

Proposition 1.5.3 (Equilibrium expected profits). *In equilibrium,*

$$\text{Informed trader's profit: } \mathbf{E}\left[\pi^I\right] = \frac{\sigma_v \sigma_z}{2} \times \theta \tag{1.5.3}$$

$$\text{Noise traders' profit: } \mathbf{E}\left[\pi^N\right] = \frac{\sigma_v \sigma_z}{2} \times \frac{-1}{\theta} \tag{1.5.4}$$

$$\text{Fast traders' total profit: } \mathbf{E}\left[\pi^F\right] = \sum_{j,k} \mathbf{E}\left[\pi^F_{j,k}\right] = \frac{\sigma_v \sigma_z}{2} \times \frac{1-\theta^2}{\theta} \tag{1.5.5}$$

where $\theta = \sqrt{\frac{1-\rho\gamma}{1+\rho\gamma}}$ *as in Theorem 1.3.1.*

Proof. See Appendix A.4.2. □

Corollary 1.5.4 (Market liquidity and fast trading). *Keep $\sigma_v\sigma_z$ fixed. In equilibrium, market is less liquid when fast traders' information becomes more accurate ($\rho \uparrow$) or speed friction goes up ($\gamma \uparrow$): temporary price impact is higher ($\lambda^T \uparrow$), the informed trader is less profitable ($\mathbf{E}[\pi^I] \downarrow$), and noise traders lose more ($\mathbf{E}[\pi^N] \downarrow$). Only fast traders' aggregate profit is higher ($\mathbf{E}[\pi^F] \uparrow$).*

Proof. Follows from Proposition 1.5.3. □

Figure 1.9 illustrates the effect of fast traders' information quality ρ and the speed friction γ on market liquidity. When fast traders levy a higher speed tax rate $\rho\gamma$ on market makers' price impact revenue $\mathbf{E}[\lambda^T y^2]$, market makers raise temporary price impact λ^T so that they can still break even. As a result, it becomes harder for the informed trader to extract rent based on the same information. Noise traders face less adverse selection because the informed trader trades less. Nonetheless noise traders suffer more losses to trade the same number of shares. Effectively, noise traders must pay information rent to the informed trader and speed rent to the fast trader. The reduction in information rent is not enough to cover the higher speed rent. Hence, the market is less liquid for the informed and noise traders.

We can also look at the impact of fast trading from a tax incidence perspective. As discussed earlier, fast trading effectively impose a speed tax on market makers' price impact revenue. In the model market makers' demand is the most elastic and noise traders' demand is the least elastic. Thus the burden of the tax is paid mostly by the noise traders, less so by the informed trader, and not by market makers.

Do fast traders provide liquidity? Based on the way they trade, it might seem that they do. Fast traders "take liquidity" during time $(t_1, 1)$ when liquidity is cheap and "provide liquidity" at time 1 to the informed and noise traders when liquidity is expensive. One might do a reduced-form counter-factual analysis and find that the price would have been much worse for the liquidity demanders if fast traders were not trading at time 1. In the context of this paper, the conclusion is incorrect

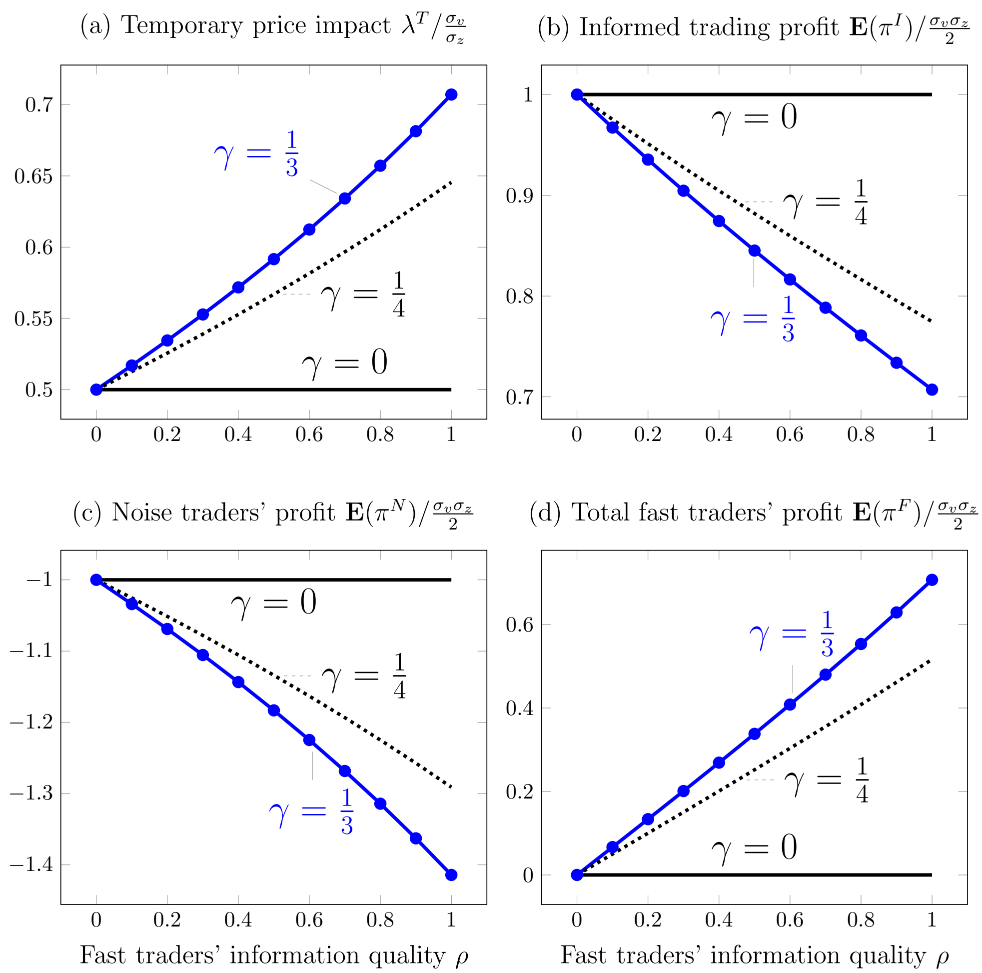

Figure 1.9: Equilibrium market liquidity.normalized by the volatility of fundamental value σ_v and volatility of noise trading σ_z. Proposition 1.5.3 imply that equilibrium liquidity is determined by $\rho\gamma$ where $\rho \in [0, 1]$ represents fast traders' information quality and $\gamma \in [0, 1/3]$ represents the level of speed friction. The informed and noise traders' expected profits are decreasing in $\rho\gamma$, whereas temporary price impact λ^T and fast traders' total expected profits are increasing in $\rho\gamma$.

because if fast traders were not present, the temporary price impact λ^T would have been much lower. Liquidity demanders, including the informed and noise traders, would have been better off without fast traders.

In sum, to improve market liquidity, one should try to lower fast traders' information quality ρ and speed friction γ.

1.6 Policy discussion

Regardless of one's desired balance of liquidity and price informativeness, when front-running fast traders have a more informative signal ρ or when speed friction γ is higher, the market quality is unambiguously worse: prices are less (ex post) informative and liquidity is more costly. The only possible social value is a more informative intermediate price which quickly becomes obsolete. Considering its short life, it hardly improves social welfare.

When considering a potential policy about high frequency trading, one should focus on gauging its impact on fast traders' information precision ρ and speed friction γ. A policy that reduces $\rho\gamma$ is going to improve ex post price informativeness and market liquidity.

Speed competitions

Reducing the entry cost to become a fast trader does not necessarily improve market quality. Not all competitions are equal. One needs to carefully induce Cournot competition and avoid Stackelberg competition among fast traders. When fast traders engage in Cournot competition on quantities, increased competition drives down their aggregate profit and enhance social welfare. When fast traders engage in Stackelberg competition on speed, increased competition may not result in lower aggregate front-running profits. Relative speed might only serve as a tiebreaker among front-runners to split the front-running profits. Under the setting of this paper, the aggregate front-running profit can even be higher when more front-runners with different speed are present. At the very least, the possibility to competite on speed weakens the competition on quantity or on price. It is puzzling to see that existing markets typically have no limit on speed competition while imposing tick size and minimum trade size to limit competition on price and quantity.

We may examine the effect of speed competition on market quality in Figure 1.10. The Cournot speed profile could be stable when the trading frequency (time granularity) is not too high. In a market where time is almost continuous, the slightest speed advantage counts. A fast trader has very strong incentive to develop and invest in new speed technologies to shave every nanosecond off her latency. Existing markets with high trading frequency are more likely to be close to the Stackelberg bound where front-running traders' profits are the highest, market liquidity is the most expensive, and ex post information efficiency is the lowest.

Periodic batch auctions

A natural way to deter the arms race in speed is to convert a continuous market

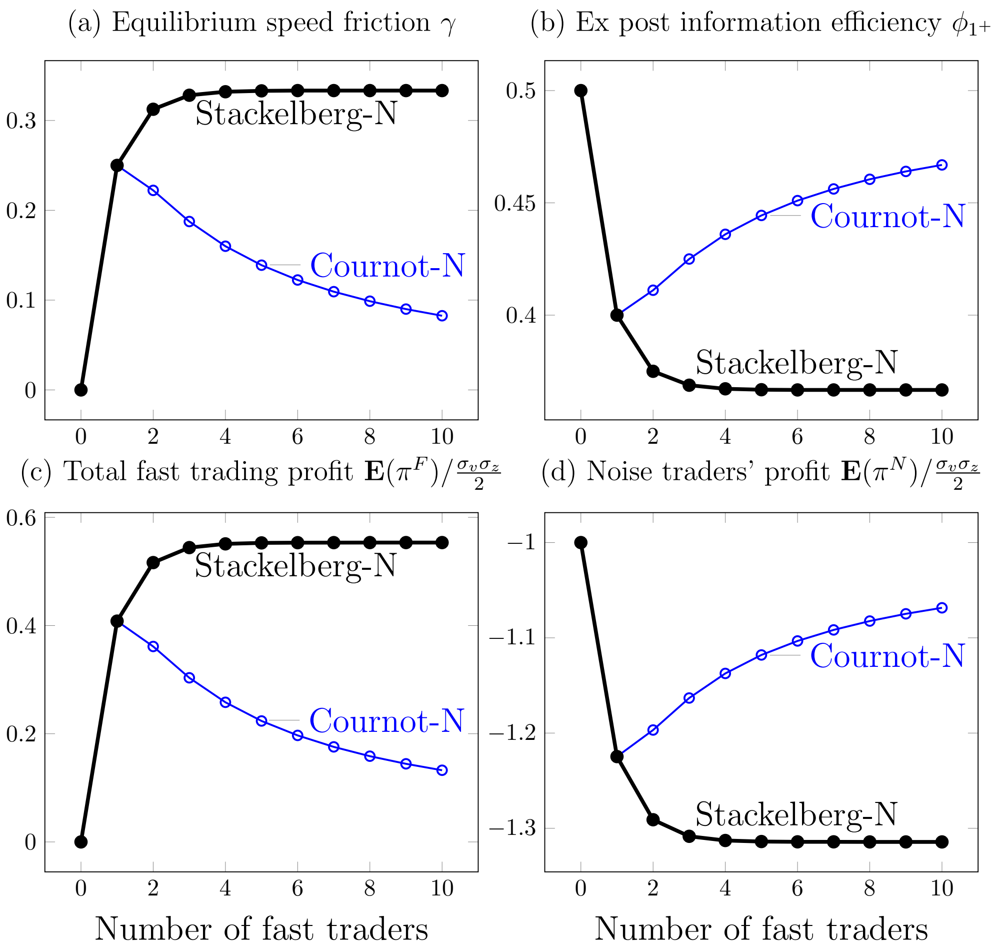

Figure 1.10: Market quality and speed competition. Fast traders information precision $\rho = 0.8$. **(a)** Speed friction increases (decreases) with number of fast traders if fast traders follow the Stackelberg-N (Cournont-N) speed profiles. **(b)** Ex post price informativeness decreases (increases) with number of fast traders if fast traders follow the Stackelberg-N (Cournont-N) speed profiles. **(c)** Total fast trading profit increases (decreases) with number of fast traders if fast traders follow the Stackelberg-N (Cournont-N) speed profiles. **(d)** Expected noise traders' profits decrease (increases) with number of fast traders if fast traders follow the Stackelberg-N (Cournont-N) speed profiles.

to a periodic uniform price auctions market as proposed by Budish et al. (2013).[25] Effectively, by eliminating the time priority of orders arriving withing a batching interval, the periodic auction enforces a Cournot speed profile. Given the information structure of this model, such a market design reduces the speed friction γ caused by front-running fast traders.

The result of this paper could shed light on the choice of the optimal batching interval Δ_t. To be consistent with the model of this paper, suppose that at the beginning of each batching interval, a pricing function is announced. Then, traders are allowed to submit market orders. Orders are accumulated until time Δ_t and executed at the same price according to the pricing function.

The batching interval Δ_t should be long but it does not need to be very long. Once several fast traders fall into the same batching interval, knowing that they would all get the same price, they follow the Cournot strategy. Speed friction γ goes down from $\frac{1}{3}\left(1-\frac{1}{4^N}\right)$ to $\frac{N}{(N+1)^2}$ where N is the number of fast traders in the same batch. Figure 1.10 shows that the improvement on market quality is substantial when Δ_t is long enough to batch the 5 fastest traders together. Considering the speed of existing high frequency traders, a batching interval of 1 second would probably make most of them barely profitable. The only cost is that we do not observe intermediate price updates during the one second interval and some liquidity demanders need to wait one second to fill their orders.

[25] Goldman and Sosin (1979) show that when speculators have a convex payoff, they tend to over speculate and cause price overshooting. Thus a market with finite trading frequency could have more “efficient” prices than a continuous market because less price overshooting.

Random order delays and latency floors

Recently, some trading venues have implemented innovative rules to curb the speed advantage of high frequency traders. They relax the time priority rule in various ways. In April 2013, a new foreign exchange trading platform ParFX adds a 20-80 millisecond random delays to orders arriving at the matching engine; in August, a major foreign exchange trading platform EBS introduced a "latency floor" on trades of AUD/USD: orders are first bundled within one to three milliseconds and then randomly placed in the queue.[26]

It might seem that these rules would have similar effects to those under periodic batch auctions. Surprisingly, however, they are far less effective and when the random delays or the latency floors are not long enough, these measures could even make market quality worse than the Stackelberg case.

The random order delays or latency floors turn a deterministic speed advantage into a random advantage. For the ParFX case, the fastest trader still has a speed advantage in probability because other traders are also subject to random delays. For the EBS case, each fast traders arrive during the same 1-3 millisecond interval has the same probability of being the first. Let's analyze the EBS's floor latency as an example.

Assume that N fast traders fall into the same batch and they all estimate that the incoming order flow of the *next* batch is $\hat{y}$. EBS randomly shuffles their positions in the queue so that each fast trader has a $\frac{1}{N}$ probability of being placed at each position j for all $1 \leq j \leq N$. To be comparable with the previous results, suppose all

[26] Retrieved from Reuters.com http://goo.gl/30SiwH

fast traders exit together in the next batch when y arrives.[27] We have the following results.

Proposition 1.6.1 (Speed friction under the EBS latency floor). *In a symmetric equilibrium where all N fast traders trade the same, each fast trader trades:* $u = \frac{2}{N+3}\hat{y}$ *and equivalent speed friction* $\gamma = \frac{4N}{(N+3)^2}$

Proof. See Appendix A.5. $\square$

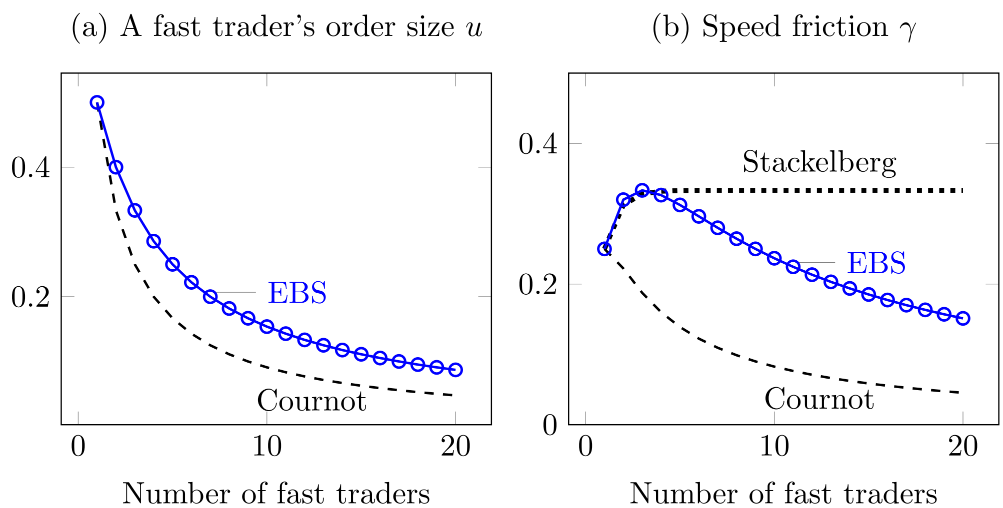

Figure 1.11: Speed friction and each fast trader's order size under the EBS latency floor rule.

The proposition is illustrated in Figure 1.11. Under the latency floor rule, speed friction is always higher than the Cournot case $\frac{N}{(N+1)^2}$ for all N. When $N \leq 3$, it could be even higher than speed friction of the Stackelberg case $\frac{1}{3}(1 - \frac{1}{4^N})$.

[27] Like before, a fast trader j could add $-u_j$ to the limit order book. In the next batch, the order of y shares arrives and is executed at $p_1 = p_0 + \lambda^T \sum_j u_j + \lambda^T(y - \sum_j u_j) = p_0 + \lambda^T y$.

The impact of the latency floor rule on fast traders is subtle. On one hand, the rule equalizes the speed of fast traders in the same batching interval in probability. So each fast trader submits an order of the same size. On the other hand, unlike in the periodic uniform price auctions, fast traders still receive different prices under the latency floor rule.

Compared with the case in which all fast traders have deterministically different speeds, the latency floor rule induces fast traders to choose different order sizes. Under the latency floor rule, the fast trader who ends up being the first in the queue makes less profit because her trade size is too small ($\frac{2}{N+3} \leq \frac{1}{2}$). This leaves more profits to the second and third fast traders. The fast trader who ends up being the last in the queue, however, makes less profits because her trade size is too large $\left(\frac{2}{N+3} \geq \frac{1}{2^N}\right)$. Hence, the impact of the latency floor rule on the aggregate profits of fast traders depends on the length of the randomizing interval.[28]

When the randomizing interval is short ($N \leq 3$), the speed friction γ can even be higher than the Stackelberg case. Although the first fast trader's expected profit is lower, the second and third fastest traders make more profits. As a result the total profit for the top three fast traders are higher. Intuitively, when the randomizing interval is short, the latency floor effectively turns a race to be the first into a race to be the top three.

When the randomizing interval is long, the number of fast traders in the same interval N increases. In the limit when $N \to \infty$, the total fast trading volume

[28] Under the latency floor rule, when fast traders collude, each fast trader chooses an order size of $\hat{y}/(N+1)$. See Appendix A.7.

$Nu = \frac{2N}{N+3}\hat{y} \to 2\hat{y}$. Fast traders who end up in the first half of the queue make profits and fast traders in the second half of the queue suffer losses. Their profits and losses almost cancel and speed friction γ converges to zero. The rate of convergence, however, is much slower than under the uniform price periodic auction (Cournot), as illustrated in the panel (b) of Figure 1.11.

In short, the well intentioned latency floor could worsen the market quality when the randomizing interval is short. Even if the interval is long enough, market quality under the latency floor is still worse than a market with uniform price batch auctions.

Minimum order life time and order cancellation fee

Some high frequency traders cancel an excessively high fraction of the limit orders they submit. For example, Gai et al. (2012) find that order cancellation/execution ratio is around 30:1 on the NASDAQ exchange in two weeks of year 2010. Although the motive of such behavior is not well understood,[29] it constitutes a cost to exchanges and raises concerns of manipulative strategies like "quote stuffing". Exchanges and regulators have been discussing and experimenting measures, such as a minimum order life and order cancellation fees, to curb the high message volume.

In this paper, front-running HFTs are assumed to trade with market orders. To fully understand the effect of these rules, we need a more complete model of both

[29] Baruch and Glosten (2013) show that liquidity providers submit flickering quotes to play a mixed strategy. Hasbrouck and Saar (2009) provide evidence suggesting that short-lived quotes are used to search for latent liquidity.

aggressive and passive HFTs, which is outside of the scope of this paper. I provide some analysis based on the likely impact on the aggressive front-running HFTs.

First, between the two measures, a minimum order life time has the additional drawback of creating more room for the front-running HFTs. It makes the temporary price impact factor λ^T less flexible for longer durations. Second, reduced message volume might improve the quality ρ of front-running fast traders' information. With fewer updates to the limit order book, front-runners potentially have less noise to filter out and large trades become easier to detect. Third, liquidity providers would need to demand high compensation for posting limit orders due to the higher direct fees or the higher indirect speed tax levied by front-running HFTs. The extra cost is shared among all market participants; traders whose demands are most inelastic pay a higher fraction of the cost.

Hence, we could observe a less liquid market following the implementation of such a rule as found by Malinova, Park and Riordan (2012). The effect could be partially due to the heightened impact of aggressive front-running HFTs ex post.

1.7 Empirical implications

Empirical studies, for example Baron et al. (2012) and Hagströmer and Nordén (2013), have revealed that HFTs use diverse strategies. This paper mainly models the impact of front-running HFTs. It is an empirical challenge to identify front-runners. Some HFTs initiate most of their trades rather than passively absorb others' orders. These HFTs are more likely to be front-runners. The empirical implications are likely to be more relevant for the aggressive HFTs than the passive ones. It should be noted that front-running is not limited to aggressive traders; passive HFTs also effectively front-run when they detect a large incoming order and cancel their limit orders.

1.7.1 Market capacity

Prediction 1.7.1. *In a continuous market, only a few front-running HFTs can survive if they trade on the same order flow information.*

Prediction 1.7.2. *In a continuous market, the entry of a fastest front-running HFT reduces trading volume and profits of all existing HFTs. It might not significantly affect aggregate trading volume, aggregate HFT profits, or market quality if the existing market is almost saturated.*

In a continuous market, practically all traders have different speeds. As is shown in panel (c) of Figure 1.10, the aggregate fast trading profit flattens out quickly along the Stackelberg bound when all fast traders have different speeds. Considering the high costs of staying relatively fast, only a very limited number of HFTs are likely to make net profits.[30]

After the entry of a fastest front-running HFT, all existing HFTs drop one spot in the rankings of relative speeds. They capture fewer shares at worse prices than before. Thus their trading volume and profit decline. Because the market capacity

[30] This prediction is contingent on the assumption that all front-runners have the same order flow information. More HFTs can survive if they specialize in predicting different components of the order flow.

for front-running HFTs is limited, the entry of a faster HFT might simply crowd out slower existing HFTs. The total trading volume and HFT profts are almost unaffected by the entry. This is consistent with the findings of Breckenfelder (2013) and Budish et al. (2013).

1.7.2 Market quality

Prediction 1.7.3. *A new policy does not affect market quality significantly if it does not change front-running HFTs' information quality about the incoming order flow or the heterogeneity of HFTs' relative speeds.*

Gai et al. (2012) find that after the NASDAQ reduces trading latency from microsecond to nanosecond, there is no significant change in market quality. The findings are consistent with my paper because such a reduction does not change HFTs' *relative* speeds, especially in the short run. The reduction could, however, trigger a new round of socially wasteful investment in speed technology among HFTs due to a finer time granularity.

Impact of technology shocks on speed friction γ could be difficult to determine. Roughly speaking, speed friction increases with the heterogeneity of fast traders' relative speeds. For example, introduction of co-location service could reduce speed heterogeneity and thus reduces the speed friction γ. Without co-location service provided by the exchange, HFTs rent rooms nearby the exchange to reduce latencies. With co-location, all HFTs are able to place their computers in the same room. Hence co-location could reduce the speed differences among HFTs and improve market quality. Boehmer et al. (2012) and Frino, Mollica and Webb (2013) find that the introduction of co-location improves market liquidity.

1.7.3 Volatility, momentum, and reversal

Prediction 1.7.4 (Volatility)**.** *Other things equal, more severe HFT front-running ($\rho\gamma \uparrow$) increases short-term volatility and reduces long-term volatility.*

Prediction 1.7.5 (Momentum and reversal)**.** *Other things equal, more severe HFT front-running ($\rho\gamma \uparrow$) causes stronger short-term momentum and reversal.*

Front-running fast traders' impact is higher when they have better information ρ or when the speed friction γ is higher. Here, "more severe HFT front-running" means that $\rho\gamma$ is higher.

The two predictions are both related to the divergence of temporary price impact λ^T and permanent price impact λ^P. The higher short-term price volatility is caused by a higher temporary price impact λ^T. Long-term volatility, for example volatility calculated with daily closing prices, is mainly determined by the permanent price impact λ^P, which is determined by the trading intensity of the informed trader. Faced with a higher price impact λ^T, the informed trader trades slower. Daily volatility might not be affected when the informed trader finishes trading within a day. In the long run, informed trader produces less information, resulting in lower daily price volatilities.

Short run price momentum is caused by fast traders trading on the similar information; short run price reversal is caused by the price adjustment after fast traders have exited. As illustrated in Figure 1.4, fast traders with different speeds trade in the same direction and march along the supply curve of the steeper slope λ^T. After they have exited, price reverse to the efficient level implied by the flatter slope λ^P. Market makers amplify λ^T/λ^P when front-running traders extract more profits. We can characterize short-term price reversal with $-\mathbf{C}\text{ov}[p_1 - p_0, p_{1+} - p_1] = \frac{\sigma_v^2}{4}\frac{2\rho\gamma}{1-\rho\gamma}$. The magnitude increases with front-runners' impact $\rho\gamma$ and fundamental volatility σ_v^2.

1.7.4 Trading volume

Proposition 1.7.1 (Trading volume)**.** *Suppose fast traders (HFT) exit with market orders. Then,*

$$\textit{Volume initiated by non-HFT: } \mathbf{E}\left[|y|\right] = \sqrt{\frac{2}{\pi}}\sqrt{1+\theta^2}\sigma_z$$

$$\textit{Volume initiated by HFT: } \mathbf{E}\left[2|S_J|\right] = \sqrt{\frac{2}{\pi}}\sqrt{1+\theta^2}\sigma_z 2\sqrt{\rho}\cdot\Omega$$

$$\textit{Fraction initiated by HFTs :} \frac{\mathbf{E}\left[2|S_J|\right]}{\mathbf{E}\left[|y|+2|S_J|\right]} = \frac{2\sqrt{\rho}\cdot\Omega}{1+2\sqrt{\rho}\cdot\Omega}$$

where $\Omega = 1 - \prod_{i=0}^{J}(1-\alpha_i)$.

Proof. See Appendix A.6. □

Prediction 1.7.6. *When there is more noise trading ($\sigma_z \uparrow$), the market becomes more liquid ($\lambda^T \downarrow$), the informed trades more ($\beta \uparrow$), and HFTs initiate higher trading volume ($\mathbf{E}[|S_J|] \uparrow$).*

The prediction is based on Theorem 1.3.1 and Proposition 1.7.1. The observed positive correlation between HFT volume and market liquidity could be induced by time varying levels of noise trading.

Prediction 1.7.7. *Keeping volatility of noise trading σ_z constant, normal-speed traders initiate less volume when there is more front-running high frequency trading ($\rho\gamma \uparrow$).*

A higher total trading volume $\mathbf{E}[|y| + 2|S_J|]$ does not mean that the market is more liquid. Volume could be generated by intermediaries including HFTs. Other things equal, volume initiated by long-term buy-side traders (proxy of $\mathbf{E}[|y|]$), however, is a good indicator of market liquidity.

Buy side liquidity demanders can monitor the costs of their own trades and choose the trading venue in response to their trading costs. The increasing popularity of dark pools among institutional traders suggests that trading on exchanges has become expensive relative to trading in dark pools. Recently, Tong (2013) uses a dataset on institutional trades and finds that HFTs increases the execution shortfalls of traditional institutional traders.

Prediction 1.7.8. *Keep the number of front-running high frequency traders and their relative speeds fixed. The fraction of volume initiated by these high frequency traders is determined by their information quality ρ.*[31]

More specifically, if there are N front-runners. The fraction of volume initiated by HFTs $\frac{\mathbf{E}[2|S_J|]}{\mathbf{E}[|y|+2|S_J|]}$ is $\frac{2\sqrt{\rho}\cdot N/(N+1)}{1+2\sqrt{\rho}\cdot N/(N+1)}$ for the Cournot case and $\frac{2\sqrt{\rho}\left(1-2^{-N}\right)}{1+2\sqrt{\rho}(1-2^{-N})}$ for the Stackelberg case. Both ratios approximate $\frac{2\sqrt{\rho}}{1+2\sqrt{\rho}}$ when N is reasonably large, especially for the Stackelberg case.[32] Empirically, we can use the ratio to back out

[31] If HFTs are more risk averse than other investors, the ratio would be lower when HFTs' risk exposure is higher. See Appendix A.1.4.

[32] For example, if 5 HFTs having different speeds are in the market, the entry of a new HFT

fast traders' information quality ρ in cross-section and in time series. For example, Brogaard et al. (2013a) find that on days when the fastest co-located traders initiate a larger fraction of the volume, market is less liquid. This finding is consistent with my model prediction that front-running HFTs initiate a larger fraction of the volume when they predict the order flow more accurately and this renders the market less liquid.

Even when HFTs only have a very noisy signal about the order flow, they could generate a high fraction of the trading volume. For example, let's assume 5 aggressive HFTs initiate 30% of the volume. Then, fast traders' information quality is $\rho = \left(\frac{1}{2}\frac{0.3}{1-0.3}\frac{1}{1-2^{-5}}\right)^2 \approx 0.05$. It implies that the probability for fast traders to trade in the right direction is 58%, only slightly higher than 50%.

1.7.5 Profits and inventory management

Prediction 1.7.9. *Front-running HFTs' profits increase with fundamental uncertainty σ_v, noise trading σ_z, HFTs' information quality ρ, and speed friction γ.*

Prediction 1.7.10. *Aggregate front-running HFTs' profits represent a higher fraction of noise traders' implementation shortfall when front-running is more severe ($\rho\gamma \uparrow$).*

Proposition 1.5.3 shows that fast traders' profit $\mathbf{E}[\pi^F] = \frac{\sigma_v\sigma_z}{2}\left(\frac{1}{\theta} - \theta\right)$ where $\theta = \sqrt{\frac{1-\rho\gamma}{1+\rho\gamma}}$. If we use implementation shortfall to proxy the trading loss of noise traders $-\mathbf{E}[\pi^N] = \frac{\sigma_v\sigma_z}{2}\frac{1}{\theta}$, then $\mathbf{E}[\pi^F]/\mathbf{E}[-\pi^N] = 1-\theta^2$. The effect of $\sigma_v\sigma_z$ is canceled

with a different speed would barely change the ratio because 2^{-5} and 2^{-6} are too close.

once we take the ratio. Empirically, we can use the ratio to measure the overall impact of all front-running HFTs on market quality θ.

