High-Frequency Trading, Endogenous Capital Commitment and Market Quality

Yenan Wang*

Jan. 2021 **Latest Version**

Abstract

I study market quality implications of the competition between traditional market making and high-frequency trading. A long-run traditional market maker responds to the competition from high-frequency traders by reducing both the spread and the amount of capital committed in market making. While a lower spread level is beneficial, less capital commitment deteriorates market quality. Specifically, the market's capacity to satisfy large demand is impaired. My model provides a more comprehensive illustration of high-frequency trading's implications on market quality by integrating both price and quantity effects. I further use this framework to analyze implications of different high-frequency trading regulatory measures.

Key words: High-Frequency Trading, Capital Commitment, Market Quality

JEL Code: G14 G18

*Wang: Fuqua School of Business, Duke University, 100 Fuqua Drive, Durham, NC 27708, yenan.wang@duke.edu. The auther thanks Ravi Bansal, Bruno Biais(Discussant), Briana Chang, Simon Gervais, Pete Kyle, Mark Loewenstein, Song Ma, Claudia Moise, Narayan Naik, Adriano Rampini, Todd Sarver, Batchimeg Sambalaibat, John Shim (Discussant), Felipe Varas, S. Viswanathan, Yajun Wang, Pierre-Olivier Weill, Brian Weller, Randall Wright, Ming Yang, Hongda Zhong, participants of SaMMF Workshop for Job Market Candidates, SFS Cavalcade 2020, Fuqua Brownbag and Fuqua finance Theory Student Group for helpful comments and discussions. All remaining errors are mine.

1 Introduction

Over the past decade, high-frequency trading has become increasingly prevalent worldwide. According to O'Hara (2015), high-frequency traders (henceforth HFTs) contribute more than half of market trading volume. This growing trend of high-frequency trading has led to a policy debate over proper regulatory measures to adapt to this change. Clearly, policy makers have yet to reach a consensus over this issue as different countries are implementing regulations with opposing intended effects.[1] Most European countries have carried out strict rules to reduce high-frequency trading and "level the playing field" while some Asian countries such as Japan and Singapore embrace high-frequency trading by providing systematic support including introducing co-location service and rebating high-frequency trades.

Extant empirical research has documented that the presence of HFTs leads to lower spreads in the market. Some papers take this as direct evidence that high-frequency trading improves market quality.[2] There are essentially two rationales behind this claim. First, lower spreads indicate less information asymmetry. Second, lower spreads enhance market efficiency by facilitating assets moving to agents with higher valuations.

However, an implicit market clearing assumption lies behind the second claim. That is, at each instant, the asset price is determined by a centralized planner, who receives all market participants' supply and demand schedules, to clear the market. Although it is a reasonable assumption for analyzing the long-run behavior of the market, it is a strong assumption in modeling high-frequency trading for two reasons. First, since trading happens very fast, it is unlikely that each market participant has time to submit a sequence of limit orders to form a demand or supply schedule in each trade. Second, even if there is a planner with all the information, the price may not be adjusted quickly enough to clear the market at each instant. Without the market clearing assumption, the one to one link between price and quantity breaks; i.e., a lower spread level no longer indicates a larger trading volume. Specifically, fac-

[1] For a comprehensive survey of the global high-frequency trading regulation environment, see Bell and Searles (2014)

[2] See Hendershott, Jones, and Menkveld (2011), Boehmer, Fong, and Wu (2018), Brogaard, Hendershott, and Riordan (2014), Hendershott, Jones, and Menkveld (2011), Boehmer, Fong, and Wu (2018), Hendershott and Riordan (2013), Hasbrouck and Saar (2013), Brogaard, Hagströmer, Nordén, and Riordan (2015), Conrad, Wahal, and Xiang (2015) and Conrad and Wahal (2018), among others.

ing competition from HFTs, a market maker might reduce his capacity in absorbing market imbalance as well as the spread since market making becomes less profitable. On the other hand, HFTs' abilities to provide liquidity are constrained by market conditions and might be insufficient to fill the gap left by tte market maker. The decrease of market making capacity would lead to lower trading volume and deteriorate market quality. Indeed, Chordia, Roll, and Subrahmanyam (2011), O'Hara, Yao, and Ye (2014) and Korajczyk and Murphy (2019a) show that the average order size becomes smaller and investors have difficulties executing large orders.

I consider a model where the market maker and the HFT compete to sell shares to a potential buyer in each period.[3] For clarity, I use female pronouns for the HFT and male pronouns for the market maker and the buyer. The market maker contracts with the exchange to provide liquidity and is obliged to post quotes in the market.[4] As a firm, the market maker can either commit his capital in market making, i.e, buying shares from an inter-dealer market for sale, or paying out dividend to investors.[5] The amount of capital committed in market making is endogenously determined by equalizing the marginal value of market making and the marginal value of paying dividend. When no HFT exists, the market maker is modeled as a monopolist due to the market power he enjoys from advantageous terms provided by the exchange.[6] Under this circumstance, making the market is highly profitable and the market maker commits a large amount of capital in market making.

In my model, the HFT enters the market with exogenous probability and holding. This means to capture the reality that the HFT makes profit by anticipating the arrival of future orders.[7] If the HFT detects a buying order, she tries to quickly buy cheaper shares from other channels and sell to the buyer at a slightly higher price. This way of operation makes the HFT's presence and the amount of shares supplied highly depend on exogenous market conditions. The competition from the HFT affects the market maker's pricing and capital commitment decisions. The market maker

[3]This model bears similarities to Kreps and Scheinkman (1983).

[4]In practice, the market maker in my model can be considered as a designated market maker in NYSE or a specialist in NASDAQ.

[5]Alternatively, I can assume that there is a risk-free asset with unlimited supply the market maker can invest in.

[6]This differs from the competitive market making assumption in Kyle (1985) and other models in market micro-structure.

[7]There are certainly other types of high-frequency traders. For instance, some market makers nowadays adopt advanced technology for trader. In my model, they fall into the market maker category.

may tighten the spread to compete with the HFT. This reduces buyers' transaction costs and improves market quality. On the other hand, market making becomes less attractive because of the competition and the market maker would reduce his capital commitment in market making. This weakens the market's capacity to satisfy large demands and effectively leads to a shallower market.[8]

I first consider the setting where the HFT possesses superior trading technology relative to the market maker. It enables the HFT to observe both the market maker's capital commitment and spread before making her pricing decision. In other words, the market maker and the HFT set spreads *sequentially*. The market maker faces a trade-off. If the market maker sets a high spread aiming to achieve a high expected payoff when the HFT does not enter, upon entering, the HFT would undercut and the market maker would only receive the residual demand. If the market maker sets a low spread, he sacrifices some profit when the HFT does not enter. Yet a low spread protects the market maker from the HFT's undercut. In the steady state, the market maker posts a high (low) spread if the HFT's entry probability is low (high). In other words, competition from the HFT has a positive price effect on market quality but reduces the return of market making. Thus, the market maker's steady state capital commitment is (weakly) decreasing in the HFT's entry probability. This deteriorates market quality. I use liquidity, the expected shares sold to the buyer, as a proxy of market quality to measure the aggregate effect of high-frequency trading. Importantly, under mild assumptions, liquidity is not changing monotonically with respect to the HFT's entry probability. This lack of monotonicity has two implications. First, using linear regression to analyze high-frequency trading's market and welfare effects may lead to erroneous conclusions. Second, past observations on high-frequency trading's effects on financial markets may not be sufficient to guide policy making, which would change the market condition faced by HFTs dramatically.

I then analyze the setting where the market maker and the HFT's trading technologies are "head to head". The HFT and the market maker in this setting set spreads *simultaneously*. This corresponds to realistic situations with high-frequency market making or limitations on maximum trading speed. In the equilibrium, the market maker and the HFT both use mixed pricing strategies. The market maker's expected

[8]In my model, this corresponds to the buyer leaving the market with a smaller portion of his fulfilled. In practice, this may corresponds to the buyer purchasing a large portion of shares from other liquidity providers at a higher price.

payoffs are the same setting spreads sequentially and simultaneously. However, the HFT's expected payoff is (weakly) lower when submitting spreads simultaneously.[9] Specifically, when the HFT's entry probability is low, the HFT has incentive to acquire superior trading technology at a low cost but the market maker would have no incentive to match the technology level. This is detrimental to market quality.

In both settings, two regimes of equilibrium (the wide spread region and the tight spread region) exist depending on the HFT entry probability. In the wide spread region with low HFT entry probability, the market maker sets a high spread and his capital commitment is decreasing in HFT entry probability. An increase in HFT entry in this region has ambiguous effects on market quality since it increases liquidity supplied by the HFT but decreases liquidity supplied by the market maker. In the tight spread region where the HFT entry probability is high, the market maker sets a low spread and his capital commitment is not changing in the HFT's entry probability. Thus, more HFT entry leads to better market quality in this region. Moreover, in the wide spread region, equalizing trading technologies of the market maker and the HFT improves market quality. This is because by switching from sequential pricing to simultaneous pricing, the average spread becomes lower while the market maker's capital commitment remains the same.

My model differs from the existing theory in two ways.[10] First, I explicitly consider the market maker's capital commitment decision, which has critical implications for market quality. Second, liquidity suppliers in my model face asymmetric constraints. Specifically, the market maker has an affirmative obligation to provide liquidity and faces a trade off between committing capital in market making and paying dividend. On the contrary, the HFT's entry and the amount of liquidity supplied (extensive and intensive margin) depend on exogenous market conditions. Although market making is profitable for the HFT, these constraints limit the HFT's ability to fill the gap when the market maker commits less capital. Contrary to conventional wisdom, competition does not necessarily lead to better markets when there is asymmetry among liquidity suppliers.

I further consider three extensions. In the first extension, I endogenize the HFT's entry probability by imposing a fixed high-frequency trading participation cost. The

[9]This is in line with the evidence in Baron, Brogaard, Hagströmer, and Kirilenko (2018) that faster HFTs achieve higher payoffs.

[10]For examples, see Goettler, Parlour, and Rajan (2009), Budish, Cramton, and Shim (2015), Biais, Foucault, and Moinas (2015) and Foucault, Hombert, and Roşu (2016), etc.

HFT needs to pay the cost to enter the market with an exogenous probability.[11] If the cost is high, the exogenous entry probability is low or the market is competitive, the HFT would rationally not participate in high-frequency trading. The market maker in this extension enjoys an additional strategic advantage. He can make the market competitive by setting a low spread to deter the HFT from participating in trading. This deterring spread is increasing with the HFT's participation cost. The equilibrium outcome depends on the magnitude of the participation cost. When the participation cost is low, market quality is the same as in the baseline model since deterring the HFT's participation is too costly for the market maker. Conversely, with a high cost, the HFT may not participate in high-frequency trading and the market maker's spread and capital commitment increase with the participation cost and eventually converge to the monopolistic levels. The overall effect of the participation cost on market quality is ambiguous. Yet it is certain that high participation cost harms the market.

The second extension considers flipping. That is, the HFT can purchase shares from the market maker and re-supply them at a higher spread. With high HFT entry probability, the market maker sets a low spread to induce flipping. When the HFT flips shares, market quality appears to be good since the expected trading volume is high and the average spread is low. However, these indicators are not characterizing market quality faithfully under this situation for two reasons. First, most of the cheaper shares are purchased by the HFT rather than the liquidity buyer. Second, the trading volume is "double-counted". The actual volume sold to the buyer is much lower. This extension demonstrates the importance of separating trades between liquidity suppliers and trades from liquidity suppliers to other investors to avoid over-estimating market quality.

In the third extension, the market maker can post a supply schedule to sell shares at different spreads.[12] With no HFT, the market maker chooses to sell all shares at the monopolistic spread. However, facing competition from the HFT, the market maker would sell shares at a continuum of spreads. I characterize conditions that determine the market maker's pricing strategy and capital commitment at the steady state and discuss implications for market quality. Furthermore, this extension illustrates how

[11] For example, EU's trading tax on both executed and canceled orders can be considered as a cost of this type.

[12] In the baseline model, I assume the market maker has to sell all shares at one spread.

competition between the market maker and the HFT determines the shape of limit order book.

My model contributes to the theoretical literature on high-frequency trading by exploring how high-frequency trading affects market quality via the capital commitment channel. Competition from the HFT leads the market maker to commit less capital in market making. This effect dampens the benefit brought by pricing competition, and, if large enough, the presence of a potential HFT might even deteriorate market quality. Ait-Sahalia and Saglam (2017a) and Han, Khapko, and Kyle (2014) also consider market quality implications with competition between the HFT and the market maker. However, in these papers, the size of orders is fixed. This assumption constrains these models' abilities to capture how capital commitment of the market maker affects market quality.[13] In my model, it is possible that a market with wide spread has better quality than a market with tight spread. The reason is that in the latter market, the market maker commits much less capital in market making.

The implications of my model are consistent with the following empirical findings in the literature: (1) High-frequency trading leads to lower average spreads in the market; (2) the average trade size becomes smaller; (3) market makers commit less capital in market making; (4) Large orders might face higher trading costs with the presence of HFTs; (5) market quality improves when all market participants have similar trading speeds. My model also provides several insights for future studies. First, the price information alone does not provide a complete characterization of market quality. The volume information is equally important. Second, market quality may not change monotonically with increasing HFT presence. In this sense, we cannot only rely on linear regression for accurate welfare implications of high-frequency trading. Third, when the HFT can flip orders, it is important to differentiate trades between liquidity providers and trades from liquidity providers to other investors. Otherwise, the data cannot faithfully reflect market quality since HFTs would exploit most of the cheaper orders with superior trading technology.

This paper also generates important insights for HFT regulations. The model suggests that if high-frequency trading is prevalent in the market, encouraging high-frequency trading benefits liquidity. On the other hand, when high-frequency trading

[13]In Ait-Sahalia and Saglam (2017a), the HFT, as a long run market maker, also holds inventory. However, since the supply is fixed to one, the inventory does not have a quantity effect. Instead, it has a price effect due to the inventory aversion assumption.

is less prevalent, more HFT's presence drives out the market maker's capital and has ambiguous effects on market quality. Second, when the HFT's entry probability is low, equalizing the trading speeds of the HFT and the market maker improves market quality. When the HFT's entry probability is high, it benefits mid-valuation buyers yet hurts low-valuation buyers. I also analyze implications of high-frequency trading participation cost. A low participation cost does not affect the market quality while a high participation cost increases market maker's capital commitment but also drives up the spread. The aggregate effect is ambiguous.

Finally, my paper complements the literature of limit order book formation by illustrating the effect of asymmetric competition between the market maker and the HFT over limit order book shape.[14] Specifically, with no HFT, the market maker would sell all shares at the monopolistic spread. Facing the competition from the HFT, the market maker sets a non-degenerate limit order book to avoid the HFT's undercutting. My model predicts a downward sloping limit order book at lower spreads and a large volume supplied at the monopolistic spread.

The rest of the paper is organized as follows. Section 2 reviews related literature. Section 3 presents baseline models. Section 4 analyzes baseline models. Section 5 considers the costly participation extension. Section 6 uses results developed in Sections 3, 4 and 5 to discuss market quality implications of various high-frequency trading regulations. Section 7 considers the flipping extension. Section 8 considers the extension in which the market maker can submit supply schedules. Section 9 concludes.

2 Related Literature

2.1 HFT Behavior

An existing theory literature analyzes how high-frequency trading effects market quality from the information perspective.[15] Han, Khapko, and Kyle (2014) demonstrate how adverse selection problem arising from fast order cancellation leads to wide

[14]For other papers on limit order book formation, see Glosten (1994), Chakravarty and Holden (1995), Seppi (1997), Biais, Martimort, and Rochet (2000), Viswanathan and Wang (2002) Parlour and Seppi (2003), Foucault, Kadan, and Kandel (2005), Roşu (2009), Back and Baruch (2013), Baruch and Glosten (2019), etc.

[15]For a comprehensive survey, see Menkveld (2016).

spreads when the HFT enters the market with probability between 0 and 1. Budish, Cramton, and Shim (2015) show how mechanical arbitrage in high-frequency time horizon hurts liquidity and propose frequent batch auctions mechanism as a solution. Biais, Foucault, and Moinas (2015) endogenize investment decisions on fast trading technology and show that equilibrium investment level on fast trading is higher than the social optimal level because high-frequency trading has a negative externality. Van Kervel (2015) analyzes the link between high-frequency trading and order cancellations across trading venues. Foucault, Hombert, and Roşu (2016) analyzes news trading by fast speculators and its implications for trading volume and asset price. Ait-Sahalia and Saglam (2017a) and Ait-Sahalia and Saglam (2017b) analyze high-frequency market making and show that the faster market maker provides more liquidity. Baldauf and Mollner (2019) consider the informational implication of high-frequency trading and conclude that the bid-ask spread narrows yet the information production also diminishes. Budish, Lee, and Shim (2019) consider a model where exchanges capture economic rents by selling speed technologies to discuss exchanges' incentives to adopt new market designs. Li, Wang, and Ye (2020) model competition between slower execution algorithms and high-frequency traders featuring the implication of the tick size. My model differs from the existing literature by explicitly considering the market maker's capital commitment decision facing competition from HFT and its implications for market quality.

Many empirical papers test high-frequency trading's impact on liquidity. Research generally documents an increase in liquidity with high-frequency trading. For instance, Hendershott, Jones, and Menkveld (2011), Hendershott and Riordan (2013), Hasbrouck and Saar (2013), Conrad, Wahal, and Xiang (2015) and Conrad and Wahal (2018),[16] using spread as a proxy for liquidity, conclude that liquidity is improved by high-frequency trading. Brogaard, Hendershott, and Riordan (2014), using order flow data, conclude that HFT is liquidity improving around macroeconomic news since liquidity supply is greater than liquidity demand. Boehmer, Fong, and Wu (2018) using execution shortfalls as a proxy, reach the similar conclusion. My model does not contradict these evidences. However, it does suggest that some important quantity aspects of market quality cannot be captured by these proxies. Specifically,

[16]Hasbrouck and Saar (2013) also examines number of shares displayed on the order book as a proxy for depth. One concern is that since HFTs can cancel orders with fast speed, this *NearDepth* might not able to capture real market depth.

spread measures might not capture the quantity information related to the market maker's capital commitment. The execution shortfall can better capture the price change facing large demand. Yet even the execution shortfall does not incorporate information about unexecuted and canceled orders. Moreover, order flow as a proxy of liquidity often includes trades between HFTs. This might lead to an over estimate of market quality. The extension on flipping directly addresses this concern. Recently, Korajczyk and Murphy (2019b) and Korajczyk and Murphy (2019a) document that less high-frequency trading is associated with higher transaction costs for small trades and lower transaction costs for large trades. Hu (2019) shows that market quality improves when IEX, an institute implementing a trading speed bumps to all participant, became a national securities exchange. These findings are in line with predictions in my model.

Some empirical papers focus on characteristics of traditional market makers and HFTs. Kirilenko, Kyle, Samadi, and Tuzun (2017) document that, different from traditional market makers, HFTs behaviors during the flash crash are more consistent with the latency arbitrage theory. Hirschey (2018) shows that HFTs can anticipate and trade ahead of other investors' order flow. Baron, Brogaard, Hagströmer, and Kirilenko (2018) find that faster HFTs gain higher payoffs. This is in line with the prediction of my model that small HFTs has incentive to upgrade trading technology to be able to undercut the market maker. Van Kervel and Menkveld (2019) document that HFTs initially lean against institutional orders but eventually trade along long-lasting orders since they are likely to be information-motivated.[17] Yao and Ye (2018) document that HFTs provide more liquidity for stocks with higher relative tick size. Clark-Joseph, Ye, and Zi (2017) use data of two trading halts to show that designated market makers' participation has important liquidity implications. This clearly shows that designated market makers and HFTs operate on different business models. Bessembinder, Hao, and Zheng (2019) also highlight the importance of designated market makers by showing that an improving of contract terms for designate market makers in NYSE improves market quality. This is consistent to the prediction of my model. If the market maker receives extra rebate on each share, he will commit more capital in market making and posts a lower spread.[18]

[17] This finding is consistent with my assumption that the HFT acts as a liquidity provider. However, my model is silent on the HFT trading alone the information-motivated orders since my model does not consider informed trading.

[18] Bessembinder, Hao, and Zheng (2019) also document the spillover effect in market quality

2.2 Capital Constraint and Capital Commitment

Many models explore the link between capital constraints of intermediaries and liquidity provision. Kyle and Xiong (2001) describe the situation that when convergence traders lose capital, their liquidation leads to excess volatility and more correlation among different markets. Gromb and Vayanos (2002) show that constrained arbitrageurs might provide too much or too little liquidity compare to the social optimal level, depending on their initial investment positions. Weill (2007) and Brunnermeier and Pedersen (2008) both demonstrate that insufficient capital of the market maker would lead to lower liquidity provision then the optimal level. In Weill (2007), lack of capital prevents the market maker to absorb enough order imbalance when the economy is recovering from a negative shock. In Brunnermeier and Pedersen (2008), traders' lack of funding and market liquidity deterioration reinforce each other and let to "liquidity spiral". My paper contributes to this strand of literature by showing that, even when the market maker is not constrained, his capital commitment decision plays an important role to market quality when facing competition from high-frequency trading.

A relatively small empirical literature examines the capital commitment of market makers. Hameed, Kang, and Viswanathan (2010) show that negative market return decreases liquidity asymmetrically. The authors attribute the decrease to the market maker's capital constraint. Comerton-Forde, Hendershott, Jones, Moulton, and Seasholes (2010) find a similar result using data on NYSE specialist positions and revenues. Bessembinder, Jacobsen, Maxwell, and Venkataraman (2018) document that capital commitments of corporate bond dealers are decreasing overtime, specifically in markets with more electronically facilitated trades. The authors interpret this as a result of electronic trading reducing search cost and required capital. This model suggests an alternative explanation. The decrease of capital commitment might due to the growing entry of HFTs facilitated by electronic trading. Brogaard and Garriott (2019) document similar capital commitment decreases of market makers in the stock market.

improvement because of the strategic complementary effect in market making. My model is silent on this aspect because I assume a deep inter-dealer market.

