High Frequency Trading and Fragility*

Giovanni Cespa† and Xavier Vives‡

June 2016

Abstract

We show that market fragmentation, induced by an informational friction resulting from high frequency trading, can induce strategic complementarities between liquidity consumption and provision: traders consume *more* liquidity when the cost of liquidity provision *increases*, which in turn jacks up the cost of liquidity provision. This can generate market instability, where an initial dearth of liquidity degenerates into a liquidity rout (as in a flash crash). While in a transparent market, liquidity is increasing in the proportion of high frequency traders, in an opaque market strategic complementarities can make liquidity U-shaped in this proportion.

Keywords: Market fragmentation, high frequency trading, flash crash, asymmetric information.

JEL Classification Numbers: G10, G12, G14

*A previous version of this paper circulated with the tile "The welfare impact of high frequency trading." For helpful comments we thank Bruno Biais, Evangelos Benos, Thierry Foucault, Denis Gromb, Pete Kyle, Albert Menkveld, Sophie Moinas, Andreas Park, Joël Peress, Liyan Yang, Bart Yueshen, and seminar participants at INSEAD, HEC (Paris), Rotterdam School of Management, the 9th Annual Central Bank Workshop on Microstructure (Frankfurt, 9/13), the conference on High Frequency Trading at Imperial College, Brevan Howard Centre (London, 12/14), the Workshop on Microstructure Theory and Applications (Cambridge, 3/15), the third workshop on Information Frictions and Learning (Barcelona, 6/15), the Bank of England, and the AFA (San Francisco, 1/16). Cespa acknowledges financial support from the Bank of England (grant no. RDC151391). This paper has been prepared by Vives under the Wim Duisenberg Research Fellowship Program sponsored by the ECB. Any views expressed are only those of the authors and do not necessarily represent the views of the ECB or the Eurosystem.

†Cass Business School, City University London, CSEF, and CEPR. 106, Bunhill Row, London EC1Y 8TZ, UK. e-mail: `giovanni.cespa@gmail.com`

‡IESE Business School, Avinguda Pearson, 21 08034 Barcelona, Spain.

"The report describes how on October 15, some algos pulled back by widening their spreads and other reduced the size of their trading interest. Whether such dynamic can further increase volatility in an already volatile period is a question worth asking, but a difficult one to answer." (Remarks Before the Conference on the Evolving Structure of the U.S. Treasury Market (Oct. 21, 2015), Timothy Massad, Chairman, CFTC.)

1 Introduction

Concern for crashes has recently revived, in the wake of the sizeable number of "flash events" that have affected different markets. For futures, in the 5-year period from 2010, more than a 100 flash events have occurred (see Figure 1). For other contracts, the list of events where markets suddenly crash and recover is by now quite extensive.[1]

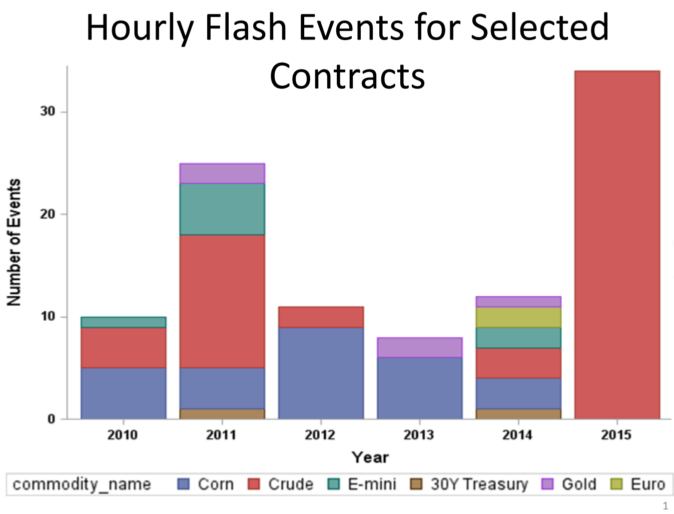


Figure 1: Number of *flash events* in futures contracts from 2010 to 2015. A flash event is an episode in which "the price of a contract moved at least 200 basis points within a trading hour but returned to within 75 basis points of the original or starting price within that same hour." (Source: Remarks Before the Conference on the Evolving Structure of the U.S. Treasury Market (Oct. 21, 2015), Timothy Massad, Chairman, CFTC.).