Prediction 1.7.11. *Faster front-running HFTs tend to trade more shares, have higher inventory levels, hold inventory for longer time periods, and make larger profit per share.*

The fastest HFTs do not necessarily trade more frequently. When trading on similar signals, a higher speed allows an HFT to acquire more shares at better prices. Larger inventory and longer holding horizons could result from better market timings. Consistent with this prediction, Brogaard et al. (2013a) recently find that after a subgroup of HFTs upgrade to the fastest co-location service, these HFTs hold larger inventory for longer durations.

1.8 Conclusion

I analyze the implications of traders' speed differences in a strategic model of asymmetrically informed traders. In this model, front-running HFTs use their speed advantage to extract rents from normal-speed traders and the extracted rents are allocated among HFTs according their relative speeds.

A higher trading frequency makes it less likely for HFTs to compete on quantity and more likely to compete on speed. Unlike price or quantity competitions, speed competition may not benefit normal-speed traders. Even when infinitely many front-running HFTs compete against one another on speed, their negative market impact in general does not vanish because the fastest HFTs still make positive profits. At the very least, higher trading frequency weakens the effectiveness of pro-competition policies designed to mitigate the negative impact of HFTs.

We already limit competition on price and quantity by imposing rules of the minimum price variation (one tick) and the minimum quantity variation (one share). Findings of this paper suggest that imposing an upper limit on trading frequency is equally, if not more, justifiable because such a limit would deter the socially wasteful competition on speed.

Inevitably, I have made simplifying assumptions. Nevertheless, the key results seem to be robust in more general settings. First, speeds are exogenous in the paper. When speed is costly, a higher granularity of time would still encourage investment in speed because being relatively faster generates high payoffs.

Second, I assume that each fast trader can only trade twice in each trading round. As a result, only relative speed matters. If a fast trader can trade multiple times before the next fast trader arrives, both absolute and relative speed differences would affect their profits. The faster one would have more trading opportunities in addition to the advantage of trading earlier. In the limit when the fastest trader can trade infinitely many times before the second fastest trader, only the fastest makes positive profits. Nevertheless, lowering the trading frequency can still reduce the aggregate front-running profits because it reduces the number of trading opportunities of the fastest front-runners.

Finally, this paper focuses on the front-running HFTs. Many other HFT strategies can have benign or beneficial effects on the market. It seems, however, that lowering the frequency of trading would also help many liquidity enhancing HFTs because they could spend less resource on protecting themselves against front-runners.

A lower trading frequency, of course, has its drawbacks. For example, the noise traders would have to wait longer before their liquidity demands are met. In addition, in the next Chapter, I show that a higher trading frequency enables the patient informed traders to lower their price impact and extract more profits from their private information. The results of this paper suggest that a higher trading frequency is not always socially beneficial. Policy makers need to carefully weigh the costs and benefits of imposing no limit on the frequency of trading.

Chapter 2: The Fast and the Faster: Trading Frequency and Market Quality

2.1 Introduction

Financial markets have undergone considerable changes in recent years. In particular, many securities are being traded almost continuously. In this paper, I investigate the impact of higher trading frequencies on the quality of financial markets.

First, I develop a dynamic model in which traders with different trading motives interact. All traders act at the same frequency. Using a variant of the dynamic model of Kyle (1985), I show that when trading frequency is higher, prices become more informative. The informed trader produces more private information and trades more aggressively. As a result, the market incorporates more information faster. I also find that when the trading frequency is higher, market is less liquid in the beginning but the illiquidity decays more quickly because the information asymmetry decays faster. Overall, a higher trading frequency benefits the informed trader and leads noise traders to lose more.

Second, I extend the benchmark model to include high frequency traders

(HFTs) who anticipate other traders' order flow and quickly trade to profit from the order flow information. I embed HFT front-runners as modeled in the first chapter of this dissertation. When the trading frequency is finite, I show that a higher HFT intensity reduces price efficiency and market liquidity. The informed trader trades more slowly and market prices incorporate his private information less quickly. Such negative impact of HFT on price efficiency decreases with the expected lifetime of the private information. An informed trader with long-lived private information is less affected by HFTs.

Third, I show that when the trading frequency goes to infinity, the impact of HFTs on the informed trader vanishes. The informed trader trades in the same way and makes the same amount of profits for any HFT intensity. Consequently, price efficiency is unaffected by HFTs in the continuous time limit. HFT front-runners make all their profits from noise traders who demand immediacy.

2.2 Trading game

In this section, trading is modeled as a dynamic game in which traders with different trading motives interact. In the benchmark model, everyone operates at the same frequency. The time interval between two adjacent trading opportunities is Δ_t. As Δ_t decreases, everyone acts at a higher and *equal* frequency. In the limit, as $\Delta_t \to 0$, trading can be approximated by a continuous time game. In the extended model, I add HFTs who act at a higher frequency than other traders.

2.2.1 Model setup

2.2.1.1 Assets and agents

Two assets are traded. The risk free asset has a fixed value of 1. The risky asset's fundamental value v is normally distributed as

$$v \sim \mathcal{N}(0, \sigma_v^2) \tag{2.2.1}$$

The distribution and volatility σ_v of v are common knowledge to all agents.

Three types of risk neutral agents anonymously trade in the market: (1) One informed trader has monopolistic access to a costly technology that can generate private information about v. The informed trader's time discount factor equals the risk free rate 0.[1] (2) A continuum of noise traders trade for liquidity reasons that are unrelated to fundamentals of the risky asset. They demand immediate execution.[2] (3) A continuum of perfectly competitive risk neutral market makers set prices and absorb the net order flow imbalance coming from other traders.[3]

---

[1] The model can be extended such that the informed trader's time discount factor is greater than the risk free rate. It represents the impatience of the monopolistic informed trader and may be due to his higher cost of capital or margin requirement. The informed trader would then choose to produce information that decays at the same rate as his discount factor.

[2] Noise traders demand immediacy at the cost of expected trading losses. The losses may be offset by gains outside of the trading game.

[3] Risk sharing motives are not modeled explicitly. We can consider the risk neutrality of market makers as a good approximation when many market makers compete to absorb a very small amount of risk transferred in each transaction.

2.2.1.2 Timeline

The game starts at time 0 when the monopolistic informed trader chooses an *observable* level of accuracy of his signal and produces private information about the risky asset's value v.

Starting from time Δ_t, periodic batch auctions are held and the time interval between any two adjacent auctions is Δ_t. The first trade occurs at Δ_t. The fair value v is revealed to the public at a random time T. At time T, the game ends. All traders liquidate their positions at the fair price v.

Right before time $n\Delta_t$, the inventory of the informed trader is x_{n-1}, the aggregate inventory of all noise traders is z_{n-1}, and the cumulative shares sold by the market makers is y_{n-1}. The initial inventories of all traders are zero ($x_0 = y_0 = z_0 = 0$).

At time $n\Delta_t$, if news has not been announced ($T > n\Delta_t$), the informed traders submit an order of $\Delta x_n = x_n - x_{n-1}$ shares, the noise trader submit orders of $\Delta z_n = z_n - z_{n-1}$ shares, and the market makers set a price p_n and sell $\Delta y_n = y_n - y_{n-1}$ shares. The market clearing condition requires that

$$\Delta y_n = \Delta x_n + \Delta z_n. \tag{2.2.2}$$

Noise traders' order flow follows the process

$$\Delta z_n = z_n - z_{n-1} = \sigma_z (B_{n\Delta_t} - B_{(n-1)\Delta_t}) \tag{2.2.3}$$

where B_t is a standard Brownian motion independent from the risky asset's fundamental value v. The volatility σ_z is common knowledge to all agents.

The announcement time T is exponentially distributed with probability density function as follows

$$f_T(t) = \eta e^{-\eta t}, \quad t \geq 0 \tag{2.2.4}$$

At any trading time $n\Delta_t$ before the public announcement T, it is possible that the risky asset's value v is announced in the interval $T \in (n\Delta_t, (n+1)\Delta_t]$ such that the game ends before time $(n+1)\Delta_t$ and all traders liquidate their positions at the price of $p_T = v$. The probability of this event is

$$\Pr\{T < (n+1)\Delta_t | T > n\Delta_t\} = 1 - e^{-\eta\Delta_t} \tag{2.2.5}$$

The expected lifetime of the informed trader's signal is $\mathbf{E}[T] = 1/\eta$. A longer lived information has a smaller information arrival rate η. For a given η, the likelihood of news announcement occuring in the next interval $\eta\Delta_t$ is lower when trading frequency $1/\Delta_t$ is higher

This paper investigates the impact of trading frequency, lifetime of private information, and front-running HFTs on the quality of the financial market.

2.2.1.3 Information

The informed trader produces one noisy signal I_v about the fundamental value v only at time 0. The precision of this signal is observable but the realization I_v is private information to the informed trader. No further information is produced after time 0. Starting from time Δ_t, the informed trader trades on this signal until time T when the fair value v is revealed to the public. It is common knowledge that the signal I_v satisfies

$$I_v = v + e, \quad e \sim \mathcal{N}(0, \sigma_e^2) \tag{2.2.6}$$

The Gaussian noise term e is independent from v. Quality of the informed trader's signal I_v is captured by ρ_i, the squared correlation coefficient of S and v.

$$\rho_i = \mathbf{C}\mathrm{orr}^2(I_v, v) = \frac{\sigma_v^2}{\sigma_v^2 + \sigma_e^2} \tag{2.2.7}$$

ρ_i is in bounded in $[0, 1]$. A more informative signal I_v has a larger ρ_i. When $\rho_i = 0$, the signal is entirely noise; when $\rho_i = 1$, the signal equals v with probability 1. From the projection theorem of normally distributed random variables,

$$\mathbf{E}[v|I_v] = \frac{\mathbf{C}\mathrm{ov}(v, I_v)}{\mathbf{V}\mathrm{ar}(I_v)} I_v = \rho_i I_v \tag{2.2.8}$$

$$\mathbf{V}\mathrm{ar}[v|I_v] = \mathbf{V}\mathrm{ar}[v] - \frac{\mathbf{C}\mathrm{ov}^2(v, I_v)}{\mathbf{V}\mathrm{ar}(I_v)} = (1 - \rho_i)\sigma_v^2 \tag{2.2.9}$$

It is assumed that information production is observable and thus the market makers do not need to estimate the informed trader's information quality ρ_i. Knowing ρ_i does not eliminate the adverse selection problem because the signal I_v is private to the informed trader and trading is anonymous.[4]

Price history $\{p_n\}$ is public information. The informed trader observes the signal I_v and his own inventory history $\{x_n\}$. In equilibrium, due to the monotonicity of the pricing rule, the informed trader can also perfectly infer the history of the noise

[4] The informed trader possibly has an incentive to hide his information production effort so that the market makers may underestimate ρ_i. There may be a "forecast the forecasts of others" problem because the informed trader then has to estimate the market makers' estimate of ρ_i. Extending the model to allow for hidden information production can be explored in the future.

trades $\{z_n\}$. Denote the expectation taken over the informed trader's information set by

$$\mathbf{E}_n^I[\cdot] = \mathbf{E}\left[\cdot \left| I_v, x_n, \mathcal{F}\left\{\{p_m, x_m\}_{0\leq m<n}\right\}\right.\right] \tag{2.2.10}$$

where $\mathcal{F}\{\cdot\}$ is the σ-algebra generator. Notice that at time $n\Delta_t$, the informed trader does not observe price p_n and liquidity trades z_n.

The market makers observe price history $\{p_n\}$ and the order flow imbalance process $\{y_n = x_n + z_n\}$. Because trading is anonymous, market makers cannot differentiate informed trading x_n and liquidity demand z_n. Denote the expectation taken over the market makers' information by

$$\mathbf{E}_n^M[\cdot] = \mathbf{E}\left[\cdot \left| y_n, \mathcal{F}\left\{\{p_m, y_m\}_{0\leq m<n}\right\}\right.\right] \tag{2.2.11}$$

where $\mathcal{F}\{\cdot\}$ is the σ-algebra generator. The market makers set the price p_n conditional on their information set. The degree of information asymmetry between the informed trader and market makers can be measured in

$$\Sigma_n = \mathbf{V}\mathrm{ar}_n^M[\hat{v}] \tag{2.2.12}$$

2.2.2 Trading with equal frequency

From the informed trader's perspective, this is a two stage game. (1) In the information production stage, he chooses the quality of information ρ_i and produce a signal I_v with quality ρ_i. (2) In the trading stage, he choose a trading strategy to maximize his expected trading profit conditioning on the signal I_v. The two stages are separable because once trading starts the informed trader receives no additional information.

In this section I solve the trading game after the informed trader has chosen the quality ρ_i of the signal I_v. As introduced before, market makers observes the quality of the informed trader's signal ρ_i but not the signal I_v itself. The informed trader's estimate of the liquidation value v is

$$\hat{v} = \mathbf{E}[v|I_v] = \rho_i I_v \tag{2.2.13}$$

2.2.2.1 Continuous time trading

To provide some intuition, let's first examine the limiting case where interval between trades Δ_t converges to 0 and assets are traded in continuous time. Notations are adapted accordingly. The expectation taken with respect to the insider's information and market makers' information are

$$\mathbf{E}_t^I[\cdot] = \mathbf{E}\left[\cdot \left| I_v, x_t, \mathcal{F}\left\{\{x_s, p_s\}_{s<t}\right\}\right.\right] \tag{2.2.14}$$

$$\mathbf{E}_t^M[\cdot] = \mathbf{E}\left[\cdot \left| y_t, \mathcal{F}\left\{\{y_s, p_s\}_{s<t}\right\}\right.\right] \tag{2.2.15}$$

and the informed trader's perceived pricing error of the market price p_{t^-} is

$$D_t = \hat{v} - p_{t^-}. \tag{2.2.16}$$

Magnitude of D_t reflects the advantage of the informed over the public information. And the degree of information asymmetry is represented by

$$\Sigma_t = \mathbf{V}\mathrm{ar}_t^M[\hat{v}] \tag{2.2.17}$$

At time t, the informed trades dx_t shares, the noise traders dz_t shares, and the market makers clear the market at price p_t after observing the aggregate order

$dx_t + dz_t$. In equilibrium, market price is semi-strong efficiency and the informed maximize expected current and aggregate future profits. I assume that the informed trader's value function exists and it is defined as follows:

$$V(t, D_t) := \max_{dx_t} \left\{ \mathbf{E}_t^I \left[\int_t^T (v - p_t) dx_t \right] \right\} \tag{2.2.18}$$

In this paper, I only consider the linear equilibrium as defined below.

Definition 2.2.1 (Continuous time trading game equilibrium)**.** The informed trader and market makers follow strategies characterized by β_t and λ_t:[5]

$$\text{Informed's trade size: } dx_t = \beta_t dt \tag{2.2.19}$$

$$\text{Price updating rule: } dp_t = \lambda_t (dx_t + dz_t) \tag{2.2.20}$$

with the constraint[6] that

$$\mathbf{E}_0^I \left[\int_0^T (v - p_t) dx_t \right] < \infty. \tag{2.2.21}$$

In equilibrium, prices are semi-strong form efficient and the informed trader maximizes

[5] In the conjectured equilibrium, informed order flow dx_t has no diffusion term. Back (1992) proves it's not optimal for the insider order flow to have a diffusion term even when the insider can effectively observe the noise trading in the continuous trading limit. In Foucault et al. (2012) high frequency news traders' (HFNT) order flow does have a Brownian motion term. It comes from a piece of information that is going to expired at the next instant and thus HFNT trades very aggressively despite the high price impact. The trading game is essentially a repeated static game of the Kyle (1985) model. The intuition is similar to Chau and Vayanos (2008) and Li (2013).

[6] The constraint rules out the doubling strategies as in Back and Baruch (2004).

expected aggregate trading profits from time t on

$$p_t = \mathbf{E}_t^M[v] = \mathbf{E}_t^M[\hat{v}] \tag{2.2.22}$$

$$\mathbf{E}_t^I\left[\int_t^T (v - p_t)\beta_t dt\right] \geq \mathbf{E}_t^I\left[\int_t^T (v - p_t)\beta_t' dt\right] \quad \forall \beta_t' \tag{2.2.23}$$

Theorem 2.2.1 (Continuous time trading game equilibrium)**.** *There exists a linear equilibrium where*

$$\beta_t = 2\eta \frac{D_t}{\lambda_t} = \sqrt{\frac{2\eta}{\rho_i}} \frac{\sigma_z}{\sigma_v} e^{\eta t} D_t \tag{2.2.24}$$

$$\lambda_t = \lambda_0 e^{-\eta t} = \sqrt{2\eta\rho_i} \frac{\sigma_v}{\sigma_z} e^{-\eta t} \tag{2.2.25}$$

The information asymmetry

$$\Sigma_t = \mathbf{V}\mathrm{ar}_t^M[\hat{v}] = \rho_i \sigma_v^2 e^{-2\eta t}, \quad t \geq 0 \tag{2.2.26}$$

and the informed trader's value function is

$$V(t, D_t) = \frac{D_t^2}{2\lambda_t} + \frac{\lambda_t}{4\eta}\sigma_z^2 = \left(\frac{D_t^2}{\rho_i \sigma_v^2} e^{\eta t} + e^{-\eta t}\right) \frac{1}{2} \sqrt{\frac{\rho_i}{2\eta}} \sigma_v \sigma_z, \quad t \geq 0 \tag{2.2.27}$$

Proof. See Appendix B.1. □

As pointed out by Back (1992), there are in general multiple optima for the informed trader in the continuous time limit. When price impact λ_t is set as above, the informed trader can achieve maximum expected profits with multiple trading strategies. For example, an alternative expression of the value function is

$$V(t, \hat{v} - p_t) = \frac{(\mathbf{E}[v|I_v] - p_t)^2}{2\lambda_t} + \frac{1}{2}\mathbf{E}\left[\int_{t^+}^T dp_s dz_s \middle| I_v\right] \tag{2.2.28}$$

The first term is the profit of pushing price to equal $\hat{v}$ at time t; the second term is the noise traders' expected loss from time t^+ to the ending time T when the informed trader follows the given equilibrium strategy from t^+ on.

So informed trader's trading intensity β_t cannot be uniquely determined. The above equilibrium is chosen such that the market makers' belief updating is correct at any time instant t not only over the interval $[0, T]$. This equilibrium refinement implies that the continuous equilibrium is an approximation of the discrete time reality where Δ_t is small yet still strictly positive.

Corollary 2.2.2. *[Equilibrium properties of β_t, λ_t, and Σ_t] Price impact λ_t and decay exponentially at the rate $-\eta$, i.e.,*

$$\frac{d\lambda_t}{\lambda_t} = -\eta dt, \tag{2.2.29}$$

The expected pricing error $\mathbf{E}_t^M[D_t] = \mathbf{E}_t^M[\hat{v} - p_t]$ and the information asymmetry measured in variance $\Sigma_t = \mathbf{V}\mathrm{ar}_t^M[\hat{v}]$ both decay exponentially at the rate of -2η, i.e.,

$$\frac{\mathbf{E}_t^M[dD_{t^+}]}{D_t} = \frac{d\Sigma_t}{\Sigma_t} = -2\eta dt, \tag{2.2.30}$$

Informed trader's trade size $\beta_t dt$ is proportional to pricing error normalized by price impact D_t/λ_t, i.e.,

$$dx_t = \frac{\beta_t dt}{D_t/\lambda_t} = 2\eta dt \tag{2.2.31}$$

Initially, the price impact λ_0 and information asymmetry $\Sigma_0 = \mathbf{V}\mathrm{ar}_0^M[\hat{v}]$ are

$$\lambda_0 = \sqrt{2\eta\rho_i}\frac{\sigma_v}{\sigma_z}, \quad \Sigma_0 = \rho_i\sigma_v^2 \tag{2.2.32}$$

Proof. Trivial. □

Key feature of the equilibrium is that the price impact factor λ_t decays exponentially, unlike in the continuous time model of Kyle (1985) where λ_t is constant. In this model, the informed trader is faced with the risk that their private information may expire at the next instant with probability $e^{-\eta dt}$. Hence, the informed trader trades more when he still can. He does not trade away all his information in one trade because the price impact is lower at the next time instant.

Since the noise traders do not change their trading intensity σ_z, the fraction of volume initiated by the informed trader is higher in early trading periods. Hence, market makers learn more and the information asymmetry reduces by more $(2\eta\Sigma_t)$ in early trading periods when Σ_t is higher.

2.2.2.2 Market quality of the continous time equilibrium

Market liquidity and information efficiency are two important measures of the quality of a financial market. In this paper, market liquidity is measured by the price impact factor λ_t and information efficiency is measured by the level of residual information asymmetry $\Sigma_t = \mathbf{V}\mathrm{ar}_t^M[\hat{v}]$. Market is more liquid when price impact λ_t is lower and prices are more informationally efficient if Σ_t is lower.

Several factors given in the following proposition only affect the initial market quality, captured by λ_0 and Σ_0, when trading starts at time 0.

Proposition 2.2.3. *[Effect of σ_z,σ_v, and ρ_i] Initial price impact λ_0 is higher when there is more total fundamental uncertainty ($\sigma_v \uparrow$), the informed trader's information is more accurate ($\rho_i \uparrow$), or there is less noise trading ($\sigma_z \downarrow$). Initial information asymmetry is higher ($\Sigma_0 \uparrow$) when there is more total fundamental uncertainty ($\sigma_v \uparrow$) or the informed trader's information is more accurate ($\rho_i \uparrow$).*

Proof. Follows from Corollary 2.2.2. □

In the continuous time equilibrium, the only parameter that affects the *dynamics* of the equilibrium is η, the arrival rate of the public information. The public news announcement follows a Poisson arrival process. Due to the memoryless property of Poisson process, at each instant, the game is essentially the same after proper normalization. For the informed trader, the only relevant state variable is D_t/λ_t, the number of shares one can trade to drive the pricing error D_t to zero. In equilibrium, the informed trader trades a fraction $2\eta dt$ of D_t/λ_t at every instant. As a result, the pricing error D_t on average is reduced by a fixed fraction $2\eta dt$. The price impact factor λ_t is also reduced by a fixed fraction ηdt.

Proposition 2.2.4 (**Effect of lifetime of private information** $1/\eta$). *When the public news arrives faster ($\eta \uparrow$) or equivalently the expected lifetime of the private information is shorter ($1/\eta \downarrow$), the informed trader trades away a higher fraction of the residual information ($\frac{\beta_t dt}{D_t/\lambda_t} \uparrow$). Consequently,*

1. *the initial market liquidity is worse ($\lambda_0 \uparrow$) but the speed of liquidity improvement is higher ($-d\lambda_t/\lambda_t \uparrow$);*

2. *the initial information asymmetry (Σ_0) is unchanged and the speed of information revelation through prices is faster ($-d\Sigma_t/\Sigma_t \uparrow$).*

Proof. Follows from Corollary 2.2.2. □

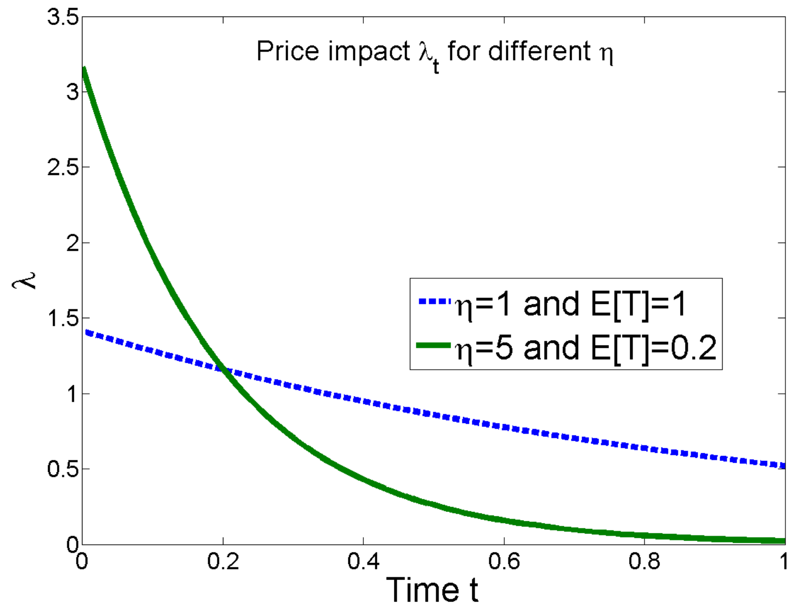

Figure 2.1: Dynamics of price impact λ_t and public information arrival rate η

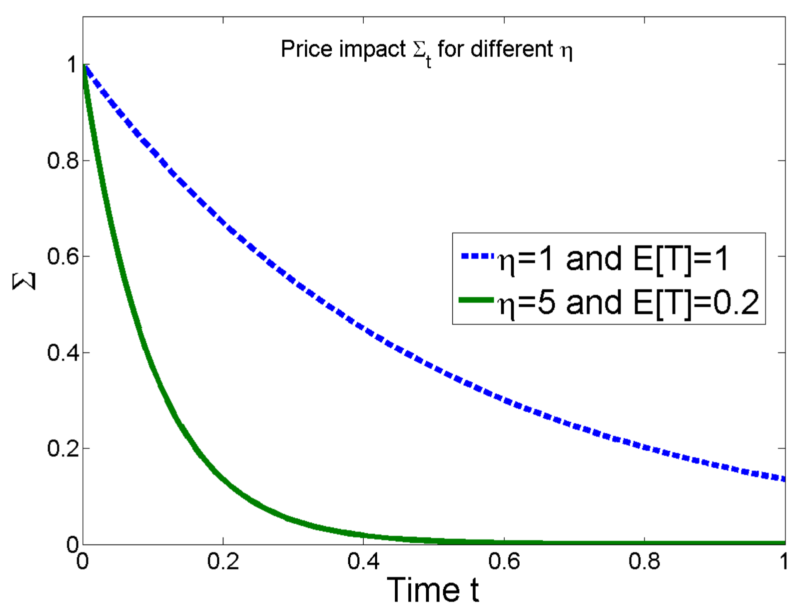

Figure 2.2: Dynamics of information asymmetry Σ_t and public information arrival rate η

Arrival speed of the public news (η), or equivalently the expected lifetime of the informed trader's private information ($1/\eta$), has ambiguous effect on market liquidity as illustrated in Figure 2.1. An important observation is that η affects both the initial price impact factor λ_0 and the speed of the decay of price impact $d\lambda_t/\lambda_t$. When the public information arrives faster and the expected lifetime of the private information is shorter ($\eta \uparrow$), the initial price impact λ_0 is higher and its speed of decay ηdt is also higher. Intuitively, when private information expires faster ($\eta \uparrow$), the informed trader chooses to trade more aggressively at the beginning. Hence, market is less liquid initially ($\lambda_0 \uparrow$) due to higher intensity of informed trader.

On the other hand, a faster arrival rate of public news ($\eta \uparrow$) improves information efficiency of market prices as illustrated in Figure 2.2. The variance $\Sigma_n = \mathbf{V}\mathrm{ar}_t^M[\hat{v}]$ measures the information asymmetry between the informed trader and the market maker. Its initial value Σ_0 is determined by the amount of information the informed trader choose to produce and is unaffected η. Its rate of decay $-\Sigma_t/\Sigma_t$, however, is higher when the informed trades more aggressively ($\eta dt \uparrow$). Hence, the residual information asymmetry Σ_t is lower at any time t when public information arrives faster ($\eta \uparrow$).

2.2.2.3 Discrete time trading

The continuous time equilibrium illustrates the impact of expected lifetime of a private information ($1/\eta$) when the trading frequency is infinitely high. This section investigates the discrete time trading game where trading frequency $1/\Delta_t$ is finite.

Right before time $n\Delta_t$, the informed trader believes that the error of the last traded price is

$$D_n = \hat{v} - p_{n-1}. \tag{2.2.33}$$

Since the news announcement time T is exponentially distributed, the conditional probability of T occurring in the next Δ_t interval is time invariant and always equals $1 - \exp(-\eta\Delta_t)$.

In this paper, I consider linear equilibrium of the following form

$$\Delta x_n = x_n - x_{n-1} = \frac{D_n}{\lambda_n} \cdot \beta\Delta_t, \quad \beta\Delta_t \in [0, 1] \tag{2.2.34}$$

$$\Delta p_n = p_n - p_{n-1} = \lambda_n(\Delta x_n + \Delta z_n) = \lambda_n \Delta y_n \tag{2.2.35}$$

In the continuous time equilibrium, $\frac{dx_t}{D_t/\lambda_t} = \beta dt$ is time-invariant. Hence, I conjecture that in the discrete time equilibrium $\frac{\Delta x_n}{D_n/\lambda_n}$ is also time invariant. $\beta\Delta_t$ is bounded in $[0, 1]$ so that the informed trader does not expected to incur losses in each trade.[7] Given the linearity of the informed trader's strategy and the normality of the random variables, the market makers' pricing rule is also linear.

Given such a conjectured equilibrium, at the beginning of each interval, the structure of the game is essentially the same after proper normalization. The only difference to the informed trader is $\frac{D_n}{\lambda_n}$, which measures the expected number of shares he can submit to push the execution price to $\hat{v}$.

The informed trader's expected terminal profit from trading from time $n\Delta_t$ to

[7] I do not rule out the possibility of other forms of equilibria.

the news announcement time T is

$$\Pi(\beta, n, D_n) = \mathbf{E}_n^I \left[\sum_{i=n}^{\lfloor T/\Delta_t \rfloor} (v - p_n) \cdot \Delta x_i \right] \tag{2.2.36}$$

where $\lfloor \frac{T}{\Delta_t} \rfloor$ equals the smallest integer no less than $\frac{T}{\Delta_t}$.

Definition 2.2.2. A linear equilibrium of the trading game is defined as a pair of β and $\{\lambda_n\}$ such that the two following conditions hold.