3 Model Setting

3.1 The Setup

Consider a game with infinite many periods and three (kinds of) players: a long-run market maker, a short-run HFT and a short-run buyer. The market maker's discount rate is δ and has net worth w_0 in period 0. In each period, the market maker can either pay dividend d or acquire shares from a inter-dealer market at the fair price 1 for market making.[19] The market maker maximizes $E_0(\sum_{t=0}^{\infty} \delta^t d_t)$, the expected dividend payout. In each period, a short-run HFT enters the market with probability π. Upon entering, the HFT holds q_h shares and aims at maximizing her expected profit. The market maker and the HFT are both sellers and compete to provide liquidity for the short-run buyer. Due to liquidity or hedging needs, the buyer is willing to pay $v > 1$ for each share and demands q_b shares; i.e., he is willing to pay a premium $v - 1$ for each share within his demand q_b.

The sequence of events in a single period, illustrated in Figure 1, can be specified as follows: Let w_t be the market maker's net worth at the beginning of period t. The market maker first chooses a non-negative dividend level d_t. He then commits the remaining capital, $w_t - d_t$, to purchase $q_{m,t} = w_t - d_t$ shares from the inter-dealer market at the fair price 1.[20] The market maker then posts a spread $x_{m,t}$, committing to sell all shares at the ask price $1 + x_{m,t}$. After the market maker sets his spread, a short-run HFT holding q_h shares enters the market with probability π. If the HFT's trading technology is superior to the market maker, she observes the market maker's capital commitment $q_{m,t}$ and spread $x_{m,t}$ before setting her spread x_h (the sequential pricing game). Otherwise, the HFT only observes the market maker's capital commitment $q_{m,t}$ (the simultaneous pricing game). After the market maker and the HFT determine their spreads, the short-run buyer arrives with random demand q_b and random buying threshold $v > 1$. After the buyer finishes buying, the market maker and the HFT (if enters) may sell the remaining shares at the fair price 1 back to the inter-dealer market. This concludes a period.

I make the following assumptions on the distributions of q_b and v: $v - 1$ follows

[19] Another interpretation can be that the market maker invests some capital into a safe asset and deposits the rest of capital into a margin account to cover the cost of potential short selling.

[20] It is without loss of generality to assume that the market maker commits all remaining net worth in market making. If he chooses to commit less, he may raise his dividend payout to achieve a higher payoff.

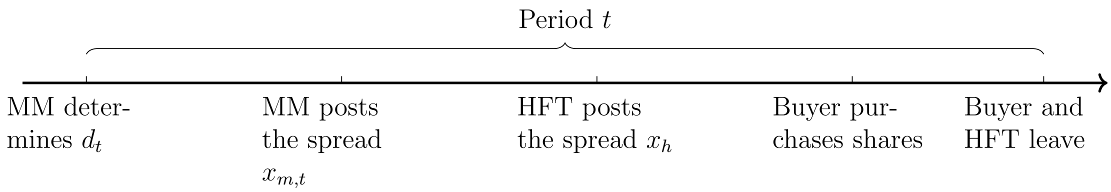


Figure 1: A Concise Time-line

a distribution supported on $[0, \hat{x}]$ with CDF F. q_b follows a distribution with finite expectation, positive support and CDF G. F and G are independent and continuously differentiable. I further assume that F has non-decreasing hazard rate; i.e., $\frac{f(x)}{1-F(x)}$ is non-decreasing, or equivalently, f is log-concave.

Several specific assumptions are worth more discussion. First, the buyer's demand q_b is inelastic when spreads are lower than $v-1$. In practice, this corresponds to the buyer posting a limit order with quantity q_b at price v. Since the market maker and the HFT do not observe q_b when setting spreads, higher spreads reduce the probability of trade. Thus, although the demand curve of each buyer is inelastic, from the market maker's and the HFT's perspectives, the demand curve is downward sloping. Second, in this model, the HFT is a short-run player with an exogenous entry probability π and a fixed shareholding q_h. This assumption by no means denies the possibility of the HFT being a long term market participant in practice. Instead, it means to reflect two features of high-frequency trading: (1) The HFT's entry and holding heavily depend on exogenous market conditions; (2) the HFT focuses on short term trading and only carries positions for a short period of time. Third, I only consider a one-sided market; i.e., the market maker and the HFT only sell shares to other investors. This is without loss of generality given that the market maker can adjust his position with no cost in the inter-dealer market. Considering a two-sided market setting leads to similar qualitative predictions.

3.2 Measures of Market Quality

Liquidity is one of the most important indicators of market quality. In this section, I define liquidity, the main measure of market quality in this model, and briefly discuss other market quality measures. Formally, L_t, liquidity in period t, is the expected number of shares sold to the buyer in period t. Since I focus on the steady state, where the market maker's pricing and capital commitment decisions are time invariant, I drop the time subscript and define liquidity (in the steady state) to be

$$L = \pi \underbrace{E(min(q_b, q_m I_{\{x_m \leq v-1\}} + q_h I_{\{x_h \leq v-1\}}))}_{\text{Expected selling volume with HFT}} + (1-\pi) \underbrace{E(min(q_b, q_m I_{\{x_m \leq v-1\}}))}_{\text{Expected selling volume without HFT}} .$$

Further define $L(v)$ to be the expected number of shares sold to the buyer with buying threshold v. It captures the market's capacity to satisfy buyers with buying thresholds higher than v. Specifically, the fill rate at buying threshold v can be measured by $L(v)/E(q_b)$. It is also worthwhile to examine the average spread. Define it to be the expected profit of liquidity providers (the market maker and the HFT) divided by liquidity.

Several features of this liquidity definition worth discussing. First, this definition incorporates both price and quantity information of the market. If spreads are high, the buyer's buying probability would be low. Then even with a large supply, liquidity would be low due to the lack of buyer. On the other hand, low spreads alone does not imply high liquidity. If the aggregate supply is small due to the low profit margin, liquidity would still be low since only a small portion of the buyer's demand can be satisfied. Second, this measure is closely related to (the buyer's) welfare.[21] Since the buyer has a higher valuation for each share, holding everything else equal, higher liquidity indicates better welfare. This definition differs from the buyer's surplus, by putting equal weights on each share sold. These two measures bear similarity in the sense that the buyer's surplus and liquidity almost always change in the same direction in the comparative statics. Moreover, liquidity is more feasible than the buyer's surplus as a market quality measure since trading volume is easier to observe in practice.

3.3 Equilibrium Definition

Two facts suggest that the market maker's net worth, w, should be considered as the state variable. First, net worth constraint is the only constraint faced by the market maker. Second, given the market maker's strategy, the HFT has no incentive to relate her action to the history of the game. Thus, equilibrium can be defined as follows:

[21]Here I follow the tradition of the literature by regarding the market maker and the HFT as integrated parts of the financial market and focusing on the buyer's welfare. The discussion on the social planner's problem is delegated to the Appendix.

Definition 1 *Consider a infinite horizon game* (w_0, q_h, π) *where the market maker starts with net worth* w_0 *and the HFT enters the market with probability* π *and* q_h *shares.*

1. *An equilibrium in a sequential pricing game is a triple* $(q_m(w), x_m(q_m(w)), x_h(q_m, x_m))$ *such that: (i) Given* q_m *and* x_m, $x_h(q_m, x_m)$ *maximizes the expected payoff of the HFT. (ii) Given* $x_h(q_m, x_m)$, $\{q_{m,t} = w_t - d_t\}_{t=0}^{\infty}$ *and* $\{x_{m,t} = x_m(q_{m,t})\}_{t=0}^{\infty}$ *maximize* $E_0(\sum_{t=0}^{\infty} \delta^t d_t)$.[22] *(iii)* $0 \leq q_{m,t} \leq w_t$ *for all* t.
2. *An equilibrium in a simultaneous pricing game is a triple* $(q_m(w), x_m(q_m(w)), x_h(q_m))$ *such that: (i) Given* q_m, $x_h(q_m)$ *maximizes the expected payoff of the HFT. (ii) Given* $x_h(q_m)$, $\{q_{m,t} = w_t - d_t\}_{t=0}^{\infty}$ *and* $\{x_{m,t} = x_m(q_{m,t})\}_{t=0}^{\infty}$ *maximize* $E_0(\sum_{t=0}^{\infty} \delta^t d_t)$. *(iii)* $0 \leq q_{m,t} \leq w_t$ *for all* t.

I focus on the steady state capital commitment and spread to characterize the long term market quality. The formal definition of a steady state equilibrium is as follows:

Definition 2 *An equilibrium is a steady state equilibrium if there exists* q_m, x_m *and* x_h *such that* $q_{m,t} = q_m$, $x_{m,t} = x_m$ *and* $x_{h,t} = x_h$ *for all* t.[23]

Intuitively, in a steady state equilibrium, the market maker's capital commitment $q_{m,t}$, spread $x_{m,t}$ and the HFT's spread $x_{h,t}$ are time invariant. Since the focus of this paper is on capital commitment rather than capital constraint, I assume that the market maker always starts the game with a sufficiently large net worth w_0.

[22] Notice that the distribution of w_{t+1} can be uniquely determined by w_t and the equilibrium strategies. Given w_0, the dynamic of w_t is well-defined.

[23] In the simultaneous moving game, x_m and x_h might be distributions rather than numbers.

4 Baseline Models

4.1 Benchmark Case with No HFT

First consider the situation with no HFT (or equivalently, $\pi = 0$). The market maker's value function satisfies the following equation:

$$V(w) = max_{d,x_m} d + \delta \underbrace{F(x_m)V(w-d)}_{\text{Buyer not buying}} + \delta \underbrace{[(1-F(x_m)) \int_0^{w-d} V(w-d+x_m q)g(q)dq}_{\text{Buyer buying with small demand}} + \underbrace{(1-F(x_m))(1-G(w-d))V((1+x_m)(w-d))}_{\text{Buyer buying with large demand}}] \quad (1)$$

with the budget constraint

$$0 \leq d \leq w \; . \quad (2)$$

Let

$$k(s) = E_G(min(q_b, s)) \; .$$[24]

This function measures expected trading volume when s shares are within the buying threshold. Since the buyer's demand is random, this function is strictly concave and the marginal value of capital commitment is decreasing to zero.[25] In other words, at a certain point, the market maker would find it more profitable to payout dividend. Specifically, there exists a steady state equilibrium with the capital commitment $q_m = \bar{q}$ and the spread $x_m = x^*$. The market maker pays out $w_0 - \bar{q}$ in period 0 and his profit in subsequent periods as dividend. The steady state can be characterized by the following theorem:

Theorem 1 *With no HFT, there exists a unique steady state equilibrium where the market maker set $q_{m,t} = \bar{q}$ ($d_t = w_t - \bar{q}$) and $x = x^*$ for all t. x^* satisfies*

$$x^* = argmax_x (1 - F(x))x \; .$$

[24] The subscript emphsizes that the expectation is over the buyer's demand.

[25] Alternatively, if q_b is deterministic, when the market maker's capital commitment is lower than q_b, at any fixed spread, each unit of capital committed has the same marginal value. Then the capital commitment problem become trivial since the market maker would choose to commit either q_b or 0.

$\bar{q}$ satisfies

$$\frac{\delta}{1-\delta}(1-F(x^*))x^*(1-G(\bar{q})) = 1 \ .^{26}$$

The market maker's expected payoff is

$$V(w_0) = \frac{\delta}{1-\delta}(1-F(x^*))x^*k(\bar{q}) + (w_0 - \bar{q}) \ .$$

Liquidity at the steady state is

$$L = (1-F(x^*))k(\bar{q}) \ .$$

The average spread is x^*.

Proof. See Appendix. ■

This theorem has a clear economic interpretation. Since the buyer's demand is random, each additional share is less likely to be sold at any given spread. Thus, the market maker's capital commitment has decreasing marginal value. Conversely, the marginal value of dividend payout is constant. This implies that in the equilibrium, the market maker would commit capital up to a unique level where the marginal value of capital commitment equals the marginal value of dividend payout. In the steady state, the market maker maintains his capital commitment level and pays out the profit. This makes him act like a short-run monopolist, setting the spread to maximize the expected profit.

With no HFT, the market has high supply at a high spread. Indeed, from the pricing perspective, x^* is the highest possible spread set by any rational liquidity supplier.[27] From the capital commitment perspective, the marginal value of capital commitment is the highest for the market maker facing no competition from the HFT. With HFT's presence, the market maker's steady state capital commitment is always lower than $\bar{q}$.

4.2 Sequential Pricing Game

In the sequential pricing game, the HFT observes the market maker's shareholding q_m and spread x_m before posting her spread x_h. In practice, this corresponds to the

[26] If no such $\bar{q}$ exists, the optimal strategy is to liquidate ($d = w_0$) at $t = 0$.

[27] At a spread higher than x^*, the loss from selling with lower probability dominates the benefit from selling at a higher price.

situation where the HFT has a superior trading technology and can undercut the market maker before the market maker is able to adjust the spread.

To characterize the steady state, it is helpful to first consider a one-shot game with fixed capital commitment/shareholding. The reason is clear: In the steady state, the market maker's capital commitment is constant over time and he pays out his profit from the previous period as dividend. Thus, in the steady state, the market maker would set spread as if he is a one-shot profit maximizer.

Consider a one-shot game (q_m, q_h, π) where the market maker holds q_m shares and the HFT enters with probability π holding q_h shares. The market maker sets spread x_m first and the HFT, if enters, sets spread x_h after observing x_m. Each player aims for maximizing his/her expected profit and can sell shares back to the inter-dealer market at the end of the game at price 1. Equilibrium of this one-shot game can be defined as follows:

Definition 3 *An equilibrium of a one-shot sequential pricing game* (q_m, q_h, π) *is a pair* $(x_m, x_h(x_m))$. *Given the market maker's spread* x_m, *the HFT's spread* $x_h(x_m)$ *maximizes her expected payoff. Given the HFT posting her spread according to* $x_h(x_m)$, *the market maker's spread* x_m *maximizes his expected payoff.*

First consider the HFT's pricing problem. If the HFT sets her spread $x_h \leq x_m$, her shares would be purchased first but at a lower price. Conversely, if $x_h > x_m$, the HFT would earn higher profit per share sold. Yet she would only receive the residual demand. Within each pricing region ($x_h \leq x_m$ or $x_h > x_m$), the HFT only faces the trade-off between earning higher unit profit and losing the buyer with low valuation.[28] This trade-off is characterized by the term $(1 - F(x))x$, the expected marginal value of supplying a share at spread x within the buyer's demand. Since I assume F, the CDF of the buyer's buying threshold, has non-decreasing hazard rate, $(1 - F(x))x$ is increasing in x for $x \leq x^*$ and decreasing in x for $x \geq x^*$. This simplifies the HFT's optimal pricing strategy.

Lemma 1 *Given the market maker's capital commitment* q_m *and spread* x_m, *the HFT's optimal spread is either* $x_h = x_m$ *or* $x_h = x^*$.

[28] Notice that for a buyer with buying threshold higher than $1 + x_h$, his expected demand does not depend on x_h. This relies on the independence assumption of the buyer's buying threshold and demand.

Proof. See Appendix. ■

Next consider the market maker's pricing problem. If the market maker sets a spread such that the HFT chooses $x_h = x_m$, the market maker would be better off setting $x_m = x^*$. If the market maker sets a spread such that the HFT chooses $x_h = x^*$. Since $(1 - F(x))x$ is increasing in x for $x \leq x^*$, the market maker would optimally set $x_m = \underline{x} < x^*$ such that the HFT is indifferent between setting $x_h = \underline{x}$ and setting $x_h = x^*$. All other pricing strategies are dominated by either of the aforementioned two strategies. To simplify the notation, define

$$a(x) = \frac{(1 - F(x))x}{(1 - F(x^*))x^*} \leq 1 \text{ .}$$[29]

The following lemma characterizes the market maker's optimal pricing strategy:

Lemma 2 *The market maker's optimal spread is either* $x_m = x^*$ *or* $x_m = \underline{x} < x^*$. $\underline{x}$ *is pinned down by the HFT's indifference condition*

$$a(\underline{x})k(q_h) = k(q_m + q_h) - k(q_m) \text{ .}$$

Proof. See Appendix. ■

By Lemma 1 and 2, I can pin down the equilibrium by comparing the market maker's payoffs with pricing strategies $x_m = \underline{x}$ and $x_m = x^*$.

Proposition 1 *If* $k(q_m) > \pi k(q_h)$*, the unique equilibrium is* $x_m = x_h = x^*$*. If* $k(q_m) < \pi k(q_h)$*, the unique equilibrium is* $x_m = \underline{x}$ *and* $x_h = x^*$*. When* $k(q_m) = \pi k(q_h)$*, both equilibria exist.*

Proof. See appendix. ■

By Proposition 1, the market maker has two possible pricing strategies against the potential HFT. I name $x_m = x^*$ to be the *wide* spread strategy. This strategy yields a high expected profit without HFT presence. When the HFT enters, however, the market maker will be undercut and only receives the residual demand. The effectiveness of this strategy depends on the HFT's entry probability π and shareholding q_h. $x_m = \underline{x}$ is the *tight* spread strategy. Under this strategy, the market maker receives a lower expected profit when the HFT does not enter. Yet when the market maker

[29]Note that $x^* = argmax_x (1 - F(x))x$.

uses the *tight* spread strategy, it is unprofitable for the HFT to undercut the market maker. Thus, the buyer would always buy shares from the market maker first and the HFT's entry does not affect the market maker's expected profit.

Another observation is that the HFT always sets spread $x_h = x^*$ in the equilibrium. However, this does not imply that the HFT always sells shares at a higher spread. When the market maker is using the wide spread strategy, the HFT's pricing strategy should be understood as undercutting the market maker at $x_h = x^* - \epsilon$ with an infinitesimally small ϵ.[30]

4.2.1 Steady State Characterization

In this section, I solve for the steady state equilibrium of the infinite period game. Let $M(q)$ be the market maker's expected profit in the one-shot game with $q_m = q$. Let $\hat{x}_m(q)$ and $\hat{x}_h(q)$ correspond to the market maker and the HFT's equilibrium spreads in the one-shot game.[31] If the game reaches a steady state in period 0 with capital commitment q, the market maker's expected payoff is

$$\frac{\delta}{1-\delta} M(q) + (w_0 - q) \ .$$

$\frac{\delta}{1-\delta} M(q)$ is the present value of a perpetuity paying out the market maker's expected profit starting from period 1. $w_0 - q$ is the market maker's dividend payout in period 0 to reach the steady state. An obvious candidate of the market maker's steady state capital commitment is

$$q_m = argmax_{q \in [0, \bar{q}]} \frac{\delta}{1-\delta} M(q) + (w_0 - q) \ .$$

The following theorem validates that q_m is indeed the market maker's capital commitment in the steady state equilibrium.

Theorem 2 *Let* $q_m = argmax_{q \in [0,\bar{q}]} \frac{\delta}{1-\delta} M(q) + (w_0 - q)$.

1. $q_{m,t} = q_m$, $x_m = \hat{x}_m(q_m)$, $x_h = \hat{x}_h(q_m)$ *consists a steady state equilibrium. The*

[30] If a minimum tick size exists, then the HFT would post a lower spread than the market maker in this situation.

[31] I suppress the dependency of these functions on q_h and π.

market maker's expected payoff in the equilibrium is

$$V(w_0) = \frac{\delta}{1-\delta} M(q_m) + (w_0 - q_m) \ .$$

2. *If the market maker uses the wide spread strategy in the equilibrium, market liquidity is*

$$L = (1 - F(x^*))[\pi k(q_m + q_h) + (1-\pi)k(q_m)] \ .$$

   The average spread is x^*.

3. *If the market maker uses the tight spread strategy in the equilibrium, market liquidity is*

$$L = (1 - F(x_m))k(q_m) + \pi(F(x^*) - F(x_m))(k(q_m + q_h) - k(q_m)) \ .$$

   The average spread is lower than x^*.

Proof. See appendix. ■

I now discuss some important corollaries.

Corollary 1 *For* $\pi > 0$, $q_m < \bar{q}$.

Corollary 1 states that the market maker commits less capital facing competition from the HFT. The competition reduces the marginal value of the market maker's capital commitment. This is either due to the HFT's undercut or the market maker using a lower spread. Lower marginal value leads to less capital commitment in the equilibrium.

Corollary 2 *If* $\bar{q} > 0$, $q_m > 0$. *In other words, the market maker never fully exit the market in the steady state equilibrium. Moreover,* q_m*, the market maker's steady state capital commitment, satisfies the following conditions:*

1. *If the market maker uses the wide spread strategy,*

$$\frac{\delta}{1-\delta}(1 - F(x^*))x^*[(1-\pi)(1 - G(q_m)) + \pi(1 - G(q_m + q_h))] = 1 \ .$$

2. *If the market maker uses the tight spread strategy,*

$$\frac{\delta}{1-\delta}(1 - F(\underline{x}))\underline{x}(1 - G(q_m)) > 1 \ .$$

Proof. See appendix. ■

Corollary 2, derived from first order conditions of the market maker, is useful for comparative statics in π. Using the wide spread strategy, the market maker's marginal value of committing capital equals to the marginal value of dividend payment fixing x_h. On the contrary, under the tight spread strategy, the market maker's marginal value of committing capital is larger fixing x_h. This means using the tight spread strategy, the market maker refrains from committing more capital because the market maker needs to maintain a low spread to prevent the HFT from undercutting.

4.2.2 Comparative Statics on π

In this section, I analyze how the steady state equilibrium and market quality change with π, the HFT's entry probability. Higher π indicates fiercer competition from the HFT. The market maker would adjust his capital commitment and pricing strategies accordingly and thus changes market quality.