A common trait of these episodes seems to be the apparent jamming of the "rationing" function of market illiquidity. Indeed, in normal market conditions, traders perceive a lack of liquidity as a cost, which in turn leads them to limit their demand for immediacy.[2] This eases the pressure on liquidity suppliers, thereby producing a stabilizing effect on the market.

[1]Starting with the May 6, 2010 U.S. "flash-crash" where U.S. equity indices dropped by 5-6% and recovered within half an hour; moving to the October 15, 2014 Treasury Bond crash, where the yield on the benchmark 10-year U.S. government bond, dipped 33 basis points to 1.86% and reversed to 2.13% by the end of the trading day; to end with the August 25, 2015 ETF market freeze, during which more than a fifth of all U.S.-listed exchange traded funds and products were forced to stop trading. More evidence of flash events is provided by NANEX.

[2]To minimize market impact and the associated trading costs they incur e.g., by using algorithms that parcel out their orders.

However, during a crash, a liquidity drought, short of limiting traders' demand for immediacy, produces the *opposite* effect: traders attempt to place orders *despite* the liquidity shortage. In these conditions, a bout of illiquidity no longer has a stabilizing impact, and can instead foster a disorderly "run for the exit," that is conducive to a rout. What can account for such a dualistic feature of market illiquidity?

In this paper, we argue that an important ingredient in the answer to this question is represented by the fragmentation of liquidity supply induced by computerized trading.[3] Indeed, the automation of the trading process fosters liquidity supply fragmentation in that it limits the participation of some liquidity suppliers (Duffie (2010) and SEC (2010)), as well as some traders' access to reliable and timely market information (Ding et al. (2014)).[4] This, in turn, creates an informational friction that can be responsible for the type of behavior we described above.

We analyze a model in which two classes of risk-averse dealers provide liquidity to two cohorts of risk-averse, short-term traders who receive a common endowment shock, in a two-period market. Traders, thus, enter the market to partially hedge their exposure to the risky asset. In the first round of trade both dealers' types absorb the (market) orders of the first traders' cohort. In the second trading round, only one class of dealers, named 'full,' is able to participate. Full dealers, like stylized high frequency traders (HFTs), are continuously in the market and can therefore accommodate the reverting orders of the first traders' cohort, as well as those of the incoming second cohort who observe an imperfect signal about the first period order imbalance.[5]

A central finding of our analysis is that dealers' limited market participation favors the propagation of the endowment shock across time. This is because when first period traders load their positions, a part of their orders is absorbed by standard dealers. These agents, however, are not in the market in the second period, when first period traders unwind. As a consequence, an order imbalance (induced by first period traders' unwinding orders and) affecting the second period price, arises. As standard dealers are unable to rebalance in the second period, they require a larger price concession to absorb traders' orders. This implies that as liquidity dries up, standard dealers absorb more of the imbalance, magnifying the propagation effect.

We first study a benchmark market in which second period traders have access to a perfect signal on the first period imbalance. This situation is likely to arise at low trading frequencies (e.g., intradaily), or in a transparent setup where all market participants have access to the same type of feed, even at high frequencies. In this case we show that first period traders' demand

[3]Automated trading is by now pervasive across different markets. For financial futures, automated trading accounts for about two-thirds of the activity in Eurodollars and Treasury contracts (Source: Keynote Address of CFTC Commissioner J. Christopher Giancarlo before the 2015 ISDA Annual Asia Pacific Conference).