1. Semi-strong market efficiency. For a given β, $\{\lambda_n\}$ are chosen such that

$$p_n = \mathbf{E}_n^M[v] = \begin{cases} 0 & n = 0 \\ p_{n-1} + \lambda_n(\Delta x_n + \Delta z_n) & n > 0 \end{cases} \tag{2.2.37}$$

2. Profit maximization. For given $\{\lambda_n\}$, β is chosen such that

$$\mathbf{E}_n^I \left[\Pi(\beta, n, D_n)\right] \geq \mathbf{E}_n^I \left[\Pi(\beta', n, D_n)\right] \quad \forall \beta' \tag{2.2.38}$$

The semi-strong market efficiency condition is natural given the assumption of a continuum of perfectly competitive market makers.[8] It implies that both the returns and the order flows are unpredictable based only on the public information.

$$\mathbf{E}_{n-1}^M[\Delta p_n] = \mathbf{E}_{n-1}^M[p_n] - p_{n-1} = \mathbf{E}_{n-1}^M[\mathbf{E}_n^M[v]] - \mathbf{E}_{n-1}^M[v] = 0 \tag{2.2.39}$$

$$\mathbf{E}_{n-1}^M[\Delta x_n + \Delta z_n] = \mathbf{E}_{n-1}^M \left[\beta \frac{D_n}{\lambda_n} \Delta_t\right] = \frac{\beta}{\lambda_n} \Delta_t \, \mathbf{E}_{n-1}^M[\hat{v} - p_{n-1}] = 0 \tag{2.2.40}$$

Hence, $\{p_n\}$ and $\{y_n\}$ are martingales adapted to the information set of the market makers. The profit maximization condition is forward looking. It ensures that trade

[8] Bernhardt and Hughson (1997), Guo and Kyle (2009), and Liu and Wang (2010) relax the assumption of perfectly competitive risk neutral Bertrand market makers in different ways.

at any instant optimally balance the instantaneous expected profit and the impact to all subsequent periods.

Assume that there exists an optimal strategy $\{x_n\}$ for the informed trader.

Definition 2.2.3 (Value function of the informed trader)**.** The informed trader trades optimally in the time interval of $[n\Delta_t, T)$ and the expected profit equals

$$V(n, D_n) = \max_{\Delta x_n} \left\{ \mathbf{E}_n^I \left[\sum_{i=n}^{\lfloor T/\Delta_t \rfloor} (v - p_i)\Delta x_i \right] \right\}$$

Theorem 2.2.5 (Continuous time trading game equilibrium)**.** *There exists an equilibrium where the equilibrium trading intensity* $\beta\Delta_t \in [0, 1/2]$ *is the unique solution to the equation*

$$\frac{1 - 2\beta\Delta_t}{(1 - \beta\Delta_t)^{3/2}} = e^{-\eta\Delta_t} \tag{2.2.41}$$

where $\Pr\{T < (n+1)\Delta_t | T \geq n\Delta_t\} = 1 - e^{-\eta\Delta_t} \in [0, 1]$*. The price impact coefficients are*

$$\lambda_0 = \frac{\sigma_v}{\sigma_z}\sqrt{\rho_i}\sqrt{\beta} \tag{2.2.42}$$

$$\lambda_n = \lambda_0 \cdot (1 - \beta\Delta_t)^{n/2} \tag{2.2.43}$$

The informed trader's trade sizes are

$$\Delta x_n = \beta\Delta_t \frac{D_n}{\lambda_n} \tag{2.2.44}$$

and the informed trader's value function at time $n\Delta_t$ *is*

$$V(n, D_n) = (1 - \beta\Delta_t)\frac{D_n^2}{2\lambda_n} + (1 - 2\beta\Delta_t)\frac{\sigma_z^2}{2\beta}\lambda_n \tag{2.2.45}$$

The information asymmetry is

$$\Sigma_0 = \rho_i \sigma_v^2 \tag{2.2.46}$$

$$\Sigma_n = \Sigma_0 \cdot (1 - \beta\Delta_t)^n, \quad n \geq 1. \tag{2.2.47}$$

Proof. See Appendix B.2. □

Theorem 2.2.6. *As trading frequency goes to infinity ($\Delta_t \to 0$), the discrete time equilibrium of Theorem 2.2.5 converges to the continuous time equilibrium of Theorem 2.2.1.*

Proof. See Appendix B.3. □

The discrete time equilibrium is characterized by β, which is determined by the public information arrival rate η and inter-trade interval length Δ_t. Specifically, (1) the decaying speeds of price impact λ_n and information asymmetry Σ_n are determined by $\beta\Delta_t$ and (2) the initial price impact λ_0 depends on β. The following result summarizes the effects of public news arrival rate η and duration between trades Δ_t.

Proposition 2.2.7. *The probability of public news announcement in the next Δ_t interval equals $1 - e^{-\eta\Delta_t}$. Equation (2.2.41) defines a function $g(\cdot)$ implicitly such that*

$$\beta\Delta_t = g(1 - e^{-\eta\Delta_t}), \quad \eta, \Delta_t \geq 0. \tag{2.2.48}$$

Then, the function $g(\cdot)$ is increasing and concave ($g'(\cdot) > 0$ and $g''(\cdot) < 0$) when $\eta > 0$ and $\Delta_t > 0$. In addition, $\frac{\partial(\beta\Delta_t)}{\partial\eta} > 0$, $\frac{\partial(\beta\Delta_t)}{\partial\Delta_t} > 0$, $\frac{\partial\beta}{\partial\eta} > 0$ and $\frac{\partial\beta}{\partial\Delta_t} < 0$.

Proof. See Appendix B.4. □

2.2.2.4 Market quality of the discrete time equilibrium

Similar to Section 2.2.2.2, the quality of the market in a discrete time setting is also captured by the price impact factor λ_n and information asymmetry Σ_n. As in Proposition 2.2.3 of the continuous time game, the levels of the fundamental uncertainty σ_v, noise trading σ_z, and the informed trader's information production ρ_i only affect the initial market liquidity and information efficiency.

Proposition 2.2.8. *[Effect of σ_z,σ_v, and ρ_i] Initial price impact λ_0 is higher when there is more total fundamental uncertainty ($\sigma_v \uparrow$), the informed trader's information is more accurate ($\rho_i \uparrow$), or there is less noise trading ($\sigma_z \downarrow$). Initial information asymmetry is higher ($\Sigma_0 \uparrow$) when there is more total fundamental uncertainty ($\sigma_v \uparrow$) or the informed trader's information is more accurate ($\rho_i \uparrow$).*

Proof. Follows from Theorem 2.2.5. □

The *dynamics* of equilibrium is reflected by the speeds of decay of price impact λ_n and information asymmetry Σ_n. In the continuous time, the arrival speed of public news η determines these speeds. In the discrete time, the trading frequency $1/\Delta_t$ also affects the dynamics of the equilibrium.

In a discrete time game, we need to measure the speed of decay carefully. For example, when we reduce Δ_t, the number of trades per unit of time increases. Hennce the information asymmetry Σ_n decays less *per trade* but decays more *per unit time*. The speed of decay per unit of time is captured by the parameter's half-life.

Definition 2.2.4 (Half life of information asymmetry Σ_n and price impact λ_n)**.** Half life of information asymmetry T^Σ is the solution to $(1-\beta\Delta_t)^{\frac{T^\Sigma}{\Delta_t}} = \frac{1}{2}$ and half life of liquidity improvement T^λ is the solution to $(\sqrt{1-\beta\Delta_t})^{\frac{T^\lambda}{\Delta_t}} = \frac{1}{2}$. Hence,

$$T^\lambda = 2T^\Sigma = \frac{\Delta_t \log 2}{-\log(1-\beta\Delta_t)} \tag{2.2.49}$$

Proposition 2.2.9. *[Effect of η and Δ_t on half-life of information asymmetry Σ_n and illiquidity λ_n] When public information arrives faster ($\eta \uparrow$) or when trading frequency increases ($\frac{1}{\Delta_t}\uparrow$), information asymmetry Σ_n and price impact λ_n decay faster per unit of time and have shorter half-lives. Formally speaking, $\frac{\partial T^\lambda}{\partial \eta} = 2\frac{\partial T^\Sigma}{\partial \eta} < 0$ and $\frac{\partial T^\lambda}{\partial \Delta_t} = 2\frac{\partial T^\Sigma}{\partial \Delta_t} > 0$ for $\beta\Delta_t \in [0, 1/2]$ and $\Delta_t > 0$. In the continuous limit as $\Delta_t \to 0$, $T^\lambda = 2T^\Sigma = \frac{\log 2}{\eta} \approx 0.69\,\mathbf{E}[T]$.*

Proof. * See Appendix B.5. □

Proposition 2.2.10 (Effect of lifetime of private information $1/\eta$ on liquidity and efficiency)**.** *When the lifetime of the private information is shorter ($\frac{1}{\eta}\downarrow$), initial price impact is higher ($\lambda_0\uparrow$), initial information asymmetry (Σ_0) is unchanged, price impact and information asymmetry decay faster per trade* $(1-\beta\Delta_t)\downarrow$ *and per unit of time ($T^\lambda\downarrow$ and $T^\Sigma\downarrow$).*

Proof. Follows from Proposition 2.2.7 and Proposition 2.2.9. □

The effects of η here are similar to its effects in the continuous time limit.

Proposition 2.2.11 (Effect of trading frequency $1/\Delta_t$ on liquidity and efficiency). *When trading frequency is higher ($\frac{1}{\Delta_t}\uparrow$) and time duration between trades is lower ($\Delta_t\downarrow$), initial price impact is higher ($\lambda_0\uparrow$), initial information asymmetry*

(Σ_0) is unchanged, price impact and information asymmetry decay less per trade $(1 - \beta\Delta_t) \uparrow$ and decay more per unit of time ($T^{\lambda} \downarrow$ and $T^{\Sigma} \downarrow$).

Proof. Follows from Proposition 2.2.7 and Proposition 2.2.9. □

The effect of trading frequency $1/\Delta_t$ on liquidity is illustrated in Figure 2.3. A higher trading frequency results in a higher initial price impact λ_0, a slower decay of λ_n per trade, and a faster decay of λ_n per unit of time. Intuitively, when the informed trader has more trading opportunity, he can extract more profits from the price deviation generated by the randomly arrived noise trades. Market makers initially set a higher price impact λ_0 so that they can offset the increased loss to the informed trader. Although the informed trades less in each round and reveals less information, he reveals more information per unit of time because he trades more often.

A higher trading frequency $1/\Delta_t$ does not affect the initial information asymmetry (Σ_0) and improves the speed of information revelation through prices. The effect is illustrated in Figure 2.4 in which the expected public information arrival time is $\mathbf{E}[T] = 1/\eta = 5$. By the time $t = 4$, about 80% of the private information is revealed if $\Delta_t = 0.01$, and about 60% of the private information is revealed if $\Delta_t = 1$.

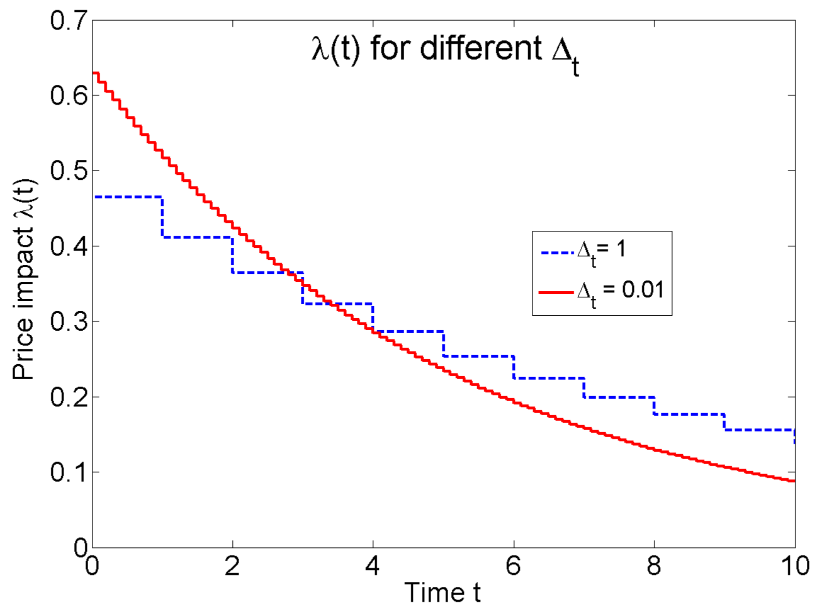

Figure 2.3: Price impact λ_t for different Δ_t when expected lifetime of the private information is $\mathbf{E}[T] = 1/\eta = 5$.

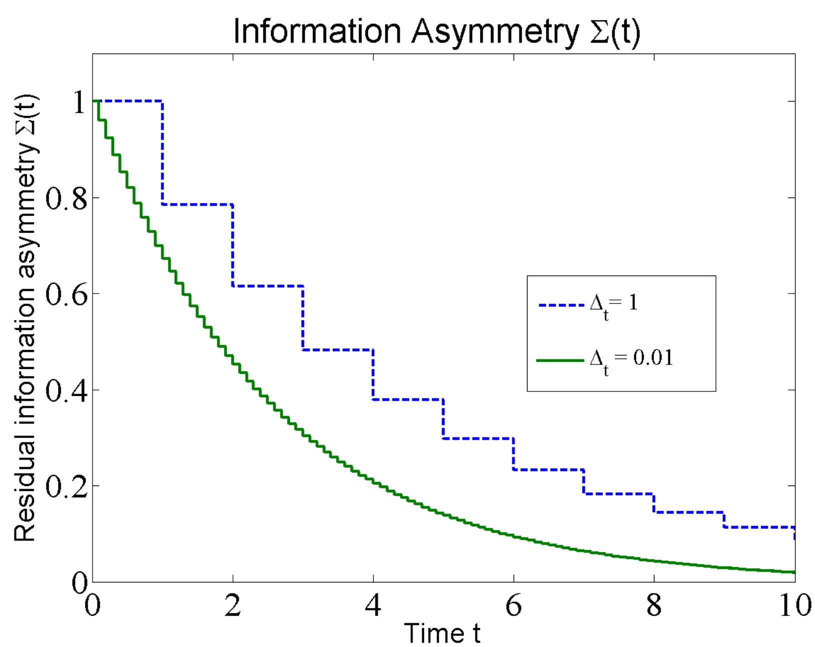

Figure 2.4: Information asymmetry Σ_t for different Δ_t when expected lifetime of the private information $\mathbf{E}[T] = 1/\eta = 5$.

2.3 Trading with high frequency traders

In the benchmark model of last section, we investigate the impact of trading frequency on the financial market when every trader acts at the same trading frequency.

High frequency traders (HFT), however, are generally perceived as those who can trade and quote *much faster* than others. HFTs use a variety of strategies and this paper does not attempt to be comprehensive. In this section, I extend the benchmark model to include HFTs as modeled in Chapter 1.

HFTs in this paper, as in Chapter 1, use their speed advantage and advance information of others' order flow to front-run slower traders. To be more specific, I assume that HFTs anticipate the incoming order flow within a small time window. They cannot differentiate informed trading Δx_n from noise trading Δz_n. HFTs only have noisy information about the size of the total incoming order of $\Delta y_n = \Delta x_n + \Delta z_n$. The continuum of market makers in aggregate effectively runs a market with the pricing rule $\Delta p_n = \lambda_n \Delta y_n$. HFTs first trades in the same direction of Δy_n and then reverses the trade.

The net effect of HFTs on other traders is captured by the divergence of temporary price impact $\hat{\lambda}$ and permanent price impact λ. When the informed and noise traders trade, they pay the temporary price impact $\hat{\lambda}$; market makers, however, only update prices with permanent price impact λ. The temporary price impact $\hat{\lambda}$ is higher than the permanent price impact λ such that the market makers can still break even when HFTs are present. For example, when buying Δy_n shares, liquidity demanders pay the price of $p_{n-1} + \hat{\lambda}_n \Delta y_n$ per share while the market makers set the end of period price as $p_{n-1} + \lambda_n \Delta y_n$. HFTs make low risk profits in a small time window around every trading instant. They cannot, however, compound the profits because the capacity of their strategy is limited by the liquidity demander's trading volume.

Definition 2.3.1 (HFT Intensity)**.** HFT intensity ξ_n is defined as the amplifying factor of temporary price impact caused by HFTs:

$$\xi_n := \frac{\hat{\lambda}_n}{\lambda_n} - 1 \tag{2.3.1}$$

As is shown in Chapter 1, ξ_n is determined by the quality of HFTs' information ρ and the degree of their competition γ. Assume that ρ and γ are constant over time, then the HFT intensity is also constant and[9]

$$\xi := \xi_n \in [0, 1] \tag{2.3.2}$$

Here HFT intensity ξ is exogenously given. It can be endogenous when we know the costs functions of HFTs' information and speed. A more complete model would consider the HFTs' technology, capital constraint, risk aversion, information precision, and belief about the market condition. A given ξ can be think of an equilibrium level resulting from the complete model.

Definition 2.3.2. A linear equilibrium of the trading game with HFT is a pair of informed trading intensity $\hat{\beta}$ and market makers' price updating rule $\hat{\lambda}_n$ satisfying

[9] Using the notations of Chapter 1, $\frac{\hat{\lambda}}{\lambda} = \frac{\lambda^T}{\lambda^P} = \frac{1+\theta^2}{2\theta^2} = \frac{1}{1-\rho\gamma}$ where $\rho \in [0, 1]$ and $\gamma \in [0, 0.5]$. Therefore, $\xi = \frac{\lambda^T}{\lambda^P} - 1 = \frac{\rho\gamma}{1-\rho\gamma} \in [0, 1]$

1. Information efficiency

$$p_n = \mathbf{E}_n^M[\hat{v}] = p_{n-1} + \hat{\lambda}_n \Delta y_n \tag{2.3.3}$$

2. Profit maximization

$$\hat{\beta} = \underset{\beta}{\text{argmax}} \quad \mathbf{E}_n^I \left[\sum_{i=n}^{\lfloor T/\Delta_t \rfloor} (v - p_{i-1} - (1+\xi)\hat{\lambda}_i)\Delta x_i \right]$$

where $\Delta x_i = \beta \frac{D_i}{\lambda_i} \Delta_t$.

Theorem 2.3.1. *[Equilibrium with HFTs] There exists a linear equilibrium where the informed trading intensity $\hat{\beta}$ is the solution to the equation*

$$e^{-\eta\Delta_t} = \frac{1 - 2\hat{\beta}(1+\xi)\Delta_t}{\sqrt{1-\hat{\beta}\Delta_t}(1-\hat{\beta}(1+2\xi)\Delta_t)} \tag{2.3.4}$$

where $\eta\Delta_t \in (0,1]$ and $\hat{\beta}\Delta_t \in \left(0, \frac{1}{2(1+\xi)}\right)$. In addition, for $n \geq 0$,

$$\textit{Informed trading size: } \Delta x_n = \frac{D_n}{\lambda_n}\hat{\beta}\Delta_t \tag{2.3.5}$$

$$\textit{Information asymmetry: } \Sigma_n = \rho_i \sigma_v^2 (1 - \hat{\beta}\Delta_t)^n \tag{2.3.6}$$

$$\textit{Permanent price impact: } \lambda_n = \sqrt{\hat{\beta}}\sqrt{\rho_i}\frac{\sigma_v}{\sigma_z}(1-\hat{\beta}\Delta_t)^{n/2} \tag{2.3.7}$$

$$\textit{Temporary price impact: } \hat{\lambda}_n = (1+\xi)\lambda_n \tag{2.3.8}$$

And the informed trader's value function is

$$V(n, D_n) = (1-(1+2\xi)\hat{\beta}\Delta_t)\left(\frac{D_n^2}{2\lambda_n} + \frac{\lambda_n \sigma_z^2}{2\hat{\beta}} \frac{1-2(1+\xi)\hat{\beta}\Delta_t}{1-\hat{\beta}\Delta_t}\right) \tag{2.3.9}$$

Proof. See Appendix B.6. □

When HFT intensity $\xi = 0$, the equilibrium reduces to the discrete time equilibrium of Theorem 2.2.5.

Corollary 2.3.2. *Let* $\hat{\beta}\Delta_t = \hat{g}(\eta\Delta_t)$*. Then,* $\hat{g}(\cdot) > 0$ *and* $\hat{g}''(\cdot) < 0$*. We can also write* $\hat{\beta} = \hat{g}(\eta\Delta_t)/\Delta_t$*. We have that* $\frac{\partial\hat{\beta}}{\partial\eta} > 0$*,* $\frac{\partial\hat{\beta}}{\partial\Delta_t} < 0$*, and* $\frac{\partial\hat{\beta}}{\partial\xi} < 0$*.*

Proof. Omitted. See Figure 2.5 and Figure 2.6. □

Corollary 2.3.3. *The half life of information asymmetry* Σ_t *and price impact* λ_t *is increasing in HFT intensity* ξ*.*

Proof. It follows directly from $\frac{\partial\hat{\beta}}{\partial\xi} < 0$ and the Theorem 2.3.1. □

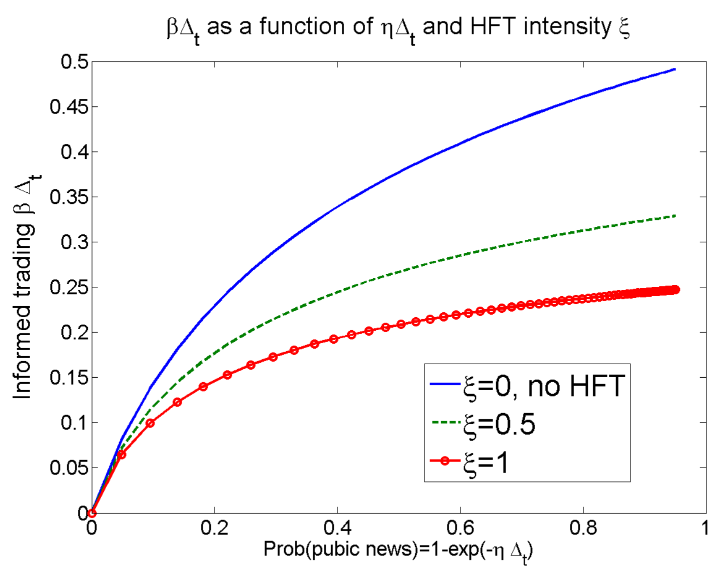

Figure 2.5: $\beta\Delta_t$ as a function of $\eta\Delta_t$ and HFT intensity ξ

In the discrete time equilibrium, when HFT intensity is higher ($\xi \uparrow$), the informed agent trade less $\beta\Delta_t$ on given information advantage D_n/λ_n. As a result, the market learns about the private information slower and the illiquidity decays slower.

HFTs do not affect all informed traders equally. Figure 2.6 illustrates that the impact of HFTs on informed traders is lower when the private information has a

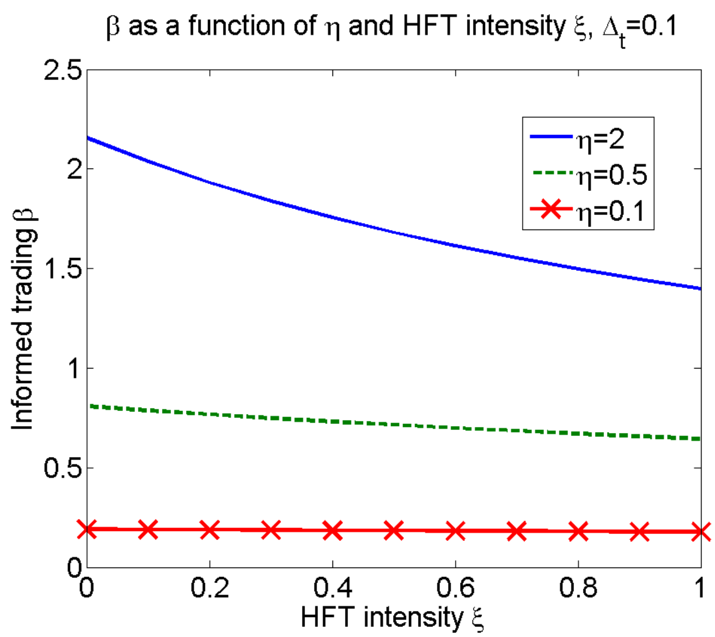

Figure 2.6: Informed trading intensity β as a function of HFT intensity ξ and public news arrival rate η

longer expected lifetime ($\mathbf{E}[T] = 1/\eta \uparrow$). When $\eta = 0.1$, the expected number of trading opportunity is $\mathbf{E}[T]/\Delta_t = 100$. The informed's trading intensity β is almost unaffected by HFT intensity ξ. By contrast, when $\eta = 2$ and the informed expects to trade $\mathbf{E}[T]/\Delta_t = 5$ times, informed trading intensity is significantly mitigated when HFT intensity ξ is higher.

Figure 2.7 illustrates this more clearly. As is shown in the figure, as the expected number of trading opportunities goes up ($\frac{\mathbf{E}[T]}{\Delta_t} = \frac{1}{\eta\Delta_t} \uparrow$), the effect of HFT on the trading intensity of the informed trader is reduced.

Keep the lifetime of information η fixed. The impact of HFT on the informed trader can also be reduced when the trading frequency is increased ($\Delta_t \downarrow$). In fact, the following theorem result shows that in the continuous time limit when $\Delta_t \to 0$,

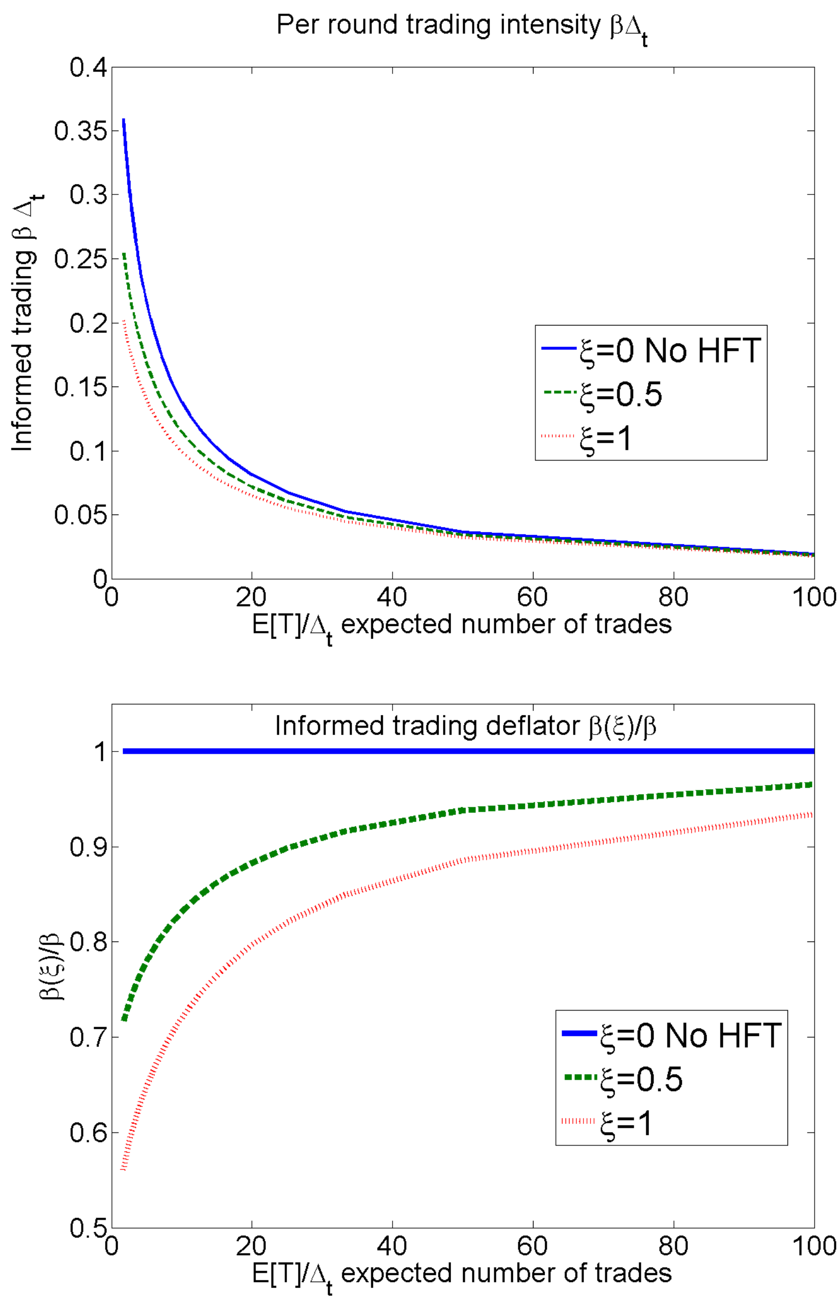

Figure 2.7: Impact of HFT on informed trading intensity

informed trader's trading intensity β becomes independent of HFT intensity ξ.

Proposition 2.3.4 (Continuous time limit informed trading)**.** *In the limit when $\Delta_t \to 0$, the informed trading intensity $\hat{\beta} \to 2\eta$ and his value function $V(n, D_n) \to \frac{D_n^2}{2\lambda_n} + \frac{\lambda_n \sigma_z^2}{4\eta}$.*

Proof. See Appendix B.7. □

Corollary 2.3.5 (Market quality in the continuous time limit with HFTs)**.** *In the continuous time limit when $\Delta_t \to 0$, the information efficiency Σ_t and the permanent price impact λ_t are unaffected by HFT intensity ξ. The temporary price impact $\hat{\lambda}_t = (1+\xi)\lambda_t$.*

We see that the impact of HFT front-running on the informed trader vanishes as trading frequency goes to infinity. The informed trader trades in exactly the same way as if there is no HFT. His value function is also exactly the same regardless of the level of HFT intensity ξ. This result might seem contradictory because the informed pays a higher price impact $\hat{\lambda}_t = (1+\xi)\lambda_t$ when HFT ξ is positive. Yet this higher cost of trading has no impact on the informed trader's strategy or his profits.

The result is due to the fact that the informed trader perfectly smooth out his trades in the continuous time limit. As long as his information has a positive expected lifetime $\eta < \infty$, in the continuous time limit, the informed trader expects to have infinitely many trading opportunities ($\frac{1}{\eta dt} \to \infty$). At any time instant, the informed's order size has the magnitude of dt while the noise traders' order size has the magnitude of $\sqrt{dt}$. Thus, the fraction of volume from the informed at any instant is zero. The price change dp_t is completely driven by noise trading dz_t.

Noise traders also trade the same amount for any HFT intensity ξ. Hence, the information content of the order flow is thus the same regardless of the intensity of HFT ξ. Market makers use the same permanent price impact factor λ_t to update the prices from period to period.

Note that the informed trader chooses trading size dx_t based on the permanent price impact λ_t rather than the temporary price impact $\hat{\lambda}_t$. This suggests that in the continuous time limit, the informed trader maximizes his trading profits when he chooses the optimal rate to use his information. Although the actual price impact costs $\hat{\lambda}_t$ is affected by HFT intensity ξ, the rate of information decay is not. From this perspective, this result resonates the original Kyle (1985) model where the informed trades trades away the same amount of information no matter how higher is σ_v/σ_z.

Noise traders suffer extra losses because they do not smooth out their trades and pay a higher price impact cost $\hat{\lambda}_t = (1+\xi)\lambda_t$ when HFT intensity ξ is higher. Noise traders' extra losses equal exactly the profits made by the HFTs. The informed trader is indifferent.