First, consider the one-shot game. Importantly, the HFT's pricing decision does not depend on π. Thus, regardless of π, the market maker's candidates for the optimal spread, i.e., x^* and $\underline{x}$, are the same. Furthermore, if $x_m = x^*$, the market maker's expected payoff is decreasing in π due to the HFT's undercut. Conversely, the market maker's expected payoff does not depend on π when $x_m = \underline{x}$. Consequently, the tight spread strategy becomes more attractive with higher HFT entry probability. The comparative statics for one-shot games can be characterized by the following proposition:

Proposition 2 *Consider two one-shot games (q_m, q_h, π_1) and (q_m, q_h, π_2) with $\pi_2 > \pi_1$.*

1. *If the market maker adopts the tight spread strategy in the equilibrium in game (q_m, q_h, π_1), then he would also adopt the tight spread strategy in game (q_m, q_h, π_2). His expected profits in two games are the same.*

2. *If the market maker adopts the wide spread strategy in the equilibrium in game (q_m, q_h, π_2), then he would also adopt the wide spread strategy in game (q_m, q_h, π_1). His expect payoff is higher in game (q_m, q_h, π_1).*

Proof. Since q_m and q_h are fixed, the equilibrium strategy choices are implied by proposition 1.

Note that the tight spread $\underline{x}$ is determined by the equation

$$k(q_h + q_m) - k(q_m) = a(\underline{x})k(q_h) ,$$

which does not depend on π. Thus, the market maker's expected payoff when adopting the tight spread strategy, $(1 - F(\underline{x}))\underline{x}k(q_m)$, does not depend on π.

The market maker's expected net profit of adopting the wide spread strategy is

$$(1 - F(x^*)x^*)[\pi(k(q_h + q_m) - k(q_h)) + (1 - \pi)k(q_m)] .$$

This quantity is decreasing in π since

$$k(q_h + q_m) < k(q_h) + k(q_m) .$$

■

Now consider the infinite period game in two markets with different HFT entry probabilities. If the market maker uses the tight spread strategy in both markets, then he would make identical pricing and capital commitment decisions and enjoy the same expected payoffs. On the other hand, by Corollary 2, if the market maker uses the wide spread strategy in both markets, in the market with high HFT entry probability, he commits less capital and achieves a lower expected payoff. Combining these observations leads to the following result:

Theorem 3 *There exists $\hat{\pi} \in (0, 1]$ such that in the steady state equilibrium, $x_m = x^*$ when $\pi < \hat{\pi}$ and $x_m = \underline{x}$ when $\pi > \hat{\pi}$. Denote $[0, \hat{\pi})$ to be the wide spread region and $(\hat{\pi}, 1]$ to be the tight spread region.*

1. *In the wide spread region, the market maker's expected payoff $V(w_0)$ and equilibrium capital commitment q_m is decreasing in π; liquidity L's change in π is ambiguous.*

2. *In the tight spread region, the market maker's expected payoff $V(w_0)$ and equilibrium capital commitment q_m remain constants; Liquidity L is increasing in π.*

3. *The market maker's equilibrium capital commitment is smaller in the tight spread region comparing to any equilibrium capital commitment in the wide spread region.*

4. *In the wide spread region, the average spread is x^*. In the tight spread region, the average spread is lower than x^* and increasing in π.*

Proof. See appendix. ■

By Theorem 3, the steady state equilibrium can be categorized into two regimes depending on π. In the wide spread region, the market maker sets the monopolistic spread $x_m = x^*$ and responds to the competition by cutting capital commitment. In this region, the competition between the market maker and the HFT does not benefit low-valuation buyers since both the market maker and the HFT set the monopolistic spread. Instead, when the HFT enters the market, she improves market quality by increasing the market's capacity to satisfy high-valuation buyers' demands. Conversely, when the HFT does not enter, the market's capacity to satisfy large demand is lower and decreasing in π since the market maker's capital commitment is decreasing in π in this region.

In the tight spread region, low-valuation buyers benefit from the competition since the market maker's spread is lower then the monopolistic spread. However, to deter the HFT from undercutting, the market maker keeps his capital commitment at a lower level. This impairs the market's capacity to satisfy large demands and the market becomes shallower. Indeed, although shares become cheaper, the supply is limited. When the buyer's demand is large, either the price per share would jump to the monopolistic price with the HFT's presence or no enough supply exists to fulfill the order.[32] Moreover, in this region, an increase in the HFT's entry probability improves market quality since the market maker's capital commitment and spread are not changing in π. A higher HFT entry probability increases the market's capacity to satisfy buyers with large demands.

This theorem also demonstrates why the average spread and the implementation shortfall may fail to faithfully characterize market quality. Since higher average spread indicates higher implementation shortfall in this model, I only focus on the average spread in the following discussion. In the wide spread region, although liquidity (and thus the buyer's welfare) is changing with π, the average spread remains the same since both the market maker and the HFT set the monopolistic spread x^*. In the tight spread region, higher π leads to better market quality. Yet the average spread is also increasing because the HFT's spread is higher. With a higher HFT entry

[32] In this model, I do not consider other liquidity providers. Yet in reality it can be the case that the rest of the order are fulfilled by other suppliers at a higher price.

probability, a larger proportion of shares are sold at the higher spread. This drives up the average spread.

How liquidity changes with π is ambiguous in the wide spread region. Under some mild assumptions, more competition from the HFT is not always beneficial to the market. When the wide spread region is large enough, there is always a region where the liquidity is decreasing with the level of competition.

Proposition 3 *Suppose the wide spread region is* $[0, 1]$*; i.e., the market maker uses the wide spread strategy when* $\pi = 1$*. Then either there exists a region where* L *is strictly decreasing in* π *or* L *is constant over* $[0, 1]$*.*

Proof. See appendix. ■

The reason behind this result is simple. If the market maker uses the wide spread strategy at $\pi = 1$, from the first order condition, $q_m + q_h = \bar{q}$. In other words, from a buyer's perspective, the market is identical to the monopolistic market and thus has the same liquidity. By continuity, if L is not constant over π, there exists a region where L is strictly decreasing in π. Importantly, when the HFT's shareholding q_h is small, the assumption of Proposition 3 holds. Intuitively, with low q_h, the HFT's undercut is not much of a concern for the market maker. It is optimal for the market maker to set the monopolistic spread regardless of the HFT's entry probability. Thus, when q_h is low enough, there is always a region where liquidity is decreasing in π.

Assumptions may also be imposed on the distribution of the buyer's demand q_b. If G follows the exponential distribution (which has constant hazard rate), liquidity is not changing in π over the wide spread region. If G has increasing hazard rate (or equivalently, g is log-concave),[33] there always exists a region where L is decreasing in π.

Proposition 4 *If* G *follows an exponential distribution, liquidity is a constant with respect to* π *in the wide spread region.*

Proof. See appendix. ■

The discussion above leads to the following theorem regarding the non-monotonicity of liquidity L over the level of competition π:

[33] Many distributions satisfy this property including uniform distribution, gamma distribution with $\alpha > 1$, truncated normal distribution, etc.

Theorem 4 *If G has increasing hazard rate or q_h is small, L is non-monotonic with respect to π on* $[0, 1]$.

Proof. See appendix ■

This theorem, albeit simple, bears important implications for both empirical analysis and policy debate over high-frequency trading. In many high-frequency trading empirical research, when market quality as a welfare indicator is the dependent variable, there is an independent variable highly correlated to the HFT entry probability. For example, it can be high-frequency trading volume, frequency of order submission and cancellation, etc. If liquidity, as a measure of market quality, is not changing monotonically with respect to the HFT entry probability, the linear regression model might not deliver accurate prediction over high-frequency trading's effects over market quality.

From the policy making perspective, this theorem suggests that policy makers cannot only rely on past observations of how high-frequency trading changes the market to predict the welfare and market quality effects of high-frequency trading regulation. The reason is that regulations' would have huge effects on the HFT entry probability. Without monotonicity, the welfare and market quality effects might "flip signs". A theoretical framework is necessary to achieve a critical stance over high-frequency trading policy making.

4.3 Simultaneous Pricing Game

In this section I analyze the situation where the HFT only observes q_m (but not x_m) before setting her spread x_h. This corresponds to the market maker and the HFT having similar trading technologies and the HFT cannot undercut the market maker easily. This is related to two real world scenarios. First, some HFTs may become designated market makers.[34] With a better trading technology, the market maker can flicker quotes fast enough to avoid the HFT's detection. Second, the HFT might be constrained by exchange policies or regulation requirements such that she can no longer observe the price information ahead of other traders or undercut other traders easily.

[34]Actually, two out of four NYSE's major designated market makers, Citadel Securities LLC and Virtu Americas LLC, are considered also as high-frequency trading firms.

We first analyze a one-shot simultaneous pricing game (q_m, q_h, π). In this game, the market maker's shareholding is q_m and the HFT enters the market holding q_h shares with probability π. Similar to the sequential pricing game, the buyer would purchase shares from the HFT first if the HFT and the market maker post the same spread.[35]

Definition 4 *An equilibrium of a one-shot simultaneous pricing game (q_m, q_h, π) is a pair of cumulative distribution function (H_m, H_h) such that x_m has CDF H_m and x_h has CDF H_m. Let the support of x_m (x_h) be a measurable set X_m (X_h). The equilibrium satisfies following conditions:*

1. *Given that the HFT posts spreads according to H_h, the market maker posting spreads according to H_m maximizes his expected payoff.*

2. *Given that the market maker posts spreads according to H_m, the HFT posting spreads according to H_h maximizes her expected payoff.*

3. *Given H_h, any $x_m \in X_m$ yields the same expected payoff for the market maker; this expected payoff is weakly higher than the expected payoff by posting a spread $x_m \notin X_m$.*

4. *Given H_m, any $x_h \in X_h$ yields the same expected payoff for the market maker; this expected payoff is weakly higher than the expected payoff by posting a spread $x_h \notin X_h$.*

The following proposition characterizes candidates of equilibrium.

Proposition 5 *No pure strategy equilibrium exists. Let the infimum of $X_m(X_h)$ be $\underline{x}_m(\underline{x}_h)$ and the supremum of $X_m(X_h)$ be $\bar{x}_m(\bar{x}_h)$. In any mixed strategy equilibrium, $\underline{x}_m = \underline{x}_h = \underline{x}$, $\bar{x}_m = \bar{x}_h = x^*$. X_m and X_h are dense in $[\underline{x}, x^*]$. There exists no $x_m(x_h) \in [\underline{x}, x^*)$ such that $x_m(x_h)$ is posted with positive probability in the equilibrium.*

Proof. See appendix. ■

By Proposition 5, without loss of generality, I consider equilibrium where X_m and X_h are intervals. The equilibrium can be pinned down by the market maker and the HFT's indifference conditions.

[35]The only purpose of this assumption is to make the simultaneous pricing case comparable to the sequential pricing case. The specific tie-breaking rule does not matter.

Proposition 6 *There exists a unique equilibrium in the one-shot game* (q_m, q_h, π) *satisfying the following conditions:*

1. *If* $k(q_m) \geq \pi k(q_h)$*, in the equilibrium the market maker posts spread* $x_m = x^*$ *with positive probability* $\bar{P}_m = 1 - \frac{\pi k(q_h)}{k(q_m)}$.

   $\underline{x}$ *is uniquely determined by*

$$(1-\pi)k(q_m) + \pi(k(q_m+q_h) - k(q_h)) = a(\underline{x})k(q_m) \ . \quad (3)$$

   The market maker's mixed strategy satisfies

$$H_m(x) = (1 - \frac{a(\underline{x})}{a(x)}) \cdot \frac{k(q_h)}{k(q_m) + k(q_h) - k(q_m+q_h)} \quad \forall x \in [\underline{x}, x^*) \ . \quad (4)$$

   H_m *satisfies* $H_m(\underline{x}) = 0$, $\lim_{x \to x^{*-}} H_m(x) = 1 - \bar{P}_m$.

   The HFT's mixed strategy satisfies

$$H_h(x) = \frac{1}{\pi}(1 - \frac{a(\underline{x})}{a(x)}) \cdot \frac{k(q_m)}{k(q_m) + k(q_h) - k(q_m+q_h)} \quad \forall x \in [\underline{x}, x^*) \ . \quad (5)$$

   H_h *satisfies* $H_h(\underline{x}) = 0$, $\lim_{x \to x^{*-}} H_h(x) = 1$.

2. *If* $k(q_m) \leq \pi k(q_h)$*, in the equilibrium the HFT posts spread* $x_h = x^*$ *with positive probability* $\bar{P}_h = 1 - \frac{k(q_m)}{\pi k(q_h)}$.

   $\underline{x}$ *is uniquely determined by*

$$k(q_m+q_h) - k(q_m) = a(\underline{x})k(q_h) \ . \quad (6)$$

   H_m *satisfies Equation (4). Moreover,* $H_m(\underline{x}) = 0$, $\lim_{x \to x^{*-}} H_m(x) = 1$.

   H_h *satisfies Equation (5). Moreover,* $H_h(\underline{x}) = 0$, $\lim_{x \to x^{*-}} H_h(x) = 1 - \bar{P}_h$.

Proof. By Proposition 5, X_m and X_h are dense in $[\underline{x}, x^*]$. Thus, in any "regular" equilibrium, $(\underline{x}, x^*) \in X_m$; $(\underline{x}, x^*) \in X_h$. Then the uniqueness naturally follows from the equilibrium construction.

I only prove the first part of the theorem here since the calculation for the second part is similar. The only difference is that x^* is not in the support of X_m since the payoff of posting x^* is strictly lower than posting $x^* - \epsilon$ for a small ϵ.

The HFT's indifference condition implies

$$(1-\bar{P}_m)(k(q_m+q_h)-k(q_m))+\bar{P}_m k(q_h)=a(\underline{x})k(q_h)\ . \tag{7}$$

The market maker's indifference condition implies

$$(1-\pi)k(q_m)+\pi(k(q_m+q_h)-k(q_h))=a(\underline{x})k(q_m)\ . \tag{8}$$

By equation (7) and (8),

$$\bar{P}_m=\frac{a(\underline{x})k(q_h)+k(q_m)-k(q_m+q_h)}{k(q_h)+k(q_m)-k(q_m+q_h)}=1-\frac{\pi k(q_h)}{k(q_m)}\ . \tag{9}$$

H_m can be pinned down by the HFT's indifference condition:

$$a(x)[H_m(x)(k(q_m+q_h)-k(q_m))+(1-H_m(x))k(q_h)]=a(\underline{x})k(q_h)\quad \forall x\in[\underline{x},x^*)\ . \tag{10}$$

H_h can be pinned down by the market maker's indifference condition:

$$a(x)\{(1-\pi)k(q_m)+\pi[H_h(x)(k(q_m+q_h)-k(q_h))+(1-H_h(x))k(q_m)]\}=a(\underline{x})k(q_m)\ \ \forall x\in[\underline{x},x^*)\ . \tag{11}$$

Notice that $a(x)$ is increasing with x for $x\in[0,x^*]$ and $k(q_m+q_h)<k(q_h)+k(q_m)$. Thus, existence and uniqueness of H_m and H_h is guaranteed by the intermediate value theorem. For the market maker (HFT), the indifference condition guarantees any strategy in support X_m (X_h) yields the same expect profit. From the proof of Proposition 5, no player has incentive to deviate to a spread smaller than $\underline{x}$ or larger than x^*. ■

An important corollary of Proposition 6 is that the market maker's expected payoffs are the same in both the sequential pricing game and the simultaneous pricing game. Since the market maker acts as if a short term payoff maximizer in the steady state, the same one-shot payoff induces the same capital commitment decision. This observation simplifies the comparison of market quality under two settings.

Corollary 3 *In any one-shot game (q_m,q_h,π), the market maker's expected profits are the same under sequential pricing and simultaneous pricing.*

Proof. If $k(q_m)>\pi k(q_h)$, the market maker would use the wide spread strategy in the sequential pricing game with expected profit

$$(1 - F(x^*))x^*[(1 - \pi)k(q_m) + \pi(k(q_m + q_h) - k(q_h))] .$$

This equals the expected profit in the simultaneous pricing game when $k(q_m) > \pi k(q_h)$.

If $k(q_m) < \pi k(q_h)$, in the sequential pricing game, the market maker would use the tight spread strategy to achieve the expected payoff $(1 - F(\underline{x}))\underline{x}k(q_m)$ where the tight spread $\underline{x}$ is determined by

$$k(q_m + q_h) - k(q_m) = a(\underline{x})k(q_h) .$$

This equals the expected profit in the simultaneous pricing game when $k(q_m) < \pi k(q_h)$. ■

4.3.1 Steady State Characterization

The following theorem relates equilibria in one-shot games to the steady state equilibrium of the infinite period game. Moreover, this theorem offers comparison over the market maker and the HFT's expected payoffs in the sequential pricing game and the simultaneous pricing game.

Theorem 5 *Let* $q_m = argmax_{q \in [0,\bar{q}]} \frac{\delta}{1-\delta} M(q) + (w_0 - q)$.

1. *Let* $x_m(q_m)$ *and* $x_h(q_m)$ *follow the mixed strategy defined in Proposition 6. Then* q_m, $x_m(q_m)$ *and* $x_h(q_m)$ *determines a steady state equilibrium.*[36] *In this equilibrium, the market maker's expected payoff is*

$$V_m(w_0) = \frac{\delta}{1-\delta} M(q_m) + (w_0 - q_m) .$$

2. *The market maker's expected payoffs and steady state capital commitments are the same in both sequential pricing and simultaneous pricing games.*

[36] When G has non-decreasing hazard rate, this equilibrium can be micro-founded by considering a model where the HFT does not observe δ and the market makers signals δ with capital commitment. There exists a perfect Bayesian equilibrium that shares the same on path property as this steady state equilibrium.

3. *The HFT is strictly better off in the sequential pricing game if π is in the wide spread region. The HFT's expected payoffs are the same under both settings if π is in the tight spread region.*

4. *In a simultaneous pricing game, the steady state liquidity is*

$$
\begin{aligned}
L =&(1-F(x^*))[\pi k(q_m+q_h)+(1-\pi)k(q_m)]+\pi\int_{\underline{x}}^{x^*}[H_m(z)H_h(z)k(q_m+q_h) \\
&+(1-H_m(z))H_h(z)k(q_h)+H_m(z)(1-H_h(z))k(q_m)f(z)dz \\
&+(1-\pi)\int_{\underline{x}}^{x^*}H_m(z)k(q_m)f(z)dz]\ .
\end{aligned}
$$

Proof. See appendix. ■

It is informative to compare market qualities under the sequential pricing game and the simultaneous pricing game. By Theorem 5, the market maker's equilibrium capital commitments are the same in both settings. Pricing decisions of the market maker and the HFT drive the difference in market qualities. The following proposition summarizes liquidity comparison results.

Proposition 7 *Denote the steady state liquidity in the sequential pricing game and the simultaneous pricing game to be L_{se} and L_{sim}.*

1. *$L_{sim} > L_{se}$ if π is in the wide spread region.*

2. *$L_{sim} - L_{se}$ is constant for any π in the tight spread region.*

3. *L_{sim} and L_{se} is increasing in π in the tight spread region.*

Proof. See Appendix. ■

To understand the intuition, first consider the wide spread region. In the sequential pricing game, all shares are supplied at the monopolistic spread x^* while in the simultaneous pricing game, spreads are lower than x^* with positive probability. Since the market maker makes the same capital commitment decisions, liquidity is higher in the simultaneous pricing game. Moreover, since the HFT cannot undercut the market maker at the spread x^* in the simultaneous pricing game, the HFT's expected payoff is lower. In other words, the HFT in the simultaneous pricing game is willing to pay a small cost to trade faster than the market maker. Conversely, since the market maker's expected payoffs are the same in both settings, in the sequential pricing game, the market maker has no incentive to upgrade his technology to trade at the same speed as the HFT. This means the HFT has stronger incentive to upgrade trading technology than the market maker. Yet as discussed above, this incentive is detrimental to market quality.

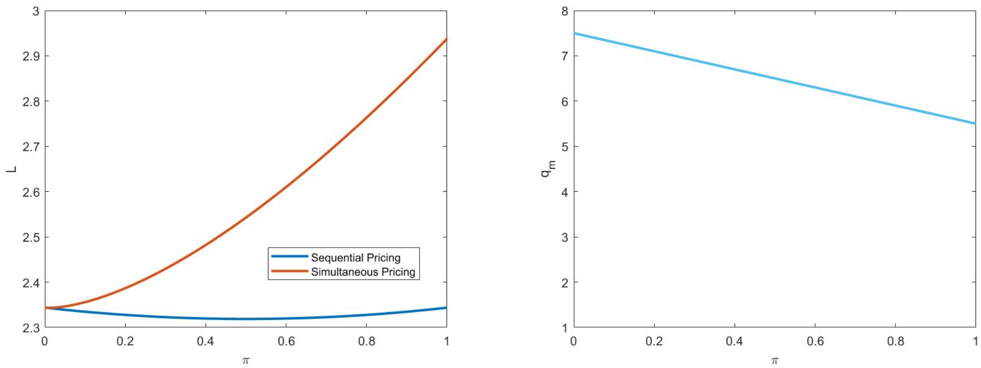
In the tight spread region, liquidity comparison between two settings is ambiguous. In the sequential pricing game, more shares are supplied at a low price because the market maker fixes a tight spread. Yet for a buyer with buying threshold between $1+\underline{x}$ and $1+x^*$, only the market maker's supply is available. On the other hand, in a simultaneous pricing game, a buyer with buying threshold $1+\underline{x}$ will not purchase any share with probability one. Yet for a buyer with buying threshold slightly lower than $1+x^*$, in expectation he would be able to purchase more shares in a simultaneous pricing game. This ambiguity does not impose much difficulties in the quantitative analysis. I show that the liquidity difference between the sequential pricing game and the simultaneous pricing game is not changing in π in the tight spread region. With specific assumptions on distributions of the buyer's buying threshold and demand, I can achieve a clear comparison over liquidity under two settings in the tight spread region.

4.4 Numerical Examples

This section contains numerical examples to visualize results in sections 4.2 and 4.3. In all examples, the buyer's buying threshold v follows a uniform distribution. The difference lies in the distribution of the buyer's demand q_b and the magnitude of HFT's shareholding q_h.

Figure 2 depicts liquidity and the market maker's equilibrium capital commitment under different HFT entry probabilities when the buyer's demand q_b follows a uniform distribution and the HFT's shareholding q_h is small. With small q_h, even when $\pi = 1$, the market maker still sets the monopolistic spread in the equilibrium; i.e., the wide spread region is $[0, 1]$. As shown in Figure 2b, with no regime change, the market maker's equilibrium capital commitment is decreasing continuously with π.