[4]Ding et al. (2014) argue that in the U.S. "...not all market participants have equal access to trade and quote information. Both physical proximity to the exchange and the technology of the trading system contribute to the latency."

[5]In a companion paper, we then embed the baseline model in a simple platform competition setup in which exchanges compete in the supply of trading services (co-location capacity). In this framework we endogenize the decision of a dealer to acquire the technology to be continuously in the market, and the number of exchanges supplying trading services.

for liquidity is a decreasing function of illiquidity (i.e., the compensation that dealers demand to hold the asset inventory in equilibrium): the less liquid is the market, the higher is the cost these traders incur to reduce exposure, and the less aggressive is their liquidity consumption (the closer to zero is their hedging aggressiveness). Conversely, illiquidity is decreasing in traders' hedging aggressiveness. This is because lower aggressiveness limits liquidity consumption, which in turn shrinks dealers' inventory, allowing for cheaper liquidity provision. Thus, illiquidity in this case has a direct, "rationing" effect on traders' liquidity consumption, and a unique equilibrium arises. Furthermore, along this equilibrium, small shocks to the model's parameters have a minimal impact on market liquidity.

In contrast, when access to imbalance information is impaired, the market is opaque, and illiquidity also displays a feedback, liquidity consumption "expanding" effect. This can create a self-sustaining loop where a liquidity evaporation, short of curtailing traders' liquidity demand, fosters a stronger liquidity consumption. As a consequence, the demand for liquidity can become *increasing* in illiquidity, and multiple equilibria can arise. To see this, note that due to propagation, second period traders speculate against the imbalance generated by their first period peers the more, the stronger is such propagation. Suppose now that liquidity evaporates in the first period market. As a consequence, standard dealers intermediate more of the outstanding imbalance, magnifying the propagation of the first period endowment shock, and leading second period traders to trade more aggressively against it. However, as information on the first period imbalance is noisy, these trades increase the first period uncertainty about the second period price. This can lead first period traders to consume more liquidity (as holding exposure to the asset becomes riskier), and liquidity suppliers to charge more to absorb the order imbalance (as their inventory of the risky asset increases), eventually reinforcing the initial, negative shock to market liquidity.

Equilibrium multiplicity induces three levels of liquidity that can be ranked in an increasing order (low, intermediate, and high). At the low (respectively, intermediate, and high) liquidity equilibrium, volatility and liquidity consumption are high (respectively, intermediate, and low). Thus, our paper highlights a channel through which the *combined* effect of a heightened demand for liquidity, and a reduced liquidity provision conjure to increase market volatility, providing a positive answer to this paper's opening quotation.

The liquidity consumption ranking across equilibria is a further manifestation of the fact that opaqueness jams the direct, rationing effect of illiquidity, while it strengthens its feedback, liquidity consumption enhancing effect. The end result is that traders aim to hedge the largest portion of their endowment, at the equilibrium where the cost of trading is at its highest. Importantly, we also find that: (i) depending on parameters' values, uniqueness obtains *either* at an equilibrium with high or one with low liquidity (corresponding respectively to the high and low liquidity equilibrium when multiplicity arises), and (ii) that the comparative statics properties of these equilibria differ. For instance, when the market hovers along an equilibrium with low liquidity, illiquidity can be hump-shaped in the proportion of fast dealers, something that does not happen in an equilibrium with high liquidity.

The strategic complementarity loop arising with market opaqueness implies that liquidity can be "fragile" in our setup. We show this with two types of examples. In the first one, we exploit equilibrium multiplicity and illustrate how a small shock to some parameter values can produce a switch from the high liquidity equilibrium to an equilibrium with low liquidity. In particular, we focus on the consequence of a shock that disconnects a small mass of full dealers from the market (a technological 'glitch'). We then analyze the effect of a positive shock to the volatility of first and second period traders' demand. These are meant to capture, respectively, an increase in the probability of a large order hitting the first period market (which is consistent with some narratives of the flash crash, see e.g. Easley et al. (2011)), and an increase in the uncertainty first period traders face on their endowment value. In all these examples *small* parameter shocks produce *large* liquidity withdrawals.