Result in this section suggests that in the current market where trading frequency is extremely high, HFTs are effectively collecting a "transaction tax" from noise traders who cannot smooth out their trades. For the informed trader who can perfectly smooth out his trading, the impact of such front-running HFTs with extremely short holding horizons is minimal.

Information efficiency of prices is unaffected in the continuous time limit because the informed is not affected. Hence, a higher trading frequency tends to encourage production of information, especially short-run information, even when there are front-runners.

Once information production is costly and trading frequency is finite, HFTs would tend to reduce the information efficiency of prices since HFTs reduces the marginal profit of information production. The effects of endogenous information production is explored in Appendix B.8.

2.4 Conclusion

In this paper, I show that a higher trading frequency has mixed effects on the financial market.

First, when all traders share the same trading frequency, the market price is unambiguously more informative. The informed trader's private information gets revealed faster and the informed trader is encouraged to produce more information. The impact of a higher trading frequency on liquidity, however, is less clear. The market is less liquid in the beginning but the illiquidity decays faster. The total trading losses of the liquidity driven noise traders are higher. Effectively, noise traders who trade in the beginning subsidize both the informed traders and the liquidity traders who trade later. The cost of a more informative price is borne by noise traders who have to trade when the informed asymmetry is most severe.

Second, when there exists high frequency traders who can front-run others, the market price is less informative because the informed trader trades less aggressively and produces less information. In the continuous time limit, however, the negative effect of HFT front-running on price informativeness vanishes. The patient informed trader perfectly smooths out his trading and HFTs effectively only front-run the impatient noise traders who are unable to trade smoothly.

Hence, HFT front-running reduces market liquidity and tends to reduce information efficiency. Such negative impact on price efficiency is mitigated when the trading frequency is high.

Results of this paper suggest an additional benefit of trading smoothly. The patient traders are less affected by short-horizon front-running HFTs. Impatient noise traders are most susceptible to short-horizon front-running because they have to trade quickly.

Chapter A: Appendices to Chapter 1

A.1 Proofs of Section 1.2

Section A.1.1 states a more general Theorem A.1.1 where the fast trader is risk averse with an exponential utility functions $-\exp(-A\pi)$. Then, section A.1.2 proves several necessary lemmas. Section A.1.3 proves the general theorem. Theorem 1.2.2 is a special case where the risk aversion coefficient is $A = 0$.

A.1.1 A generalized theorem

The equilibrium condition Equation (1.2.3) is generalized to allow risk aversion.

Definition A.1.1. Fast trader **utility** maximizing. Given the pricing function set by market makers $P(\cdot)$, the informed trader's strategy $X(\cdot)$, and a signal about the incoming order flow $I_y = x + z + e_y$ the fast trader's **utility** is maximized if she trades u^* shares at time 1^- and $-u^*$ shares at time 1, i.e.,

$$u^* = U(I_y; P(\cdot), X(\cdot)) = \underset{u}{\text{argmax}} \quad \mathbf{E}\left[U(\pi^F)\,|I_y, P(\cdot), X(\cdot)\right] \tag{A.1.1}$$

where $\pi^F = u(p_1 - p_{1^-})$ and $U(\pi) = -\exp(-A\pi)$.

Theorem A.1.1 (Equilibrium when the fast trader is risk averse). *Given Assumption 1, there is a unique equilibrium where the four strategy functions $X(\cdot)$,*

$U(\cdot)$, $P(\cdot)$, and $Q(\cdot)$ are

$$\text{Informed trading size: } x^* = X(v; p_0, \lambda^T) = \beta\,(v - p_0) \tag{A.1.2}$$

$$\text{Fast trading size: } u^* = U(I_y; p_0, \lambda^T) = \alpha\rho I_y \tag{A.1.3}$$

$$\text{Market order pricing: } p_{t_j} = P(y_j, p_{t_{j-1}}) = p_{t_{j-1}} + \lambda^T y_j \tag{A.1.4}$$

$$\text{Final quote: } p_{1+} = Q(u, y-u) = v_0 + \lambda^P y \tag{A.1.5}$$

The four endogenous parameters β, α, λ^T, and λ^P are:

$$\beta = \frac{\sigma_z}{\sigma_v}\theta, \quad \alpha = \frac{1}{2 + (1-\rho)A\sigma_v\sigma_z\frac{\theta^2+1}{2\theta}}, \quad \lambda^T = \frac{\sigma_v}{\sigma_z}\frac{1}{2\theta}, \quad \lambda^P = \frac{\sigma_v}{\sigma_z}\frac{\theta}{1+\theta^2}$$

where $\theta = \sqrt{\Theta}$ and Θ is the unique positive root in the range $\left[\frac{1-\rho/4}{1+\rho/4}, 1\right]$ of the cubic equation

$$\begin{aligned} 0 &= \left(A\sigma_v\sigma_z(1-\rho)\right)^2(\Theta+1)^2(\Theta-1) + A\sigma_v\sigma_z(1-\rho) \\ &\times(\Theta+1)\left(\Theta - \frac{1-\rho/4}{1+\rho/4}\right)(4+\rho) + 4\Theta\left(\Theta - \frac{1-\rho/4}{1+\rho/4}\right)(4+\rho) \end{aligned} \tag{A.1.6}$$

Corollary A.1.2. *If the fast trader is risk neutral ($A = 0$) the unique positive root of Equation (A.1.6) is $\theta^2 = \Theta = \frac{1-\rho/4}{1+\rho/4}$ and we have Theorem 1.2.2.*

Proof. The first two terms on the right hand side of the equation equal 0 if $A = 0$ or $1 - \rho = 0$. Only the third term remains. Eliminate the case where $\Theta = 0$ and $\lambda^T = \infty$. Then the unique equilibrium is $\theta = \sqrt{\frac{1-\rho/4}{1+\rho/4}}$ and it reduces to Theorem 1.2.2. □

A.1.2 Strategies of each trader

Lemma A.1.3. *Given the pricing function of Assumption 1, the fast trader's optimal trade size is $u^* = \alpha\hat{y}$ where $\alpha = \frac{1}{2+A\lambda^T(1-\rho)\sigma_y^2}$ and her maximized expected profit at* *time* 1^- *is* $\mathbf{E}\left[\pi^{F*}\middle|\,\lambda^T, I_y\right] = \alpha(1-\alpha)\lambda^T\hat{y}^2$ *where* $\hat{y} = \mathbf{E}[y|I_y] = \rho I_y$. *In particular, when* $A = 0$, *the fast trader is risk neutral and* $\alpha = \frac{1}{2}$ *and* $\mathbf{E}[\pi^{F*}] = \frac{1}{4}\lambda^T\hat{y}^2$.

Proof. market makers absorb residual order of size $y - u$ at time 1. Given the linear pricing function $p_{t_j} = p_{t_{j-1}} + \lambda^T y_{t_j}$, the difference of between the fast trader's exit price and entry price is $p_1 - p_{1^-} = \lambda^T(y - u)$. The fast trader's expected profit is $\mathbf{E}[\pi^F|p_0, \lambda^T, I_y, u] = \mathbf{E}\left[u\lambda^T(y-u)\middle|\, I_y, u\right] = u\lambda^T(\hat{y} - u)$ and its variance $\mathbf{V}\mathrm{ar}[\pi^F|p_0, \lambda^T, I_y, u] = (u\lambda^T)^2\,\mathbf{V}\mathrm{ar}[y - u|I_y, u] = (u\lambda^T)^2(1-\rho)\sigma_y^2$. The fast trader chooses u to maximize $\mathbf{E}[\pi^F|p_0, \lambda^T, I_y, u] - \frac{A}{2}\,\mathbf{V}\mathrm{ar}[\pi^F|p_0, \lambda^T, I_y, u]$. Solve the first order condition to get $u^* = \alpha\hat{y}$ and $\pi^{F*} = \alpha(1-\alpha)\lambda^T\hat{y}^2$. □

Remark A.1.1. The fast trader's trading intensity α is independent from price impact factor λ^T and information quality ρ only if she is risk neutral $A = 0$ or her information is perfect $\rho = 1$.

Lemma A.1.4. *Given Assumption 1, the informed trader's optimal size of trade is* $x^* = \frac{1}{2\lambda^T}(v - p_0)$ *and her maximized expected profit at time* 0^+ *is*

$$\mathbf{E}\left[\pi^{I*} = x^*(v - p_1)\middle|\, v, p_0, \lambda^T\right] = \frac{1}{4\lambda^T}(v - p_0)^2.$$

Proof. Given the pricing function $p_{t_j} = p_{t_{j-1}} + \lambda^T y_{t_j}$, the informed trader estimates that $\mathbf{E}[p_1|v, p_0, \lambda^T] = p_0 + \lambda^T\,\mathbf{E}[x + z] = p_0 + \lambda^T x$ Her expected profit is then $\mathbf{E}[\pi^I|v, p_0, \lambda^T] = x(v - p_0 - \lambda^T x)$. Solve the first order condition to find x^* and π^{I*}. □

Lemma A.1.5. *Assuming that the informed trader choose* $x = \beta(v - p_0)$ *and the fast trader chooses* $u = \alpha\hat{y}$, *the informationally efficient quotes* $p_0 = \mathbf{E}[v|\mathcal{F}_0] = v_0$ *and* $p_{1^+} = \mathbf{E}[v|\mathcal{F}_1] = v_0 + \lambda^P(x + z)$ *where* $\lambda^P = \frac{\beta\sigma_v^2}{\beta^2\sigma_v^2 + \sigma_z^2}$.

Proof. First, p_0 equals the ex ante expected value v_0. Second, the fast trader's trade size $u = \alpha\hat{y} = \alpha\rho I_y = \alpha\rho(x + z + e_y)$. Given $\alpha\rho$, u is informationally equivalent to $s_y = x + z + e_y$, a noisy observation of $y = x + z$. It has no extra information about v when $y = x + z$ is observed. Market makers can find $x + z$ by summing up two observed orders u and $x + z - u$ at time 1. From the projection theorem of normally distributed random variables,

$$p_{1^+} = \mathbf{E}[v|\mathcal{F}_1] = \mathbf{E}[v|y] = \mathbf{E}[v|\mathcal{F}_0] + \lambda^P y = v_0 + \frac{\mathbf{C}\text{ov}(y, v)}{\mathbf{V}\text{ar}[y]}(x + z)$$

Given the assumed normality and independence of z and v, we find that

$$\lambda^P = \frac{\mathbf{C}\text{ov}(y, v)}{\mathbf{V}\text{ar}[y]} = \frac{\mathbf{C}\text{ov}\left[\beta(v - p_0) + z, v\right]}{\mathbf{V}\text{ar}[\beta(v - p_0) + z]} = \frac{\beta\sigma_v^2}{\beta^2\sigma_v^2 + \sigma_z^2}$$

□

Lemma A.1.6. *Given Assumption 1 and the informed trader chooses $x = \beta(v - p_0)$ and the fast trader chooses $u = \alpha\hat{y}$, market makers makes zero expected profit if they set $\lambda^T = \frac{\lambda^P}{1-\rho(1-\alpha)\alpha}$, $\quad \rho \in [0, 1]$.*

Proof. Follows from Lemma A.2.5 on Page 122 □

A.1.3 Proof of the generalized theorem of A.1.1

Lemma A.1.3,A.1.4, A.1.5, and A.1.6 imply that equilibrium strategies are all linear given Assumption 1. And the four strategy functions are fully characterized by four parameters α, β, λ^P, and λ^T. In equilibrium, the following system of equations must hold.

$$\begin{aligned}\beta &= \frac{1}{2\lambda^T}, \quad \alpha = \frac{1}{2 + A\lambda^T(1-\rho)(\beta^2\sigma_v^2 + \sigma_z^2)}, \\ \lambda^P &= \frac{\beta\sigma_v^2}{\beta^2\sigma_v^2 + \sigma_z^2} \quad \lambda^T = \frac{\lambda^P}{1-\rho\alpha(1-\alpha)} \end{aligned} \tag{A.1.7}$$

There are three exogenous parameters: volatility of the fundamental value of the risky asset σ_v, volatility of noise trading σ_z, and the fast trader's information quality ρ as defined in Equation (1.2.1).

Define two unitless parameters as follows: $\theta := \beta\frac{\sigma_v}{\sigma_z}$ and $\eta := A\sigma_v\sigma_z$. θ is endogenous and it measures the informed trader's trading intensity. We will see that η and ρ are the only relevant exogenous parameters in equilibrium and they are also unitless. Changing the unit of v and z would have no effect on the exogenous parameters η and ρ and the endogenous parameter θ.

Then, we express β, α, λ^P, and λ^T all in terms of θ, η, and ρ as follows.

$$\beta = \frac{\sigma_z}{\sigma_v}\theta, \quad \lambda^T = \frac{\sigma_v}{\sigma_z}\frac{1}{2\theta} \quad \lambda^P = \frac{\sigma_v}{\sigma_z}\frac{\theta}{1+\theta^2}, \quad \alpha = \frac{1}{2 + (1-\rho)\eta\frac{\theta^2+1}{2\theta}}$$

The system of equations can be reduced to one equation of θ in terms of the exogenous variable $\eta = A\sigma_v\sigma_z$ and ρ.

$$\lambda^P\beta = \lambda^T\beta\left(1-\rho(1-\alpha)\alpha\right)$$

$$\Rightarrow \frac{\theta^2}{1+\theta^2} = \frac{1}{2}\left(1-\rho(1-\alpha)\alpha\right) \tag{A.1.8}$$

$$\Rightarrow \frac{2\theta^2}{1+\theta^2} = 1 - \frac{(1+\eta(1-\rho)\frac{1+\theta^2}{2\theta})\rho}{(2+\eta(1-\rho)\frac{1+\theta^2}{2\theta})^2} \tag{A.1.9}$$

Equation (A.1.9) is a sixth order polynomial equation of b:

$$0 = \theta^6\eta^2(1-\rho)^2 + 2\theta^5(4+\rho)\eta(1-\rho) + \theta^4\left(\eta^2(1-\rho)^2 + 4(\rho+4)\right)$$
$$+ 4\theta^3\eta(1-\rho)\rho - \theta^2\left(\eta^2(1-\rho)^2 - 4(\rho-4)\right) + 2\theta(\rho-4)\eta(1-\rho)$$
$$- \eta^2(1-\rho)^2 \tag{A.1.10}$$

It can be reduced to a cubic equation of θ^2. Let us define

$$U_1 := \eta(1-\rho) \geq 0, \quad U_2 := \frac{1-\rho/4}{1+\rho/4} \in [0.6, 1] \tag{A.1.11}$$

U_1 and U_2 are bounded since $\rho \in [0,1]$. Then, Equation (A.1.10) can be rewritten as a cubic equation of θ^2

$$U_1^2(\theta^2+1)^2(1-\theta^2) = U_1(\theta^2+1)(\theta^2-U_2)(4+\rho) + 4\theta^2(\theta^2-U_2)(4+\rho) \tag{A.1.12}$$

We can express θ^2 in closed-form as a function of ρ and η. The following Lemma suffice to conclude the proof of Theorem A.1.1.

Lemma A.1.7. *For $\eta \geq 0$ and $\rho \in [0,1]$, the cubic equation* (A.1.12) *has an unique real root of θ^2 in the range $[U_2, 1]$.*

Proof. For $\eta \geq 0$ and $\rho \in [0,1]$, $U_1 = \eta(1-\rho) \geq 0$ and $U_2 \in [0.6, 1]$. Define $\Theta = \theta^2 > 0$. Let $g(\Theta)$ denote the right hand side of Equation (A.1.12), i.e.,

$$g(\Theta) = U_1^2(\Theta+1)^2(\Theta-1) + U_1(\Theta+1)(\Theta-U_2)(4+\rho) + 4\Theta(\Theta-U_2)(4+\rho)$$

1. $g(\Theta) < 0$ if $0 < \Theta < U_2 \leq 1$ because the first and the second term of $g(\Theta)$ are weakly negative and the third term is strictly negative.

2. $g(\Theta) > 0$ if $\Theta > 1 \geq U_1$ because the first two terms of $g(\Theta)$ are weakly positive and the third term is strictly positive.

3. $g(\Theta) = 0$ has at least one root in $[U_2, 1]$ since $g(U_2) \leq 0 \leq g(1)$ and $g(\Theta)$ is continuous.

$$g(U_2) = U_1^2(U_2+1)^2(U_2-1) \leq 0, \quad g(1) = 2(U_1+2)(1-U_2)(4+\rho) \geq 0$$

4. The root in $[U_2, 1]$ is unique because $g(\Theta)$ is increasing in $[U_2, 1]$. The second and third terms of $g(\Theta)$ are increasing in Θ. The first term is also increasing in Θ for $\Theta > \frac{1}{3}$ because

$$\frac{d}{d\Theta}(\Theta+1)^2(\Theta-1) = 3\Theta^2 + 2\Theta - 1 > 0 \quad \text{if } \Theta > \frac{1}{3} \tag{A.1.13}$$

   Since $\Theta \geq U_2 \geq 0.6$, all three terms of $g(\Theta)$ are increasing for $\Theta \in [U_2, 1]$.

Therefore, there is a unique real $\Theta \in [U_2, 1]$ that satisfies $g(\Theta) = 0$. □

Corollary A.1.8. *If $A = 0$ or $\rho = 1$, $U_1 = 0$ and $\Theta = U_2$ is the unique positive root. If $\rho = 0$, $U_2 = 1$ and $\Theta = 1$ is the unique positive root.*

A.1.4 Equilibrium impact of fast trader's risk aversion

As illustrated in Figure A.1, risk aversion mitigates the impact of the fast trader on equilibrium when her information is noisy $\rho < 1$. The fast trader's trading intensity α is now increasing with her information quality ρ and decreasing with her risk exposure $A\sigma_v\sigma_z$.

As the fast trader becomes more risk averse ($A \uparrow$), the fast trader reduces trading intensity ($\alpha \downarrow$), market makers amplify temporary price impact less $\frac{\lambda^T}{\lambda^P} \downarrow$, and the informed trader increases trading intensity ($\beta \uparrow$).

Proposition A.1.9. $\theta^2 \in \left[\frac{1-\rho/4}{1+\rho/4}, 1\right]$ *increases with* $\eta = A\sigma_v\sigma_z$ *and decreases with* ρ. *Particularly, as* $\eta \to 0$ *or* $\rho \to 1$, $U_1 \to 0$ *and* $\theta^2 \to U_2 = 0.6$*; as* $\eta \to \infty$ *or* $\rho \to 0$, $\theta^2 \to U_2 = 1$.

Proof. Omitted. It can be derived from the closed-form representation of θ^2 as a root of the cubic equation or the implicit function theorem. □

Proposition A.1.9 shows that when $A > 0$, $\lim_{\eta\to\infty} \theta = 1$. When the fast trader is risk neutral ($A = 0$), θ is unaffected by $\sigma_v\sigma_z$. This highlights the importance of risk aversion especially when risk exposure $A\sigma_v\sigma_z$ is high. When the fast trader is risk averse, her impact becomes less important as her risk exposure $A\sigma_v\sigma_z \to \infty$.

For example, Proposition 1.5.3 implies that the ratio of the fast trader's expected profit and the informed trader's expected profit is $\frac{\mathbf{E}[\pi^F]}{\mathbf{E}[\pi^I]} = \frac{1-\theta^2}{\theta^2}$. Assume that the fast trader's information is noisy ($\rho < 1$). Then,

$$\text{When } A > 0: \quad \lim_{\sigma_v\sigma_z\to\infty} \frac{\mathbf{E}[\pi^F]}{\mathbf{E}[\pi^I]} = \lim_{\sigma_v\sigma_z\to\infty} \frac{1-\theta^2}{\theta^2} = 0 \tag{A.1.14}$$

$$\text{When } A = 0: \quad \frac{\mathbf{E}[\pi^F]}{\mathbf{E}[\pi^I]} = \frac{1-\theta^2}{\theta^2} = \frac{\rho/2}{1-\rho/4} > 0 \tag{A.1.15}$$

Keeping the fast trader's risk aversion $A > 0$ fixed and $\rho < 1$, the fast trader's expected profit $\mathbf{E}[\pi^F]$ still increases with $\sigma_v\sigma_z$ but increases slower than $\mathbf{E}[\pi^I]$. In the limit when $\sigma_v\sigma_z \to \infty$, the fast trader's profit is negligible compared to the informed trader's profit.

Intuitively, the fast trader is more risk averse than the informed trader. As $\sigma_v\sigma_z$ increases, the risk exposure is higher. The fast trader reduces trading intensity more than the informed trader. In the limit, $\sigma_v\sigma_z \to \infty$, the fast trader stop trading

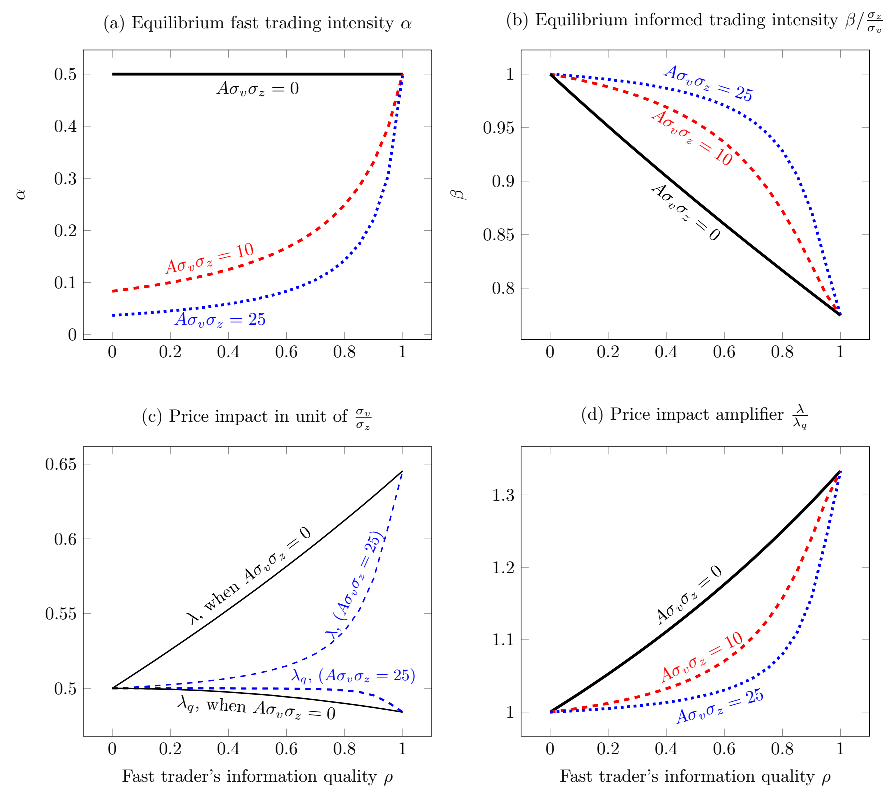

Figure A.1: Equilibrium of model with a risk averse fast trader. A is the fast trader's risk aversion. ρ is the fast trader's information quality. $A\sigma_v\sigma_z$ measures the fast trader's risk exposure. ρ and $A\sigma_v\sigma_z$ are both unitless.

(a) Fast trading intensity α increases with her information quality ρ and decreases with the risk exposure $A\sigma_v\sigma_z$. **(b)** Normalized informed trading intensity $\beta \cdot \frac{\sigma_v}{\sigma_z}$ decreases with the fast trader's information quality ρ and increases with risk exposure $A\sigma_v\sigma_z$. **(c)** Normalized temporary price impact per share λ^T increases with the fast trader's information quality ρ and decreases withe $A\sigma_v\sigma_z$. Normalized permanent price impact per share λ^P decreases with ρ and increases with $A\sigma_v\sigma_z$. **(d)** Temporary price impact amplifier $\frac{\lambda^T}{\lambda^P}$ measures the extra friction caused by the fast trader. It increases with ρ and decreases with $A\sigma_v\sigma_z$.

due to high risks and all profits go to the informed trader.

A.2 Proofs of Section 1.3

A.2.1 Trader strategies with multiple fast traders

Lemma A.2.1 (Prices)**.** *The order flow is* $\{u_1, \cdots, u_J, y - S_J\}$*. The linear pricing function specified in Assumption 1 implies that the traded prices are*

$$p_{t_j} = p_{t_{j-1}} + \lambda^T u_j = p_0 + \lambda^T S_j, \quad j = 1, 2, \cdots J \tag{A.2.1}$$

$$p_1 = p_{t_J} + \lambda^T (y - S_J) = p_0 + \lambda^T y \tag{A.2.2}$$

By equilibrium condition, market makers set initial and final quotes to their conditional expectation of v*.*

$$p_0 = \mathbf{E}[v] = v_0, \quad p_{1^+} = \mathbf{E}[v|\mathcal{F}_1] = \mathbf{E}[v\,|u_1, u_2, \cdots, u_J, y - S_J] \tag{A.2.3}$$

where $\mathcal{F}_t$ *denotes the information of market makers at time* t*.*

Lemma A.2.2 (Each fast trader's order size)**.** *Given all others' strategy* β*,* λ^T*,* λ^P*,* $\sum_{j,l\neq k} u_{j,l}$*, the last traded price* $p_{t_{j-1}}$*, and the signal* I_y*, fast trader* (j,k)*'s optimal trader size* $u^*_{j,k} = \frac{1}{2}\left(\hat{y} - S_{j-1} - \sum_{l\neq k} u_{j,l}\right)$*.*

Proof. Fast trader (j,k)'s expected profit after submitting $u_{j,k}$ is

$$\begin{aligned}\mathbf{E}[\pi^F_{j,k}|I_y, p_{t_{j-1}}] &= \mathbf{E}\left[u_{j,k}\left(p_1 - p_{t_j}\right)|I_y, p_{t_{j-1}}\right] = u_{j,k}\,\mathbf{E}\left[p_1 - p_{t_{j-1}} - \lambda^T u_j|I_y, p_{t_{j-1}}\right]\\ &= \lambda^T u_{j,k}\,\mathbf{E}\left[(y - S_{J-1}) - \sum_{l\neq k} u_{j,l} - u_{j,k}\middle| I_y, p_{t_{j-1}}\right]\\ &= \lambda^T u_{j,k}\left(\hat{y} - S_{J-1} - \sum_{l\neq k} u_{j,l}\right) - \lambda^T u^2_{j,k} \end{aligned}\tag{A.2.4}$$

First order condition then gives us $u^*_{j,k} = \frac{1}{2}\left(\hat{y} - S_{j-1} - \sum_{l \neq k} u_{j,l}\right)$. $\square$

Lemma A.2.3 (Fast trading size in symmetric equilibrium)**.** *In the symmetric equilibrium where fast traders arriving at time j submit the same order, $u^*_{j,k} = \frac{1}{n_j} u^*_j = \frac{1}{1+n_j}(\hat{y} - S_{y-1})$.*

Proof. Sum $u^*_{j,k}$ of Lemma A.2.2 over $k \in \{1, 2, \cdots, n_j\}$.

$$u^*_j = \sum_{k=1}^{n_j} u^*_{j,k} = \frac{n_j}{2}(\hat{y} - S_{y-1}) - \frac{n_{j-1}}{2} \sum_{k=1}^{n_j} u^*_{j,k}$$

$$\Rightarrow u^*_j = \frac{n_j}{1+n_j}(\hat{y} - S_{y-1}) \quad u^*_{j,k} = \frac{1}{n_j} u^*_j = \frac{1}{1+n_j}(\hat{y} - S_{y-1}) \tag{A.2.5}$$

The last step uses the symmetry. $\square$

Lemma A.2.4 (Fast trading size u_j and S_j)**.** *Assume that the aggregate fast trading at time t_j is $u_j = \alpha_j(\hat{y} - S_{j-1})$, then for $1 \leq j \leq J$*

$$u_j = \hat{y}\alpha_j \prod_{i=0}^{j-1}(1 - \alpha_i), \quad S_j = \sum_{i=0}^{j} u_i = \hat{y}\left(1 - \prod_{i=0}^{j}(1 - \alpha_i)\right) \tag{A.2.6}$$

Proof. Proof by induction. $S_0 = u_0 = \alpha_0 = 0$ by definition. $u_1 = \alpha_1(\hat{y} - S_0) = \alpha_1 \hat{y}$ and $S_1 = u_0 + u_1 = u_1 = \alpha_1 \hat{y}(1 - (1 - \alpha_1)(1 - \alpha_0))$. Suppose the Lemma holds for $j \leq K$. Then,

$$u_{K+1} = \alpha_{K+1}(\hat{y} - S_K) = \alpha_{K+1}\left(\hat{y} - \hat{y}\left(1 - \prod_{i=0}^{K}(1 - \alpha_i)\right)\right) = \hat{y}\alpha_{K+1} \prod_{i=0}^{K}(1 - \alpha_i)$$

and

$$
\begin{aligned}
S_{K+1} &= u_{K+1} + S_K = \alpha_{K+1}(\hat{y} - S_K) + S_K \\
&= \alpha_{K+1}\hat{y} + (1 - \alpha_{K+1})S_K \\
&= \alpha_{K+1}\hat{y} + (1 - \alpha_{K+1})\hat{y}\left(1 - \prod_{i=0}^{K}(1 - \alpha_i)\right) \\
&= \hat{y}\left(1 - (1 - \alpha_{K+1})\prod_{i=0}^{K}(1 - \alpha_i)\right) \\
&= \hat{y}\left(1 - \prod_{i=0}^{K+1}(1 - \alpha_i)\right)
\end{aligned}
$$

□

Lemma A.2.5 (Equation of λ^T and λ^P)**.** *Given Assumption 1 and assume that the aggregate fast trading at time t_j is $u_j = \alpha_j(\hat{y} - S_{j-1})$, the following equation must hold if market makers make 0 expected profit at time 0:*

$$\lambda^P = \lambda^T(1 - \rho\gamma), \quad \alpha_0 = 0, \quad \gamma = \sum_{j=1}^{J}\left(\alpha_j(1 - \alpha_j)\prod_{i=0}^{j-1}(1 - \alpha_i)^2\right) \tag{A.2.7}$$

Proof. From Lemma A.2.1 and A.2.4, $p_{t_j} = p_0 + \lambda^T S_j = p_0 + \lambda^T \hat{y}\left(1 - \prod_{i=0}^{j}(1 - \alpha_i)\right)$. The final quote is $p_{1^+} = p_0 + \lambda^P y$. At time 1, market makers absorb the residual order flow $y - \sum_{i=0}^{J} u_i$. Their expected profit ex ante of the trade is

$$
\begin{aligned}
\mathbf{E}\left[\left(y - \sum_{i=0}^{J} u_i\right)(p_1 - p_{1^+})\right] &= \mathbf{E}\left[\left(y - \hat{y}\left(1 - \prod_{i=0}^{J}(1 - \alpha_i)\right)\right)(\lambda^T - \lambda^P)y\right] \\
&= \sigma_y^2(\lambda^T - \lambda^P)\left(1 - \rho\left(1 - \prod_{i=0}^{j}(1 - \alpha_i)\right)\right)
\end{aligned}
$$

market makers expect to lose by trading with fast traders at each time t_j. Their expected loss ex ante is

$$\begin{aligned}
\mathbf{E}\left[\mathbf{E}\left[u_j\left(p_{1^+}-p_{t_j}\right)|I_y\right]\right] &= \mathbf{E}\left[u_j\left(\lambda^P\hat{y}-\lambda^T\hat{y}\left(1-\prod_{i=0}^{j}(1-\alpha_i)\right)\right)\right] \\
&= \mathbf{E}\left[\hat{y}^2\right]\alpha_j\prod_{i=0}^{j-1}(1-\alpha_i)\left(\lambda^P-\lambda^T+\lambda^T\prod_{i=0}^{j}(1-\alpha_i)\right) \\
&= \rho\sigma_y^2\alpha_j\prod_{i=0}^{j-1}(1-\alpha_i)\left(\lambda^P-\lambda^T+\lambda^T\prod_{i=0}^{j}(1-\alpha_i)\right)
\end{aligned}$$

Their total expected loss is then

$$\begin{aligned}
&\sum_{j=1}^{J}\rho\sigma_y^2\alpha_j\prod_{i=0}^{j-1}(1-\alpha_i)\left(\lambda^P-\lambda^T+\lambda^T\prod_{i=0}^{j}(1-\alpha_i)\right) \\
=&\rho\sigma_y^2\sum_{j=1}^{J}\left(\left(\lambda^P-\lambda^T\right)\alpha_j\prod_{i=0}^{j-1}(1-\alpha_i)+\lambda^T\alpha_j(1-\alpha_j)\prod_{i=0}^{j-1}(1-\alpha_i)^2\right) \\
=&\rho\sigma_y^2\left(\left(\lambda^P-\lambda^T\right)\sum_{j=1}^{J}\alpha_j\prod_{i=0}^{j-1}(1-\alpha_i)+\lambda^T\sum_{j=1}^{J}\alpha_j(1-\alpha_j)\prod_{i=0}^{j-1}(1-\alpha_i)^2\right) \\
=&\rho\sigma_y^2\left(\left(\lambda^P-\lambda^T\right)\left(1-\prod_{j=1}^{J}(1-\alpha_j)\right)+\lambda^T\sum_{j=1}^{J}\alpha_j(1-\alpha_j)\prod_{i=0}^{j-1}(1-\alpha_i)^2\right)
\end{aligned}$$

The last step uses Lemma A.2.4. The 0 profit condition requires that the market makers' total expected loss equal their expected profit. Hence,

$$\begin{aligned}
&\sigma_y^2(\lambda^T-\lambda^P)\left(1-\rho\left(1-\prod_{i=0}^{j}(1-\alpha_i)\right)\right) \\
&\quad=\rho\sigma_y^2\left(\left(\lambda^P-\lambda^T\right)\left(1-\prod_{j=1}^{J}(1-\alpha_j)\right)+\lambda^T\sum_{j=1}^{J}\alpha_j(1-\alpha_j)\prod_{i=0}^{j-1}(1-\alpha_i)^2\right)
\end{aligned}$$

Eliminate $\sigma_y^2>0$ from both sides and rearrange the terms and we have

$$\lambda^P=\lambda^T\left(1-\rho\sum_{j=1}^{J}\left(\alpha_j(1-\alpha_j)\prod_{i=0}^{j-1}(1-\alpha_i)^2\right)\right),\quad \alpha_0=0$$

□

Lemma A.2.6. *Suppose all fast traders choose* $u_{j,k} = \alpha_{j,k}(\hat{y} - S_{j-1})$ *and the informed trader chooses* $x = \beta(v - p_0)$. *The final quote is* $p_{1^+} = \mathbf{E}[v|\mathcal{F}_1^V] = v_0 + \lambda^P(x+z)$ *where* $\lambda^P = \frac{\beta\sigma_v^2}{\beta^2\sigma_v^2+\sigma_z^2}$.