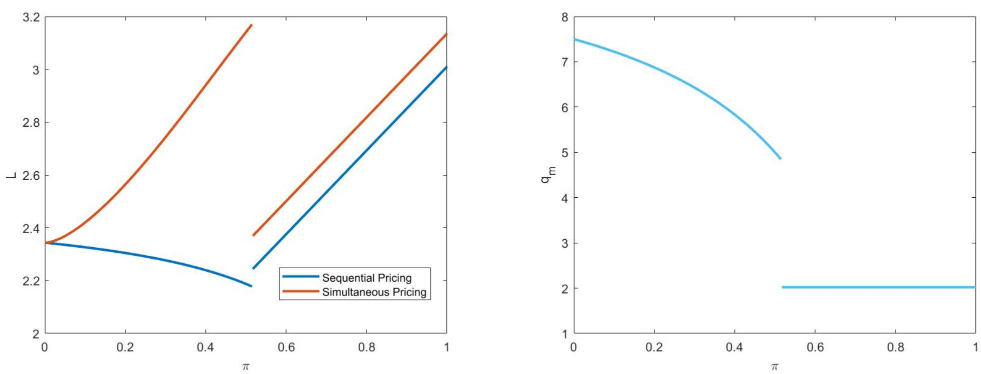
The blue line in Figure 2a shows how steady state liquidity changes with π in the sequential pricing game. There exists a region where liquidity is decreasing in π. In this example, the region is $\pi \in [0, \frac{1}{2}]$. The red line in Figure 2a shows how liquidity



(a) Liquidity (b) Capital Commitment

Figure 2: Uniform Demand with Small HFT

changes with π in the simultaneous pricing game. As predicted by Proposition 7, liquidity in the simultaneous pricing game is higher.



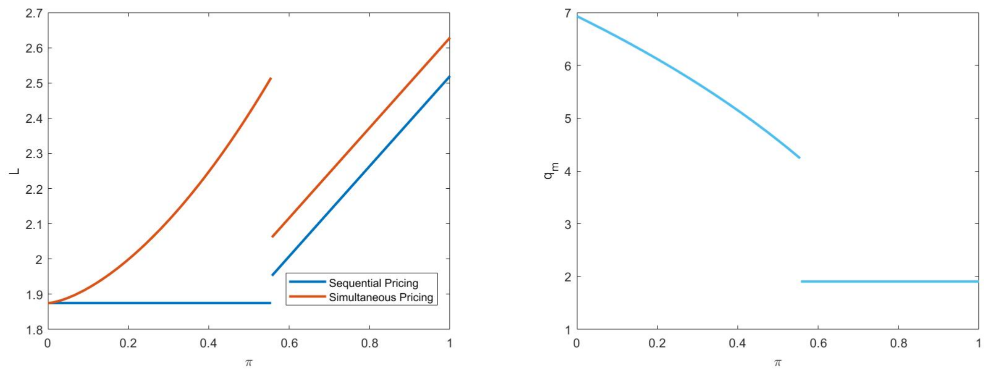
(a) Liquidity (b) Capital Commitment

Figure 3: Uniform Demand with Large HFT

Figure 3 shows liquidity and the market maker's capital commitment when q_b follows a uniform distribution and q_h is large. When π is large, the market maker would use the tight spread strategy in the equilibrium. This leads to the liquidity jump in Figure 3a and the capital commitment jump in Figure 3b. Since the market maker secures his payoff against the HFT entry in the tight spread region, the equilibrium capital commitment is not changing in π.

Another important observation can be made by comparing liquidity with $\pi \in [0.5, 0.6]$ and liquidity with $\pi = 0$ under the sequential pricing setting. Obviously, the average spread is lower in the tight spread region than in the monopolistic market.

However, the liquidity when $\pi \in [0.5, 0.6]$ is lower. The reason is that the market maker cut capital commitment facing the HFT's competition. This implies that pricing information alone cannot fully reflect market quality.



(a) Liquidity

(b) Capital Commitment

Figure 4: Exponential Demand

Figure 4 shows liquidity and the market maker's capital commitment when the buyer's demand follows an exponential distribution. This serves as a robustness check by demonstrating a similar comparative statics. The only difference is that liquidity remains constant in the wide spread region in the sequential pricing game. This follows from the constant hazard rate property of the exponential distribution.

5 Costly High-Frequency Trading Participation

In this section, I consider an extension where the HFT needs to pay a fixed cost C to participate in high-frequency trading. Specifically, after observing the market maker's capital commitment q_m (and spread x_m in the sequential pricing game), the HFT chooses whether to participate in high-frequency trading with cost C. If the HFT participates, she successfully enters the market with probability π. The cost C is paid regardless of the HFT successfully entering the market or not.[37] The HFT's

[37]Another way to model costly participation is to assume that the HFT only pays the cost C upon successfully entering the market. Yet assuming the HFT always pays the cost is in line with the regulatory measures taken in practice. For instance, the German High Frequency Trading Act of 2013 requires exchanges to charge excessive system usage fees, including both order amendments and order cancellations. France and EU also have similar requirements on charging order cancellation fee. For examples of exchange policies complying these regulations, see Eurex. (2016) and Eurex. (2019).

profit is normalized to zero if she does not participate. This extension partially endogenizes the HFT's entry decision.[38]

5.1 Sequential Pricing Game

In the sequential pricing game, the HFT observes the market maker's shareholding q_m and spread x_m before making the entry decision. Consider a one-shot game with high frequency trading cost C. Since the HFT observes the market maker's shareholding and spread before posting her spread, I focus on the pure strategy equilibrium.

Definition 5 *An equilibrium of a one-shot sequential pricing game* (q_m, q_h, π, C) *is a triple* (x_m, η, x_h)*;* $\eta \in \{0, 1\}$ *indicates the HFT's participation decision. The HFT's participation (non-participation) of high-frequency trading is denoted by* $\eta = 1$ *(*$\eta = 0$*).*

1. *Given the market maker's spread* x_m *and holding* q_m*,* x_h *maximizes the HFT's expected payoff.* $\eta = 1$ *if and only if the HFT's expected payoff is greater than* C*.*

2. *Given the HFT posts spreads according to* $x_h(x_m)$ *and makes entry decisions according to* η*,* x_m *maximizes the market maker's expected payoff.*

It is useful to compare one-shot games with and without participation cost. Condition on participating in trading, the HFT's optimal pricing strategies and thus the market maker's pricing strategies in two games are the same. On the other hand, with participation cost, the HFT takes her entry probability π into account. Specifically, the HFT would lose money if she participates in high-frequency trading but cannot enter the market. This gives the market maker an additional strategic advantage.

Let the deterring spread $\underline{x}^d \le x^*$ satisfy

$$\pi(1 - F(\underline{x}^d))\underline{x}^d k(q_h) = C \ .$$

If the market maker sets the deterring spread $\underline{x}^d$, participating in trading and undercutting the market maker is (weakly) not the optimal strategy for the HFT since doing so cannot cover the participation cost C. From the market maker's perspective,

[38]The endogenous entry is a special case of this setting with $\pi = 1$.

he would set $x_m = \underline{x}^d$ only when $\underline{x}^d \geq \underline{x}$. Facing the tight spread $\underline{x}$, the HFT is indifferent between setting the wide spread x^* and undercutting the market maker. Thus, when the market maker optimally sets $x_m = \underline{x}^d > \underline{x}$, it must be that participating in trading and setting $x_h = x^*$ cannot cover the participation cost, either. Thus, when the participation cost is high, the market maker may deter the HFT from participating by setting the deterring spread. Given the equilibrium strategy in a one-shot game without participation cost, the only additional decision for the market maker in the similar game with $C > 0$ is whether to post the deterring spread $\underline{x}^d$. This strategy becomes more profitable with higher participation cost C. Formally, the market maker and the HFT's pricing decisions in a one-shot game (q_m, q_h, π, C) can be characterized as follows:

Proposition 8 *Consider a one-shot game (q_m, q_h, π, C). Let*

$$\bar{C}(\pi) = \pi(1 - F(x^*))x^* k(q_h) \ .$$

If $C \geq \bar{C}$ the market maker posts $x_m = x^*$ and the HFT does not participate in high-frequency trading ($\eta = 0$). For $C < \bar{C}$:

1. *If **(i)** $k(q_m) < \pi k(q_h)$ and*

$$C > \pi(1 - F(x^*))x^*[k(q_m + q_h) - k(q_m)] \ ,$$

*or **(ii)** $k(q_m) > \pi k(q_h)$ and*

$$C > \frac{\pi k(q_h)}{k(q_m)}(1 - F(x^*))x^*[\pi(k(q_m + q_h) - k(q_h)) + (1 - \pi)k(q_m)] \ ,$$

the market maker posts the deterring spread $\underline{x}^d$ and the HFT does not participate in high-frequency trading ($\eta = 0$).

2. *If $k(q_m) < \pi k(q_h)$ and*

$$C \leq \pi(1 - F(x^*))x^*[k(q_m + q_h) - k(q_m)] \ ,$$

the market maker posts the tight spread $\underline{x}$ and the HFT participates ($\eta = 1$). Upon a successful entry, the HFT sets $x_h = x^*$.

3. *If* $k(q_m) > \pi k(q_h)$ *and*

$$C \leq \frac{\pi k(q_h)}{k(q_m)}(1 - F(x^*))x^*[\pi(k(q_m + q_h) - k(q_h)) + (1 - \pi)k(q_m)] \ ,$$

the market maker posts the wide spread and the HFT participates ($\eta = 1$)*. Upon a successful entry, the HFT posts* $x_h = x^*$ *to undercut the market maker.*

Proof. See appendix. ■

Now consider the steady state in the infinite period game. A similar analysis guarantees the existence of a steady state equilibrium. The following result considers the comparative statics on C.

Theorem 6 *There exists* $\hat{C}(\pi, q_h) \in (0, \bar{C})$ *such that:*

1. *For* $0 < C \leq \hat{C}$*, the steady state equilibrium is the same as the steady state equilibrium with no participation cost* ($C = 0$)*.*

2. *For* $\hat{C} < C \leq \bar{C}$*, the market maker sets the deterring spread* $x_m = \underline{x}^d$ *and the equilibrium capital commitment satisfying*

$$\frac{\delta}{1 - \delta}(1 - F(x_m))x_m(1 - G(q_m)) = 1 \ .$$

The HFT does not participate in high-frequency trading.

3. *For* $C > \bar{C}$*, the steady state equilibrium is the same as the monopolistic steady state equilibrium. The HFT does not participate in high-frequency trading.*

Proof. See appendix. ■

This result is intuitive. When the participation cost is low, it is unprofitable for the market maker to deter the HFT from participating in trading.[39] In this case, the HFT's expected payoff is larger than the participation cost C. Thus, the HFT always participates and the steady state equilibrium is the same as the equilibrium with no participation cost. If the participation cost is high enough, the market maker deters the HFT's participation with the deterring spread. Moreover, the market maker

[39] The market maker may still prevent the HFT from undercutting with a tight spread strategy as in the baseline model

optimally commits capital to the level such that the marginal value of capital commitment equals 1, the marginal value of dividend payout. The HFT in this situation does not participate in high-frequency trading. Finally, with an extremely high participation cost $C > \bar{C}$, the HFT never breaks even participating in high-frequency trading regardless of the market maker's spread. The market maker becomes a monopolist.

5.2 Simultaneous Pricing Game

In the simultaneous pricing game, the HFT only observes q_m, the market maker's shareholding, before making the participation decision. Consider a one-shot game (q_m, q_h, π, C). Similar to the simultaneous pricing game with no participation cost, no pure strategy equilibrium exists. A mixed strategy equilibrium can be defined as follows.

Definition 6 *An equilibrium of a one-shot simultaneous pricing game (q_m, q_h, π, C) is a triple (H_m, η, H_h). $\eta \in [0, 1]$ is the HFT's participation probability. x_m follows CDF H_m and x_h follows CDF H_h. Let the support of $x_m(x_h)$ be $X_m(X_h)$. The equilibrium satisfies the following conditions:*

1. *Given that the HFT posts spreads according to CDF H_h and tries to enter according to η, the market maker posting spreads according to CDF H_m maximizes his expected payoff.*

2. *Given that the market maker posts spreads according to CDF H_m, the HFT posting spreads according to CDF H_h and tries to enter according to η maximizes her expected payoff.*

3. *Given H_h and η, any $x_m \in X_m$ yields the same expected payoff for the market maker; this expected payoff is weakly higher than the expected payoff by posting a spread $x_m \notin X_m$.*

4. *Given H_m, any $x_h \in X_h$ yields the same expected payoff for the market maker; this expected payoff is weakly higher than the expected payoff by posting a spread $x_h \notin X_h$.*

To find out the equilibrium pricing strategy of the one-shot game (q_m, q_h, π, C), consider $(q_m, q_h, \pi, 0)$, a one-shot game with no participation cost. If the HFT's expect profit in the equilibrium of game $(q_m, q_h, \pi, 0)$ is greater than C, in the game

(q_m, q_h, π, C), the HFT participates with probability 1 and both players use the same pricing strategy as in game $(q_m, q_h, \pi, 0)$. Conversely, if the HFT's expected equilibrium profit in game $(q_m, q_h, \pi, 0)$ is lower than C, she would mix in participation decision. This mixing has two effects. First, it reduces the expected participation cost. Second, by entering the market with a lower probability, the HFT improves her strategic position against the market maker in the pricing game. The participating probability η can be uniquely determined by the HFT's indifference condition over participation.

Proposition 9 *Consider a one-shot simultaneous pricing game (q_m, q_h, π, C). Define $a(\underline{x})(\pi)$ as in Proposition 6. That is, if $k(q_m) \geq \pi k(q_h)$,*

$$a(\underline{x})(\pi) = 1 - \pi + \pi \frac{k(q_m + q_h) - k(q_h)}{k(q_m)} \ ;$$

if $k(q_m) < \pi k(q_h)$,

$$a(\underline{x})(\pi) = \frac{k(q_m + q_h) - k(q_m)}{k(q_h)} \ .$$

1. *If*

$$\pi(1 - F(x^*))x^* a(\underline{x})(\pi) k(q_h) \geq C \ ,$$

the HFT chooses $\eta = 1$. The equilibrium of game (q_m, q_h, π, C) coincides with the equilibrium of game $(q_m, q_h, \pi, 0)$ characterized in Proposition 6.

2. *If*

$$\pi(1 - F(x^*))x^* a(\underline{x})(\pi) k(q_h) < C \ ,$$

there exists a unique $\eta \in (0, 1)$ such that

$$\pi(1 - F(x^*))x^* a(\underline{x})(\eta\pi) k(q_h) = C \ .$$

In the equilibrium, the HFT participates with probability η and receives zero expected payoff if enters. The equilibrium of game (q_m, q_h, π, C) coincides with the equilibrium of game $(q_m, q_h, \eta\pi, 0)$.

Proof. See appendix. ■

An important implication of this proposition is as follows:

Corollary 4 *For any game (q_m, q_h, π, C), the market maker's equilibrium payoffs are the same under both the sequential pricing and the simultaneous pricing settings.*

Proof. See appendix. ■

Since the market maker receives the same expected payoffs in game (q_m, q_h, π, C) in the sequential pricing game and the simultaneous pricing game, the market maker's steady state capital commitments is both games are the same.

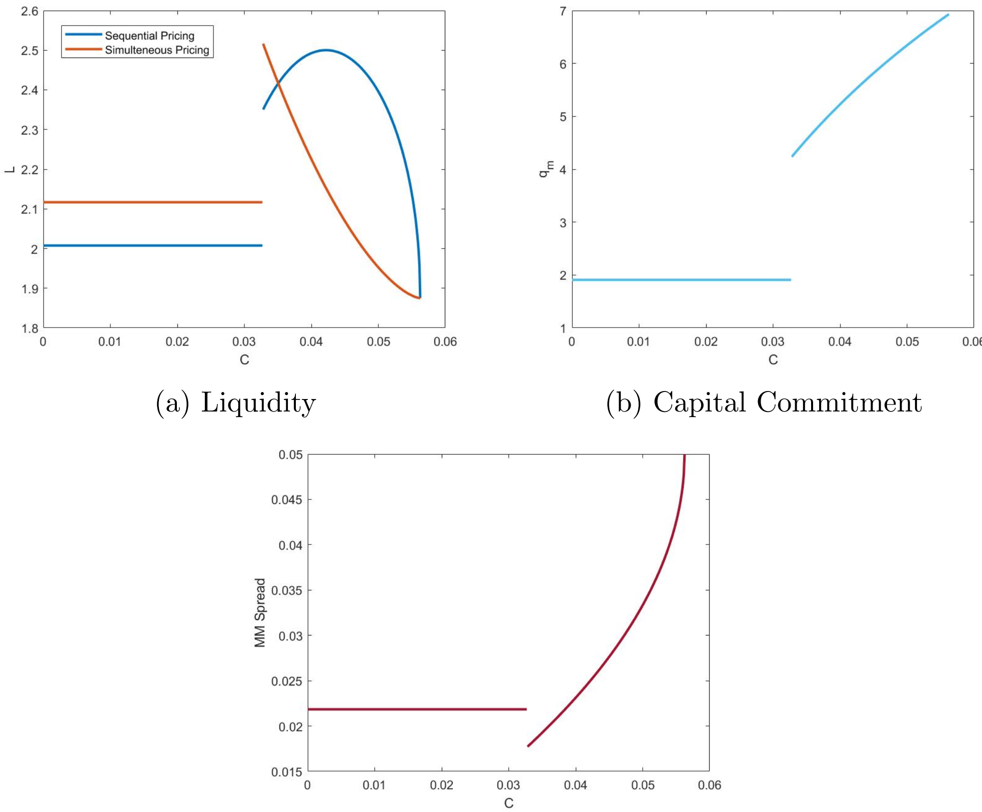
Proposition 10 *In the steady state, the market maker commits the same amount of capital in both the sequential and the simultaneous pricing game.*

Proof. This proof is similar to the no participation cost case and is thus omitted. ■

5.3 Numerical Examples

This section presents a numerical example to illustrate how market quality changes with the HFT's participation cost. In this example, the HFT's entry probability π is fixed. The buyer's buying threshold v follows a uniform distribution while his demand q_b follows an exponential distribution. The market maker uses the tight spread strategy in the steady state when $C = 0$.

The equilibrium can be divided into three regions. With a low participation cost, it is profitable for the HFT to participate with probability one. Thus, the market is the same to a market with no participation cost. As the participation cost increases, the market maker's deterring strategy becomes more profitable. Moreover, the marginal value of capital commitment also increases. Thus, the market maker's capital commitment is increasing with participation cost. One observation is that in the sequential game, the market maker's spread jumps downward when transiting into the deterring region. The reason is that when using the tight spread strategy, the spread is decreasing with the market maker's capital commitment. On the other hand, when the market maker is deterring the HFT with the deterring spread, this effect does not exist. Thus, when the market maker is indifferent between using the tight spread strategy and the deterring spread strategy, the deterring spread must be smaller. Finally, with a high participation cost, the market becomes a monopolistic market since it is never profitable for the HFT to participate.



(a) Liquidity

(b) Capital Commitment

(c) MM's Spread (Sequential Game)

Figure 5: Comparative Statics on Participation Cost

6 Policy Implications

In this section, I collect results developed in previous sections to discuss effects of regulations over high-frequency trading. Taking the sequential game as the benchmark, this paper examines three types of regulations: altering the HFT's entry probability π, leveling the trading technology difference between the HFT and the market maker and imposing high-frequency trading participation cost.

6.1 Altering the HFT's Entry Probability

In practice, the HFT's entry probability hinges on the HFT's ability to detect other investor's orders and acquire shares in a timely manner. Regulations changing the HFT's detecting and purchasing capacities affect the HFT's entry probability. For instance, banning co-location and integrating financial markets would decrease the

HFT's entry probability. Upgrading exchange's trading system without further restricting high-frequency trading would increase the HFT's entry probability.

This model predicts that in a market with high HFT entry probability (tight spread region), encouraging the HFT's entry is beneficial to market quality. The reason is that the market maker is setting a tight spread facing a fierce competition and the HFT is fulfilling the residual demand. An increase in HFT's entry probability leads to more liquidity supply from the HFT without changing the market maker's incentive to commit capital. On the other hand, in a market with low HFT entry probability, the market maker responses to the competition by committing less capital in market making. Liquidity would increase with the HFT's entry probability only if the benefit from more HFT supply outweighs the market maker's cut in capital commitment. Moreover, this model predicts that banning high-frequency trading does not necessarily deteriorates liquidity. Yet the spread would become higher due to the lack of competition.

6.2 Leveling the Trading Technology

This type of regulation "levels the playground" by making the market maker's trading technology comparable to the HFT's. For instance, the regulator can encourage HFTs to become designated market makers or incentivize existing market makers to upgrade their trading technologies. The batch auction proposed by Budish, Cramton, and Shim (2015) also achieves this goal since the market maker would have a chance to revise his order.

This model predicts that this policy is beneficial when the HFT's entry probability is low (wide spread region). Without a superior technology, the HFT mixes in posting spreads rather than undercuts the market maker at the monopolistic spread. This drives down the average price and improves market quality. When the HFT's entry probability is high, this model predicts that leveling the trading technology leads to less shares for low-valuation buyers (because the market maker now mixes rather than stick to the tight spread) and more shares for high-valuation buyers (because the HFT now mixes rather than stick to the monopolistic spread). The overall effect can be ambiguous.

6.3 Imposing High-frequency Trading Participation Cost

The third type of regulation imposes a participation cost over high-frequency trading. For example, regulations in France and Germany require a fee to be charged based on both executed and canceled orders. Regulation in Germany further requires all traders to tag algorithm generated orders. These regulations essentially induce participation costs on high-frequency trading.

This model predicts that low participation cost would not change the market quality. On the other hand, if the cost is high, the HFT would (at least partially) exit the market. The market maker's spread increases with the cost but he also commits more capital in market making. The directional change of liquidity depends on which effect dominates. Yet it is certain that extremely high participation cost always hurts the market.