In the second type of example we review the impact of the glitch, but in this case leveraging on the hump-shaped relationship between illiquidity and full dealers' participation that can obtain along an equilibrium with low liquidity. Based on this finding, we show that a high level of liquidity can suddenly evaporate because of a reduction in full dealers' participation *along the same* equilibrium.

This paper is related to four strands of the literature. First, equilibrium multiplicity, liquidity complementarities, and liquidity fragilities are known to obtain in economies where asset prices are driven by fundamentals information and noise trading (see, e.g., Cespa and Foucault (2014), Cespa and Vives (2015), Goldstein et al. (2014), and Goldstein and Yang (2015)). In this setup, in contrast, asset prices are exclusively driven by endowment shocks. However, the demand of all the traders is responsive to the volatility of the price at which these agents unwind their positions. In turn, such volatility depends on traders' demand. As we argued above, in an opaque market this two-sided loop—which in a noise traders' economy cannot possibly arise—is responsible for the multiplicity result. Other authors obtain multiple equilibria in setups where order flows are driven by only one type of shock (see, e.g., Spiegel (1998)). However, multiplicity there arises from the bootstrap nature of expectations in the steady-state equilibrium of an overlapping generations (OLG) model in which investors live for two periods. Our setup, in contrast, considers an economy with a finite number of trading rounds.

Second, this paper adds to the theoretical literature on the impact of high frequency trading (HFT) on market performance, by showing that an informational friction arising from liquidity provision fragmentation can be responsible for liquidity fragility, and reverses the common wisdom that associates an increase in computerized trading with more liquid markets. Differently from our setup, the HFT literature has mostly concentrated on modeling risk neutral agents (e.g., Budish et al. (2015), Hoffmann (2014), Du and Zhu (2014), Bongaerts and Van Achter (2015), Foucault et al. (2015), and Menkveld and Zoican (2015); see O'Hara (2015) and Menkveld (2016) for literature surveys).[6] Easley et al. (2011, 2012), find that in the

[6]Biais et al. (2015) study the welfare implications of investment in the acquisition of HFT technology. In their model HFTs have a superior ability to match orders, and possess superior information compared to human (slow) traders. They find excessive incentives to invest in HFT technology, which, in view of the negative externality generated by HFT, can be welfare reducing.

hours preceding the flash crash, signed order imbalance for the E-mini S&P500 futures contract was unusually high. They interpret this evidence as supportive of a high order flow "toxicity," which led HFTs to flee the market, eventually precipitating the crash. As argued above, our model also predicts that large imbalances can lead to a huge liquidity withdrawal. However, the channel we highlight is not related to adverse selection, but emphasizes the multiplier effect of illiquidity on the demand for immediacy that can arise when some traders have access to opaque information on imbalances. Menkveld and Yueshen (2012) argue that market *spatial fragmentation* can be detrimental to stability. In their model, HFTs have access to a private reselling opportunity which, due to impaired intermarket connectivity, can break down. When this happens, HFTs trade among themselves, providing an 'illusion' of liquidity to traders who observe volume, which in turn fosters further liquidity demand. Our focus is on the liquidity provision fragmentation induced by an informational friction in a single, *concentrated* market, a feature that is consistent with the futures markets flash events discussed above.

Third, the paper relates to the literature that assesses the impact of limited market participation. Heston et al. (2010) and Bogousslavsky (2014) find that some liquidity providers' limited market participation can have implications for return predictability. Chien et al. (2012) focus instead on the time-series properties of risk premium volatility. Hendershott et al. (2014) concentrate on the effect of limited market participation for price departures from semi-strong efficiency. Our focus is, instead, on the destabilizing dynamics that is generated by bouts of illiquidity. In this respect, our paper is also related to Huang and Wang (2009) who show that with costly market participation, idiosyncratic endowment shocks can yield crashes. Note, however, that in our setup traders are exposed to the same shock, which yields a different mechanism for market instability.