Proof. The aggregate order flow over the entire trading round equals $y = S_J + x + z - S_J = x + z$. Fast traders choose $u_{j,k}$ based on $\hat{y} = \rho I_y$. Because $I_y = x + z + e_y$ is a noisy observation of $x + z$, observing u_j has no additional information about v once y is observed. Thus, $\mathbf{E}[v|\mathcal{F}_1] = \mathbf{E}[v|x+z] = v_0 + \frac{\beta\sigma_v^2}{\beta^2\sigma_v^2+\sigma_z^2}$ based on the projection theorem for normal random variables. $\square$

A.2.2 Proof of Theorem 1.3.1

Proof. Lemma A.1.4, A.2.3, A.2.5 and A.2.6 imply that equilibrium strategies are

$$x = \beta(v - v_0), \quad p_{1^+} = v_0 + \lambda^T y, \quad p_0 = v_0 \quad p_{t_j} = p_{t_{j-1}} + \lambda^T y_j, \quad \forall j$$

$$u_{j,k} = \frac{\alpha_j}{n_j}(\hat{y} - S_{j-1}), \quad \forall j, k$$

And the parameters satisfy

$$\alpha_j = \frac{n_j}{1+n_j}, \quad \beta = \frac{1}{2\lambda^T}, \quad \lambda^P = \frac{\beta\sigma_v^2}{\beta^2\sigma_v^2+\sigma_z^2}$$

$$\lambda^P = \lambda^T\left(1 - \rho\sum_{j=1}^{J}\alpha_j(1-\alpha_j)\prod_{i=0}^{j-1}(1-\alpha_i)^2\right), \quad \alpha_0 = 0$$

Define $\theta := \beta\frac{\sigma_v}{\sigma_z} > 0, \quad \gamma := \sum_{j=1}^{J}\alpha_j(1-\alpha_j)\prod_{i=0}^{j-1}(1-\alpha_i)^2$. Then,

$$\lambda^P\beta = \frac{\theta^2}{\theta^2+1} = \frac{1}{2}(1-\rho\gamma) \quad \Rightarrow \theta = \sqrt{\frac{1-\rho\gamma}{1+\rho\gamma}} \tag{A.2.8}$$

And it follows that

$$\beta = \frac{\sigma_z}{\sigma_v}b, \quad \lambda^T = \frac{1}{2\beta} = \frac{\sigma_v}{\sigma_z}\frac{1}{2\theta} \quad \lambda^P = \frac{1}{\beta}\frac{\theta^2}{1+\theta^2} = \frac{\sigma_v}{\sigma_z}\frac{\theta}{1+\theta^2} \tag{A.2.9}$$

$\square$

A.2.3 Proof of Proposition 1.4.1

The (j,k)-th fast trader chooses $u_{j,k} = \frac{\alpha_j}{n_j}(\hat{y} - S_{j-1})$. Her execution price is $p_{t_j} = p_0 + \lambda^T S_j$ and she expects $p_1 = p_0 + \lambda^T y$. Plus, Lemma A.2.4 implies that $\hat{y} - S_j = \hat{y}\prod_{i=0}^{j}(1-\alpha_i)$. Thus,

$$\begin{aligned}
\mathbf{E}\left[u_{j,k}(p_1 - p_{t_j})|I_y\right] &= \lambda^T \frac{\alpha_j}{n_j}(\hat{y} - S_{j-1})\,\mathbf{E}\left[y - S_j|I_y\right] = \lambda^T \frac{\alpha_j}{n_j}(\hat{y} - S_{j-1})(\hat{y} - S_j) \\
&= \lambda^T \frac{\alpha_j}{n_j}\hat{y}^2(1-\alpha_j)\prod_{i=0}^{j-1}(1-\alpha_i)^2
\end{aligned}$$

Hence, her expected profit ex ante is

$$\begin{aligned}
\mathbf{E}\left[u_{j,k}(p_1 - p_{t_j})\right] &= \mathbf{E}\left[\mathbf{E}\left[u_{j,k}(p_1 - p_{t_j})|I_y\right]\right] \\
&= \lambda^T \frac{\alpha_j(1-\alpha_j)}{n_j}\,\mathbf{E}\left[\hat{y}^2\right]\prod_{i=0}^{j-1}(1-\alpha_i)^2 = \lambda^T \rho\sigma_y^2 \frac{\alpha_j(1-\alpha_j)}{n_j}\prod_{i=0}^{j-1}(1-\alpha_i)^2 \\
&= \lambda^T \rho(\theta^2+1)\sigma_z^2 \frac{\alpha_j(1-\alpha_j)}{n_j}\prod_{i=0}^{j-1}(1-\alpha_i)^2 \\
&= \sigma_v\sigma_z\rho\frac{\theta^2+1}{2\theta}\frac{\alpha_j(1-\alpha_j)}{n_j}\prod_{i=0}^{j-1}(1-\alpha_i)^2
\end{aligned} \tag{A.2.10}$$

By the definition of γ and θ,

$$\rho\sum_{j=1}^{J}\left(\alpha_j(1-\alpha_j)\prod_{i=0}^{j-1}(1-\alpha_i)^2\right) = \rho\gamma = \frac{1-\theta^2}{1+\theta^2} \tag{A.2.11}$$

Thus, total fast trading profit equals

$$\sum_{j=1}^{J}\sum_{k=1}^{n_j}\mathbf{E}\left[u_{j,k}(p_1 - p_{t_j})\right] = \sigma_v\sigma_z\frac{\theta^2+1}{2\theta}\rho\gamma = \sigma_v\sigma_z\frac{1-\theta^2}{2\theta} \tag{A.2.12}$$

A.3 Proofs of Section 1.4

A.3.1 Basic properties of speed friction γ

Proposition A.3.1. $0 \leq \gamma < 1$ *for* $J > 0$.

Proof. $\gamma \geq 0$ since $\alpha_j \in [0, 1)$. Let $\gamma(J) = \sum_{j=1}^{J} \left(\alpha_j (1 - \alpha_j) \prod_{i=0}^{j-1} (1 - \alpha_i)^2 \right)$. We can prove by induction that $\gamma(J) < 1 - \prod_{i=0}^{J} (1 - \alpha_i)^2$.

1. $\gamma(1) = (1 - \alpha_1) - (1 - \alpha_1)^2 < 1 - (1 - \alpha_1)^2$.

2. Assume $\gamma(k-1) < 1 - \prod_{i=0}^{k-1} (1 - \alpha_i)^2$ for $k - 1 \geq 1$. Then,

$$\begin{aligned}
\gamma(k) &= \gamma(k-1) + \alpha_k (1 - \alpha_k) \prod_{i=0}^{k-1} (1 - \alpha_i)^2 \\
&< 1 - \prod_{i=0}^{k-1} (1 - \alpha_i)^2 + \alpha_k (1 - \alpha_k) \prod_{i=0}^{k-1} (1 - \alpha_i)^2 \\
&< 1 - (1 - \alpha_k)^2 \prod_{i=0}^{k-1} (1 - \alpha_i)^2 = 1 - \prod_{i=0}^{k} (1 - \alpha_i)^2
\end{aligned}$$

3. Therefore, $\gamma(J) < 1 - \prod_{i=0}^{J} (1 - \alpha_j)^2 < 1$ for all $J \geq 1$.

□

Lemma A.3.2. $\frac{\partial \gamma}{\partial \alpha_k} < 0$ *if* $\alpha_k \geq \frac{1}{2}$.

Proof. α_k is not included in the terms where $j < k$. Thus,

$$\frac{\partial \gamma}{\partial \alpha_k} = \left(\prod_{i=0}^{k-1} (1 - \alpha_i)^2 \right) \cdot \frac{\partial}{\partial \alpha_k} \left(\alpha_k (1 - \alpha_k) + \sum_{j=k+1}^{J} \left(\alpha_j (1 - \alpha_j) \prod_{l=k}^{j-1} (1 - \alpha_l)^2 \right) \right)$$

Since $\alpha_i < 1$ for all $i \le k-1$, $\prod_{i=0}^{k-1}(1-\alpha_i)^2 > 0$. $\frac{\partial \gamma}{\partial \alpha_k} < 0 \iff$

$$\frac{\partial}{\partial \alpha_k}\left(\alpha_k(1-\alpha_k) + (1-\alpha_k)^2\left(\alpha_{k+1}(1-\alpha_{k+1}) + \sum_{j=k+2}^{J}\alpha_j(1-\alpha_j)\prod_{l=k+1}^{j-1}(1-\alpha_l)^2\right)\right) < 0$$

$$\iff \left((1-2\alpha_k) - 2(1-\alpha_k)\left(\alpha_{k+1}(1-\alpha_{k+1}) + \sum_{j=k+2}^{J}\alpha_j(1-\alpha_j)\prod_{l=k+1}^{j-1}(1-\alpha_l)^2\right)\right) < 0$$

$$\iff 1-2\alpha_k < 2(1-\alpha_k)\left(\alpha_{k+1}(1-\alpha_{k+1}) + \sum_{j=k+2}^{J}\alpha_j(1-\alpha_j)\prod_{l=k+1}^{j-1}(1-\alpha_l)^2\right) \tag{A.3.1}$$

Since $\alpha_k \ge \frac{1}{2}$, the left hand side of Eq. (A.3.1) $1-2\alpha_k \le 0$. The right hand side of Eq. (A.3.1) is positive because $\alpha_j < 1$. Hence, $\frac{\partial \gamma}{\partial \alpha_k} < 0$. □

Lemma A.3.3. *Consolidating the last two groups of fast traders reduces speed friction, i.e.., $\gamma_{\{\cdots,m+n\}} < \gamma_{\{\cdots,m,n\}}$.*

Proof. First, $\gamma_{\{\cdots,m+n\}} < \gamma_{\{\cdots,m\}}$ because entry of n fast traders at the same time as the m fast traders reduces speed friction when $m > 0$; second, $\gamma_{\{\cdots,m\}} = \gamma_{\{\cdots,m,\infty\}}$ because infinite number of fast traders at the end are perfectly competitive and act like the competitive market makers; third, $\gamma_{\{\cdots,m,\infty\}} < \gamma_{\{\cdots,m,n\}}$ because reducing the number of fast traders at time t_J increases speed friction. In sum, $\gamma_{\{\cdots,m+n\}} < \gamma_{\{\cdots,m\}} < \gamma_{\{\cdots,m,\infty\}} < \gamma_{\{\cdots,m,n\}}$ and it follows that $\gamma_{\{\cdots,m+n\}} < \gamma_{\{\cdots,m,n\}}$. □

A.3.2 Proof of Proposition 1.4.6

Proof. First, if $n_j > 0$, $\alpha_j = \frac{n_j}{1+n_j} \ge \frac{1}{2}$. After a new fast trader enters, the new $\alpha_j' = \frac{n_j+1}{n_j+2} > \frac{n_j}{n_j+1} = \alpha_j$. Lemma A.3.2 then implies that γ decreases.

Then I prove entry of one fast trader with a different speed from existing risk neutral fast traders increases speed friction γ. Suppose that there are J instants that the fast traders can trade. Compare the following two speed profiles of fast traders:

$$\text{Speed profile 1: } \{n_1, n_2, \cdots, 0, n_{k+1}, \cdots, n_J\},$$
$$\text{Speed profile 2: } \{n_1, n_2, \cdots, 1, n_{k+1}, \cdots, n_J\}$$

for $1 \leq k \leq J$. Speed profile 2 is obtained if 1 fast trader enter at time t_k. The aggregate trading intensities of the two speed profiles are:

$$\text{profile 1: } \{\alpha_1, \alpha_2, \cdots, 0, \alpha_{k+1}, \cdots, \alpha_J\}$$
$$\text{and profile 2: } \{\alpha_1, \alpha_2, \cdots, \alpha_k, \alpha_{k+1}, \cdots, \alpha_J\}$$

Speed friction of profile 2 is as follows:

$$\begin{aligned}
\gamma_2 &= \sum_{j=1}^{J} \left(\alpha_j (1-\alpha_j) \prod_{i=0}^{j-1} (1-\alpha_i)^2 \right) \\
&= \sum_{j=1}^{k-1} \left(\alpha_j (1-\alpha_j) \prod_{i=0}^{j-1} (1-\alpha_i)^2 \right) + \left(\alpha_k (1-\alpha_k) \prod_{i=0}^{k-1} (1-\alpha_i)^2 \right) \\
&\quad + \sum_{j=k+1}^{J} \left(\alpha_j (1-\alpha_j) \prod_{i=0}^{j-1} (1-\alpha_i)^2 \right) \\
&= \sum_{j=1}^{k-1} \left(\alpha_j (1-\alpha_j) \prod_{i=0}^{j-1} (1-\alpha_i)^2 \right) + \left(\alpha_k (1-\alpha_k) \prod_{i=0}^{k-1} (1-\alpha_i)^2 \right) \\
&\quad + \sum_{j=k+1}^{J} \left(\alpha_j (1-\alpha_j)(1-\alpha_k)^2 \prod_{\substack{i=0 \\ i\neq k}}^{j-1} (1-\alpha_i)^2 \right)
\end{aligned}$$

Setting $\alpha_k = 0$ in γ_2 and we have speed friction of the first profile γ_1:

$$\gamma_1 = \sum_{j=1}^{k-1} \left(\alpha_j (1-\alpha_j) \prod_{i=0}^{j-1} (1-\alpha_i)^2 \right) + \sum_{j=k+1}^{J} \left(\alpha_j (1-\alpha_j) \prod_{\substack{i=0 \\ i\neq k}}^{j-1} (1-\alpha_i)^2 \right)$$

Hence, the difference in speed frictions between the two profiles are:

$$
\begin{aligned}
\gamma_2 - \gamma_1 = &\left(\alpha_k(1-\alpha_k)\prod_{i=0}^{k-1}(1-\alpha_i)^2\right) \\
+\left((1-\alpha_k)^2-1\right)\prod_{i=0}^{k-1}(1-\alpha_i)^2 &\left(\alpha_{k+1}(1-\alpha_{k+1})+\sum_{j=k+2}^{J}\alpha_j(1-\alpha_j)\prod_{i=k+1}^{j-1}(1-\alpha_i)^2\right) \\
=\alpha_k\prod_{i=0}^{k-1}(1-\alpha_i)^2 &\left((1-\alpha_k)-(2-\alpha_k)\left(\alpha_{k+1}(1-\alpha_{k+1})+\sum_{j=k+2}^{J}\left(\alpha_j(1-\alpha_j)\prod_{i=k+1}^{j-1}(1-\alpha_i)^2\right)\right)\right)
\end{aligned}
$$

Since $\alpha_k \in (0,1)$, $\alpha_k \prod_{i=0}^{k-1}(1-\alpha_i)^2 > 0$. Hence, $\gamma_2 > \gamma_1 \iff$

$$
\frac{1-\alpha_k}{2-\alpha_k} > \alpha_{k+1}(1-\alpha_{k+1}) + \sum_{j=k+2}^{J}\left(\alpha_j(1-\alpha_j)\prod_{i=k+1}^{j-1}(1-\alpha_i)^2\right) \tag{A.3.2}
$$

Because $\alpha_j \in [0,1)$, $\alpha_j(1-\alpha_j) \leq \frac{1}{4}$. Thus,

$$
\text{RHS of Eq. (A.3.2)} \leq \frac{1}{4} + \frac{1}{4}\sum_{j=k+2}^{J}\prod_{i=k+1}^{j-1}(1-\alpha_i)^2
$$

Since all existing traders are risk neutral, $\alpha_j \geq \frac{1}{2}$ for all $j \geq k+1$. Thus,

$$
\sum_{j=k+2}^{J}\prod_{i=k+1}^{j-1}(1-\alpha_i)^2 \leq \sum_{j=k+2}^{J}\prod_{i=k+1}^{j-1}\frac{1}{4} = \frac{1}{3}\left(1-\frac{1}{4^{J-k-1}}\right)
$$

Hence,

$$
\text{RHS of Eq. (A.3.2)} \leq \frac{1}{4}\left(1+\frac{1}{3}\left(1-\frac{1}{4^{J-k-1}}\right)\right) = \frac{1}{3}\left(1-\frac{1}{4^{J-k}}\right) < \frac{1}{3}
$$

Because only *one* fast trader enters at time t_k, $\alpha_k \in (0, \frac{1}{2}]$. LHS of Eq. (A.3.2) is decreasing in α_k. It follows that $\frac{1-\alpha_k}{2-\alpha_k} \in [1/3, 1/2]$. Therefore,

$$
\text{LHS of Eq. (A.3.2)} = \frac{1-\alpha_k}{2-\alpha_k} \geq \frac{1}{3} > \text{RHS of Eq. (A.3.2)} \quad \iff \gamma_2 > \gamma_1
$$

Entry of *one* fast trader at a different time t_k increases α_k from 0 to at most $\frac{1}{2}$. The entry always increases speed friction. γ. $\square$

A.3.3 Proof of Proposition 1.4.5

1. Suppose there are J time instants with positive number of fast traders. $n_j \geq 1$, hence $\alpha_j \geq \frac{1}{2}$ for all $\forall 1 \leq j \leq J$ a since all fast traders are risk neutral. Lemma A.3.2 shows that $\frac{\partial \gamma}{\partial \alpha_j} < 0$ if $\alpha_j \geq \frac{1}{2}$. Thus,

$$\gamma = \sum_{j=1}^{J} \left(\alpha_j (1-\alpha_j) \prod_{i=0}^{j-1} (1-\alpha_i)^2 \right) \leq \sum_{j=1}^{J} \left(\frac{1}{2}(1-\frac{1}{2}) \prod_{i=0}^{j-1} (1-\frac{1}{2})^2 \right) = \sum_{j=1}^{J} \left(\frac{1}{4^j} \right)$$
$$= \frac{1}{3}\left(1 - \frac{1}{4^J}\right)$$

   $J \leq N$ since each fast trader can only arrive at one time. Hence, $\gamma \leq \frac{1}{3}\left(1-\frac{1}{4^J}\right) \leq \frac{1}{3}\left(1-\frac{1}{4^N}\right)$. If N fast traders follow the Stackelberg-N speed profile, $\alpha_j = \frac{1}{2}$ for all j and γ reaches the maximum which equals $\frac{1}{3}\left(1-\frac{1}{4^N}\right)$.

2. Starting from the speed profile $\{n_1, n_2, \cdots, n_J\}$, we can obtain the speed profile $\left\{\sum_{i=1}^{J} n_i\right\}$ by recursively consolidating the last two groups as follows:

$$\{n_1, \cdots, n_{J-1}, n_J\} \rightarrow \{n_1, \cdots, n_{J-2}, n_{J-1} + n_J\}$$
$$\rightarrow \left\{n_1, \cdots, n_{J-3}, \sum_{i=j-2}^{N} n_i\right\} \cdots \rightarrow \left\{n_1, \sum_{i=2}^{J} n_i\right\} \rightarrow \left\{\sum_{i=1}^{J} n_i\right\}$$

   Lemma A.3.3 implies that speed friction is reduced after each consolidation and thus the end Cournot-N speed profile $\left\{N = \sum_{i=1}^{J} n_i\right\}$ has lower speed friction than any starting profile.

A.3.4 Proof of Proposition 1.4.7

I prove sufficiency and necessity respectively.

1. Prove $\Rightarrow$. Assume $\lim_{N\to\infty}\gamma = 0$. By definition $\gamma \geq \alpha_1(1-\alpha_1) = \frac{n_1}{(1+n_1)^2}$. It follows that $0 = \lim_{N\to\infty}\gamma \geq \lim_{N\to\infty}\frac{n_1}{(1+n_1)^2} \geq 0$ and thus $\lim_{N\to\infty}\frac{n_1}{(1+n_1)^2} = 0$. We can then prove by contradiction that $\lim_{N\to\infty} n_1 = \infty$.

2. Prove $\Leftarrow$. Assume $\lim_{N\to\infty} n_1 = \infty$. By definition,

$$\begin{aligned}\lim_{N\to\infty}\gamma &= \lim_{N\to\infty}\left(\alpha_1(1-\alpha_1) + \alpha_2(1-\alpha_2)(1-\alpha_1)^2\right)\\ &\quad + \lim_{N\to\infty}\left((1-\alpha_1)^2\sum_{j=3}^{J}\alpha_j(1-\alpha_j)\prod_{i=2}^{j-1}(1-\alpha_i)^2\right)\end{aligned}$$

We have that $\lim_{N\to\infty} 1-\alpha_1 = \lim_{N\to\infty}\frac{1}{1+n_1} = 0$ and $\sum_{j=3}^{J}\alpha_j(1-\alpha_j)\prod_{i=2}^{j-1}(1-\alpha_i)^2 < 1$ due to Proposition A.3.1. Hence, $\lim_{N\to}\gamma = 0$.

A.4 Proofs of Section 1.5

A.4.1 Proof of Proposition 1.5.1

Proof. At time t_1, order flow u_1 reveals fast traders' information I_y perfectly.

$$\mathbf{V}\text{ar}\left[v|\mathcal{F}_{t_1}\right] = \mathbf{V}\text{ar}[v|u] = \mathbf{V}\text{ar}[v|I_y] = \frac{\sigma_z^2+\sigma_e^2}{\beta^2\sigma_v^2+\sigma_z^2+\sigma_e^2}\sigma_v^2 \tag{A.4.1}$$

$$\Rightarrow \phi_{t_1} = \frac{\beta^2\sigma_v^2}{\beta^2\sigma_v^2+\sigma_z^2+\sigma_e^2} = \frac{\beta^2\sigma_v^2}{\sigma_y^2+\frac{1-\rho}{\rho}\sigma_y^2} = \rho\frac{\beta^2\sigma_v^2}{\beta^2\sigma_v^2+\sigma_z^2} = \rho\frac{\theta^2}{1+\theta^2} = \rho\frac{1-\rho\gamma}{2} \tag{A.4.2}$$

Ex post, information is revealed through the order flow $y = x + z = \beta(v - v_0) + z$. The projection theorem of normal random variables implies that $\mathbf{V}\text{ar}[v|\mathcal{F}_{1^+}] = \mathbf{V}\text{ar}[v|y = \beta(v-v_0)+z] = \frac{\sigma_v^2\sigma_z^2}{\beta^2\sigma_v^2+\sigma_z^2} = \frac{\sigma_v^2}{1+\theta^2}$. Plug in the equilibrium θ and we find that $\phi = \frac{\theta^2}{1+\theta^2} = \frac{1-\rho\gamma}{2}$. □

A.4.2 Proof of Proposition 1.5.3

Proof. Noise traders' trade z shares at price of p_1. Their profit is

$$\mathbf{E}\left[\pi^N\right] = \mathbf{E}\left[-z(p_1 - p_0)\right] = -\lambda^T\sigma_z^2 = -\frac{1}{2\theta}\sigma_v\sigma_z.$$

Fast traders' total profit is given in Proposition 1.4.1. The informed trading profit equals the difference between noise traders' loss and fast traders' profit

$$\sigma_v\sigma_z\left(\frac{1}{2\theta} - \frac{1-\theta^2}{2\theta}\right) = \frac{\theta}{2}\sigma_v\sigma_z.$$

□

A.5 Proof of Proposition 1.6.1

Proof. Suppose trader 1 chooses order size u_1. In a symmetric equilibrium, all others have the same probability being placed at each position in the queue and each of them chooses to trade u. Conditional on $\hat{y}$ and others' choice of u, trader 1 has an equal probability of being at each position. Her expected profit is

$$\begin{aligned}\mathbf{E}\left[u_1(p_1 - p_{u_1})|\hat{y}\right] &= \frac{\lambda^T}{N}\left(u_1(\hat{y} - u_1) + u_1(\hat{y} - u - u_1) + \cdots + u_1(\hat{y} - (N-1)u - u_1)\right)\\ &= \lambda^T\left(\hat{y}u_1 - u_1^2 - \frac{N-1}{2}uu_1\right)\end{aligned}$$

Alternatively, $p_{u_1} = p_0 + \lambda^T u_1 + \lambda^T ku$ where k is the number of fast traders executed before trader 1 after the shuffling. k has equal probability of being $0, 1, \cdots, N-1$. Hence, $\mathbf{E}[p_{u_1}|u_1] = p_0 + \lambda^T u_1 + \lambda^T(N-1)u/2$. The first order condition implies that $u_1^* = (\hat{y} - (N-1)u/2)/2$. Due to symmetry, $u_1^* = u$ and we have that $u^* = 2\hat{y}/(N+3)$.

The aggregate profit of fast traders conditional on I_y is

$$\sum_{i=1}^{N} u^* \lambda^T (\hat{y} - iu^*) = \lambda^T \hat{y} \left(\hat{y} - \frac{N+1}{2} u^* \right) N u^* = \lambda^T \hat{y} \left(1 - \frac{N+1}{N+3} \right) \frac{2N}{N+3}$$

$$= \lambda^T \hat{y}^2 \frac{4N}{(N+3)^2} \tag{A.5.1}$$

Therefore, the effective speed friction is $\gamma = \frac{4N}{(N+3)^2}$. If $\gamma = 1$, $\gamma = 1/4$ and if $\gamma = 2$, $\gamma = 8/25$. □

A.6 Proof of Proposition 1.7.1

Proof. Lemma A.2.4 implies that the total fast traders' order flow S_J satisfies $\mathbf{E}[S_J] = 0$ and

$$\mathbf{V}\mathrm{ar}[S_J] = \left(1 - \prod_{i=0}^{J} (1 - \alpha_i) \right)^2 \rho^2 (\sigma_y^2 + \sigma_z^2)^2 = \left(1 - \prod_{i=0}^{J} (1 - \alpha_i) \right)^2 \rho \sigma_y^2$$

$$= \left(1 - \prod_{i=0}^{J} (1 - \alpha_i) \right)^2 \rho (\theta^2 + 1) \sigma_z^2$$

Because S_J is normally distributed, trading volume $|S_J|$ follows a half-normal distribution.

$$\mathbf{E}\left[|S_J| \right] = \sqrt{\mathbf{V}\mathrm{ar}[S_J]} \frac{\sqrt{2}}{\sqrt{\pi}} = \frac{\sqrt{2}}{\sqrt{\pi}} \sqrt{\rho \left(\theta^2 + 1 \right)} \sigma_z \left(1 - \prod_{i=0}^{J} (1 - \alpha_i) \right) \tag{A.6.1}$$

The informed and noise traders order flow is $y = x + z$. y is normally distributed with $\mathbf{E}[y] = 0$ and $\mathbf{V}\mathrm{ar}[y] = (1 + \theta^2)\sigma_z^2$. Hence, trading volume $|y|$ follows a half-normal distribution.

$$\mathbf{E}\left[|y| \right] = \sqrt{\frac{2}{\pi}} \sqrt{\mathbf{V}\mathrm{ar}[y]} = \sqrt{\frac{2}{\pi}} \sqrt{1 + \theta^2} \sigma_z \tag{A.6.2}$$

□

A.7 Fast traders collusion

HFTs are unlikely to collude because trading is anonymous. Nevertheless, the collusion equilibrium could be relevant because it is the upper bound of fast traders' profit. If fast traders with the same speed collude, in aggregate they would act like a monopoly as in the benchmark model and $\gamma = 1/4$. If fast traders with different speeds collude, they would make speed friction γ even higher than the Stackelberg case.