7 Flipping

In this section, I consider the situation where the HFT can flip orders by first purchasing shares from the market maker and then reselling them at a higher spread. The HFT observes the market maker's capital commitment q_m and spread x_m before making flipping and pricing decisions. There are two implicit assumptions. First, the HFT is not capital constrained.[40] Second, the market maker does not have enough time to acquire additional shares from the inter-dealer market after the HFT purchases shares from him. For the ease of notation, I assume that G has an unbounded support. When G has a bounded support, the qualitative results are essentially the same. All proofs in this section are delegated to the appendix.

First consider the HFT's flipping and pricing decisions in a one-shot game (q_m, q_h, π). If the HFT flips shares, her spread must be higher than the market maker's spread. This implies her optimal spread is $x_h = x^*$. If the market maker holds q_m shares and his spread is $x_m < x^*$, the HFT's expected payoff when buying q_f shares from the market maker is

$$r(q_f) = (1 - F(x^*))x^*[k(q_m + q_h) - k(q_m - q_f)] - x_m q_f \ . \tag{12}$$

[40]The HFT's shareholding q_h reflects the exogenous market condition. Thus, the HFT cannot expand her shareholding even she is not capital constrained.

The first term of the right hand side is the expected gain from selling $q_h + q_f$ shares at spread x^* when the market maker is left with $q_m - q_f$ shares at a lower spread x_m. The second term of the right hand side is the premium paid by the HFT. The HFT pays $1 + x_m$ for each flipped share. If the buyer does not purchase these shares, the HFT only receives 1 by selling each share left to the inter-dealer market. By purchasing shares from the market maker, the HFT reduces the market maker's supply and thus the competition. Since the market maker's spread is lower and the demand is uncertain, the marginal benefit of flipping is increasing in q_f. Thus, the HFT would follow an "all or nothing" flipping strategy.

Proposition 11 *The HFT either purchases the market maker's entire shareholding q_m or nothing. In other words, $q_f = q_m$ or 0.*

Then consider the market maker's pricing problem. At a low enough price, by Proposition 11, the HFT would purchase all shares from the market maker upon entry. Thus, comparing to the baseline case, the market maker has an additional option to strategically lower his spread to induce flipping. The highest possible spread that induces flipping, x_m^f, can be pinned down by the HFT's indifference condition: Buying all shares from the market maker should be more profitable than the optimal pricing strategy without flipping. This can be summarized by the following lemma:

Lemma 3 *x_m^f satisfies*

$$(1 - F(x^*))x^* k(q_m) \geq x_m^f q_m \tag{13}$$

and

$$(1 - F(x^*))x^* k(q_m + q_h) \geq x_m^f q_m + (1 - F(x_m^f))x_m^f k(q_h) \ . \tag{14}$$

At least one inequality is binding. Moreover, if Inequality (14) binds, the flipping strategy dominates the tight spread strategy.

The wide and tight spread strategies are still available to the market maker. Specifically, if the market maker uses a wide spread, his expected payoff is

$$(1 - F(x^*))x^*[\pi(k(q_m + q_h) - k(q_h)) + (1 - \pi)k(q_m)] \ .$$

If $\underline{x} > x_m^f$, the market maker's expected payoff from the tight spread strategy is

$$(1 - F(\underline{x}))\underline{x}k(q_m) \ .$$

If the market maker posts x_m^f, his expected payoff is

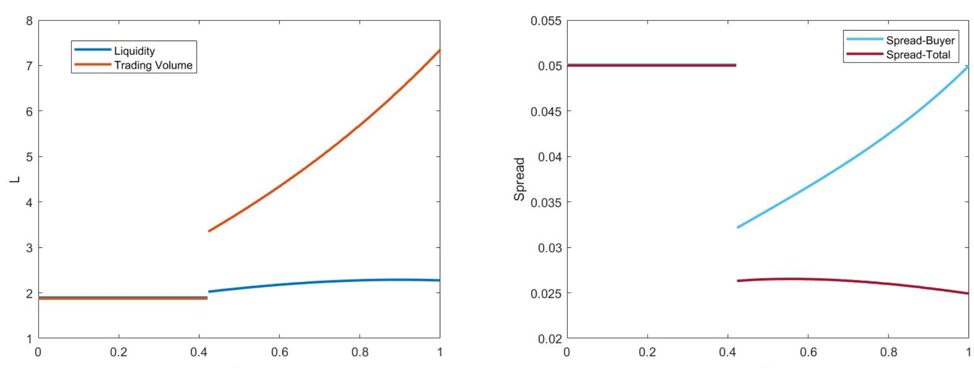
$$\pi x_m^f q_m + (1-\pi)(1-F(x_m^f))x_m^f k(q_m) \ .$$

An important observation is that, if the market maker expects the HFT to flip shares, the market maker's expected payoff is increasing in π. With flipping, the HFT is providing insurance for the market makers. When the HFT entry probability is large, the market maker would always induce flipping.

Proposition 12 *Under any q_m and q_h, if π is high enough, the market maker sets spread x_m^f in the equilibrium.*

In the infinite period game, although the market maker can be insured by the HFT, he does not have the inventive to increase capital commitment indefinitely. This is because under any capital commitment level, the expected payoff is upper-bounded by the monopolistic payoff. Moreover, when the capital commitment is large, the spread to induce flipping becomes close to zero. This implies an upper-bound exists for the market maker's capital commitment in the steady state equilibrium.[41]

Proposition 13 *For large enough w_0, a steady state equilibrium exists.*



(a) Liquidity v.s. Volume

(b) Buyers' v.s. Average Spread

Figure 6: Equilibrium Volume and Price with Flipping

Figure 6 presents a numerical example showing the equilibrium liquidity and average spread when the HFT is able to flip orders. At low HFT entry probability,

[41]The upper-bound may not be $\bar{q}$ since the market maker's expected payoff with shareholding $\bar{q}$ might be lower than the monopolistic payoff.

whether the HFT can flip shares or not does not change the equilibrium outcome. The reason is that from the market maker's perspective, the benefit of using a low spread to induce flipping is not large enough.

When the HFT's entry probability becomes large, the market maker sets a low spread to induce flipping. A large portion of transactions happens between the market maker and the HFT since the HFT purchases all low price shares upon entering. The buyer only benefits from the market maker's low spread when the HFT fails to enter the market. This suggests that it is important to separate trades between liquidity suppliers (the market maker and the HFT) and trades from liquidity suppliers to the buyer. As shown in Figure 6a, the expected trading volume and the average spread do not accurately reflect the market quality. The expected shares sold to the buyer only increase modestly in π comparing to the expected trading volume. Moreover, the average spread remains low while the buyer is facing a much higher spread increasing in π. This is because the majority of low price shares are purchased by the HFT. As the HFT becomes more likely to enter the market, the buyer becomes less likely to purchase cheap shares. When the HFT can flip shares, the entry of HFT only has limited benefits for the buyer. If we only look at the overall trading data, the benefit of high-frequency trading will be overestimated.

8 Supply Schedule and Induced Limit Order Book

Up till now I assume the market maker sells all shares at one spread. In this section, I analyze an extension where the market maker can submit a supply schedule to sell shares at different spreads. To keep the problem tractable, I maintain the assumption that the HFT sells all her shares at one spread and determines her spread after observing the market maker's capital commitment q_m and supply schedule.

Formally, given the market maker's capital commitment q_m, his pricing strategy can be represented by a supply schedule Ψ. The amount of shares supplied by the market maker with spreads less or equal to x is $q_m\Psi(x)$. In the steady state, the market maker posts the supply schedule to maximize his expected profit in each period. Thus, it is suffice to first solve for the optimal Ψ in a one-shot game under any q_m and then determine the steady state capital commitment with the market maker's indifference condition.

8.1 No HFT

Consider a one-shot game where the market maker holding q_m shares maximizes expected profit in a single period. With no HFT, even though the market maker may sell shares at different spreads, he optimally supplies all shares at the monopolistic spread x^*. Formally, we have the following proposition:

Proposition 14 *Given any q_m, in a one-shot game, the market maker would optimally set the supply schedule to be $\Psi(x) = I_{\{x \geq x^*\}}$.*

Proof. See appendix ■

A direct implication of Proposition 14 is that, in a infinite period game with no HFT, the steady state equilibrium is the same as the equilibrium in the baseline model. In other words, with no potential competition from the HFT, the market maker has no incentive to submit a non-degenerate supply schedule.

Corollary 5 *When no HFT exists, the steady state equilibrium is the same as the baseline model. Moreover, the market maker does not pay dividend when his net worth is smaller than the steady state capital commitment $\bar{w}$.*

Proof. The first statement is a straightforward result from Proposition 14. For the second statement, if the dividend payout is non-zero, the market maker can always achieve a higher payoff by refraining from paying dividend and supply the extra amount of shares at the spread x^* and payout the total return from the extra shares in the next period. ■

8.2 With HFT

When the HFT may enter the market, the market maker's pricing strategy is non-degenerate. Specifically, it is never optimal for the market maker to sell all shares at one spread. The intuition behind this result is simple. Given any single spread pricing strategy, the market maker can always sell a small amount of shares at another spread without changing the HFT's pricing strategy. If the market maker is using the wide spread strategy, he can improve his payoff by selling some shares before the HFT at a spread close to the monopolistic spread. If the market maker is using the tight spread strategy, he can sell some shares at a higher spread without the HFT undercutting him.

Proposition 15 *For any q_h and $\pi > 0$, supplying all shares at any spread x is not the optimal pricing strategy for the market maker in the steady state equilibrium.*

Proof. See appendix ■

Moreover, with the ability to flexibly sell shares, an immediate lower bound $\underline{q}$ exists for the market maker's capital commitment in the steady state. If the capital commitment level is below $\underline{q}$, the market maker can always improve his expected payoff by committing more capital and sell additional shares at the spread x^*.

Corollary 6 *The market maker would commit at least $\underline{q} > 0$ unit of capital, as long as his capital commitment with no HFT is non-zero. Specifically, $\underline{q}$ is the solution of*

$$\frac{\delta}{1-\delta}(1 - F(x^*))x^*[\pi(1 - G(\underline{q} + q_h)) + (1 - \pi)(1 - G(\underline{q}))] = 1 \ .$$

Notice that $\underline{q}$ is also the market maker's equilibrium capital commitment level in the wide spread region of the baseline model. Thus, allowing the market maker to submit a supply schedule improves liquidity in the wide spread region. Liquidity change in the tight spread region when the market maker can submit a supply schedule is ambiguous. However, the following proposition guarantees that given any specific set of parameters, the market maker's supply schedule can be easily computed. Then the change in spreads and liquidity can be characterized through numerical calculation.

Proposition 16 *The market maker's equilibrium pricing strategy $\Psi(x)$ satisfies three conditions:*

1. $\Psi(x^*) = 1$.
2. $\Psi(\cdot)$ *has no mass point for $x < x^*$.*
3. *The HFT achieves the same expected payoff by setting any $x_h \in [\underline{x}, x^*]$ where $\Psi(\underline{x}) = 0$.*[42]

Proof. See appendix ■

With this result, the market maker's equilibrium capital commitment q_m and pricing strategy $\Psi(x)$ can be numerically computed with the following algorithm: (i)

[42]Notice that the second result is implied by the third result. If there is a mass point at a spread $x < x^*$, the indifference condition cannot hold everywhere.

Fix a q_{ml}, the amount of shares sold by the market maker with spreads lower than x^*. (ii) If $q_{ml} \leq \underline{q}$, $q = \underline{q}$; i.e., the market maker sells $\underline{q} - q_{ml}$ shares at the monopolistic spread x^*. Otherwise, $q = q_{ml}$. (iii) Given q_{ml} and q, $\Psi(x)$ is pinned down by

$$(1 - F(x))x[k(\Psi(x)q + q_h) - k(\Psi(x)q)] = (1 - F(x^*))x^*[k(q_{ml} + q_h) - k(q_{ml})]$$

for $x \in [\underline{x}, x^*)$ and $\Psi(x^*) = 1$. (iv) As in the baseline case, let $M(q)$ be the expected per-period payoff of the market maker with capital commitment q. If $q = \underline{q}$, define $M(\underline{q})$ to be the maximum expected payoff for $q_{ml} \in [0, \underline{q}]$. (v) The market maker's equilibrium capital commitment is

$$q_m = max_{q \in [\underline{q}, \bar{q}]} \frac{\delta}{1 - \delta} M(q) + (w_0 - q) \ .$$

The market maker's pricing strategy is then pinned down by the procedure above.

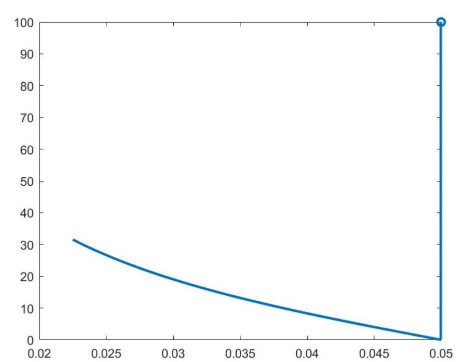


Figure 7: Supply Schedule of the Market Maker

When the buyer's demand q_b follows an exponential distribution, the market maker's supply schedule can be explicitly characterized. Specifically, let $\psi(x) = \Psi'(x)$. Then for $x \in [\underline{x}, x^*)$, $\psi(x) \propto \frac{1}{x} - \frac{f(x)}{1-F(x)}$. Figure 7 provides a visual illustration of the market maker's supply schedule under a further assumption that the buyer's spread tolerance $v - 1$ follows a uniform distribution. The x-axis represents the spread while the y-axis represents the density of the market maker's supply schedule. The density of the market maker's supply is decreasing to zero approaching the monopolistic spread x^*. Moreover, the line at rightmost of Figure 7 demonstrates that the market maker is supplying a positive quantity at the spread x^*.

8.3 Discussion

This extension analyzes the change in market quality when the market maker can sell shares at different spreads. With no HFT, the limit order book is degenerate in the sense that all shares are still supplied at the monopolistic spread x^* as in the baseline model. Conversely, when the HFT might enter the market, the market maker would supply shares at continuum of spreads. This improves liquidity in the wide spread region. The liquidity change in the tight spread region is ambiguous.

Moreover, this extension illustrates how asymmetric competition between the market maker and the HFT determines the shape of the limit order book.[43] Intuitively, fixing the HFT's pricing strategy, the market maker has incentive to increase spreads of some shares for higher expected profit. Yet to prevent the HFT from undercutting, the market maker needs to supply enough amount of shares at low spreads. This trade-off uniquely determines the shape of the limit order book. In any steady state equilibrium, the market maker would choose a supply schedule such that the HFT is indifferent between undercutting the market maker at any spread in the schedule and posting the monopolistic spread x^*.

9 Conclusion

My paper studies how high-frequency trading changes market quality through affecting the traditional market maker's capital commitment and pricing decisions. I consider a long-run market maker facing competition from short-run HFTs in providing liquidity. In the steady state, the long-run market maker responses to the competition by reducing his spread and committing less capital in market making. The latter effect impairs market quality. Thus, when taking the market maker's capital commitment channel into consideration, high-frequency trading does not necessarily improves market quality though it always (weakly) reduces the average spread. Moreover, in my model, the difference in trading technologies between the HFT and the market maker affects market quality. When the HFT's entry probability is low, "leveling the playground" by making the market maker and the HFT trade at the same speed improves market quality.

[43] Roşu (2009) analyzes a similar problem under the assumption that each market participant supplies one unit of share to the market and all market participants have the same trading speed.

I further consider three extensions. The first extension introduces high-frequency trading participation cost to endogenize the HFT's participation. When the HFT trades faster than the market maker and the participation cost is low, market quality remains the same. On the other hand, when the participation cost is high, the market maker optimally sets a spread to deter the HFT from entering the market. Although the HFT does not participate in trading when the participation cost passes a certain threshold, the cost level still affects the market quality since the market maker's deterring strategy depends on the cost. When the HFT and the market maker trade at the same speed, the model's prediction is similar except that the HFT mixes in participation facing high participation cost.

In the second extension, the HFT can "flip shares" by purchasing shares from the market maker and resupplying them at a higher spread. With high HFT entry probability, the market maker would induce flipping by posting a low spread since flipping effectively insures the market maker. Yet the buyer does not benefit much from the low spread since most of the cheaper shares are mainly acquired by the HFT. This extension demonstrates the importance to exclude the trading between liquidity suppliers when evaluating market quality. Otherwise, market quality would be overestimated with an overestimation of the expected trading volume and an underestimation of the average spread.

The third extension investigates implications on the shape of the limit order book when the market maker can sell shares at different spreads. Specifically, with no HFT, the market maker would still sell all shares at the monopolistic spread. Facing the competition from the HFT, the market maker would sell shares at a continuum of spreads. This extension demonstrates how competition between the market maker and the HFT determines the shape of the limit order book.

I want to emphasize several important insights of this model. First, the price information alone cannot fully reflect market quality; the volume information is equally important. Second, more high-frequency trading does not necessarily improves market quality since it reduces the market maker's willingness to commit capital in market making. Third, the relative trading speed between the market maker and the HFT affects market quality. When the HFT's entry probability is low, letting the market maker and the HFT trade at same speed improves market quality. Fourth, it is important to separate the trades among liquidity suppliers to avoid overestimations on market quality and the high-frequency trading's welfare effect.

References

Ait-Sahalia, Y. and M. Saglam (2017a). High frequency market making: Implications for liquidity. *Available at SSRN 2908438*.

Ait-Sahalia, Y. and M. Saglam (2017b). High frequency market making: Optimal quoting. *Available at SSRN 2331613*.

Back, K. and S. Baruch (2013). Strategic liquidity provision in limit order markets. *Econometrica 81*(1), 363–392.

Baldauf, M. and J. Mollner (2019). High-frequency trading and market performance. *Journal of Finance*.

Baron, M., J. Brogaard, B. Hagströmer, and A. Kirilenko (2018). Risk and return in high-frequency trading. *Journal of financial and quantitative analysis*, 1–32.

Baruch, S. and L. R. Glosten (2019). Tail expectation and imperfect competition in limit order book markets. *Journal of Economic Theory 183*, 661–697.

Bell, H. and H. Searles (2014). An analysis of global hft regulation: Motivations, market failures, and alternative outcomes.

Bessembinder, H., J. Hao, and K. Zheng (2019). Liquidity provision contracts and market quality: Evidence from the new york stock exchange. *The Review of Financial Studies*.

Bessembinder, H., S. Jacobsen, W. Maxwell, and K. Venkataraman (2018). Capital commitment and illiquidity in corporate bonds. *The Journal of Finance 73*(4), 1615–1661.

Biais, B., T. Foucault, and S. Moinas (2015). Equilibrium fast trading. *Journal of Financial economics 116*(2), 292–313.

Biais, B., D. Martimort, and J.-C. Rochet (2000). Competing mechanisms in a common value environment. *Econometrica 68*(4), 799–837.

Boehmer, E., K. Fong, and J. Wu (2018). International evidence on algorithmic trading.

Brogaard, J. and C. Garriott (2019). High-frequency trading competition. *Journal of Financial and Quantitative Analysis 54*(4), 1469–1497.

Brogaard, J., B. Hagströmer, L. Nordén, and R. Riordan (2015). Trading fast and slow: Colocation and liquidity. *The Review of Financial Studies 28*(12), 3407–3443.

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

Brunnermeier, M. K. and L. H. Pedersen (2008). Market liquidity and funding liquidity. *The review of financial studies 22*(6), 2201–2238.

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

Budish, E., R. Lee, and J. Shim (2019). A theory of stock exchange competition and innovation: Will the market fix the market? *NBER Working Paper 25855.*

Chakravarty, S. and C. W. Holden (1995). An integrated model of market and limit orders. *Journal of Financial Intermediation 4*(3), 213–241.

Chordia, T., R. Roll, and A. Subrahmanyam (2011). Recent trends in trading activity and market quality. *Journal of Financial Economics 101*(2), 243–263.

Clark-Joseph, A. D., M. Ye, and C. Zi (2017). Designated market makers still matter: Evidence from two natural experiments. *Journal of Financial Economics 126*(3), 652–667.

Comerton-Forde, C., T. Hendershott, C. M. Jones, P. C. Moulton, and M. S. Seasholes (2010). Time variation in liquidity: The role of market-maker inventories and revenues. *The Journal of Finance 65*(1), 295–331.

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

Conrad, J. S. and S. Wahal (2018). The term structure of liquidity provision. *Available at SSRN 2837111*.

Eurex. (2016). Hft act: Amendments to the calculation of excessive system usage fee.

Eurex. (2019). Enhancement of the excessive system usage concept: Introduction of a new limit type.

Foucault, T., J. Hombert, and I. Roşu (2016). News trading and speed. *The Journal of Finance 71*(1), 335–382.

Foucault, T., O. Kadan, and E. Kandel (2005). Limit order book as a market for liquidity. *The review of financial studies 18*(4), 1171–1217.

Glosten, L. R. (1994). Is the electronic open limit order book inevitable? *The Journal of Finance 49*(4), 1127–1161.

Goettler, R. L., C. A. Parlour, and U. Rajan (2009). Informed traders and limit order markets. *Journal of Financial Economics 93*(1), 67–87.

Gromb, D. and D. Vayanos (2002). Equilibrium and welfare in markets with financially constrained arbitrageurs. *Journal of financial Economics 66*(2-3), 361–407.

Hameed, A., W. Kang, and S. Viswanathan (2010). Stock market declines and liquidity. *The Journal of Finance 65*(1), 257–293.

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

Hasbrouck, J. and G. Saar (2013). Low-latency trading. *Journal of Financial Markets 16*(4), 646–679.

Hendershott, T., C. M. Jones, and A. J. Menkveld (2011). Does algorithmic trading improve liquidity? *The Journal of Finance 66*(1), 1–33.

Hendershott, T. and R. Riordan (2013). Algorithmic trading and the market for liquidity. *Journal of Financial and Quantitative Analysis 48*(4), 1001–1024.

Hirschey, N. (2018). Do high-frequency traders anticipate buying and selling pressure? *Available at SSRN 2238516*.

Hu, E. (2019). Intentional access delays, market quality, and price discovery: Evidence from iex becoming an exchange. *Available at SSRN 3195001*.

Kirilenko, A., A. S. Kyle, M. Samadi, and T. Tuzun (2017). The flash crash: High-frequency trading in an electronic market. *The Journal of Finance 72*(3), 967–998.