Fourth, by highlighting the first order asset pricing impact of uninformed traders' imbalance predictability, this paper shares some features of our previous work (Cespa and Vives (2012), and Cespa and Vives (2015)). In that setup, however, predictability obtained because of the assumed statistical properties of noise traders' demands, whereas in this paper it arises endogenously, because of a participation friction. A growing literature investigates the asset pricing implications of noise trading predictability. Collin-Dufresne and Vos (2015) argue that informed traders time their entry to the presence of noise traders in the market. This, in turn, implies that standard measures of liquidity (e.g., Kyle's lambda), may fail to pick up the presence of such traders. Peress and Schmidt (2015) estimate the statistical properties of a noise trading process, finding support for the presence of serial correlation in demand shocks.

The rest of the paper is organized as follows. In the next section we introduce the model, and show that with limited market participation, endowment shocks propagate across trading dates. Next, we analyze the benchmark with a transparent market. We then illustrate how the presence of an informational friction can generate strategic complementarities between traders' demand for immediacy and market illiquidity. We show that such complementarities are at the root of the loop responsible for equilibrium multiplicity and liquidity fragility. A final section contains concluding remarks. All proofs are in the appendix.

2 The model

A single risky asset with liquidation value $v \sim N(0, \tau_v^{-1})$, and a risk-less asset with unit return are exchanged in a market during two trading rounds. Three classes of traders are in the market. First, a continuum of competitive, risk-averse, High Frequency Traders (which we refer to as "Full Dealers" and denote by FD) in the interval $(0, \mu)$, are active at both dates. Second, competitive, risk-averse dealers (D) in the interval $[\mu, 1]$, are active only in the first period. Finally, a unit mass of short-term traders enters the market at date 1. At date 2, these traders unwind their position, and are replaced by a new cohort of short-term traders (of unit mass). The asset is liquidated at date 3. We now illustrate the preferences and orders of the different players.

2.1 Liquidity providers

A FD has CARA preferences (we denote by γ his risk-tolerance coefficient) and submits price-contingent orders x_t^{FD}, $t = 1, 2$, to maximize the expected utility of his final wealth: $W^{FD} = (v - p_2)x_2^{FD} + (p_2 - p_1)x_1^{FD}$.[7] A Dealer also has CARA preferences with risk-tolerance γ, but is in the market only in the first period. He thus submits a price-contingent order x_1^D to maximize the expected utility of his wealth $W^D = (v - p_1)x_1^D$. The inability of D to trade in the second period captures some liquidity suppliers' limited market participation. This friction could be due to technological reasons (as, e.g. in the case of standard dealers with impaired access to a technology that allows trading at high frequencies).

2.2 Short-term traders

In the first period a unit mass of short-term traders is in the market. A short-term trader receives a random endowment of the risky asset u_1, and posts a market order x_1^L anticipating that it will unwind its holdings in the following period, and leave the market. We assume $u_1 \sim N(0, \tau_{u_1}^{-1})$, and $\text{Cov}[u_1, v] = 0$.[8] First period traders have identical CARA preferences (we denote by γ_1^L the common risk-tolerance coefficient). Formally, a trader maximizes the expected utility of his wealth $\pi_1^L = u_1 p_2 + (p_2 - p_1)x_1^L$:

$$E\left[-\exp\{-\pi_1^L/\gamma_1^L\}|\Omega_1^L\right],$$

where Ω_1^L denotes his information set. In period 2, first period traders are replaced by a new (unit) mass of traders receiving a random endowment of the risky asset u_2, where $u_2 \sim N(0, \tau_{u_2}^{-1})$ and $\text{Cov}[u_2, v] = \text{Cov}[u_2, u_1] = 0$. A second period trader has CARA utility function with risk-tolerance γ_2^L, and submits a market order to maximize the expected utility of his wealth

[7] We assume, without loss of generality with CARA preferences, that the non-random endowment of FDs and dealers is zero. Also, as equilibrium strategies will be symmetric, we drop the subindex i.