Proposition A.7.1 (**Fast traders' collusion**). *Suppose J fast traders collude and trade at different time instants. Then,*

$$\textit{Fast trader } j \textit{ trades: } u_j = \frac{\hat{y}}{J+1} = \alpha_j \left(\hat{y} - S_{j-1}\right), \ \textit{where } \alpha_j = \frac{1}{J-j+2}$$

for $j \in \{1, 2, \cdots, J\}$. The market quality parameter $\theta = \sqrt{\frac{1-\rho\gamma}{1+\rho\gamma}}$ and the speed friction $\gamma = \frac{J}{2(J+1)}$. In the limit when $J \to \infty$, $\gamma \to \frac{1}{2}$, $\theta \to \sqrt{\frac{1-\rho/2}{1+\rho/2}}$,

Proof. Fast traders' aggregate expected profit after observing the signal I_y can be found as follows:

$$\pi_j^F = u_j \left(p_1 - p_{t_j}\right) = u_j \lambda^T (y - S_{j-1} - u_j) \tag{A.7.1}$$

$$\mathbf{E}\left[\sum_{j=1}^{J} \pi_j^F \middle| I_y\right] = \lambda^T \sum_{j=1}^{J} u_j (\hat{y} - S_{j-1} - u_j) \tag{A.7.2}$$

Hence, the maximization problem for collusive fast traders is

$$\max_{u_i, 1 \le i \le J} \quad \sum_{i=1}^{J} u_i \left(\hat{y} - \sum_{j=1}^{i-1} u_j - u_i\right) \tag{A.7.3}$$

First order conditions imply that $\forall j$, $u_j = \frac{1}{2}\left(\hat{y} - \sum_{i \neq j} u_i\right)$. It follows that $\sum_i u_i = \frac{1}{2}\left(J\hat{y} - (J-1)\sum_i u_i\right)$ and

$$\begin{aligned} &\sum_i u_i = \frac{J}{J+1}\hat{y} \Rightarrow u_j = \frac{1}{2}\left(\hat{y} - \sum_i u_i + u_j\right) = \frac{1}{2}\left(\frac{1}{J+1}\hat{y} + u_j\right) \\ &\Rightarrow u_j^* = \frac{1}{J+1}\hat{y} \quad \forall j \end{aligned} \tag{A.7.4}$$

Fast traders' trade sizes are the same as in the Cournot competition. The difference is that they submit the orders sequentially and march up the supply curve gradually. We can easily verify that the strategy is equivalent to choosing $\alpha_j = \frac{1}{J-j+2}$. Hence, the speed friction of the collusive equilibrium is

$$\begin{aligned} \gamma &= \sum_{j=1}^{J}\left(\alpha_j(1-\alpha_j)\prod_{i=0}^{j-1}(1-\alpha_i)^2\right) = \sum_{j=1}^{J}\left(\frac{J-j+1}{(J-j+2)^2}\frac{(J+2-j)^2}{(J+1)^2}\right) \\ &= \sum_{j=1}^{J}\frac{J-j+1}{(J+1)^2} = \frac{J}{2(J+1)} \end{aligned} \tag{A.7.5}$$

The collusive speed friction $\gamma = \frac{J}{2(J+1)}$ is greater than the Stackelberg speed friction $\gamma = \frac{1}{3}\left(1 - 4^{-J}\right)$ and $\lim_{J\to\infty}\gamma = 1/2$ as illustrated in Figure A.2. □

Remark A.7.1. Each of the colluding fast trader trades $\frac{1}{1+J}$ of the total estimated order $\hat{y}$. Side payments from the slower traders to the faster ones are necessary to sustain the collusion because faster front-runners make less profits in trading when they collude with slower front-runners.

Remark A.7.2. J colluding fast traders with the same speed can at best mimic a monopoly fast trader, i.e., $\gamma = 1/4$. Allowing colluding fast traders to have different speeds resulting in a much higher speed friction.

Remark A.7.3. Interpreting the result differently, J colluding fast traders with different speeds is equivalent to a monopolistic fast trader who can trade J times in a trading round. If all fast traders can trade multiple times in a trading round, the resulting speed friction curve would lie between the collusive curve and the competition curve under different speeds in Figure A.2. Competition from slower fast traders improves market quality. Even when fast traders collude, the market quality would be much better if all fast traders have the same speed. The key result of this paper still holds: higher trading frequency (or finer time granularity) allows front-runners to extract more trading profits. Lowering the frequency of periodic uniform price auction can improve market quality because it limits the number of front-runners' trading opportunities and it also tends to encourage price competition among front-runners.

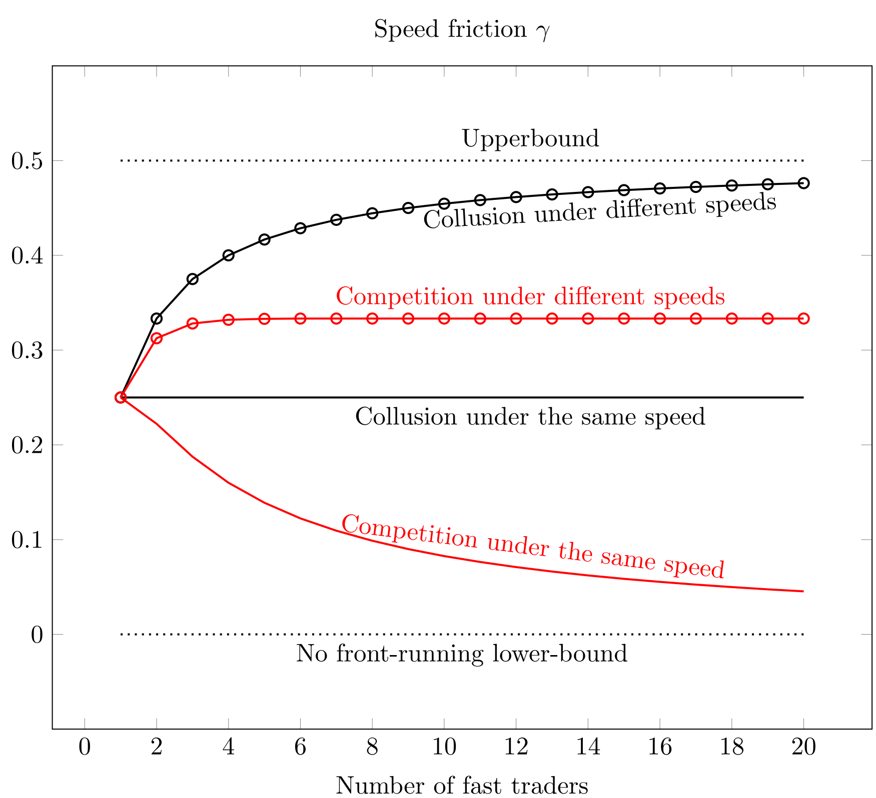

Figure A.2: Speed differences and competitions among fast traders

Chapter B: Appendices to Chapter 2

B.1 Proof of Theorem 2.2.1

The value function for the informed is

$$V(t, D_t) = \max_{\beta_t} \mathbf{E}_t^I \left[\int_t^T (v - p_t)\beta_t dt \right] = \max_{\beta_t} \mathbf{E}_t^I \left[\int_t^T (\hat{v} - p_t)\beta_t dt \right] \qquad \text{(B.1.1)}$$

where T is the random news announcement time. If the news has not been announced by time t, its probability of being announced in the $[t, t + dt)$ interval is $e^{-\eta dt}$. Thus, we can write

$$V(t) = \max_{\beta_t} \left\{ \mathbf{E}_t^I (v - p_t)\beta_t dt + e^{-\eta dt}\, \mathbf{E}_t^I\, V(t + dt) \right\} \qquad \text{(B.1.2)}$$

The Bellman's equation is

$$0 = \max \left\{ D_t \beta_t dt - \eta V_t dt + \mathbf{E}_t^I [dV_t] \right\} \qquad \text{(B.1.3)}$$

By definition,

$$dD_t = D_{t^+} - D_t = p_t - p_{t^-} = -dp_t = -\lambda_t(\beta_t dt + dz_t) \qquad \text{(B.1.4)}$$

Hence,

$$\mathbf{E}_t^I [dD_t] = \mathbf{E}_t^I [-dp_t] = -\lambda_t \beta_t dt \qquad \text{(B.1.5)}$$

$$dD_t dD_t = dp_t dp_t = \lambda_t^2 (dx_t + dz_t)^2 = \lambda_t^2 \sigma_z^2 dt \qquad \text{(B.1.6)}$$

From Ito's lemma,

$$\begin{aligned}
\mathbf{E}_t^I[dV_t] &= \frac{\partial V}{\partial t}dt - \frac{\partial V}{\partial D}\,\mathbf{E}_t^I[dp_t] + \frac{1}{2}\frac{\partial^2 V}{\partial D_t^2}(dD_t)^2 \\
&= \frac{\partial V}{\partial t}dt - \frac{\partial V}{\partial D}\lambda_t\beta_t dt + \frac{1}{2}\frac{\partial^2 V}{\partial D_t^2}\lambda_t^2\sigma_z^2 dt
\end{aligned} \tag{B.1.7}$$

Hence, the Bellman's equation implies

$$\begin{aligned}
0 &= \max_{\beta_t} \quad \left\{ D_t\beta_t - \eta V_t + \frac{\partial V}{\partial t} - \frac{\partial V}{\partial D}\lambda_t\beta_t + \frac{1}{2}\frac{\partial^2 V}{\partial D_t^2}\lambda_t^2\sigma_z^2 \right\} \\
\Rightarrow 0 &= \max_{\beta_t} \quad \left\{ \beta_t\left(D_t - \frac{\partial V}{\partial D}\lambda_t\right) + \frac{\partial V}{\partial t} + \frac{1}{2}\frac{\partial^2 V}{\partial D_t^2}\lambda_t^2\sigma_z^2 - \eta V_t \right\}
\end{aligned}$$

Because it's linear in β_t, the maximal exists and equals 0 only if

$$0 = D_t - \frac{\partial V}{\partial D}\lambda_t \tag{B.1.8}$$

$$0 = \frac{\partial V}{\partial t} + \frac{1}{2}\frac{\partial^2 V}{\partial D_t^2}\lambda_t^2\sigma_z^2 - \eta V_t \tag{B.1.9}$$

From equation (B.1.8),

$$\frac{\partial V}{\partial D} = \frac{D_t}{\lambda_t} \tag{B.1.10}$$

It implies that

$$\frac{\partial^2 V}{\partial D^2} = \frac{1}{\lambda_t}$$

Hence, equation (B.1.9) implies that

$$\frac{\partial V}{\partial t} = \eta V - \frac{\lambda_t}{2}\sigma_z^2 \tag{B.1.11}$$

Thus, we need to solve $V(t, D)$ such that

$$\frac{\partial V}{\partial D} = \frac{D}{\lambda_t} \tag{B.1.12}$$

$$\frac{\partial V}{\partial t} = \eta V - \frac{\lambda_t}{2}\sigma_z^2 \tag{B.1.13}$$

$$\frac{\partial^2 V}{\partial t \partial D} = D\left(-\frac{1}{\lambda_t^2}\frac{d\lambda_t}{dt}\right) = \eta\frac{\partial V}{\partial D} = \eta\frac{D}{\lambda_t} \tag{B.1.14}$$

Hence, we can find λ_t as follows

$$\frac{d\lambda_t}{dt} = -\eta\lambda_t \quad \Rightarrow \lambda_t = \lambda_0 e^{-\eta t} \tag{B.1.15}$$

Market makers must set an exponentially decaying price impact λ_t such that the informed trader's value function exists.

Then, we can find the value function in terms of λ_t. From equation (B.1.13), value function must have the form:

$$V(t, D) = e^{\eta t} Q(t, D) \tag{B.1.16}$$

where

$$\frac{\partial Q(t, D)}{\partial t} = -\frac{\lambda_t}{2e^{\eta t}}\sigma_z^2 \tag{B.1.17}$$

$$\Rightarrow Q(t, D) = \frac{\lambda_0 \sigma_z^2}{4\eta} e^{-2\eta t} + g(D) \tag{B.1.18}$$

In addition,

$$\frac{\partial V}{\partial D} = e^{\eta t}\frac{\partial Q}{\partial D} = \frac{D}{\lambda_t} \tag{B.1.19}$$

$$\Rightarrow g'(D) = \frac{D}{\lambda_0} \tag{B.1.20}$$

$$\Rightarrow g(D) = \frac{D^2}{2\lambda_0} + C \tag{B.1.21}$$

where C is a constant. Hence,

$$\begin{aligned} V(t, D) &= e^{\eta t} Q(t, D) \\ &= e^{\eta t}\left(\frac{\lambda_0 \sigma_z^2}{4\eta} e^{-2\eta t} + \frac{D^2}{2\lambda_0} + C\right) \\ &= \frac{D^2}{2\lambda_0 e^{-\eta t}} + \frac{\lambda_0 e^{-\eta t}}{4\eta}\sigma_z^2 + C e^{\eta t} \end{aligned} \tag{B.1.22}$$

$V(t,0)$ is the continuation value when $p_t = \mathbf{E}_I^t\, v$. It is generally not 0 because noise trading would push the price away from the correct value. If the news is going to be announced very soon, the continuation value must converge to 0.

$$0 = \lim_{\eta\to\infty} V(t,0) = \lim_{\eta\to\infty} Ce^{\eta t} \Rightarrow C = 0 \tag{B.1.23}$$

Therefore, value function can be expressed as a function of λ_t.

$$V(t, D_t) = \frac{D_t^2}{2\lambda_0 e^{-\eta t}} + \frac{\lambda_0 e^{-\eta t}}{4\eta}\sigma_z^2 = \frac{D_t^2}{2\lambda_t} + \frac{\lambda_t}{4\eta}\sigma_z^2 \tag{B.1.24}$$

We can find λ_0 from the 0 expected profit of the market makers. The expected loss of the noise trader is

$$\mathbf{E}_t^M\,[dp_t dz_t] = \mathbf{E}_t^M[\lambda_t(dx_t + dz_t)dz_t] = \lambda_t\sigma_z^2 dt \tag{B.1.25}$$

Hence, the total expected loss of the noise trader until the game ends at time T is

$$\begin{aligned}
\mathbf{E}_0^M\left[\int_0^T dp_t dz_t\right] &= \mathbf{E}_0^M\left[\int_0^T \lambda_t\sigma_z^2 dt\right] \\
&= \lambda_0\sigma_z^2\,\mathbf{E}_0^M\left[\int_0^T e^{-\eta t}dt\right] \\
&= \lambda_0\sigma_z^2\,\mathbf{E}_0^M\left[\frac{1-e^{-\eta T}}{\eta}\right] \\
&= \lambda_0\sigma_z^2\frac{1-\mathbf{E}_0^M[e^{-\eta T}]}{\eta} \\
&= \lambda_0\sigma_z^2\frac{1-\int_0^\infty e^{-\eta T}\eta e^{-\eta T}dT}{\eta} \\
&= \frac{1}{2\eta}\lambda_0\sigma_z^2
\end{aligned} \tag{B.1.26}$$

The expected gain for the informed trader at time 0 is

$$\mathbf{E}_0^M\,[V(0, D_0)] = \frac{1}{2\lambda_0}\,\mathbf{E}_0^M[D_0^2] + \frac{\lambda_0}{4\eta}\sigma_z^2 = \frac{\rho_i}{2\lambda_0}\sigma_v^2 + \frac{\lambda_0}{4\eta}\sigma_z^2 \tag{B.1.27}$$

Given the 0 profit condition,[1] we have the equation

$$\frac{1}{2\eta}\lambda_0\sigma_z^2 = \frac{\rho_i}{2\lambda_0}\sigma_v^2 + \frac{\lambda_0}{4\eta}\sigma_z^2 \Rightarrow \lambda_0 = \sqrt{\rho_i}\sqrt{2\eta}\frac{\sigma_v}{\sigma_z} \tag{B.1.28}$$

Therefore, value function is

$$V(t, D_t) = \frac{D_t^2}{2\lambda_0 e^{-\eta t}} + \frac{\lambda_0 e^{-\eta t}}{4\eta}\sigma_z^2 = \left(\frac{D_t^2}{\rho_i\sigma_v^2}e^{\eta t} + e^{-\eta t}\right)\frac{1}{2}\sqrt{\frac{\rho_i}{2\eta}}\sigma_v\sigma_z \tag{B.1.29}$$

Hence, for the market makers, the expected informed profit is

$$\mathbf{E}_0^M[V(0, D_0)] = \sqrt{\frac{\rho_i}{2\eta}}\sigma_v\sigma_z \tag{B.1.30}$$

Next, let's find the residual information asymmetry $\Sigma_t = \mathbf{V}\mathrm{ar}_t^M[\hat{v}]$. Due to orthogonality and equation (B.1.15),

$$\mathbf{V}\mathrm{ar}_0^M[\hat{v}] - \mathbf{V}\mathrm{ar}_T^M[\hat{v}] = \int_0^T \mathbf{V}\mathrm{ar}_t^M[dp_t] = \int_0^T \lambda_t^2\sigma_z^2 dt = \sigma_z^2\lambda_0^2\frac{1-e^{-2\eta T}}{2\eta}$$

Plug in λ_0 found in Equation (B.1.28),

$$\begin{aligned}
\mathbf{V}\mathrm{ar}_t^M[\hat{v}] &= \mathbf{V}\mathrm{ar}_0^M[\hat{v}] - \sigma_z^2\lambda_0^2\frac{1-e^{-2\eta t}}{2\eta} \\
&= \mathbf{V}\mathrm{ar}_0^M[\hat{v}] - \sigma_z^2\rho_i 2\eta\frac{\sigma_v^2}{\sigma_z^2}\frac{1-e^{-2\eta t}}{2\eta} \\
&= \sigma_v^2\rho_i e^{-2\eta t}
\end{aligned} \tag{B.1.31}$$

Therefore,

$$\Sigma_t = \Sigma_0 e^{-2\eta t} \tag{B.1.32}$$

$$\Sigma_0 = \sigma_v^2\rho_i \tag{B.1.33}$$

$$\frac{d\Sigma_t}{dt} = -2\eta \tag{B.1.34}$$

[1] Note here the zero profit condition is imposed on the [0,T] interval. The condition cannot be imposed for each dt interval because the informed trader's strategy is indeterminate. There are multiple β_t strategies that would reach the value function.

Finally, let's find the informed trader's trading intensity β_t. Note that the informed trader have multiple optimal strategy as in Kyle (1985) and Back (1992). We look for the equilibrium where the market maker's price updating rule is correct at each instant. The candidate strategy has the form

$$dx_t = \beta_t dt = \bar{\beta}_t(\hat{v} - p_t)dt$$

where $\hat{\beta}_t := \beta_t/(\hat{v} - p_t)$. Then, the order flow

$$dy_t = dx_t + dz_t = \bar{\beta}_t D_t dt + \sigma_z dB_t$$

Since $D_t = \hat{v} - p_t = \hat{v} - \mathbf{E}_t^M[\hat{v}]$, we have that

$$\mathbf{E}_t^M\left[dy_t \cdot \hat{v}\right] = \mathbf{E}_t^M\left[\bar{\beta}_t D_t dt \cdot \hat{v}\right] = \bar{\beta}_t \, \mathbf{V}\mathrm{ar}_t^M[\hat{v}]dt \tag{B.1.35}$$

$$\mathbf{E}_t^M\left[dy_t\right] \mathbf{E}_t^M\left[\hat{v}\right] = \mathbf{E}_t^M\left[\bar{\beta}_t D_t dt\right] \mathbf{E}_t^M[\hat{v}] = 0 \cdot p_t = 0 \tag{B.1.36}$$

Using the standard filtering theorem (see Back (2004) for an introduction),

$$\begin{aligned}
\mathbf{E}_t^M[dv] = \mathbf{E}_t^M[d\hat{v}]& \\
&= \frac{\mathbf{C}\mathrm{ov}_t^M(dy_t, \hat{v})}{\mathbf{V}\mathrm{ar}_t^M(dy_t)}(dy_t - \mathbf{E}_{t-}^M[dy_t]) \\
&= \left(\mathbf{E}_t^M\left[\bar{\beta}_t D_t dt \cdot \hat{v}\right] - \mathbf{E}_t^M\left[\bar{\beta}_t D_t dt\right] \mathbf{E}_t^M[\hat{v}]\right) \frac{1}{\mathbf{V}\mathrm{ar}_t[dy_t]}(dy_t - \mathbf{E}_{t-}^M[dy_t]) \\
&= \bar{\beta}_t \, \mathbf{V}\mathrm{ar}_t^M[\hat{v}] \frac{1}{\sigma_z^2} dy_t
\end{aligned} \tag{B.1.37}$$

Therefore,

$$dp_t = \lambda_t dy_t = \mathbf{E}_t^M[dv] = \frac{\bar{\beta}_t}{\sigma_z^2} \, \mathbf{V}\mathrm{ar}_t^M[\hat{v}] dy_t$$

$$\bar{\beta}_t = \frac{\lambda_t \sigma_z^2}{\mathbf{V}\mathrm{ar}_t^M[\hat{v}]} = \frac{\lambda_0 e^{-\eta t} \sigma_z^2}{\sigma_v^2 \rho_i e^{-2\eta t}} = \frac{2\eta}{\lambda_t} \tag{B.1.38}$$

The trading intensity can also be measured in

$$\beta = \frac{dx_t}{(D_t/\lambda_t)dt} = 2\eta \tag{B.1.39}$$

In sum, the equilibrium is

$$\text{price updating rule: } dp_t = \lambda_t dy_t = \sqrt{2\eta\rho_i}\frac{\sigma_v}{\sigma_z}e^{-\eta t}(dx_t + dz_t) \tag{B.1.40}$$

$$\text{informed trading size: } dx_t = 2\eta\frac{\hat{v} - p_t}{\lambda_t}dt = \sqrt{\frac{2\eta}{\rho_i}}\frac{\sigma_z}{\sigma_v}e^{\eta t}(\hat{v} - p_t)dt \tag{B.1.41}$$

$$\text{informed value function: } V(t, D_t) = \left(\frac{D_t^2}{\rho_i\sigma_v^2}e^{\eta t} + e^{-\eta t}\right)\frac{1}{2}\sqrt{\frac{\rho_i}{2\eta}}\sigma_v\sigma_z \tag{B.1.42}$$

$$\text{total expected information rent: } \mathbf{E}_0^M\left[V(0, D_0)\right] = \sqrt{\frac{\rho_i}{2\eta}}\sigma_v\sigma_z \tag{B.1.43}$$

$$\text{residual information: } \mathbf{V}\text{ar}_t^M[\hat{v}] = \sigma_v^2\rho_i e^{-2\eta t} \tag{B.1.44}$$

B.2 Proof of Theorem 2.2.5

Lemma B.2.1. *Given the conjectured equilibrium form specified in equations* (2.2.34) *and* (2.2.35),

$$\Sigma_n = \begin{cases} \rho_i\sigma_v^2 & n = 0 \\ (1 - \beta\Delta_t)^n\Sigma_0 & n \geq 1 \end{cases} \tag{B.2.1}$$

$$\lambda_n = \begin{cases} \sqrt{\rho_i}\sqrt{\beta}\frac{\sigma_v}{\sigma_z} & n = 0 \\ (1 - \beta\Delta_t)^{n/2}\lambda_0 & n \geq 1 \end{cases} \tag{B.2.2}$$

Σ_n *decays as fast as* λ_n^2 *at rate of*

$$\frac{\Sigma_n}{\Sigma_{n-1}} = \frac{\lambda_n^2}{\lambda_{n-1}^2} = 1 - \beta\Delta_t \in [0, 1], \quad n \geq 1 \tag{B.2.3}$$

Proof. Suppose that the informed trader follows a linear strategy

$$\Delta x_t = \beta \frac{D_n}{\lambda_n} \Delta_t \tag{B.2.4}$$

Then,

$$p_n = \mathbf{E}\left[\hat{v} \,|p_{n-1}, \Delta y_n = \Delta x_n + \Delta z_n\right] \tag{B.2.5}$$

$$= p_{n-1} + \mathbf{E}[\hat{v} - p_{n-1} | \Delta y_n = \Delta x_n + \Delta z_n, p_{n-1}] \tag{B.2.6}$$

$$= p_{n-1} + \mathbf{E}\left[D_n \left| \frac{\lambda_n}{\beta \Delta_t} \Delta y_n = D_n + \frac{\lambda_n}{\beta \Delta_t} \Delta z_n \right.\right] \tag{B.2.7}$$

$$= p_{n-1} + \frac{\Sigma_{n-1}}{\Sigma_{n-1} + \lambda_n^2 \sigma_z^2 \Delta_t / (\beta \Delta_t)^2} \frac{1}{\beta \Delta_t} \cdot \lambda_n \Delta y_n \tag{B.2.8}$$

Thus,

$$1 = \frac{\Sigma_{n-1}}{\Sigma_{n-1} + \lambda_n^2 \sigma_z^2 \Delta_t / (\beta \Delta_t)^2} \frac{1}{\beta \Delta_t} \tag{B.2.9}$$

Solve for λ_n^2 and we have

$$\lambda_n^2 \sigma_z^2 \Delta_t = \beta \Delta_t (1 - \beta \Delta_t) \Sigma_{n-1} \tag{B.2.10}$$

Because σ_z^2 and $\beta \Delta_t (1 - \beta \Delta_t)$ are time invariant, Σ_n is decreases at the same rate as λ_n^2.

In addition, Δp_n is orthogonal to p_{n-1}. Due to normality,

$$\Sigma_{n-1} - \Sigma_n = \mathbf{V}\mathrm{ar}_{n-1}^M[\Delta p_n] = \mathbf{V}\mathrm{ar}_{n-1}^M[\beta D_n \Delta_t + \lambda_n \Delta z_n] \tag{B.2.11}$$

$$= \beta^2 \Sigma_{n-1} (\Delta_t)^2 + \lambda_n^2 \sigma_z^2 \Delta_t \tag{B.2.12}$$

From equation (B.2.10), $\Sigma_{n-1} - \Sigma_n = \beta \Sigma_{n-1} \Delta_t$. Hence,

$$\frac{\Sigma_n}{\Sigma_{n-1}} = \frac{\lambda_n^2}{\lambda_{n-1}^2} = 1 - \beta \Delta_t \tag{B.2.13}$$

Equation (2.2.13) that $\hat{v} = \rho_i S$. Hence,

$$\Sigma_0 = \mathbf{V}\text{ar}_0^M[\hat{v}] = \mathbf{V}\text{ar}_0^M[\rho_i(v+e)] = \rho_i^2(\sigma_v^2 + \sigma_e^2) = \rho_i \sigma_v^2 \tag{B.2.14}$$

$$\Sigma_n = \Sigma_0(1-\beta\Delta_t)^n, \quad n \geq 1 \tag{B.2.15}$$

It follows that

$$\lambda_1^2 \sigma_z^2 \Delta_t = \beta\Delta_t(1-\beta\Delta_t)\Sigma_0 \tag{B.2.16}$$

$$\Rightarrow \lambda_0^2 = \frac{\beta\Delta_t(1-\beta\Delta_t)\Sigma_0}{\sigma_z^2(1-\beta\Delta_t)\Delta_t} = \frac{\beta\rho_i\sigma_v^2}{\sigma_z^2} \tag{B.2.17}$$

$$\Rightarrow \lambda_0 = \sqrt{\rho_i}\sqrt{\beta}\frac{\sigma_v}{\sigma_z} \tag{B.2.18}$$

$$\lambda_n = \lambda_0(1-\beta\Delta_t)^{n/2}, \quad n \geq 1 \tag{B.2.19}$$

□

Informed trader's expected profit from trading starting at time n is

$$V(n, D_n) = \max \quad \mathbf{E}_n^I \left[\sum_{i=n}^{\lfloor T/\Delta_t \rfloor} \Delta x_n (v - p_n) \right] = \max \quad \mathbf{E}_n^I \left[\sum_{i=n}^{\lfloor T/\Delta_t \rfloor} \Delta x_n (\hat{v} - p_n) \right]$$

The Bellman's equation of this dynamic programming problem is

$$\begin{aligned} &V(n, D_n) \\ &= \max_{\Delta x_n} \left\{ \mathbf{E}_n^I[\Delta x_n(\hat{v} - p_{n-1} - \lambda_n \Delta x_n - \lambda_n \Delta z_n)] + e^{-\eta\Delta_t}\, \mathbf{E}_n^I\left[V(n+1, D_{n+1})\right] \right\} \\ &= \max_{\Delta x_n} \left\{ D_n \Delta x_n - \lambda_n (\Delta x_n)^2 + e^{-\eta\Delta_t}\, \mathbf{E}_n^I[V(n+1, D_n - \lambda_n(\Delta x_n + \Delta z_n)] \right\} \end{aligned}$$

The first order condition is

$$0 = D_n - 2\lambda_n \Delta x_n - \lambda_n e^{-\eta\Delta_t}\, \mathbf{E}_n^I \left[\frac{\partial V}{\partial D_{n+1}} \right] \tag{B.2.20}$$

and the Envelope theorem implies that

$$\frac{\partial V}{\partial D_n} = \Delta x_n + e^{-\eta\Delta_t}\, \mathbf{E}_n^I \left[\frac{\partial V}{\partial D_{n+1}} \right] \tag{B.2.21}$$

Together, it follows that

$$0 = D_n - 2\lambda_n \Delta x_n - \lambda_n \left(\frac{\partial V}{\partial D_n} - \Delta x_n \right) \tag{B.2.22}$$

$$\Rightarrow \Delta x_n = \frac{D_n}{\lambda_n} - \frac{\partial V}{\partial D_n} \tag{B.2.23}$$

Given the conjectured form $\Delta x_n = \beta \frac{D_n}{\lambda_n} \Delta_t$. We can write that

$$\frac{\partial V}{\partial D_n} = (1 - \beta \Delta_t) \frac{D_n}{\lambda_n}$$

$$\Rightarrow V(n, D_n) = (1 - \beta \Delta_t) \frac{D_n^2}{2\lambda_n} + f(n) \tag{B.2.24}$$

where $\frac{\partial f(n)}{\partial D_n} = 0$. Then, from equation (B.2.21)

$$\begin{aligned}
\Delta x_n &= \frac{\partial V}{\partial D_n} - e^{-\eta \Delta_t} \, \mathbf{E}_n^I \left[\frac{\partial V}{\partial D_{n+1}} \right] \\
\Rightarrow \beta \Delta_t \frac{D_n}{\lambda_n} &= (1 - \beta \Delta_t) \frac{D_n}{\lambda_n} \\
&\quad - e^{-\eta \Delta_t} (1 - \beta \Delta_t) \, \mathbf{E}_n^I \left[\frac{D_n - \lambda_n \Delta x_n - \lambda_n \Delta z_n}{\lambda_{n+1}} \right] \\
\Rightarrow (1 - 2\beta \Delta_t) \frac{D_n}{\lambda_n} &= e^{-\eta \Delta_t} (1 - \beta \Delta_t) \frac{D_n (1 - \beta \Delta_t)}{\lambda_{n+1}} \\
\Rightarrow \frac{\lambda_{n+1}}{\lambda_n} &= \frac{e^{-\eta \Delta_t} (1 - \beta \Delta_t)^2}{1 - 2\beta \Delta_t}
\end{aligned}$$

From Lemma B.2.1, we also know that $\frac{\lambda_{n+1}}{\lambda_n} = \sqrt{1 - \beta \Delta_t}$. Therefore, the equilibrium informed trading intensity β solves the equation

$$\sqrt{1 - \beta \Delta_t} = \frac{e^{-\eta \Delta_t} (1 - \beta \Delta_t)^2}{1 - 2\beta \Delta_t}$$

$$\Rightarrow e^{-\eta \Delta_t} = \frac{1 - 2\beta \Delta_t}{(1 - \beta \Delta_t)^{3/2}} \tag{B.2.25}$$

As illustrated in Figure B.3, $\beta \Delta_t$ is a concave and increasing function of $\eta \Delta_t$. It can be easily proved using calculus.