Korajczyk, R. A. and D. Murphy (2019a). Do high-frequency traders improve your implementation shortfall? *Journal of Investment Management 18*, 18–33.

Korajczyk, R. A. and D. Murphy (2019b). High-frequency market making to large institutional trades. *The Review of Financial Studies 32*(3), 1034–1067.

Kreps, D. M. and J. A. Scheinkman (1983). Quantity precommitment and bertrand competition yield cournot outcomes. *The Bell Journal of Economics*, 326–337.

Kyle, A. S. (1985). Continuous auctions and insider trading. *Econometrica: Journal of the Econometric Society*, 1315–1335.

Kyle, A. S. and W. Xiong (2001). Contagion as a wealth effect. *The Journal of Finance 56*(4), 1401–1440.

Li, S., X. Wang, and M. Ye (2020). Who provides liquidity, and when? *NBER Working Paper* (w25972).

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

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

O'Hara, M., C. Yao, and M. Ye (2014). What's not there: Odd lots and market data. *The Journal of Finance 69*(5), 2199–2236.

Parlour, C. A. and D. J. Seppi (2003). Liquidity-based competition for order flow. *The Review of Financial Studies 16*(2), 301–343.

Roşu, I. (2009). A dynamic model of the limit order book. *The Review of Financial Studies 22*(11), 4601–4641.

Seppi, D. J. (1997). Liquidity provision with limit orders and a strategic specialist. *The Review of Financial Studies 10*(1), 103–150.

Van Kervel, V. (2015). Competition for order flow with fast and slow traders. *The Review of Financial Studies 28*(7), 2094–2127.

Van Kervel, V. and A. J. Menkveld (2019). High-frequency trading around large institutional orders. *The Journal of Finance*.

Viswanathan, S. and J. J. Wang (2002). Market architecture: limit-order books versus dealership markets. *Journal of Financial Markets 5*(2), 127–167.

Weill, P.-O. (2007). Leaning against the wind. *The Review of Economic Studies 74*(4), 1329–1354.

Yao, C. and M. Ye (2018). Why trading speed matters: A tale of queue rationing under price controls. *The Review of Financial Studies 31*(6), 2157–2183.

A Base Case Proofs and Claims

A.1 Useful Results

Lemma 4 *$(1 - F(x))x$ is unimodal.*

Proof. Note that

$$\begin{aligned}[(1 - F(x))x]' &= 1 - F(x) - xf(x) \\ &= [1 - F(x)](1 - x\frac{f(x)}{1 - F(x)}) \ .\end{aligned}$$

$1 - F(x) \geq 0$ and $1 - x\frac{f(x)}{1-F(x)}$ is continuous and decreasing. Thus, either there exists a unique x^* such that $1 - x^*\frac{f(x^*)}{1-F(x^*)} = 0$ or $1 - x\frac{f(x)}{1-F(x)} > 0$ for all $x \in [0, \hat{x}]$. In the latter case let $x^* = \hat{x}$. Easy to see that for $x > x^*$, $[(1 - F(x))x]' < 0$; for $x < x^*$, $[(1 - F(x))x]' > 0$. ■

A.2 No HFT

A.2.1 Proof of Theorem 1

Proof. First consider a relaxed problem with $d = w - q \in [-\bar{q}, w]$. Conjecture that the optimal policy is $d_t = w_t - \bar{q}$ and $x_t = x^*$ where $x^* = argmax(1 - F(x))x$, forall t. If this policy is indeed the optimal policy for this relax problem, then for $w_0 \geq \bar{q}$, this optimal policy is applicable and thus also optimal for the original more constrained problem. This proposition also implies that the market maker's payoff is linear in w_0 with $w_0 \geq \bar{q}$.

We use a method similar to one-shot deviation principle to establish the optimality of proposed policy. Notice that although the market maker discounts future dividends, the per-period dividend does not necessarily have a uniform bound. Thus, I directly check that this problem is continuous at infinity.

Consider two dividend and pricing policies $\{d_t, x_t\}_{t=0}^{\infty}$ and $\{\tilde{d}_t, \tilde{x}_t\}_{t=0}^{\infty}$. d_t, x_t, $\tilde{d}_t$, $\tilde{x}_t$ are functions of h^t, the history of the first $t-1$ periods.[44] We suppress the dependence for the ease of notation. Consider the case when $d_t = \tilde{d}_t$ and $x_t = \tilde{x}_t$ for $t \leq T$. Define the absolute value of the difference in expected payoffs between two policies to be D_T.

[44] We define $h^0 = \emptyset$.

We have

$$\begin{aligned}
D_T &= |E_0(\sum_{i=T+1}^{\infty} \delta^i (d_i - \tilde{d}_i))| \\
&\leq |E_0(\sum_{i=T+1}^{\infty} \delta^i c_i)| + \delta^{T+1}\frac{1}{1-\delta}\bar{q} \\
&\leq \delta^{T+1}E_0(w_{T+1}) + \sum_{i=T+1}^{\infty} \delta^i \bar{x} E_G(q) + \delta^{T+1}\frac{1}{1-\delta}\bar{q} \\
&= \delta^{T+1}E_0(w_{T+1}) + \delta^{T+1}\frac{1}{1-\delta}\bar{x}E_G(q) + \delta^{T+1}\frac{1}{1-\delta}\bar{q} \ .
\end{aligned}$$

The first inequality is because the worst dividend plan after period T is to pay $-\bar{q}$ for all periods. The second inequality is because for any period t, the expect profit is $(1 - F(x_t))x_t E_G(min(q, w_t - d_t))$.[45] This is uniformly bounded by $\bar{x}E_G(q)$. Thus, in each period, the expected dividend is bounded by $\bar{x}E_G(q)$ plus part of the market maker's net worth in period $T + 1$. Notice that commit more shares cannot improve the expected dividend bound since $E_G(min(q, w)) \leq E_G(q)$. Thus, the expected discounted dividend payout is bounded by the case when the market maker pays dividend equal to the entire net worth in period $t = T + 1$ and pays the upper bound of expected profit in each period.

Notice that $\delta^{T+1}\frac{1}{1-\delta}\bar{x}E_G(q) \to 0$ and $\delta^{T+1}\frac{1}{1-\delta}\bar{q} \to 0$ as $T \to \infty$. Moreover, $E_t(w_{t+1}) \leq w_t + \bar{x}E_G(q) + \bar{q}$. This implies that

$$\delta^{T+1}E_0(w_{T+1}) \leq \delta^{T+1}[w_0 + (T + 1)(\bar{x}E_G(q) + \bar{q})] \ . \tag{15}$$

Thus, $\delta^{T+1}E_0(w_{T+1}) \to 0$ as $T \to \infty$. Thus, for any two policies that different only after period T, as $T \to \infty$, $D_T \to 0$.

Since this game is continuous at infinity, if there exists a profitable deviation, then there exists a profitable deviation such that the deviating policy is different from the candidate policy for finite periods. Consider a deviation where the the deviating policy is different from the candidate policy for n periods. For $t \geq n$, the deviating policy switches back to the candidate policy $\hat{d}_t = w_t - \bar{q}$ and $\hat{x}_t = x^*$. Consider the market maker in period $t = n$ with net worth w_n. Suppose the deviating policy specifics $\hat{d}_n = w_n - \hat{w}$ and $x_n = \hat{x}_n$. Then in period n, the difference between expected

[45] E_G means q follows distribution G, I suppress the time notation because demands are *i.i.d.*

payoffs of two policies is

$$\begin{aligned}E_n(d_n - \hat{d}_n + \delta d_{n+1} - \delta \hat{d}_{n+1}) &= \hat{w} - \bar{q} + \delta(1 - F(x^*))x^* E_G(min(q, \bar{q})) \\ &\quad - \delta(1 - F(\hat{x}_n))\hat{x}_n E_G(min(q, \hat{w})) - \delta(\hat{w} - \bar{q}) \\ &\geq (1 - \delta)(\hat{w} - \bar{q}) + \delta(1 - F(x^*))x^*[E_G(min(q, \bar{q})) - E_G(min(q, \hat{w}))] \ .\end{aligned}$$

The inequality follows from $(1 - F(\hat{x}_n))\hat{x}_n \leq (1 - F(x^*))x^*$.

Define

$$A(y) = (1 - \delta)(y - \bar{q}) + \delta(1 - F(x^*))x^*[E_G(min(q, \bar{q})) - E_G(min(q, y))] \ .$$

Then

$$A^{'}(y) = 1 - \delta - \delta(1 - F(x^*))x^*(1 - G(y)) \ ,$$

$$A^{''}(y) = g(y) > 0 \ .$$

Since $A^{'}(y)$ is monotone, $A^{'}(y) = 0$ has at most one solution and upon which $A(y)$ achieves minimum. Note that $A^{'}(y) = 0$ implies

$$\frac{\delta}{1 - \delta}(1 - F(x^*))x^*(1 - G(y)) = 1 \ .$$

Thus, $A(y)$ achieves minimum at $y = \bar{q}$ and $A(\bar{q}) = 0$. Thus,

$$E_n(d_n - \hat{d}_n + \delta d_{n+1} - \delta \hat{d}_{n+1}) \geq 0 \ . \tag{16}$$

This implies that if there exists a profitable deviation such that the deviating policy differs from the candidate policy for n periods, then in period n, the market maker should adopt the candidate policy. Same reasoning then shows that the market maker should adpot the candidate policy in period $n - 1$. The backward induction goes back to period 1. Since n is arbitrary and this problem is continuous at infinity, no profitable deviation exists and the candidate policy is optimal.

■

A.2.2 Existence of Value Function

Proposition 17 *There's a unique V such that it is continuous and strictly increasing in w.*

Proof. We focus on $V(w)$ for $w \in [0, \hat{w}]$. Moreover, since $V(w) = w - \bar{q} + V(\bar{q})$ Define operator T to be

$$\begin{aligned}(Tl)(w) =& sup_{d,x} d + \delta\{F(x)l(w-d) \\ & + (1-F(x))[\int_0^{w-d} (l(min(\bar{q}, w-d+xq)) + max(0, w-c+xq-\bar{q})g(q)dq + \\ & (1-G(w-d))(l(min(\bar{q},(1+x)(w-d))) + max(0,(1+x)(w-d)-\bar{q}))]\}\end{aligned} \tag{17}$$

satisfying $c \in [0, w]$.

First check that for large enough $\bar{K}$, $l(w) \leq \bar{K} \implies Tl(w) \leq \bar{K}$. Thus, the value function is bounded and Blackwell condition is applicable. Easy to check T satisfies monotonicity and discounting.

By contract mapping theorem, operator T has a unique fixed point V. Easy to see T maps increasing functions to strictly increasing functions. This implies V must be increasing. ■

A.3 Sequential Pricing

A.3.1 Proof of Lemma 1

Proof. If $x_m > x^*$, since $argmax_x (1-F(x))x = x^*$, the HFT's optimal strategy is to set $x_h = x^*$. Consider the situation when $x_m \leq x^*$. For $x_h \leq x_m$, the HFT's expected net profit is

$$(1 - F(x_h))x_h k(q_h) ,$$

which attains maximum at $x_h = x_m$ by lemma A.1. For $x_h > x_m$, the HFT's expected net profit is

$$(1 - F(x_h))x_h[k(q_h + q_m) - k(q_m)] ,$$

which attains maximum at $x_h = x^*$. ■

A.3.2 Proof of Lemma 2

Proof. First notice that $x_m > x^*$ cannot be optimal. If $x_m > x^*$, the HFT's best response is to set $x_h = x^*$ and the market maker's expected net profit is

$$(1-F(x_m))x_m[\pi(k(q_h+q_m)-k(q_h))+(1-\pi)k(q_m)] < (1-F(x^*))x^*[\pi(k(q_h+q_m)-k(q_h))+(1-\pi)k(q_m)] \ .$$

This implies the market maker will be better off by setting $x_m = x^*$.

Next, there is a unique $\underline{x} < x^*$ such that if $x_m = \underline{x}$, the HFT is indifferent between $x_h = x^*$ and $x_h = x_m$. For any $x_m < x^*$, the HFT's expected net profit with $x_h = x^*$ is

$$(1-F(x^*))x^*[k(q_h+q_m)-k(q_m)] \ ;$$

the HFT's expected net profit with $x_h = x_m$ is

$$(1-F(x_m))x_m k(q_h) \ .$$

Since $(1-F(x))x$ is increasing for $x \in [0, x^*]$ and $k(q_h+q_m)-k(q_m) < k(q_h)$, there exists a unique $\underline{x} \in (0, x^*)$ such that

$$(1-F(\underline{x}))\underline{x}k(q_h) = (1-F(x^*))x^*[k(q_h+q_m)-k(q_m)] \ ,$$

or equivalently,

$$a(\underline{x})k(q_h) = k(q_h+q_m)-k(q_m) \ .$$

Finally, check that any other pricing strategy of the market maker is dominated either by $x_m = x^*$ or $x_m = \underline{x}$. If $x_m \in (\underline{x}, x^*)$, the HFT would set $x_h = x_m$. The market maker's expected net profit is

$$(1-F(x_m))x_m[\pi(k(q_h+q_m)-k(q_h))+(1-\pi)k(q_m)] < (1-F(x^*))x^*[\pi(k(q_h+q_m)-k(q_h))+(1-\pi)k(q_m)] \ .$$

Thus, he would be better off switch to $x_m = x^*$. For $x_m \in (0, \underline{x})$, the HFT would set $x_h = x^*$. The market maker's expected net profit is

$$(1-F(x_m))x_m k(q_m) < (1-F(\underline{x}))\underline{x}k(q_m) \ .$$

This suggests that he would be better off to set $x_m = \underline{x}$. ∎

A.3.3 Proof of Proposition 1

Proof. For any q_m, the tight spread can be determined by the equation

$$a(\underline{x}(q_m)) = \frac{k(q_m + q_h) - k(q_m)}{k(q_h)} \ . \tag{18}$$

The tight spread strategy strategy is optimal if

$$a(\underline{x}(q_m))k(q_m) \geq \pi[k(q_m + q_h) - k(q_h)] + (1 - \pi)k(q_m) \ . \tag{19}$$

Subtract $k(q_m)$ from both sides,

$$\frac{k(q_m + q_h) - k(q_m) - k(q_h)}{k(q_h)}k(q_m) \geq \pi[k(q_m + q_h) - k(q_h) - k(q_m)] \ . \tag{20}$$

Since $k(q_m + q_h) - k(q_h) - k(q_m) < 0$ for $q_m > 0$, $q_h > 0$, we have

$$\frac{k(q_m)}{k(q_h)} \leq \pi \ . \tag{21}$$

■

A.3.4 Proof of Theorem 2

Proof. Consider a relaxed problem where $d_t \in [-\bar{q}, w_t]$. Given HFT's best response , this problem can be reduced to a decision problem of the market maker. Suppose the policy proposed in this theorem is not optimal. Using the same argument as in the proof of theorem 1, this game is continuous at infinity. Thus, I can focus on considering a finite period deviation. Consider a better policy with deviation for at most n periods. In period n, I only need to consider the difference of dividends in period n and $n+1$. If $d_n \neq w_n - q_m$, by Proposition 1, the market maker's optimal strategy is to set $x_m = \hat{x}_m(q_m)$ and get expected net profit $M(q_m)$. This is exactly the original policy. Suppose $d_n = w_n - \hat{w}$. Since the market maker's maximum expected profit in period n is $M(\hat{w})$,

$$w_n - \hat{w} + \delta[M(\hat{w}) + (\hat{w} - q_m)] > w_n - q_m + \delta M(q_m) \ .$$

This implies

$$\frac{\delta}{1-\delta}M(\hat{w}) - \hat{w} > \frac{\delta}{1-\delta}M(q_m) - q_m \ .$$

Since $q_m = argmax_{w\in[0,\bar{q}]}\frac{\delta}{1-\delta}M(w)+(w_0-w)$, if such $\hat{w}$ exists, it must be $\hat{w} > \bar{q}$. Since $M(w) = max((1-F(x^*))x^*[k(w+q_h)-k(q_h)], (1-F(\underline{x}(w))\underline{x}(w))k(w))$ is continuous and differentiable almost everywhere. Easy to see that $\frac{\delta}{1-\delta}M'(w) \leq 1$ for $w \geq \bar{q}$. Thus, if $\hat{w} > \bar{q}$,

$$\frac{\delta}{1-\delta}M(\bar{q}) - \bar{q} \geq \frac{\delta}{1-\delta}M(\hat{w}) - \hat{w} \ .$$

This is because

$$\frac{\delta}{1-\delta}M(\hat{w}) = \frac{\delta}{1-\delta}M(\bar{q}) + \int_{\bar{q}}^{\hat{w}} \frac{\delta}{1-\delta}M'(x)dx \ .$$

This implies that any n period deviation can be dominated by a $n-1$ period deviation for all n. Repeating this argument implies that no finite period deviation exists and establishes the optimality of the proposed policy. Since $w_0 > \bar{q}$, the proposed policy is implementable in the original problem and is thus optimal. The HFT's optimality condition is satisfied since the HFT always plays the best response. ■

A.3.5 Proof of Corollary 2

Proof. Two conditions are derived from the first order condition of $max_{w\in[0,\bar{q}]}\frac{\delta}{1-\delta}M(w)+(w_0-w)$. To see the market maker never fully exit the market, notice that $\underline{x} \to x^*$ when $x_m \to 0$. Since $\bar{q} > 0$, $\frac{\delta}{1-\delta}(1-F(x^*))x^* > 1$. Then there always exists a $q_m > 0$ such that $\frac{\delta}{1-\delta}(1-F(x^*))x^*(1-G(q_m)) = 1$. ■

A.3.6 Proof of Theorem 3

Proof. For the ease of notation, let q_m^π be the equilibrium capital commitment of the market maker when the HFT's entry probability is π. Let $\underline{x}(q)$ be the tight spread when the market maker's shareholding is q. Notice that $\underline{x}$ does not depend on π. Consider a sequential pricing game with $\pi = 1$. If in the steady state equilibrium, the market maker uses the wide spread strategy with shareholding q_m^1, then by Theorem 2, for any $q \in [0, \bar{q}]$,

$$\frac{\delta}{1-\delta}(1-F(x^*))x^*[k(q_m^1+q_h)-k(q_h)]+(w_0-q_m^1) \geq \frac{\delta}{1-\delta}(1-F(\underline{x}(q)))\underline{x}(q)k(q)+(w_0-q) \ .$$

That is, adopting the wide spread strategy with shareholding q_m^1 is better than using the tight spread strategy at any level of shareholding. Then for $\pi < 1$ and any q,

$$
\begin{aligned}
&\frac{\delta}{1-\delta}(1-F(x^*))x^*\{\pi[k(q_m^\pi+q_h)-k(q_h)]+(1-\pi)k(q_m^\pi)\}+(w_0-q_m^\pi)\\
>&\frac{\delta}{1-\delta}(1-F(x^*))x^*[k(q_m^1+q_h)-k(q_h)]+(w_0-q_m^1)\\
\geq&\frac{\delta}{1-\delta}(1-F(\underline{x}(q)))\underline{x}(q)k(q)+(w_0-q)\ .
\end{aligned}
$$

Thus, for $\pi < 1$, the market maker's equilibrium strategy must still be the wide spread strategy. This corresponds to the case where $\hat{\pi} = 1$.

If the market maker is using the tight spread strategy at a $\pi_1 < 1$, then for $\pi_2 > \pi_1$, by a similar argument with Proposition 2, the market maker would still use the tight spread strategy. Moreover, $q_m^{\pi_1} = q_m^{\pi_2}$ and thus $\underline{x}(q_m^{\pi_1}) = \underline{x}(q_m^{\pi_2})$ and the market maker has the same equilibrium payoff. Denote this equilibrium payoff when the market maker is using a tight spread strategy by V^{tight}. Define

$$
V_\pi^{wide} = \frac{\delta}{1-\delta}(1-F(x^*))x^*[\pi(k(q+q_h)-k(q_h))+(1-\pi)k(q)]+(w_0-q)
$$

where q satisfies

$$
\frac{\delta}{1-\delta}(1-F(x^*))x^*[1-\pi G(q+q_h)-(1-\pi)G(q)] = 1\ .
$$

V_π^{wide} is the equilibrium payoff for the market maker if the wide spread strategy is adopted in the equilibrium. V_π^{wide} is continuous and decreasing with respect to π. Moreover, V_0^{wide} goes to the monopolistic payoff. Since $V^{tight} > V_1^{wide}$ and V^{tight} is bounded away from the monopolistic payoff, there exist $\hat{\pi} \in (0,1)$ such that $V_{\hat{\pi}}^{wide} = V^{tight}$. By previous argument, the market maker adopts the wide spread strategy if $\pi < \hat{\pi}$ and tight spread strategy if $\pi > \hat{\pi}$.

In the tight spread region,

$$
L = (1-F(x_m))k(q_m)+\pi(F(x^*)-F(x_m))(k(q_m+q_h)-k(q_m))\ .
$$

Since in the tight spread region, q_m and $x_m = \underline{x}(q_m)$ is not changing with respect to π, L is increasing in π.

For the third statement, consider a game at $\pi = \hat{\pi} < 1$. Two equilibrium shareholdings for market maker, w_m^{tight} and w_m^{wide} both exist. If the market maker chooses shareholding q_m^{tight} (q_m^{wide}), he will play the tight (wide) spread strategy in the equilibrium. By Proposition 1, $k(q_m^{wide}) \geq \hat{\pi}k(q_h) \geq k(q_m^{tight})$. This implies $q_m^{wide} \geq q_m^{tight}$. For $\pi < \hat{\pi}$, $q_m > q_m^{wide} \geq q_m^{tight}$. This establishes that the market maker always have a higher equilibrium shareholding in the wide spread region. ■

A.3.7 Proof of Proposition 3

Proof. Notice that in the wide spread region, L is continuous in π. Moreover, if the wide spread region is $[0, 1]$, liquidity is the same at $\pi = 0$ and $\pi = 1$. These two observations imply the proposition. ■

A.3.8 Proof of Proposition 4

Proof. For any $w \geq 0$, given G is an exponential distribution, $k(s) = E_G(min(q, s)) = E_G(q)G(s)$. By theorem 1, when no HFT exists, the market maker's capital commitment $\bar{q}$ satisfies $\frac{\delta}{1-\delta}(1 - F(x^*))x^*(1 - G(\bar{q})) = 1$. By corollary 2, when the market maker posts a wide spread in the equilibrium, his capital commitment satisfies $\frac{\delta}{1-\delta}(1 - F(x^*))x^*[(1 - \pi)(1 - G(q_m)) + \pi(1 - G(q_m + q_h))] = 1$. Thus, $G(\bar{q}) = \pi G(q_m + q_h) + (1 - \pi)G(q_m)$.

Then,

$$\begin{aligned} k(\bar{q}) &= E_G(q)G(\bar{q}) \\ &= E_G(q)(\pi G(q_m + q_h) + (1 - \pi)G(q_m)) \\ &= \pi k(q_m + q_h) + (1 - \pi)k(q_m)\ . \end{aligned} \tag{22}$$

This implies that liquidity does not depend on π in the wide spread region and is equal to the liquidity in a monopolistic market. ■

A.3.9 Proof of Theorem 4

Proof. Let's consider the first statement. Since I take other parameters as fixed and only change π, I represent liquidity by $L(\pi)$ and the market maker's capital commitment by $q_m(\pi)$ to make their dependences on π explicit while suppressing all other dependences.

As $\pi \to 0$, the market maker's payoff by posting the wide spread converges to the monopolistic payoff. By continuity of the market maker's payoff, for π small enough, the market maker would post a wide spread in the steady state equilibrium. In the wide spread region, the market maker's capital commitment $q_m(\pi)$ satisfies

$$\frac{\delta}{1-\delta}(1-F(x^*))x^*[(1-\pi)(1-G(q_m(\pi)))+\pi(1-G(q_m(\pi)+q_h))]=1\ . \tag{23}$$

Take derivative with respect to π,

$$G(q_m(\pi))-G(q_m(\pi)+q_h)-\pi g(q_m(\pi)+q_h)q_m^{'}(\pi)-(1-\pi)g(q_m(\pi))q_m^{'}(\pi)=0\ . \tag{24}$$

Collecting terms to get

$$q_m^{'}(\pi)=\frac{G(q_m(\pi))-G(q_m(\pi)+q_h)}{\pi g(q_m(\pi)+q_h)+(1-\pi)g(q_m(\pi))}\ . \tag{25}$$

In the wide spread region, $L(\pi)=(1-F(x^*))[(1-\pi)k(q_m(\pi))+\pi k(q_m(\pi)+q_h)]$. Then

$$\begin{aligned}\frac{1}{1-F(x^*)}L^{'}(\pi)=&k(q_m(\pi)+q_h)-k(q_m(\pi))+\pi(1-G(q_m(\pi)+q_h))q_m^{'}(\pi)\\&+(1-\pi)(1-G(q_m(\pi)))q_m^{'}(\pi)\ .\end{aligned} \tag{26}$$

Easy to see this function is continuous in π. Consider $L^{'}(\pi)$ at $\pi=0$. Since $q_m(0)=\bar{q}$,

$$\frac{1}{1-F(x^*)}L^{'}(0)=k(\bar{q}+q_h)-k(\bar{q})+(1-G(\bar{q}))\frac{G(\bar{q})-G(\bar{q}+q_h)}{g(\bar{q})}\ . \tag{27}$$

$L^{'}(0)<0$ if and only if

$$\frac{G(\bar{q}+q_h)-G(\bar{q})}{k(\bar{q}+q_h)-k(\bar{q})}>\frac{g(\bar{q})}{1-G(\bar{q})}\ . \tag{28}$$

Use integration by parts,

$$
\begin{aligned}
k(s) &= s(1-G(s)) + \int_0^s qg(q)dq \\
&= s(1-G(s)) + sG(s) - \int_0^s G(q)dq \\
&= s - \int_0^s G(q)dq \\
&= \int_0^s (1-G(q))dq \ .
\end{aligned}
\tag{29}
$$

Thus, $L^{'}(0) < 0$ if and only if

$$
\frac{\int_{\bar{q}}^{\bar{q}+q_h} g(q)dq}{\int_{\bar{q}}^{\bar{q}+q_h} (1-G(q))dq} > \frac{g(\bar{q})}{1-G(\bar{q})} \ . \tag{30}
$$

Let

$$
I(x) = \int_{\bar{q}}^{\bar{q}+x} g(q)dq - \frac{g(\bar{q})}{1-G(\bar{q})} \int_{\bar{q}}^{\bar{q}+x} (1-G(q))dq \ .
$$

Inequality (30) holds if and only if $I(q_h) > 0$. Notice that $I(0) = 0$. Moreover,

$$
I^{'}(x) = g(\bar{q}+x) - \frac{g(\bar{q})}{1-G(\bar{q})}(1-G(\bar{q}+x)) \ .
$$

Since $\frac{g(x)}{1-G(x)}$ is increasing, for $x > 0$, $I^{'}(x) > 0$. Thus, $I(q_h) > 0$ and $L^{'}(0) < 0$. Then by continuity of $L^{'}(\pi)$, there exists a small region around 0 such that liquidity is decreasing in π.

Notice that the calculation above works for the situation when $\bar{q} + q_h$ is in the support of G. If $\bar{q}+q_h$ is not in the support of G, replace $\bar{q}+q_h$ with the upper-bound of G's support yields the same result.

For the increasing part, it is suffice to consider the situation where $\pi = 1$ is in the tight spread region. Since liquidity is increasing with π in the tight spread region, there exists $\tilde{\pi}$ such that liquidity is increasing for $\pi \in [\tilde{\pi}, 1]$. This finish the proof of the first statement.

Now I consider the second statement. Fix $\pi = 1$. Notice that for any fixed $q_m > 0$,

$$
a(\underline{x}) = \frac{k(q_m + q_h) - k(q_m)}{k(q_h)} \to 1 - G(q_m) < 1 \text{ as } q_h \to 0 \ .
$$

This implies that the market maker's payoff by using the tight spread strategy is bounded away from the monopolistic payoff as $q_h \to 0$. On the other hand, if the market maker uses the wide spread strategy, easy to see as $q_h \to 0$, the expected payoff converges to the monopolistic payoff. Thus, for small enough q_h, the market maker would use the wide spread strategy at the steady state even when $\pi = 1$. This finish the proof of the second statement. ■

A.4 Simultaneous Pricing

A.4.1 Proof of Proposition 5

Proposition 5 can be divided into following claims.

Claim 1 *Players never propose spreads greater than x^*.*

Proof. If a player propose a spread greater than x^*, regardless of the other player's strategy, switching to proposing x^* yields a strictly larger payoff. ■

Claim 2 *Neither players would use pure strategies in an equilibrium.*

Proof. Suppose the market maker posts spread $x_m = x$ in a equilibrium. The HFT's optimal strategy would be posting $x_h = x^*$, $x_h = x$ or a mix between these two price. Then the market maker would achieve higher payoff by undercutting the HFT's lowest possible price for a small enough ϵ. Contradiction

Suppose the HFT post spread $x_h = x$ in an equilibrium. Then in an equilibrium the market maker can only post x^*. (Undercutting will lead to no equilibrium because the payoff of the market maker is not continuous at x.) This implies $x_h \neq x^*$ in the equilibrium. However, if $x_h < x^*$, given the market maker is posting $x_m = x^*$, the HFT would be better off posting $x_h = x^*$. Contradiction. ■

Suppose there exists a mixed strategy equilibrium. Denote the infimum and supremum of the spread posted by the market maker (HFT) by $\underline{x}_m(\underline{x}_h)$ and $\bar{x}_m(\bar{x}_h)$.

Claim 3 *$\underline{x}_m = \underline{x}_h$ and neither the market maker nor the HFT would post this spread with positive probability in an equilibrium.*

Proof. If $\underline{x}_m \neq \underline{x}_h$, the player with smaller spread lower-bound could raise the lower-bound by a small enough amount to achieve a higher payoff. Denote this common lower-bound by $\underline{x}$. If the HFT posts this spread with positive probability, rather than posting $\underline{x}$, the market maker would be strictly better off undercutting the HFT for a small amount.

Suppose the market maker posts $\underline{x}$ with positive probability. Let $B(x, r)$ be a open ball centered at x with radius r. First note that $\forall \epsilon > 0$, $\exists x_h \in B(\underline{x}, \epsilon)$ such that x_h is in HFT's mixed strategy's support. If not, since $\underline{x}$ is posted by the HFT with zero probability, the market maker can increase $\underline{x}_m$ by ϵ to achieve higher payoff. Then for small enough ϵ, HFT's profit of posting $\underline{x} + \epsilon$ is strictly smaller than posting $\underline{x}$. Contradiction. ∎

Claim 4 *(No Holes) $\nexists a, b \in (\underline{x}, \bar{x}_m)$, $a < b$ such that $(a, b) \bigcap X_m = \emptyset$. A similar claim holds for X_h.*

Proof. Suppose this claim is false. Without loss of generality, let (a, b) be the maximum interval satisfying the claimed property. That is, $(a, b) \bigcap X_m = \emptyset$ and for any $a^{'} < a$ and $b^{'} > b$, $(a^{'}, b) \bigcap X_m \neq \emptyset$ and $(a, b^{'}) \bigcap X_m \neq \emptyset$.

By claim 1, $\bar{x}_m, \bar{x}_h \leq x^*$. Notice that if $(a, b) \notin X_m$, then $(a, b) \notin X_h$. This is because if $x \in (a, b)$ and $x \in X_h$, the HFT may increase x by a small amount to increase her payoff.

Then notice that $a \notin X_m$. This is because posting $x_m \in (a, b)$ will achieve a higher payoff given $(a, b) \notin X_h$. Moreover, $a \notin X_h$ by a similar argument.

Given that spread a is not posted by the HFT and the market maker with positive probability, when $x_m \to a$ from below, the payoff goes to the payoff of posting $x_m = a$ by continuity, which is smaller than posting $x_m \in (a, b)$. Since (a, b) is a maximum interval satisfying $(a, b) \bigcap X_m = \emptyset$, $\forall \epsilon > 0$, $B(a, \epsilon) \bigcap X_m \neq \emptyset$. This contradicts the equilibrium definition that $x_m \in X_m$ is a best response to the HFT's pricing strategy.
∎

Claim 5 $\bar{x}_m = \bar{x}_h = x^*$.

Proof. Suppose that $\bar{x}_m < \bar{x}_h$. Then $(\bar{x}_m, \bar{x}_h) \bigcap X_h = \emptyset$ since posting $x_h = \bar{x}_h$ yields a higher payoff. This contradicts Claim 4. Similarly, it is impossible that $\bar{x}_m > \bar{x}_h$. If $\bar{x}_m = \bar{x}_h < x^*$, $\bar{x}_m \notin X_m$ since $x_m = x^*$ would yield higher payoff. Since $\bar{x}_m \notin X_m$, by the same argument, $\bar{x}_h \notin X_h$. However, then by the continuity argument, for small enough ϵ, $x_m \in B(\bar{x}_m, \epsilon)$ will be dominated by posting $x_m = x^*$. Contradiction. ∎

Claim 6 *For all $x \in (\underline{x}, x^*) \bigcap X_m$, x is not proposed by the market maker with positive probability in an equilibrium. For all $x \in (\underline{x}, x^*) \bigcap X_h$, x is not proposed by the HFT with positive probability in an equilibrium.*

Proof. We prove by contradiction. Suppose that the market maker posts spread x with positive probability. Then by claim 4, $\forall \epsilon > 0$, $B(x+\epsilon, \epsilon) \bigcap X_h \neq \emptyset$. However, by continuity, when ϵ is small, the payoff posting that spread is dominated by posting x. Contradiction. If the HFT posts spread x, note that the market maker's profit when posting a spread approaching x from the left is larger than the profit when posting a spread approaching x from the right. This leads to a contradiction. ■

A.4.2 Proof of theorem 5

Proof. The proof of the first part is the same as the proof of Theorem 2. For the second and the third statement, note that expected payoffs of the market maker are the same in all one-shot games. Thus, in the equilibrium the market maker commits the same amount of capital to the market. The HFT's payoffs can be calculated from the corresponding one-shot game. ■

A.4.3 Proof of proposition 7

Proof. For the first statement, notice that

$$L_{se} = (1 - F(x^*))[\pi k(q_m + q_h) + (1 - \pi)k(q_m)] \ .$$

Compare this to L_{sim} in Theorem 5 to reach the conclusion.

Notice that I have shown that L_{se} is increasing in π. Thus, the third statement is merely a corollary of the second statement. If π is in the tight spread region, in equilibrium, $k(q_m) \leq \pi k(q_h)$ and $a(\underline{x})$ is not changing with π. Moreover, q_m also remains constant with respect to π. Then by the market maker's indifference condition, for all $x \in (\underline{x}, x^*)$,

$$a(x)\{(1 - \pi)k(q_m) + \pi[H_h(x)(k(q_m + q_h) - k(q_h)) + (1 - H_h(x))k(q_m)]\} \qquad (31)$$

is constant for all π in the tight spread region. This implies for any given x, $\pi H_h(x)$ is constant for all π in the tight spread region. This together with Theorem 5 implies that $L_{sim} - L_{se}$ is constant. It also implies that in the tight spread region, increase in π only benefits buyers with buying thresholds higher than $1 + x^*$. ■

B Extension: Costly Entry

B.1 Sequential Pricing

B.1.1 Proof of Proposition 8

Proof. If $C \geq \bar{C} = \pi(1 - F(x^*))x^* k(q_h)$, the expected return of the HFT cannot cover the cost even when the HFT undercuts the market maker at spread x^*. Thus, the HFT will not enter the market regardless of the market maker's spread. In equilibrium, the market maker would choose $x_m = x^*$.

Now consider the situation where $C < \bar{C}$. In this case, if the market maker posts the wide spread x^*, the HFT would attempt to enter the market and undercut the market maker upon entry. Moreover, the HFT would not choose to enter and undercut the market maker if the market maker posts the deterring spread x satisfying $\pi(1-F(x))xk(q_h) = C$. If the market maker posts a spread higher than the deterring spread x, the HFT will always enter since she can always undercut the market maker and earn a expected payoff higher than C.

If $k(q_m) < \pi k(q_h)$, given the HFT chooses to enter the market, the market maker's optimal spread is the tight spread satisfying $(1-F(x))xk(q_h) = (1-F(x^*))x^*[k(q_m + q_h) - k(q_m)]$. Moreover, as long as the HFT does not undercut the market maker, the market maker always prefers to set the spread x_m higher (given $x_m \leq x^*$). Thus, in equilibrium, the market maker will compare the tight spread and the deterring spread and pick the greater one. Specifically, if $C > \pi(1 - F(x^*))x^*[k(q_m + q_h) - k(q_m)]$, posting the deterring spread is more profitable. Otherwise, posting the tight spread is more profitable. Furthermore, when facing the tight spread, the HFT is indifferent between posting the monopolistic spread and undercutting the market maker. Then when the market maker posts the deterring spread, upon entering, the HFT is better off undercutting the market maker. This implies that when the market maker posts the deterring spread, the HFT will choose not to try to enter the market. The discussion for $k(q_m) > \pi k(q_h)$ follows the similar logic and is thus omitted. ■

B.1.2 Proof of Theorem 6

Proof. Let $\underline{x}^d$ satisfies $\pi(1-F(\underline{x}^d))\underline{x}^d k(q_h) = C$ for $C \in [0, \bar{C}]$. Let q_m^d satisfies $\frac{\delta}{1-\delta}(1-F(\underline{x}^d))\underline{x}^d(1-G(q_m^d)) = 1$. This is the equilibrium capital commitment if the market maker uses a deterring entry strategy. The equilibrium payoff is $V_C(w_0) = \frac{\delta}{1-\delta}(1-F(\underline{x}^d))\underline{x}^d k(q_m^d) + (w_0 - q_m^d)$. Easy to see that this quantity is increasing in C. Easy to see that when $C \geq \bar{C}$, this quantity becomes monopolistic payoff. Let the market maker's equilibrium payoff when $C = 0$ be $V_0(w_0)$. There exist a unique $\hat{C}$ such that $V_{\hat{C}}(w_0) = V_0(w_0)$. Thus, for $C > \hat{C}$, the market maker is using the deterring strategy in the equilibrium.

When the market maker is using the deterring strategy, suppose the HFT chooses to participate, then she optimally set $x_h = x^*$. Since (1) the HFT is not undercutting the market maker, and (2) when the HFT participates, her optimal pricing strategy does not depend on C, when $C = 0$, the market maker can use the same equilibrium strategy to achieve a higher expected payoff. Contradiction. Thus, the HFT does not choose to participate. ■

B.2 Simultaneous Pricing

B.2.1 Proof of Proposition 9

Proof. First consider the case where $C > \bar{C}$. In this case, the HFT's expect profit can never cover the cost regardless of the market maker's pricing strategy. Thus, $\eta = 0$ and the market maker sets $x_m = x^*$.

Now consider the situation when $C \in [0, \bar{C}]$. Suppose the HFT chooses $\eta = 1$ and plays a mixed pricing strategy as in a game $(q_m, q_h, \pi, 0)$. By Proposition 6, the HFT's expected profit is $\pi(1-F(x^*))x^* a(\underline{x})(\pi)k(q_h)$. If $\pi(1-F(x^*))x^* a(\underline{x})(\pi)k(q_h) \geq C$, since C is paid at the end of the period, the equilibrium characterized by Proposition 6 still holds.

If $\pi(1-F(x^*))x^* a(\underline{x})(\pi)k(q_h) < C < \bar{C}$, note that $\eta \neq 0$ in the equilibrium. This is because if $\eta = 0$, the market maker would post $x_m = x^*$. The HFT has incentive to deviate to $\eta = 1$. Thus, I need to consider an equilibrium where the HFT mixes between participating. In other words, $\eta \in (0, 1)$. η can be pinned down by the indifference condition that the HFT earns zero profit when trying to enter the market.

First consider the situation $k(q_m) \geq \pi k(q_h)$. By Proposition 6, if the HFT tries to enter with probability η, $\underline{x}$ is determined by

$$(1-\eta\pi)k(q_m) + \eta\pi(k(q_m+q_h) - k(q_h)) = a(\underline{x})k(q_m) \ . \tag{32}$$

Notice that $\underline{x}$ is decreasing in η and $\underline{x} \to x^*$ as $\eta \to 0$. Thus, there exist a unique $\eta \in (0,1)$ such that $\eta\pi(1-F(x^*))x^*a(\underline{x})(\eta\pi)k(q_h) = \eta C$ where $\underline{x}$ is the lower-bound of the mixed strategy in the game $(q_m, q_h, \eta\pi, 0)$. If the HFT participates with probability η and posts spread according to H_h in the game $(q_m, q_h, \eta\pi, 0)$, the market maker has no incentive to deviate from posting spread according to H_m in the game $(q_m, q_h, \eta\pi, 0)$. If the market maker sets price according to H_m, upon entering, the HFT has no incentive to deviate from posting spread according to H_h. Moreover, the HFT earns zero expected profit for trying to enter. Thus, the HFT has no incentive to deviate from η.

Next consider the situation $k(q_m) < \pi k(q_h)$. Notice that $\underline{x}$ remains constant in this region. Let $\bar{\eta}$ satisfies $k(q_m) = \bar{\eta}\pi k(q_h)$. By the same argument, there exists a unique $\eta \in (0, \bar{\eta})$ such that $k(q_m) > \pi\eta k(q_h)$ and $\pi(1-F(x^*))x^*a(\underline{x})(\pi)k(q_h) = C$ where $\underline{x}$ is the lower-bound of the mixed strategy in the game $(q_m, q_h, \eta\pi, 0)$. The rest of the verification is the same. ■

B.2.2 Proof of Corollary 4

Proof. This proof essentially involves only comparing the market maker's payoffs under two settings with different parameter values. Fix a game (q_m, q_h, π, C). First consider the case when $k(q_m) \geq \pi k(q_h)$. In the one-shot simultaneous pricing game,

$$a(\underline{x})(\pi) = 1 - \pi + \frac{k(q_m+q_h) - k(q_h)}{k(q_m)}\pi \ .$$

If $\pi(1-F(x^*))x^*a(\underline{x})(\pi)k(q_h) \geq C$, by Proposition 9, in the simultaneous pricing game, the HFT participates in high-frequency with probability 1. The market maker enjoys the same expected payoff as in the simultaneous pricing one-shot game $(q_m, q_h, \pi, 0)$, which equals to $(1-\pi)k(q_m) + \pi(k(q_m+q_h) - k(q_h))$. By Proposition 8, the market maker receives the same expected payoff in the sequential pricing one-shot game . For $\pi(1-F(x^*))x^*a(\underline{x})(\pi)k(q_h) < C$, in the simultaneous pricing game, the market maker receives payoff $(1-F(x^*))x^*a(\underline{x})(\eta\pi)k(q_m)$ by the indifference condition where

$$a(\underline{x})(\eta\pi) = \frac{C}{\pi(1 - F(x^*))x^* k(q_h)} \ .$$

Thus, the market maker's expected payoff is $\frac{C}{\pi k(q_h)} k(q_m)$, which equals to the expected payoff in a one-shot sequential pricing game by Proposition 8..

Next consider the case when $k(q_m) < \pi k(q_h)$. In a one-shot simultaneous pricing game,

$$a(\underline{x})(\pi) = \frac{k(q_m + q_h) - k(q_m)}{k(q_h)} \ .$$

If $\pi(1 - F(x^*))x^* a(\underline{x})(\pi)k(q_h) = \pi(1 - F(x^*))x^*(k(q_m + q_h) - k(q_m)) \geq C$, in a simultaneous pricing game, the market maker's expected payoff is $(1 - F(x^*)x^*)a(\underline{x})k(q_m)$. This is the same as the expected payoff in a sequential pricing game. If $\pi(1 - F(x^*))x^* a(\underline{x})(\pi)k(q_h) < C$, in a simultaneous pricing game, the market maker receives payoff $(1 - F(x^*))x^* a(\underline{x})(\eta\pi)k(q_m)$ where

$$a(\underline{x})(\eta\pi) = \frac{C}{\pi(1 - F(x^*))x^* k(q_h)} \ .$$

Thus, the market maker's expected payoff is $\frac{C}{\pi k(q_h)} k(q_m)$, which equals to the expected payoff in the sequential pricing game. ■

C Flipping

C.1 Proof of Proposition 11

Proof. Notice that

$$r^{'}(q_f) = (1 - F(x^*))x^*(1 - G(q_m - q_f)) - x_m \ ,$$

$$r^{''}(q_f) = (1 - F(x^*))x^* g(q_m - q_f) > 0 \ .$$

This implies the maximum is achieved at the boundary $q_f = 0$ or $q_f = q_m$. ■

C.2 Proof of Lemma 3

Proof. Inequality (13) guarantees that the HFT is better off with $q_f = q_m$ than with $q_f = 0$ when setting $x_h = x^*$. Inequality (14) guarantees that the HFT is better off with $w = q_m$ than undercutting the market maker. Since the market maker is better off choosing the highest possible spread given the HFT is flipping orders, one of the inequalities must be binding.