Next, we need to find the second term of the value function $f(n)$. When $D_n = 0$, the informed trading at this instant $\Delta x_n = 0$ and $D_{n+1} = -\lambda_n \Delta z_n$. In addition, from equation (B.2.24),

$$
\begin{aligned}
\frac{f(n)}{e^{-\eta\Delta_t}} = \frac{V(n,0)}{e^{-\eta\Delta_t}} &= \mathbf{E}_n^I[V(n+1, -\lambda_n \Delta z_n)] \\
&= \mathbf{E}_n^I\left[(1-\beta\Delta_t)\frac{\lambda_n^2(\Delta z_n)^2}{2\lambda_{n+1}} + f(n+1)\right] \\
&= (1-\beta\Delta_t)\frac{\lambda_n^2\sigma_z^2\Delta_t}{2\lambda_{n+1}} + f(n+1) \\
&= \frac{\sigma_z^2\Delta_t\lambda_{n+1}}{2} + f(n+1)
\end{aligned}
$$

Multiply both side by $e^{-(n+1)\eta\Delta_t}$ and rearrange the terms.

$$
\begin{aligned}
e^{-(n+1)\eta\Delta_t} f(n+1) &= e^{-n\eta\Delta_t} f(n) - \frac{\sigma_z^2\Delta_t}{2}\lambda_0(1-\beta\Delta_t)^{\frac{n+1}{2}} e^{-(n+1)\eta\Delta_t} \\
&= e^{-n\eta\Delta_t} f(n) - \frac{\sigma_z^2\Delta_t}{2}\lambda_0\left(\frac{1-2\beta\Delta_t}{1-\beta\Delta_t}\right)^{n+1} \qquad \text{(B.2.26)}
\end{aligned}
$$

Take the summation of the above equation over n from 0 to $k-1$.

$$
\begin{aligned}
\sum_{n=0}^{k-1} e^{-(n+1)\eta\Delta_t} f(n+1) &= \sum_{n=0}^{k-1} e^{-n\eta\Delta_t} f(n) - \frac{\sigma_z^2\Delta_t\lambda_0}{2}\sum_{n=0}^{k-1}\left(\frac{1-2\beta\Delta_t}{1-\beta\Delta_t}\right)^{n+1} \\
\Rightarrow e^{-k\eta\Delta_t} f(k) &= f(0) - \frac{\sigma_z^2\Delta_t\lambda_0}{2}\sum_{n=0}^{k-1}\left(\frac{1-2\beta\Delta_t}{1-\beta\Delta_t}\right)^{n+1} \\
&= f(0) - \frac{\sigma_z^2\lambda_0\Delta_t}{2}\frac{\frac{1-2\beta\Delta_t}{1-\beta\Delta_t}}{1-\frac{1-2\beta\Delta_t}{1-\beta\Delta_t}}\left(1-\left(\frac{1-2\beta\Delta_t}{1-\beta\Delta_t}\right)^k\right) \\
&= f(0) - \frac{\sigma_z^2\lambda_0}{2\beta}(1-2\beta\Delta_t)\left(1-\left(\frac{1-2\beta\Delta_t}{1-\beta\Delta_t}\right)^k\right) \qquad \text{(B.2.27)}
\end{aligned}
$$

The transversality condition $\lim_{k\to\infty} e^{-k\eta\Delta_t} f(k) = 0$ implies that

$$
f(0) = \frac{\sigma_z^2\lambda_0}{2\beta}(1-2\beta\Delta_t) \qquad \text{(B.2.28)}
$$

Thus,

$$
f(k) = f(0)e^{k\eta\Delta_t}\left(\frac{1-2\beta\Delta_t}{1-\beta\Delta_t}\right)^k = f(0)(1-\beta\Delta_t)^{k/2} \qquad \text{(B.2.29)}
$$

Therefore,

$$V(n, D_n) = (1 - \beta\Delta_t)\frac{D_n^2}{2\lambda_n} + (1 - 2\beta\Delta_t)\frac{\sigma_z^2\lambda_n}{2\beta} \tag{B.2.30}$$

B.3 Proof of Theorem 2.2.6

Define $B = \beta\Delta_t \in [0, 1/2]$ and $G = 1 - e^{-\eta\Delta_t} \in [0, 1]$. We know from Theorem 2.2.5 that

$$\frac{1 - 2B}{(1 - B)^{3/2}} = 1 - G \tag{B.3.1}$$

Thus,

$$\lim_{\Delta_t \to 0} \frac{G}{B} = \lim_{B \to 0} \frac{1}{B}\left(1 - \frac{1 - 2B}{(1 - B)^{3/2}}\right) = \frac{1}{2} \tag{B.3.2}$$

$$\lim_{\Delta_t \to 0} \frac{G}{B} = \lim_{\Delta_t \to 0} \frac{1 - e^{-\eta\Delta_t}}{\beta\Delta_t} = \lim_{\Delta_t \to 0} \frac{e^{-\eta\Delta_t}\eta}{\beta} = \lim_{\Delta_t \to 0} \frac{\eta}{\beta} \tag{B.3.3}$$

Hence, $\lim_{\Delta_t \to 0} \beta = 2\eta$. Rest of the proof follows.

B.4 Proof of Proposition 2.2.7

The probability of public news announcement in the next Δ_t interval equals $1 - e^{-\eta\Delta_t}$. The function $g(\cdot)$ is implicitly defined by the equantion

$$\frac{1 - 2\beta\Delta_t}{(1 - \beta\Delta_t)^{3/2}} = 1 - (1 - e^{-\eta\Delta_t}) \tag{B.4.1}$$

such that $\beta\Delta_t = g(1 - e^{-\eta\Delta_t})$. It is easy to show that $g'(x) > 0$ and $g''(x) < 0$ for $x \in [0, 1]$ with calculus. Figure B.1 illustrates function $g(\cdot)$, which reflects the relationship between the trading intensity $\beta\Delta_t$ and probability of news announcement in the next Δ_t interval $1 - e^{-\eta\Delta_t}$. As $\eta\Delta_t \to \infty$, it's almost for sure the news is going to be announced before the next period. The trading game reduces to the one period game of Kyle (1985) where the informed trader trade away half of the information ($\beta\Delta_t = 1/2$) in one trade. When $\eta\Delta_t \to 0$, the probability of news announcement is minimal and the informed trade very patiently ($\beta\Delta_t \to 0$).

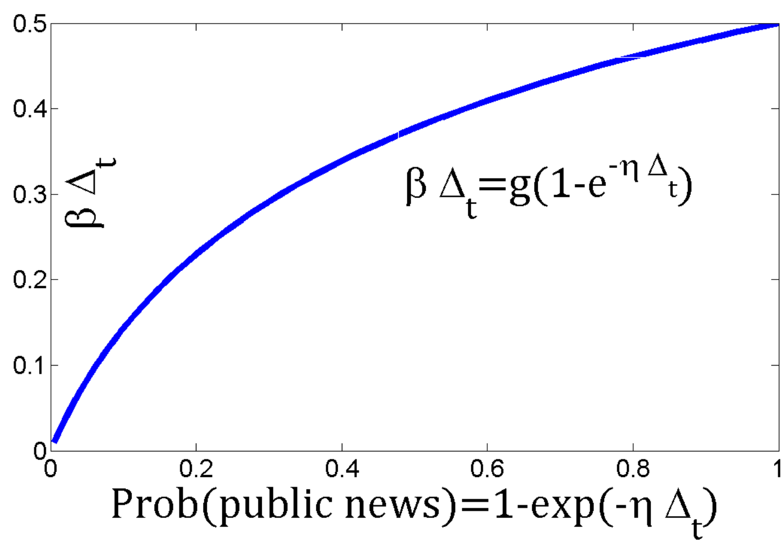

Figure B.1: $\beta\Delta_t = g(1 - e^{-\eta\Delta_t})$ where $g(\cdot) > 0$ and $g''(\cdot) < 0$.

In addition, $\frac{\partial}{\partial\eta}(1 - e^{-\eta\Delta_t}) = e^{-\eta\Delta_t}\Delta_t > 0$ and $\frac{\partial}{\partial\Delta_t}(1 - e^{-\eta\Delta_t}) = e^{-\eta\Delta_t}\eta > 0$. Hence, $\beta\Delta_t$ increases with η or Δ_t. Intuitively, when η or Δ_t is higher, it's more likely that the public news arrives before the next trading opportunity. Thus, the informed trades away more information when he still has the information advantage.

We also needs to find how public information arrival rate η and trading frequency $\frac{1}{\Delta_t}$ affect β. Figure B.3 illustrates the effects.

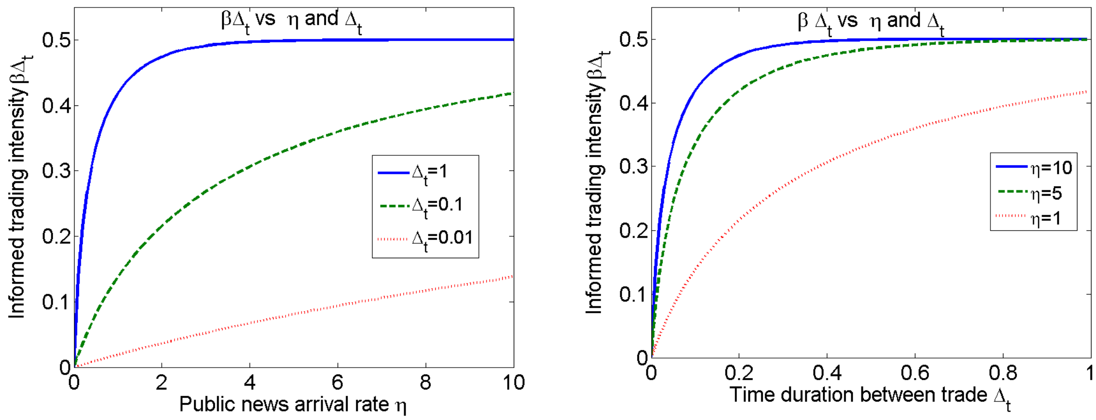

Figure B.2: $\beta\Delta_t$ increases with public information arrival rate η and time duration between trades Δ_t.

First, β increases with η because

$$\frac{\partial \beta}{\partial \eta} = \frac{g'(1-e^{-\eta\Delta_t})e^{-\eta\Delta_t}\Delta_t}{\Delta_t} = g'(1-e^{-\eta\Delta_t})e^{-\eta\Delta_t} > 0 \tag{B.4.2}$$

Second, β is higher when the trading frequency $\frac{1}{\Delta_t}$ is higher. Let $G = 1 - e^{-\eta\Delta_t}$. Then, $\beta = g(G)/\Delta_t$ and

$$\frac{\partial \beta}{\partial \Delta_t} = \frac{g'(G)\frac{\partial G}{\partial \Delta_t}\Delta_t - g(G)}{(\Delta_t)^2} = \frac{g'(G)e^{-\eta\Delta_t}\eta\Delta_t - g(G)}{(\Delta_t)^2} \tag{B.4.3}$$

It's easy to show with Taylor expansion that $e^{\eta\Delta_t} - 1 > \eta\Delta_t$. It follows that

$$e^{-\eta\Delta_t}\eta\Delta_t < e^{-\eta\Delta_t}(e^{\eta\Delta_t} - 1) < 1 - e^{-\eta\Delta_t} \tag{B.4.4}$$

Moreover, applying mean value theorem to $g(x)$ and use the concavity that $g''(\cdot) < 0$, we have that there exists for $\kappa \in [0, 1]$ such that

$$0 = g(0) = g(x) - xg'(x) + \frac{1}{2}(x)^2 g''(\kappa x) < g(x) - xg'(x) \tag{B.4.5}$$

$$\Rightarrow xg'(x) < g(x) \tag{B.4.6}$$

Therefore,

$$\frac{\partial \beta}{\partial \Delta_t} < \frac{1}{\Delta_t^2} \left(g'(G)G - g(G) \right) < 0 \tag{B.4.7}$$

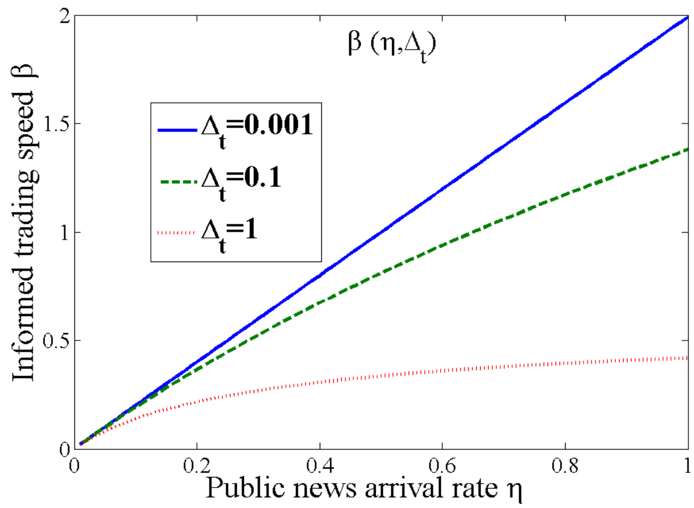

Figure B.3: $\beta = g(1 - \exp(-\eta\Delta_t))/\Delta_t$ increases with η and decreases with Δ_t.

B.5 Proof of Proposition 2.2.9

(*) It's obvious that $T^{\lambda}(\Delta_t) = 2T^{\Sigma}(\Delta_t)$ and thus we only prove it for T^{Σ}. Following the definition,

$$T^{\Sigma}(\Delta_t) = -\frac{\Delta_t \log 2}{\log(1 - \beta\Delta_t)}$$

Note that $\beta = g(\zeta\Delta_t)/\Delta_t$ is a function of Δ_t for a given ζ. Thus,

$$\frac{dT^{\Sigma}}{d\Delta_t} = \frac{-\log 2}{\log^2(1 - \beta\Delta_t)} \left(\log(1 - \beta\Delta_t) + \Delta_t \frac{\frac{\partial \beta}{\partial \Delta_t} + \beta}{1 - \beta\Delta_t} \right)$$

It's easy to check that $\log(1-\beta\Delta_t)+\frac{\beta\Delta_t}{1-\beta\Delta_t}$ is increasing in $\beta\Delta_t$ when $\beta\Delta_t \geq 0$ and $\log(1-\beta\Delta_t)+\frac{\beta\Delta_t}{1-\beta\Delta_t}=0$ when $\beta\Delta_t=0$. In addition, Proposition 2.2.7 implies that $\frac{\partial\beta}{\partial\Delta_t}>0$. Hence, when $\beta\Delta_t \geq 0$, $\left(\log(1-\beta\Delta_t)+\Delta_t\frac{\frac{\partial\beta}{\partial\Delta_t}+\beta}{1-\beta\Delta_t}\right)>0$ and $\frac{dT^\Sigma}{d\Delta_t}<0$. The continuous limit T^λ and T^Σ can be found immediately from 2.2.6 because $\lim_{\Delta_t\to 0}(1-\beta\Delta_t)=2\eta$. This result is illustrated in Figure B.4. The $T^\Sigma(\Delta_t)$ curve illustrates that the information efficiency of the trading game monotonically increases with trading frequency. As Δ_t goes from 1 to 0.01, the half life of Σ_t goes down from about 2.7 to about 1.7. Half of the private information is revealed to the market in 1.7 unit of time in the continuous time limit. Check Figure 2.4 for a different presentation of the result. The $T^\lambda(\Delta_t)$ curve illustrates that the illiquidity factor λ_t decays twice as fast as Σ_t. As Δ_t goes down, λ_t also decays faster.

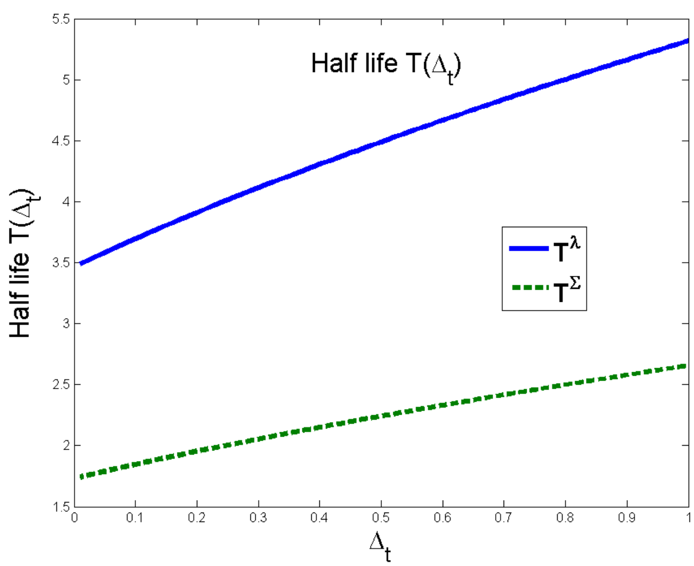

Figure B.4: Half life of information asymmetry T^Σ and price impact T^λ for different Δ_t when public information arrival rate $\eta=0.2$ and $\mathbf{E}[T]=1/\eta=5$.

B.6 Proof of Theorem 2.3.1

Assume the market makers set the end of period price at $p_n = p_{n-1} + \lambda_n(\Delta x_n + \Delta z_n)$ but the liquidity demanders get the prices $p'_n = p_{n-1} + (1+\xi)\lambda_n(\Delta x_n + \Delta z_n)$ where $\xi \in [0,1]$. Then,

Informed trader's expected profit from trading starting at time n is

$$V(n, D_n) = \max \quad \mathbf{E}_n^I \left[\sum_{i=n}^{T} \Delta x_n (v - p'_n) \right] = \max \quad \mathbf{E}_n^I \left[\sum_{i=n}^{T} \Delta x_n (\hat{v} - p'_n) \right]$$

The Bellman's equation of this dynamic programming problem is

$$\begin{aligned} V(n, D_n) = \max_{\Delta x_n} \Big\{ & \mathbf{E}_n^I [\Delta x_n (\hat{v} - p_{n-1} - (1+\xi)\lambda_n \Delta x_n - (1+\xi)\lambda_n \Delta z_n)] \\ & + e^{-\eta \Delta_t} \mathbf{E}_n^I \left[V(n+1, D_{n+1}) \right] \Big\} \\ = \max_{\Delta x_n} \Big\{ & D_n \Delta x_n - (1+\xi)\lambda_n (\Delta x_n)^2 \\ & + e^{-\eta \Delta_t} \mathbf{E}_n^I [V(n+1, D_n - \lambda_n(\Delta x_n + \Delta z_n))] \Big\} \end{aligned} \tag{B.6.1}$$

The first order condition is

$$0 = D_n - 2(1+\xi)\lambda_n \Delta x_n - \lambda_n e^{-\eta \Delta_t} \mathbf{E}_n^I \left[\frac{\partial V}{\partial D_{n+1}} \right] \tag{B.6.2}$$

and the Envelope theorem implies that

$$\frac{\partial V}{\partial D_n} = \Delta x_n + e^{-\eta \Delta_t} \mathbf{E}_n^I \left[\frac{\partial V}{\partial D_{n+1}} \right] \tag{B.6.3}$$

Then, it follows that

$$0 = D_n - 2(1+\xi)\lambda_n \Delta x_n - \lambda_n \left(\frac{\partial V}{\partial D_n} - \Delta x_n \right) \tag{B.6.4}$$

$$\Rightarrow \Delta x_n = \frac{1}{1+2\xi} \left(\frac{D_n}{\lambda_n} - \frac{\partial V}{\partial D_n} \right) \tag{B.6.5}$$

Given the conjectured form $\Delta x_n = \hat{\beta}\frac{D_n}{\lambda_n}\Delta_t$. We can write that

$$\frac{\partial V}{\partial D_n} = (1-(1+2\xi)\hat{\beta}\Delta_t)\frac{D_n}{\lambda_n} \tag{B.6.6}$$

$$\Rightarrow V(n, D_n) = (1-(1+2\xi)\hat{\beta}\Delta_t)\frac{D_n^2}{2\lambda_n} + f(n) \tag{B.6.7}$$

where $\frac{\partial f(n)}{\partial D_n} = 0$. Then, from equation (B.6.3)

$$\Delta x_n = \frac{\partial V}{\partial D_n} - e^{-\eta\Delta_t}\,\mathbf{E}_n^I\left[\frac{\partial V}{\partial D_{n+1}}\right] \tag{B.6.8}$$

$$\begin{aligned}\Rightarrow \hat{\beta}\Delta_t\frac{D_n}{\lambda_n} = (1-(1+2\xi)\hat{\beta}\Delta_t)\Bigg(&\frac{D_n}{\lambda_n}\\ &- e^{-\eta\Delta_t}\,\mathbf{E}_n^I\left[\frac{D_n - \lambda_n\Delta x_n - \lambda_n\Delta z_n}{\lambda_{n+1}}\right]\Bigg)\end{aligned} \tag{B.6.9}$$

$$\Rightarrow (1-2(1+\xi)\hat{\beta}\Delta_t)\frac{D_n}{\lambda_n} = e^{-\eta\Delta_t}(1-\hat{\beta}(1+2\xi)\Delta_t)\frac{D_n(1-\hat{\beta}\Delta_t)}{\lambda_{n+1}} \tag{B.6.10}$$

$$\Rightarrow \frac{\lambda_{n+1}}{\lambda_n} = \frac{e^{-\eta\Delta_t}(1-\hat{\beta}\Delta_t)(1-(1+2\xi)\hat{\beta}\Delta_t)}{1-2\hat{\beta}(1+\xi)\Delta_t} \tag{B.6.11}$$

From Lemma B.2.1, we also know that $\frac{\lambda_{n+1}}{\lambda_n} = \sqrt{1-\hat{\beta}\Delta_t}$. Therefore,

$$\sqrt{1-\hat{\beta}\Delta_t} = \frac{e^{-\eta\Delta_t}(1-\hat{\beta}\Delta_t)(1-\hat{\beta}(1+2\xi)\Delta_t)}{1-2(1+\xi)\hat{\beta}\Delta_t} \tag{B.6.12}$$

$$\Rightarrow e^{-\eta\Delta_t} = \frac{1-2(1+\xi)\hat{\beta}\Delta_t}{\sqrt{1-\hat{\beta}\Delta_t}(1-(1+2\xi)\cdot\hat{\beta}\Delta_t)} \tag{B.6.13}$$

Then, $\hat{\beta}\Delta_t \leq \frac{1}{2(1+\xi)}$ to ensure that the RHS is not negative. $\hat{\beta}\Delta_t$ is a concave and increasing function of $\eta\Delta_t$ as illustrated in Figure 2.5. Following the similar derivation as in Theorem 2.2.5, for $n \geq 0$,

$$\Sigma_n = \rho_i\sigma_v^2(1-\hat{\beta}\Delta_t)^n \tag{B.6.14}$$

$$\lambda_n = \sqrt{\hat{\beta}}\sqrt{\rho_i}\frac{\sigma_v}{\sigma_z}(1-\hat{\beta}\Delta_t)^{n/2} \tag{B.6.15}$$

$$\hat{\lambda}_n = (1+\xi)\lambda_n = (1+\xi)\sqrt{\hat{\beta}}\sqrt{\rho_i}\frac{\sigma_v}{\sigma_z}(1-\hat{\beta}\Delta_t)^{n/2} \tag{B.6.16}$$

$$\Delta x_n = \hat{\beta}\frac{D_n}{\lambda_n}\Delta_t \tag{B.6.17}$$

To fully determine the value function, we need to find $f(n)$ in Equation (B.6.7). When $D_n = 0$, $\Delta x_n = 0$. Thus,

$$\begin{aligned} f(n) = V(n,0) &= e^{-\eta\Delta_t}\,\mathbf{E}_n^I[V(n+1, -\lambda_n \Delta z_n)] \\ &= e^{-\eta\Delta_t}\,\mathbf{E}_n^I\left[(1-(1+2\xi)\hat{\beta}\Delta_t)\frac{(\lambda_n\Delta z_n)^2}{2\lambda_{n+1}} + f(n+1)\right] \\ &= e^{-\eta\Delta_t}\left((1-(1+2\xi)\hat{\beta}\Delta_t)\frac{\sigma_z^2\Delta_t\lambda_{n+1}\lambda_n^2}{2\lambda_{n+1}^2} + f(n+1)\right) \\ &= e^{-\eta\Delta_t}\left(\frac{1-(1+2\xi)\hat{\beta}\Delta_t}{1-\hat{\beta}\Delta_t}\frac{\sigma_z^2\Delta_t\lambda_{n+1}}{2} + f(n+1)\right) \end{aligned} \tag{B.6.18}$$

Multiply both sides by $e^{-n\eta\Delta_t}$, we have that

$$\begin{aligned} e^{-(n+1)\eta\Delta_t}f(n+1) &= e^{-n\eta\Delta_t}f(n) + e^{-(n+1)\eta\Delta_t}\frac{1-(1+2\xi)\hat{\beta}\Delta_t}{1-\hat{\beta}\Delta_t}\frac{\sigma_z^2\Delta_t\lambda_{n+1}}{2} \\ &= e^{-n\eta\Delta_t}f(n) + \frac{\lambda_0\sigma_z^2\Delta_t(1-2(1+\xi)\hat{\beta}\Delta_t)^{n+1}}{2(1-\hat{\beta}\Delta_t)(1-(1+2\xi)\cdot\hat{\beta}\Delta_t)^n} \end{aligned} \tag{B.6.19}$$

Take the summation over n from 0 to $k-1$. Then,

$$\begin{aligned} e^{-k\eta\Delta_t}f(k) &= f(0) - \frac{\lambda_0\sigma_z^2\Delta_t}{2}\frac{1-(1+2\xi)\cdot\hat{\beta}\Delta_t}{1-\hat{\beta}\Delta_t}\sum_{n=0}^{k-1}\left(\frac{1-2(1+\xi)\hat{\beta}\Delta_t}{1-(1+2\xi)\cdot\hat{\beta}\Delta_t}\right)^{n+1} \\ &= f(0) - \frac{\lambda_0\sigma_z^2\Delta_t}{2}\frac{1-(1+2\xi)\cdot\hat{\beta}\Delta_t}{1-\hat{\beta}\Delta_t}\left(\frac{1-2(1+\xi)\hat{\beta}\Delta_t}{1-(1+2\xi)\cdot\hat{\beta}\Delta_t}\right)\frac{1-\left(\frac{1-2(1+\xi)\hat{\beta}\Delta_t}{1-(1+2\xi)\cdot\hat{\beta}\Delta_t}\right)^k}{1-\left(\frac{1-2(1+\xi)\hat{\beta}\Delta_t}{1-(1+2\xi)\cdot\hat{\beta}\Delta_t}\right)} \\ &= f(0) - \frac{\lambda_0\sigma_z^2\Delta_t}{2}\frac{1-(1+2\xi)\cdot\hat{\beta}\Delta_t}{1-\hat{\beta}\Delta_t}\left(\frac{1-2(1+\xi)\hat{\beta}\Delta_t}{\hat{\beta}\Delta_t}\right) \\ &\quad\times\left(1-\left(\frac{1-2(1+\xi)\hat{\beta}\Delta_t}{1-(1+2\xi)\cdot\hat{\beta}\Delta_t}\right)^k\right) \end{aligned} \tag{B.6.20}$$

From the transversality condition that $\lim_{k\to\infty} e^{-k\eta\Delta_t}f(k) = 0$, we have that

$$\begin{aligned} f(0) &= \frac{\lambda_0\sigma_z^2\Delta_t}{2}\frac{1-(1+2\xi)\cdot\hat{\beta}\Delta_t}{1-\hat{\beta}\Delta_t}\left(\frac{1-2(1+\xi)\hat{\beta}\Delta_t}{\hat{\beta}\Delta_t}\right) \\ &= \frac{\lambda_0\sigma_z^2}{2\hat{\beta}}\frac{1-(1+2\xi)\hat{\beta}\Delta_t}{1-\hat{\beta}\Delta_t}\left(1-2(1+\xi)\hat{\beta}\Delta_t\right) \end{aligned} \tag{B.6.21}$$

Then,

$$
\begin{aligned}
f(k) &= e^{k\eta\Delta_t} f(0) \left(\frac{1 - 2(1+\xi)\hat{\beta}\Delta_t}{1 - (1+2\xi)\cdot\hat{\beta}\Delta_t} \right)^k \\
&= f(0)\left(1 - \hat{\beta}\Delta_t\right)^{k/2}
\end{aligned}
\tag{B.6.22}
$$

Therefore, the value function is

$$
\begin{aligned}
&V(n, D_n) \\
=&(1 - (1+2\xi)\hat{\beta}\Delta_t)\frac{D_n^2}{2\lambda_n} + \frac{\lambda_0\sigma_z^2}{2\hat{\beta}}\frac{1 - (1+2\xi)\hat{\beta}\Delta_t}{1 - \hat{\beta}\Delta_t} \\
&\times (1 - \hat{\beta}\Delta_t)^{n/2}\left(1 - 2(1+\xi)\hat{\beta}\Delta_t\right) \\
=&(1 - (1+2\xi)\hat{\beta}\Delta_t)\frac{D_n^2}{2\lambda_n} + \frac{\lambda_n\sigma_z^2}{2\hat{\beta}}\frac{1 - (1+2\xi)\hat{\beta}\Delta_t}{1 - \hat{\beta}\Delta_t}\left(1 - 2(1+\xi)\hat{\beta}\Delta_t\right) \\
=&(1 - (1+2\xi)\hat{\beta}\Delta_t)\left(\frac{D_n^2}{2\lambda_n} + \frac{\lambda_n\sigma_z^2}{2\hat{\beta}}\frac{1 - 2(1+\xi)\hat{\beta}\Delta_t}{1 - \hat{\beta}\Delta_t}\right)
\end{aligned}
\tag{B.6.23}
$$

B.7 Proof of Proposition 2.3.4

We can prove that $\lim_{\Delta_t\to 0}\hat{\beta} = 2\eta$ using the similar derivations as in Appendix B.3. Let $G = 1 - e^{-\eta\Delta_t}$ and $\hat{B} = \hat{\beta}\Delta_t$. Then,

$$
1 - G = \frac{1 - 2(1+\xi)\hat{B}}{\sqrt{1-\hat{B}}(1 - (1+2\xi)\hat{B})}
\tag{B.7.1}
$$

Thus,

$$
\begin{aligned}
\lim_{\Delta_t\to 0}\frac{G}{\hat{B}} &= \lim_{\hat{B}\to 0}\frac{1}{\hat{B}}\left(1 - \frac{1 - 2(1+\xi)\hat{B}}{\sqrt{1-\hat{B}}(1 - (1+2\xi)\hat{B})}\right) \\
&= \lim_{\hat{B}\to 0}\frac{\partial}{\partial\hat{B}}\left(\frac{2(1+\xi)\hat{B} - 1}{\sqrt{1-\hat{B}}(1 - (1+2\xi)\hat{B})}\right)
\end{aligned}
\tag{B.7.2}
$$

From calculus,

$$\begin{aligned}
&\frac{\partial}{\partial \hat{B}}\sqrt{1-\hat{B}}(1-(1+2\xi)\hat{B}) \\
&\qquad = \frac{-1}{2\sqrt{1-\hat{B}}}(1-(1+2\xi)\hat{B}) - \sqrt{1-\hat{B}}(1+2\xi) \\
&\qquad = \frac{-1}{2\sqrt{1-\hat{B}}}\Big((1-(1+2\xi)\hat{B} + 2(1+2\xi)(1-\hat{B})\Big) \\
&\qquad = \frac{-1}{2\sqrt{1-\hat{B}}}\Big((3+4\xi) - 3(1+2\xi)\hat{B}\Big) \qquad \text{(B.7.3)}
\end{aligned}$$

Thus,

$$\begin{aligned}
&\lim_{\Delta_t \to 0} \frac{G}{\hat{B}} = \\
&\lim_{\hat{B}\to 0} \frac{2(1+\xi)\sqrt{1-\hat{B}}(1-(1+2\xi)\hat{B}) + \frac{2(1+\xi)\hat{B}-1}{2\sqrt{1-\hat{B}}}\Big((3+4\xi) - 3(1+2\xi)\hat{B}\Big)}{\sqrt{1-\hat{B}}^2(1-(1+2\xi)\hat{B})^2} \\
&\quad = \lim_{\hat{B}\to 0} 2(1+\xi) - \frac{1}{2}(3+4\xi) \\
&\quad = \frac{1}{2} \qquad \text{(B.7.4)}
\end{aligned}$$

Also,

$$\begin{aligned}
\lim_{\Delta_t \to 0} \frac{G}{\hat{B}} &= \lim_{\Delta_t \to 0} \frac{1-e^{-\eta\Delta_t}}{\hat{\beta}\Delta_t} \\
&= \lim_{\Delta_t \to 0} \frac{1-e^{-\eta\Delta_t}}{\Delta_t} \frac{1}{\lim_{\Delta_t \to 0}\hat{\beta}} \\
&= \frac{\eta}{\lim_{\Delta_t \to 0}\hat{\beta}} \qquad \text{(B.7.5)}
\end{aligned}$$

Therefore, $\lim_{\Delta_t \to 0}\hat{\beta} = 2\eta$. Then, it is trivial to show that as $\Delta_t \to 0$, $\hat{\beta} \to 2\eta$ and the value function converges to

$$\lim_{\Delta_t \to 0} V(n, D_n) = \frac{D_n^2}{2\lambda_n} + \frac{\lambda_n \sigma_z^2}{4\eta} \qquad \text{(B.7.6)}$$

B.8 Information production

In this section, I find the optimal level of information production. The informed trader has to decide on the amount of information production ρ_i before the trading game starts. Thus, at the information production stage, he chooses a ρ_i to maximize the expected trading profit $V(0, D)$.