Moreover, if inequality (13) binds,

$$x_m^f = \frac{(1-F(x^*))x^*k(q_m)}{q_m} .$$

Otherwise, since $(1-F(x))x$ is increasing in $[0, x^*]$, there exists a unique

$$x_m^f \in (0, \frac{(1-F(x^*))x^*k(q_m)}{q_m})$$

such that

$$(1-F(x^*))x^*k(q_m+q_h) = x_m^f q_m + (1-F(x_m^f))x_m^f k(q_h) .$$

If inequality (14) binds,

$$(1-F(x^*))x^*k(q_m) > x_m^f q_m \tag{33}$$

and

$$(1-F(x^*))x^*k(q_m+q_h) = x_m^f q_m + (1-F(x_m^f))x_m^f k(q_h) . \tag{34}$$

Then,

$$(1-F(x^*))x^*[k(q_m+q_h)-k(q_m)] < (1-F(x_m^f))x_m^f k(q_h) . \tag{35}$$

Thus,

$$x_m^f > \underline{x} . \tag{36}$$

In this case, the tight spread strategy is never optimal because the market maker can raise the spread to x_m^f to achieve higher expected payoff. ■

C.3 Proof of Proposition 12

Proof. Consider the situation when $\pi = 1$. If inequality (13) is binding, the market maker's expected payoff with flipping is

$$x_m^f q_m = (1-F(x^*))x^*k(q_m) .$$

This is the highest possible payoff. If inequality (14) is binding, by Lemma 3, the tight spread strategy is dominated. Moreover,

$$x_m^f q_m = (1-F(x^*))x^* k(q_m+q_h) - (1-F(x_m^f))x_m^f k(q_h) > (1-F(x^*))x^*(k(q_m+q_h)-k(q_h)) \; .$$

Thus, setting $x_m = x_m^f$ is better than setting $x_m = x^*$. By continuity, for π large enough, it is always optimal to induce flipping. ■

C.4 Proof of Proposition 13

Proof. The proof is omitted since it is similar to the existence result proved in previous sections. ■

D Extension: Supply Schedule and Induced Limit Order Book

D.1 Proof of Proposition 14

Proof. Obviously, it is not optimal for the market maker to sell any share at a spread higher than x^*. Then without loss of generality, I only consider the situation where the market maker set spreads lower than x^*. The proof consists of two steps. I first show that if the market maker can supply shares with n spreads $x_1, ..., x_n$ with $\sum_{i=1}^n q_i = q_m$, then he should optimally set $x_1 = ... = x_n = x^*$. Then I show that the market maker's payoff under any supply schedule $\Psi(x)$ can be approximated with arbitrary precision by a n-spreads supply plan with a large enough n.

Consider the situation when $n = N$. Without loss of generality, suppose $x_1 \leq x_2 \leq ... \leq x_N \leq x^*$. Define $q_0 = 0$. The market maker's expected payoff is

$$\sum_{i=1}^{N}(1-F(x_i))x_i[k(\sum_{j=0}^{i} q_j) - k(\sum_{r=0}^{i-1} q_r)] \; .$$

Note that the market maker can increase his expected payoff by setting $x_1 = x_2$ since $(1-F(x))x$ is increasing in $x \in [0, x^*]$. This reduce the problem to $n = N-1$ situation. By induction, for arbitrary n, $x_1 = ... = x_n = x^*$ is the optimal supply schedule.

Next consider the approximation procedure under arbitrarily fixed q_m. For arbitrary $\Psi(x)$, divide its support into n intervals $\{I_1, ..., I_n\}$. The I_i interval is from $\frac{i-1}{n}th$ quantile to $\frac{i}{n}th$ quantile. Consider a new supply schedule that supply shares at n spreads. Specifically, in the new schedule, the market maker supplies q_i shares at spread x_i for $i = 1, ..., n$. Let $x_i = E_\Psi(x|x \in I_i)$; $q_i = \frac{q_m}{n}$ for all i. Under any buyer's demand and buying threshold, realized profits of this new schedule and schedule Ψ differ by at most a factor of $\frac{q_m}{n}$, which goes to 0 as $n \to \infty$. Thus, expected profit from any supply schedule Ψ can be approximated to an arbitrarily close level by a schedule with n spreads when n is large enough. This establish the fact that the optimal supply schedule is to sell all shares at the spread x^*. ■

D.2 Proof of Corollary 5

Proof. The first statement is a straightforward result from Proposition 14. For the second statement, if the dividend payout is non-zero, the market maker can always achieve a higher payoff by refraining from paying dividend and supply the extra amount of shares at the spread x^* and payout the total return from the extra shares in the next period. ■

D.3 Proof of Proposition 15

Proof. From the analysis of the baseline model, any single spread pricing strategy is dominated either by the wide spread strategy or the tight spread strategy. Thus, I only need to show that, when the market maker can submit a supply curve, using the wide spread strategy or the tight spread strategy is not optimal.

Suppose for some π and q_h there exists a steady state equilibrium with capital commitment q_m and supply schedule $\Psi(x) = I_{\{x \geq x^*\}}$. Then upon entering the market, the HFT would set spread $x_h = x^*$. The market maker's expected dividend payout each period would be

$$\pi(1 - F(x^*))x^*[k(q_m + q_h) - k(q_h)] + (1 - \pi)(1 - F(x^*))x^*k(q_m) .$$

Consider a deviation of the market maker by selling ϵ shares at the spread x_ϵ satisfying

$$(1 - F(x^*))x^*(k(\epsilon + q_h) - k(\epsilon)) = (1 - F(x_\epsilon))x_\epsilon k(q_h)$$

and $q_m - \epsilon$ shares at the spread x^*. Then the HFT would still set spread $x_h = x^*$ and the market maker's expected dividend payout would be

$$div(\epsilon) = (1-F(x_\epsilon))x_\epsilon k(\epsilon) + \pi(1-F(x^*))x^*[k(q_m+q_h) - k(q_h+\epsilon)] + (1-\pi)(1-F(x^*))x^*[k(q_m) - k(\epsilon)] \ .$$

Easy to check

$$div'(0) = (1 - F(x^*))x^* \pi G(q_h) > 0 \ .$$

Thus, the market maker can deviate in pricing to achieve a higher expected payoff. Contradiction.

For the tight spread strategy, a similar argument can show that the market maker can achieve higher expected payoff by increasing the spread of a small amount of shares. This completes the proof. ■

D.4 Proof of Proposition 16

Lemma 5 *In any steady state equilibrium, the HFT set $x_h = x^*$.*

Proof. Suppose not, then $\lim_{x \to x_h^-} \Psi(x) < 1$. The market maker would achieve a higher expected payoff by sell $q_m(\lim_{x \to x^{*-}} \Psi(x) - \lim_{x \to x_h^-} \Psi(x))$ shares at the spread x^*. ■

Proof. Proposition 16 First note that if $\Psi(x^*) < 1$, the market maker can become better off by selling all shares with spreads higher than x^* at spread x^*.

Suppose $\Psi(x)$ has a mass point at $x < x^*$. If the HFT is strictly prefers posting $x_h = x^*$, then there exists an ϵ such that the market maker can sell these shares at the spread $x + \epsilon$ to achieve higher payoff. If the HFT is indifferent, then there must exist an ϵ such that the HFT is strictly prefer setting $x_h = x^*$ than setting $x_h = x + \epsilon$. If the HFT is indifferent between posting x and x^*, since x is a mass point, there exists a $\epsilon > 0$ such that the HFT strictly prefers setting $x_h = x^*$ to setting $x_h \in (x, x + \epsilon)$. The market maker can then improve his pricing by selling all shares within the spread range $(x, x + \epsilon)$ and some shares at the spread x to the spread $x + \epsilon$.

The next step is to show that for any Ψ violating the indifference condition of the HFT, the market maker can always find a better pricing plan. Specifically, I consider this problem holding $q_m \Psi(x^{*-})$ and q_m constant. First notice that $\underline{x}$ can be uniquely pinned down by

$$(1 - F(x^*)x^*)k(q_h + q_m \Psi(x^{*-})) - k(q_m \Psi(x^{*-})) = (1 - F(\underline{x})\underline{x})k(q_h) \ .$$

Denote the pricing distribution satisfying the HFT's indifference condition by $\underline{\Psi}(x)$. Then for all $x \in [\underline{x}, x^*]$, $\Psi(x) \geq \underline{\Psi}(x)$. Otherwise the HFT will not set $x_h = x^*$ and the pricing distribution cannot be optimal at the steady state. Suppose $\Psi \neq \underline{\Psi}$, let $\acute{x} = inf_x\{\Psi(x) > \underline{\Psi}(x)\}$. Since $\Psi(x)$ does not have mass point, there exists $\xi > 0$ such that $\Psi(x) > \underline{\Psi}(x)$ for $x \in (\acute{x}, \acute{x} + \xi]$ and $\Psi(\acute{x} + \xi) > \Psi(\acute{x})$. Then by the same approximation and moving mass argument, the market maker is better off selling shares in the spread interval $(\acute{x}, \acute{x} + \xi)$ at the spread $\acute{x} + \xi$. ■

E Capital Commitment when G has Non-decreasing Hazard Rate

This section provides a detailed analysis of the market maker's capital commitment strategy when the buyer's demand G follows a distribution with increasing hazard rate. Particularly, under any fixed HFT shareholding q_h, the market maker has a unique optimal steady state tight spread strategy.

Proposition 18 *Let $B = \frac{\delta}{1-\delta}(1 - F(x^*))x^* > 1$.*[46] *If G has non-decreasing hazard rate, $max_y B\frac{k(y+q_h)-k(y)}{k(q_h)}k(y) + (w_0 - y)$ has a unique solution $q_m \in [0, \bar{q}]$.*

The connection between this proposition and the tight spread strategy is clear. Notice that $a(\underline{x}) = \frac{k(q_m+q_h)-k(q_m)}{k(q_h)}$. By posting spread x_m satisfying $(1 - F(x_m))x_m = (1 - F(x^*))x^*a(\underline{x})$, a short-run HFT with q_h shares has no incentive to undercut the market maker.

Proof. The first order condition is

$$W^{'}(y) = \frac{B}{k(q_h)}[(1-G(y))(k(y+q_h)-k(y)) - (G(y+q_h)-G(y))k(y)] - 1 = 0 \ . \quad (37)$$

When $y = 0$, $W^{'}(0) = B - 1 > 0$. When $y \geq \bar{q}$, $W^{'}(y) < B(1 - G(\bar{q}))\frac{k(y+q_h)-k(y)}{k(q_h)} - 1$. Since $B(1 - G(\bar{q})) = 1$, $W^{'}(y) < 0$. By continuity, $W^{'}(y)$ cross zero at least once for $y \in [0, \bar{q}]$. If I can show that $W^{'}$ only cross zero once, then a unique maximizer exists.

Consider any q_m such that $W^{'}(q_m) = 0$. We have

$$(1-G(q_m))[k(q_m+q_h)-k(q_m)] - (G(q_m+q_h)-G(q_m))k(q_m) = k(q_h)(1-G(\bar{q})) > 0 \ . \quad (38)$$

[46] If $B \leq 1$, the market maker would not make that market even as a monopolist.

Thus,

$$\frac{k(q_m+q_h)-k(q_m)}{k(q_m)} > \frac{G(q_m+q_h)-G(q_m)}{1-G(q_m)} \ . \tag{39}$$

Next I show that $W''(q_m) < 0$. $W''(q_m) < 0$ is equivalent to

$$g(q_m)[k(q_m+q_h)-k(q_m)]+k(q_m)[g(q_m+q_h)-g(q_m)]+2(1-G(q_m))[G(q_m+q_h)-G(q_m)] > 0 \ . \tag{40}$$

Since $G(q_m+q_h)-G(q_m) > 0$, a sufficient condition for inequality (40) is

$$g(q_m)[k(q_m+q_h)-k(q_m)] > k(q_m)[g(q_m)-g(q_m+q_h)] \ . \tag{41}$$

Since G has non-decreasing hazard rate, $\frac{g(q_m+q_h)}{1-G(q_m+q_h)} \geq \frac{g(q_m)}{1-G(q_m)}$. Thus, $g(q_m)-g(q_m+q_h) \leq \frac{G(q_m+q_h)-G(q_m)}{1-G(q_m)}g(q_m)$. This implies inequality (39) is sufficient for inequality (41).

In sum, there exists a $q_m \in [0,\bar{q}]$ such that $W'(q_m)=0$. Moreover, for any q_m such that $W'(q_m)=0$, $W''(q_m)<0$. This implies that $W(y)$ has a unique maximum. ■

Proposition 19 *Suppose G has non-decreasing hazard rate. Consider two simultaneous pricing games where the market maker has discount rate δ_1 in the first game and discount rate δ_2 in the second game. Suppose $\delta_1 > \delta_2$ and all other parameters are the same. Let q_m^1 (q_m^2) be the market maker's steady state capital commitment in the first game (the second game). Then $q_m^1 > q_m^2$.*

Proof. Let $B_1 = \frac{\delta_1}{1-\delta_1}(1-F(x^*))x^*$; $B_2 = \frac{\delta_2}{1-\delta_2}(1-F(x^*))x^*$. If the market maker is using the wide spread strategy in both games, then $q_m^1 > q_m^2$ directly follows from the first order condition. If the market maker is using the tight spread strategy in both games, then by the first order condition,

$$\frac{B_1}{k(q_h)}[(1-G(q_m^1))(k(q_m^1+q_h)-k(q_m^1))-(G(q_m^1+q_h)-G(q_m^1))k(q_m^1)]-1=0 \ .$$

Since $B_1 > B_2$, we have

$$\frac{B_2}{k(q_h)}[(1-G(q_m^1))(k(q_m^1+q_h)-k(q_m^1))-(G(q_m^1+q_h)-G(q_m^1))k(q_m^1)]-1<0 \ .$$

Then by Proposition 18, there exists a unique $q_m^2 < q_m^1$ such that

$$\frac{B_2}{k(q_h)}[(1 - G(q_m^2))(k(q_m^2 + q_h) - k(q_m^2)) - (G(q_m^2 + q_h) - G(q_m^2))k(q_m^2)] - 1 = 0 \ ,$$

and q_m^2 maximize the market maker's expected payoff given he is using a tight spread strategy in the steady state. If the market maker is using the wide spread strategy in the first game and the tight spread strategy in the second game, combine the result about with Theorem 3 yield the result that $q_m^1 > q_m^2$. This covers all situations when the market maker is using the wide spread strategy in the first game.

Now consider the situation where the market maker is using the tight spread strategy in the first game. Let q_t^1 and q_w^1 (q_t^2 and q_w^2) be the market maker's shareholding under the optimal tight and wide spread strategy in the first (second) game. Since the market maker is using the tight spread strategy in the first game, $q_m^1 = q_t^1$. By the discussion above, $q_t^1 > q_t^2$; $q_w^1 > q_w^2$. If $q_t^1 > q_w^2$, the claim is true. Thus, we only consider the case when $q_t^1 \leq q_w^2$.

From the optimality condition,

$$\frac{\delta_1}{1-\delta_1}M(q_t^1)+(w_0-q_t^1) \geq \frac{\delta_1}{1-\delta_1}M(q_w^1)+(w_0-q_w^1) > \frac{\delta_1}{1-\delta_1}M(q_w^2)+(w_0-q_w^2) \ , \quad (42)$$

where $M(\cdot)$ is the expected profit of the market maker in a one-shot game. If $M(q_t^1) > M(q_w^2)$, since $q_t^1 \leq q_w^2$, we have

$$\frac{\delta_2}{1-\delta_2}M(q_t^2) + (w_0 - q_t^2) > \frac{\delta_2}{1-\delta_1}M(q_t^1) + (w_0 - q_t^2) > \frac{\delta_2}{1-\delta_2}M(q_w^2) + (w_0 - q_w^2) \ .$$

Thus, the market maker would use the tight spread strategy in the second game and $q_m^1 > q_m^2 = q_t^2$.

If $M(q_t^1) \leq M(q_w^2)$, from equation 42 and $\frac{\delta_1}{1-\delta_1} > \frac{\delta_2}{1-\delta_2}$, we also have

$$\frac{\delta_2}{1-\delta_2}M(q_t^2) + (w_0 - q_t^2) > \frac{\delta_1}{1-\delta_2}M(q_t^1) + (w_0 - q_t^2) > \frac{\delta_2}{1-\delta_2}M(q_w^2) + (w_0 - q_w^2) \ .$$

This implies $q_m^1 > q_m^2 = q_t^2$ and concludes the proof. ■

This result is important for the simultaneous pricing game extension. Notice that the equilibrium I construct in the simultaneous pricing game might not be sub-game perfect. In the sub-game where the market maker commits less capital than the steady state level, it might not be optimal for the market maker to stick to the strategy specified in the equilibrium since net worth may have additional benefit. However, if I assume G has non-decreasing hazard rate, this is not a problem since I can consider a game where the HFT is uncertain about the market maker's discount rate δ and infers it from the market maker's capital commitment decision. This result guarantees the existence of a separating equilibrium where in equilibrium, the market maker's discount rate is perfectly signaled by his capital commitment decision.[47] In this sense, the equilibrium I propose coincide with a perfect Bayesian equilibrium in this extended game.

Proposition 20 *Suppose G has non-decreasing hazard rate. If $q_h \geq \frac{\bar{q}}{2}$, $argmax_y W(y) \in [0, \frac{\bar{q}}{2}]$.*

Proof. By Proposition 18, if $W'(\frac{\bar{q}}{2}) \leq 0$, then $argmax_y W(y) \in [0, \frac{\bar{q}}{2}]$. Thus, it is sufficient to show that for all $q_h \geq \frac{\bar{q}}{2}$, $W'(\frac{\bar{q}}{2}) \leq 0$.

This is equivalent to

$$k(q_h)(1-G(\bar{q}))+(G(\frac{\bar{q}}{2}+q_h)-G(\frac{\bar{q}}{2}))k(\frac{\bar{q}}{2})-[1-G(\frac{\bar{q}}{2})][k(\frac{\bar{q}}{2}+q_h)-k(\frac{\bar{q}}{2})] \geq 0 \ . \quad (43)$$

When $q_h = \frac{\bar{q}}{2}$, the LHS of inequality (43) becomes

$$(1 - G(\frac{\bar{q}}{2}))[2k(\frac{\bar{q}}{2}) - k(\bar{q})] \ . \quad (44)$$

This quantity is greater than zero since $2k(\frac{\bar{q}}{2}) > k(\bar{q})$. Denote the LHS of inequality (43) by $J(q_h)$. If $J(q_h)$ is increasing in q_h, the lemma is proved.

$$J'(q_h) = (1 - G(q_h))(1 - G(\bar{q})) + g(\frac{\bar{q}}{2})k(\frac{\bar{q}}{2}) - [1 - G(\frac{\bar{q}}{2})](1 - G(\frac{\bar{q}}{2} + q_h)) \ . \quad (45)$$

A sufficient condition of $J'(q_h) \geq 0$ is $\frac{1-G(\bar{q})}{1-G(\frac{\bar{q}}{2})} \geq \frac{1-G(\frac{\bar{q}}{2}+q_h)}{1-G(q_h)}$. Since $q_h \geq \frac{\bar{q}}{2}$, it is sufficient to have $\frac{1-G(\bar{q}+z)}{1-G(\frac{\bar{q}}{2}+z)}$ decreasing in z. Take derivative to get

$$-g(\bar{q}+z)(1-G(\frac{\bar{q}}{2}+z)) + g(\frac{\bar{q}}{2}+z)(1-G(\bar{q}+z)) \leq 0 \ . \quad (46)$$

[47] When the market maker's capital commitment cannot be mapped to any δ, any off path belief can be specified. For example, the HFT may assume that the market maker is maximizing the short term profit.

This condition is satisfied due to the increasing hazard rate of G. ■

F Social Planner's Perspective on Welfare

In this section, I briefly discuss the social planner's perspective on welfare. In practice, the social planner can be either a policy maker or an exchange, aiming at maximizing market participants' welfare. I assume that the social planner can control q_m, the market maker's capital commitment and π, the HFT's entry probability. In this case, the social planner provides the market maker operating capital q_m at period 0 to make the market and the market maker pays the profit from market making as dividend.[48] For simplicity, I focus on the situation where the HFT trades faster than the market maker.

If the social planner aims at maximizing liquidity, he has two possible policies. Either he relies on the market maker to supply liquidity by setting $q_m = \infty$ and $\pi = 0$.[49] In this market, $L = (1-F(x^*))E_G(q_b) = (1-F(x^*))k(\infty)$. Alternatively, the social planner can rely on both the market maker and the HFT to supply liquidity by setting $q_m = q_h$ and $\pi = 1$. In this market, $L = (1-F(x^*))k(2q_h)+[F(x^*)-F(\underline{x})]k(q_h)$ where $\underline{x}$ is uniquely pinned down by $a(\underline{x})k(q_h) = k(2q_h) - k(q_h)$. The social planner would choose the policy that provides higher liquidity. Intuitively, when q_h is large, the social planner tends to use the latter policy. In either case, the social planner is committing more capital than the profit maximizing market maker.

[48] Alternatively, I may assume the social planner set a mandatory capital commitment level for the market maker. Two settings lead to similar qualitative results.

[49] π can be any number between 0 and 1 and market liquidity will be the same.