B.8.1 Informed trader's profits

The informed trader extract "information rent" by trading on his private information. The magnitude of the information rent is affected by the parameters of the trading game.

Note that β is determined by η and Δ_t and it does not depend on ρ_i. This does not mean the informed trader does not care about ρ_i. Even though the informed trader is risk neutral, the level of information production would change λ_0 through the market makers and thus change the expected trading profit.

Proposition B.8.1. *The expected trading profit at time* 0 *as*

$$\mathcal{U}(\rho_i; \eta, \Delta_t) = \mathbf{E}\left[V(0, D_0)\right] = \left(1 - \frac{3}{2}\beta\Delta_t\right)\frac{\sqrt{\rho_i}\sigma_v\sigma_z}{\sqrt{\beta}} \tag{B.8.1}$$

where $\beta = g(1 - e^{-\eta\Delta_t})/\Delta_t$ *is the unique solution to equation* (2.2.41) *such that* $\beta\Delta_t \in [0, 1/2]$*. In the continuous time limit when* $\Delta_t \to 0$*,*

$$\mathcal{U}(\rho_i; \eta) = \frac{\sqrt{\rho_i}\sigma_v\sigma_z}{\sqrt{2\eta}} \tag{B.8.2}$$

Proof. Before receiving the signal I_v, we have that

$$\mathbf{E}[D_0^2] = \mathbf{E}\left[(\rho_i I_v)^2\right] = \rho_i^2(\sigma_v^2 + \sigma_e^2) = \rho_i \sigma_v^2 \tag{B.8.3}$$

From Theorem 2.2.1, in the continuous time equilibrium,

$$\mathcal{U}(\rho_i;\eta) = \mathbf{E}\left[V(0, D_0)\right] = \left(\frac{\mathbf{E}[D_0^2]}{\rho_i\sigma_v^2} + 1\right)\frac{1}{2}\sqrt{\frac{\rho_i}{2\eta}}\sigma_v\sigma_z = \sqrt{\frac{\rho_i}{2\eta}}\sigma_v\sigma_z \tag{B.8.4}$$

From Theorem 2.2.5, in the discrete time equilibrium,

$$\begin{aligned}
\mathcal{U}(\rho_i;\eta,\Delta_t) &= \mathbf{E}\left[V(0, D_0)\right] \\
&= \frac{1-\beta\Delta_t}{2\lambda_0}\mathbf{E}[D_0^2] + (1-2\beta\Delta_t)\frac{\sigma_z^2}{2\beta}\lambda_0 \\
&= \frac{1-\beta\Delta_t}{2\sigma_v\sqrt{\rho_i\beta}/\sigma_z}\rho_i\sigma_v^2 + (1-2\beta\Delta_t)\frac{\sigma_z^2}{2\beta}\sigma_v\sqrt{\rho_i\beta}\frac{1}{\sigma_z} \\
&= (1-\beta\Delta_t)\frac{\sqrt{\rho_i}\sigma_v\sigma_z}{2\sqrt{\beta}} + (1-2\beta\Delta_t)\frac{\sqrt{\rho_i}\sigma_v\sigma_z}{2\sqrt{\beta}} \\
&= \frac{2-3\beta\Delta_t}{2\sqrt{\beta\Delta_t}}\cdot\sqrt{\Delta_t}\cdot\sqrt{\rho_i}\sigma_v\sigma_z
\end{aligned} \tag{B.8.5}$$

When $\Delta_t \to 0$, $\beta \to 2\eta$ and

$$\lim_{\Delta_t\to 0}\mathcal{U} = \lim_{\Delta_t\to 0}\left(1 - \frac{3}{2}\Delta_t\right)\sqrt{\frac{\rho_i}{2\eta}}\sigma_v\sigma_z \tag{B.8.6}$$

So $\mathcal{U}$ becomes linear in Δ_t with a slope of $-3/2$. □

Corollary B.8.2. *The informed trader's profit is increasing in the total fundamental uncertainty σ_v, the amount of his private information ρ_i, and noise trading σ_z.*

Proof. From Theorem 2.2.5, β is determined by η and Δ_t. Then, the corollary immediately follows from Proposition B.8.1. □

The initial information asymmetry $\mathbf{V}\text{ar}[\hat{v}] = \rho_i\sigma_v^2$ captures the information advantage of the informed trader and the volatility σ_z captures the volume of noise trading. The informed expects to make more profits when he has higher information advantage or there is more noise trading volume.

In addition to the above factors, the informed trader's profits are also affected by the expected lifetime of his private information $1/\eta$ and the frequency of the trading game $1/\Delta_t$.

Proposition B.8.3. *The informed trader's expected trading profit increases the expected lifetime of the private information $1/\eta$, i.e., $\frac{\partial \mathcal{U}}{\partial \eta} < 0$. In the continuous time limit $\Delta_t \to 0$,*

$$\frac{\mathcal{U}(\rho_i, \eta_1)}{\mathcal{U}(\rho_i, \eta_2)} = \frac{\sqrt{\eta_2}}{\sqrt{\eta_1}} \tag{B.8.7}$$

When probability of public news is low ($\eta\Delta_t \ll 1$), informed trader's expected trading profit increases with trading frequency ($1/\Delta_t$).[2]

Proof. * Follows from Proposition 2.2.7 and Proposition B.8.1. □

Given the same degree of information advantage $\rho_i\sigma_v^2$ and level of noise trading σ_z, the informed expects to make more profits when his private information has a longer lifetime or when the trading frequency is higher. Figure B.5 illustrates

[2] A definitive bound is yet to be found. Informed trader's profit could increase with Δ_t when $\eta\Delta_t$ is large. This result is due to the assumption that the first trade occurs at Δ_t and the amount of noise trading at the first trading opportunity is $\sigma_z\Delta_t$, which increases with Δ_t. For example, suppose $\eta\Delta_t \gg 10$ and the probability of trade is close to 0. The benefit of smoothing out trade becomes negligible because the probability of trading more than once is extremely low. Then the informed chooses $\beta\Delta_t \approx 1/2$ and his expected profits is proportional to $1/\sqrt{\beta}$, which increases with Δ_t and is mainly driven by the level of noise trading $\sigma_z\Delta_t$. Intuitively, when $\Delta_t \to 0$, $\beta \to 2\eta$ and the informed trader's profit is approximately proportional to $1 - \frac{3}{2}2\eta\Delta_t$, which increases with $1/\Delta_t$.

$\mathcal{U}(\eta, \Delta_t)$ as a function of η and Δ_t. Informed trader's profits are more sensitive to η than to Δ_t.

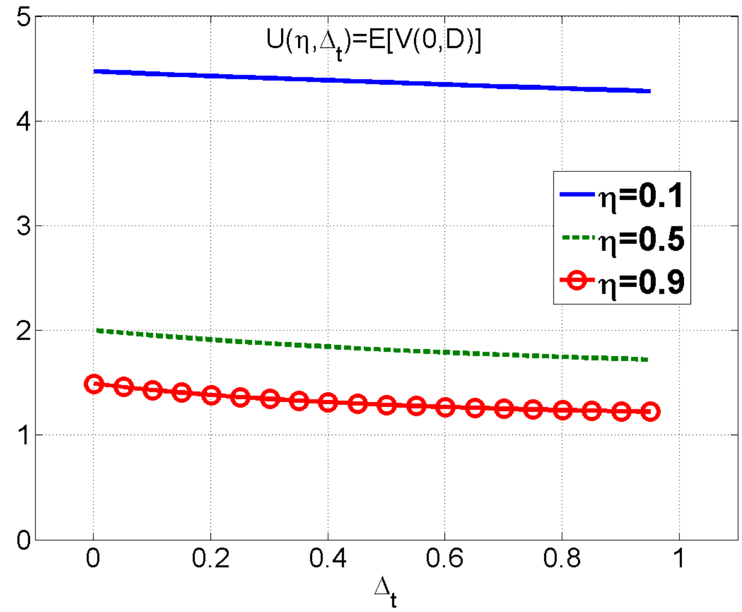

Figure B.5: Expected informed trading profit $\mathcal{U} = \mathbf{E}[V(0, D_0)]$ decreases with public information arrival speed η and time duration between trades Δ_t.

It shows that as everyone trades at a higher frequency, the ex ante expected trading profit for informed trader is higher, despite that his private information gets incorporated faster into the price and despite the higher initial market illiquidity.

Informed trader's expected trading profit increases with the expected lifetime of the private information $\mathbf{E}[T] = 1/\eta$. It seems to suggest that the informed trader would achieve infinite expected trading profit if he could produce private information that would never be revealed ($\mathbf{E}[T] \to \infty$). This unrealistic result is due to the assumption of zero discount rate. If the informed trader has a positive time discount rate, he would prefer to produce information that would be realized sooner than later so that his trading profit is discounted less. In fact, once the time discount is positive, the informed trader would prefer a private information whose speed of expiration matches his discount rate. A more patient trader would choose to produce longer lived information.

B.8.2 Optimal information production

Next, let's investigate the optimal level of information production ρ_i. To explore the effects of trading frequency $1/\Delta_t$, we need to specify the information production cost function $C(\rho_i, \omega)$.

Definition B.8.1. Aggregate expected profit combining the information production stage and the trading stage is defined

$$\mathcal{V}(\rho_i, \eta, \Delta_t, \omega) = \mathcal{U}(\rho_i, \eta, \Delta_t) - C(\rho_i, \omega), \quad \rho_i \in [0, 1] \tag{B.8.8}$$

for the discrete time model and

$$\mathcal{V}(\rho_i, \eta, \omega) = \mathcal{U}(\rho_i, \eta) - C(\rho_i, \omega), \quad \rho_i \in [0, 1] \tag{B.8.9}$$

for the continuous time model.

Definition B.8.2 (Information production equilibrium)**.** Equilibrium Information production ρ_i of the discrete time model satisfies

$$\mathcal{V}(\rho_i, \eta, \Delta_t, \omega) \geq \mathcal{V}(\rho_i', \eta, \Delta_t, \omega) \quad \forall \rho_i' \in [0, 1] \tag{B.8.10}$$

The equilibrium ρ_i is a function of ζ, Δ_t, and ω. For the continuous time model, ρ_i satisfies

$$\mathcal{V}(\rho_i, \eta, \omega) \geq \mathcal{V}(\rho_i', \eta, \omega) \quad \forall \rho_i' \in [0, 1] \tag{B.8.11}$$

The marginal expected trading profit of information is

$$\frac{\partial \mathcal{U}}{\partial \sqrt{\rho_i}} = \left(1 - \frac{3}{2}\beta\Delta_t\right)\frac{\sigma_v \sigma_z}{\sqrt{\beta}} \tag{B.8.12}$$

for the discrete time model and

$$\frac{\partial \mathcal{U}}{\partial \sqrt{\rho_i}} = \frac{\sigma_v \sigma_z}{\sqrt{2\eta}} \tag{B.8.13}$$

for the continuous time model. In addition, β is determined by η and Δ_t and not affected by ρ_i.

Assume that $C(\rho_i, \omega)$ is twice differentiable for $\rho_i \in (0,1)$ and for a parameter ω that I will specify later. Denote the marginal cost function $C_1(\rho_i, \omega) = \frac{\partial C(\rho_i,\omega)}{\partial \sqrt{\rho_i}}$, $C_2(\rho_i, \omega) = \frac{\partial C_1(\rho_i,\omega)}{\partial \sqrt{\rho_i}}$, and the marginal trading profit $\mathcal{U}_1 = \frac{\partial \mathcal{U}}{\partial \sqrt{\rho_i}}$.

Theorem B.8.4 (Equilibrium Information production ρ_i)**.** *The equilibrium ρ_i^* might be one of the following cases:*

1. *If $\lim_{\rho_i \to 0+} C_1(\rho_i, \omega) \geq \mathcal{U}_1(\rho_i, \eta, \Delta_t)$ and $C_2(\rho_i, \omega) \geq 0$, $\rho_i^* = 0$. It's optimal not to produce any information because the marginal cost of information is always higher than the marginal expected trading profit.*

2. *If $\lim_{\rho_i \to 1-} C_1(\rho_i, \omega) \leq \mathcal{U}_1(\rho_i, \zeta, \Delta_t)$ and $C_2(\rho_i, \omega) \geq 0$, $\rho_i^* = 1$. It's optimal to produce all information because the marginal cost of information is always lower than the marginal expected trading profit.*

3. *If $\lim_{\rho_i \to 0+} C_1(\rho_i, \omega) \leq \mathcal{U}_1(\rho_i, \zeta, \Delta_t) \leq \lim_{\rho_i \to 1-} C_1(\rho_i, \omega)$ and $C_2(\rho_i, \omega) \geq 0$, there exists an unique interior level of information production $\rho_i^* \in (0,1)$.*

4. *If $C_2(\rho_i, \omega) < 0$ for some ρ_i, there might exists multiple equivalent level of optimal ρ_i^* in $[0,1]$.*

The continuous time result is similar.

Proof. It follows directly from the first order and second order condition of $\mathcal{V}(\rho_i, \eta, \Delta_t, \omega)$.

$\square$

Corollary B.8.5. *Assume $C_2(\rho_i, \omega) \geq 0$, $C_1(0, \omega) = 0$ and $C_1(1, \omega) = \infty$. Under these assumptions, equilibrium information production $\rho_i^* \in (0, 1)$ is unique. Then,*

1. *If the noise trading does not affect the marginal cost of information production ($\frac{\partial C_1}{\partial \sigma_z} = 0$), the informed trader produces more information when there is more noise trading then $\frac{\partial \rho_i^*}{\partial \sigma_z} > 0$.*

2. *If the total uncertain about v does not affect the marginal cost of information production ($\frac{\partial C_1}{\partial \sigma_v} = 0$), the informed trader produces more information when there is more fundamental uncertainty $\frac{\partial \rho_i^*}{\partial \sigma_v} > 0$.*

3. *If information production becomes cheaper without affecting trading frequency $1/\Delta_t$ and public information arrival rate ζ, then the informed trader produces more information.*

4. *If trading frequency $1/\Delta_t$ does not affect the marginal cost of information production $\frac{\partial C_1}{\partial \Delta_t} = 0$, then the informed trader produces more information when trading frequency increases $\frac{\partial \rho_i^*}{\partial \Delta_t} < 0$.*

5. *If public information arrival rate η does not affect the marginal cost of information production ($\frac{\partial C_1}{\partial \eta} = 0$), then the informed trader produces more information as the expected life of the private information $\frac{1}{\eta}$ increases ($\frac{\partial \rho_i^*}{\partial \eta} < 0$).*

Proof. The first three cases are straightforward. The fourth and fifth cases follow from Lemma B.8.3 noticing that $\frac{\partial^2 \mathcal{U}}{\partial \rho_i \partial \Delta_t}$ has the same sign as $\frac{\partial \mathcal{U}}{\partial \Delta_t}$ and $\frac{\partial^2 \mathcal{U}}{\partial \rho_i \partial \eta}$ has the same sign as $\frac{\partial \mathcal{U}}{\partial \eta}$. □

Higher frequency induces the informed trader to produce more private information and especially short-lived information. When trading frequency is high, the informed would have more trading opportunity to exploit a short-lived private information.

For a given public information arrival rate η, a higher trading frequency ($\frac{1}{\Delta_t} \uparrow$) induces the informed trader to produce more information ($\rho_i \uparrow$) and to trade more aggressively ($\beta \downarrow$). Thus, more fundamental information is produced and such information is revealed to the market through trading faster. The informativeness of the market price unambiguously increases with trading frequency $\frac{1}{\Delta_t}$.

Bibliography

Angel, James J., Lawrence E. Harris, and Chester S. Spatt, "Equity Trading in the 21st Century," *The Quarterly Journal of Finance*, 2011, *1* (01), 1–53.

Back, Kerry, "Insider Trading in Continuous Time," *Review of Financial Studies*, 1992, *5* (3), 387–409.

________, "Incomplete and Asymmetric Information in Asset Pricing Theory," *Stochastic Methods in Finance*, 2004, pp. 263–268.

________ **and Shmuel Baruch**, "Information in Securities Markets: Kyle meets Glosten and Milgrom," *Econometrica*, 2004, *72* (2), 433–465.

________**, C. Henry Cao, and Gregory A. Willard**, "Imperfect Competition among Informed Traders," *Journal of Finance*, 2000, *55* (5), 2117–2155.

Baron, Matthew, Jonathan Brogaard, and Andrei Kirilenko, "The Trading Profits of High Frequency Traders," *NBER Working Paper*, 2012.

Baruch, Shmuel and Lawrence R. Glosten, "Flickering Quotes," *Working Paper*, 2013.

Bernhardt, Dan and Bart Taub, "Front-running Dynamics," *Journal of Economic Theory*, 2008, *138* (1), 288–296.

——— **and Eric Hughson**, "Splitting Orders," *Review of Financial Studies*, 1997, *10* (1), 69–101.

——— **and Jianjun Miao**, "Informed Trading When Information Becomes Stale," *Journal of Finance*, 2004, *59* (1), 339–390.

Biais, Bruno, Thierry Foucault, and Sophie Moinas, "Equilibrium Fast Trading," *Working Paper*, 2013.

Bikhchandani, Sushil, David Hirshleifer, and Ivo Welch, "Learning from the Behavior of Others: Conformity, Fads, and Informational Cascades," *The Journal of Economic Perspectives*, 1998, *12* (3), 151–170.

Boehmer, Ekkehart, Kingsley Fong, and Julie Wu, "International Evidence on Algorithmic Trading," *Available at SSRN 2022034*, 2012.

Breckenfelder, Johannes, "Competition between High-Frequency Traders and Market Quality," *Available at SSRN 2264858*, 2013.

Brogaard, Jonathan, Björn Hagströmer, Lars Norden, and Ryan Riordan, "Trading Fast and Slow: Colocation and Market Quality," *Available at SSRN 2317797*, 2013.

———**, Terrence Hendershott, and Ryan Riordan**, "High Frequency Trading and Price Discovery," *Available at SSRN: 1928510*, 2013.

Brunnermeier, Markus K., "Information Leakage and Market Efficiency," *Review of Financial Studies*, 2005, *18* (2), 417–457.

________ **and Lasse Heje Pedersen**, "Predatory Trading," *Journal of Finance*, August 2005, *60* (4), 1825–1863.

Budish, Eric, Peter Cramton, and John Shim, "The High-Frequency Trading Arms Race: Frequent Batch Auctions as a Market Design Response," *Working Paper*, 2013.

Carlin, Bruce Ian, Miguel Sousa Lobo, and S. Viswanathan, "Episodic Liquidity Crises: Cooperative and Predatory Trading," *Journal of Finance*, October 2007, *62* (5), 2235–2274.

Carrion, Allen, "Very Fast Money: High-frequency Trading on the NASDAQ," *Journal of Financial Markets*, 2013, *16* (4).

Chaboud, Alain, Ben Chiquoine, Erik Hjalmarsson, and Clara Vega, "Rise of the Machines: Algorithmic Trading in the Foreign Exchange Market," *Journal of Finance*, 2013, (Forthcoming).

Chau, Minh and Dimitri Vayanos, "Strong-form Efficiency with Monopolistic Insiders," *Review of Financial Studies*, 2008, *21* (5), 2275–2306.

Clark-Joseph, Adam D., "Exploratory Trading," *Harvard University Working Paper*, 2013.

Cohen, Samuel N. and Lukasz Szpruch, "A Limit Order Book Model for Latency Arbitrage," *Mathematics and Financial Economics*, 2012, pp. 1–17.

Cvitanic, Jaksa and Andrei Kirilenko, "High Frequency Traders and Asset Prices," *Available at SSRN 1569075*, 2010.

Egginton, Jared, Bonnie Van Ness, and Robert Van Ness, "Quote Stuffing," *Available at SSRN 1958281*, 2013.

Foster, F. Douglas and S. Viswanathan, "Strategic Trading when Agents Forecast the Forecasts of Others," *Journal of Finance*, 1996, pp. 1437–1478.

Foucault, Thierry, Johan Hombert, and Ioanid Rosu, "News Trading and Speed," *Working Paper*, 2012.

Frino, Alex, Vito Mollica, and Robert I Webb, "The Impact of Co-location of Securities Exchanges' and Traders' Computer Servers on Market Liquidity," *Journal of Futures Markets*, 2013.

Froot, Kenneth A., David S. Scharfstein, and Jeremy C. Stein, "Herd on the Street: Informational Inefficiencies in a Market with Short-term Speculation," *Journal of Finance*, 1992, pp. 1461–1484.

Gai, Jiading, Chen Yao, and Mao Ye, "The Externality of High Frequency Trading," *Working Paper*, 2012.

Gerig, Austin and David Michayluk, "Automated Liquidity Provision and the Demise of Traditional Market Making," *Available at SSRN 1639954*, 2010.

Glosten, Lawrence R. and Paul R. Milgrom, "Bid, Ask and Transaction Prices in a Specialist Market with Heterogeneously Informed Traders," *Journal of Financial Economics*, 1985, *14* (1), 71 – 100.

Goldman, M. Barry and Howard B. Sosin, "Information Dissemination, Market Efficiency and the Frequency of Transactions," *Journal of Financial Economics*, 1979, *7* (1), 29 – 61.

Guo, Ming and Albert S. Kyle, "Intertemporal Strategic Trading, Risk-averse Passive Investors, and Liquidity Dynamics," *Working Paper*, 2009.

Hagströmer, Björn and Lars Nordén, "The Diversity of High-frequency Traders," *Journal of Financial Markets*, 2013, *16* (4).

Hasbrouck, Joel and Gideon Saar, "Technology and Liquidity Provision: the Blurring of Traditional Definitions," *Journal of Financial Markets*, 2009, *12* (2), 143–172.

——— **and** ———, "Low-latency Trading," *Journal of Financial Markets*, 2013, *16* (4).

Hendershott, Terrence and Pamela C. Moulton, "Automation, Speed, and Stock Market Quality: the NYSE's Hybrid," *Journal of Financial Markets*, 2011, *14* (4), 568–604.

——— **and Ryan Riordan**, "Algorithmic Trading and the Market for Liquidity," *Journal of Financial and Quantitative Analysis, Forthcoming*, 2012.

———, **Charles M. Jones, and Albert J. Menkveld**, "Does Algorithmic Trading Improve Liquidity?," *Journal of Finance*, 2011, *66* (1), 1–33.

Hirschey, Nicholas H., "Do High-Frequency Traders Anticipate Buying and Selling Pressure?," *Working Paper*, 2013.

Hirshleifer, David, Avanidhar Subrahmanyam, and Sheridan Titman, "Security Analysis and Trading Patterns when Some Investors Receive Information before Others," *Journal of Finance*, 1994, *49* (5), 1665–1698.

Hirshleifer, Jack, "The Private and Social Value of Information and the Reward to Inventive Activity," *American Economic Review*, 1971, *61* (4), 561–574.

Hoffmann, Peter, "A Dynamic Limit Order Market with Fast and Slow traders," *Working Paper, Available at SSRN 1969392*, 2012.

Holden, Craig W. and Avanidhar Subrahmanyam, "Long-lived Private Information and Imperfect Competition," *Journal of Finance*, 1992, pp. 247–270.

Jarrow, Robert A. and Philip Protter, "A Dysfunctional Role of High Frequency Trading in Electronic Markets," *International Journal of Theoretical and Applied Finance*, 2012, *15* (03).

Jiang, George J, Ingrid Lo, and Giorgio Valente, "High Frequency Trading in the US Treasury Market: Evidence Around Macroeconomic News Announcement," *Working Paper*, 2013.

Jovanovic, Boyan and Albert J. Menkveld, "Middlemen in Limit-order Markets," *Working Paper*, 2012.

Kirilenko, Andrei, Albert S. Kyle, Mehrdad Samadi, and Tugkan Tuzun, "The Flash Crash: the Impact of High Frequency Trading on an Electronic Market," *University of Maryland Working Paper*, 2011.

Kyle, Albert S., "Continuous Auctions and Insider Trading," *Econometrica*, November 1985, *53* (6), 1315–1336.

________, **Anna A. Obizhaeva, and Yajun Wang**, "Smooth Trading with Overconfidence and Market Power," *University of Maryland Working Paper*, 2013.

Laughlin, Gregory, Anthony Aguirre, and Joseph Grundfest, "Information Transmission Between Financial Markets in Chicago and New York," *ArXiv e-prints*, February 2013.

Li, Su, "Imperfect Competition and Long-lived Flow of Private Information," *Working Paper*, 2013.

Litzenberger, Robert, Jeff Castura, and Richard Gorelick, "The Impacts of Automation and High Frequency Trading on Market Quality," *Annual Review of Financial Economics*, 2012, *4*, 59–98.

Liu, Hong and Yajun Wang, "Asymmetric Information, Endogenous Illiquidity, and Asset Pricing With Imperfect Competition," *Working Paper*, 2010.

Malinova, Katya, Andreas Park, and Ryan Riordan, "Do Retail Traders Suffer from High Frequency Traders?," *Available at SSRN 2183806*, 2012.

Martinez, Victor and Ioanid Rosu, "High Frequency Traders, News and Volatility," *Working Paper*, 2013.

Menkveld, Albert J., "High Frequency Trading and the New Market Makers," *Journal of Financial Markets*, 2013, *16* (4), 712–740.

Pagano, Marco and Ailsa Röell, "Front-running and Stock Market Liquidity," *Financial Markets' Liberalisation and the Role of Banks*, 1993, *178.*

Pagnotta, Emiliano and Thomas Philippon, "Competing on Speed," *Working Paper*, 2012.

Penalva, José and Álvaro Cartea, "Where is the Value in High Frequency Trading?," *Quarterly Journal of Finance*, 2012, *02* (03), 1250014.

SEC, "Concept Release on Equity Market Structure," *Release No. 34-61358; File No. S7-02-10.*, 2010.

Subrahmanyam, Avanidhar, "Transaction Taxes and Financial Market Equilibrium," *The Journal of Business*, 1998, *71* (1), 81–118.

Tong, Lin, "A Blessing or a Curse? The Impact of High Frequency Trading on Institutional Investors," *Available at SSRN 2330053*, 2013.

Vayanos, Dimitri, "Strategic Trading and Welfare in a Dynamic Market," *The Review of Economic Studies*, 1999, *66* (2), 219–254.

——— **and Jiang Wang**, "Market Liquidity — Theory and Empirical Evidence," *NBER Working Paper*, 2012.

Weller, Brian, "Fast and Slow Liquidity," *Working Paper*, 2013.

Yueshen, Bart Zhou, "Queuing Uncertainty," *Working Paper*, 2013.

Zhang, S. Sarah, "Need for Speed: An Empirical Analysis of Hard and Soft Information in a High Frequency World," *Working Paper, Karlsruhe Institute of Technology*, 2012.