[8] The assumption of a random endowment in the risky asset is akin to Huang and Wang (2009), and Vayanos and Wang (2012) who instead posit that traders receive an endowment in a consumption good that is perfectly correlated with the value of the risky asset at the terminal date.

$\pi_2^L = u_2 v + (v - p_2)x_2^L$:

$$E\left[-\exp\{-\pi_2^L/\gamma_2^L\}|\Omega_2^L\right],$$

where Ω_2^L denotes his information set.[9]

2.3 Information sets

We now describe the information sets of the different market participants. At equilibrium, we conjecture that a period 1 trader submits an order $x_1^L = b_1^L u_1$, where b_1^L denotes the first period "hedging" aggressiveness, to be determined in equilibrium, while a FD and a dealer respectively post a limit order $x_1^{FD} = \varphi_1^{FD}(p_1)$, $x_1^D = \varphi_1^D(p_1)$ where $\varphi_1^{FD}(\cdot)$, $\varphi_1^D(\cdot)$ are linear functions of p_1. In the second period, a FD submits a limit order $x_2^{FD} = \varphi_2(p_1, p_2)$, where $\varphi_2(\cdot)$ is a linear function of prices. A second period trader observes a signal of the first period endowment shock $s_{u_1} = u_1 + \eta$, with $\eta \sim N(0, \tau_\eta^{-1})$, and independent from all the other random variables in the model, and submits a market order $x_2^L = b_{21}^L u_2 + b_{22}^L s_{u_1}$, where b_{21}^L and b_{22}^L denote respectively the second period hedging and speculative aggressiveness. With these assumptions, we obtain

Lemma 1. *At equilibrium, p_1 is observationally equivalent to u_1, and the sequence $\{p_1, p_2\}$ is observationally equivalent to $\{u_1, x_2^L\}$.*

A first period trader observes the endowment shock u_1. Therefore, his information set coincides with the one of Ds and FDs: $\Omega_1^L = \Omega_1^{FD} = \Omega_1^D = \{u_1\}$. A second period trader receives an endowment shock u_2, and can observe a signal s_{u_1}. Thus, his information set is $\Omega_2^L = \{u_2, s_{u_1}\}$. Finally, a FD in period 2 observes the sequence of prices: $\Omega_2^{FD} = \{p_1, p_2\}$ from which he retrieves $\{u_1, x_2^L\}$.

Thus, according to our model, liquidity provision is fragmented because (i) only one class of dealers is able to participate in the second period and (ii) some traders (the second cohort of short-term traders) have access to opaque information on the first period price. This assumption is consistent with the evidence that exchanges sell fuller access to their matching engine, as well as direct feeds of their market information at a premium (see, e.g., O'Hara (2015)).[10] Figure 2 displays the timeline of the model.

[9] Our results are robust to the case in which the first period market is populated by a mass β of short-term traders, that unwind at date 2, and a mass $(1 - \beta)$ of long-term ones that hold their position until liquidation.

[10] This assumption is also similar to Foucault et al. (2015) who posit that HFTs receive market information slightly ahead of the rest of the market Ding et al. (2014) compare the NBBO (National Best Bid and Offer, which is the price feed computed by the Security Industry Processors in the US) to the fuller feeds market participants obtain via a direct access to different trading platforms. Their findings point to sizeable price differences that can yield substantial profits to HFTs. Latency in the reporting of market data can also be profitably exploited for securities with centralized trading, see "High-speed traders exploit loophole," *Wall Street Journal*, May 1, 2013.

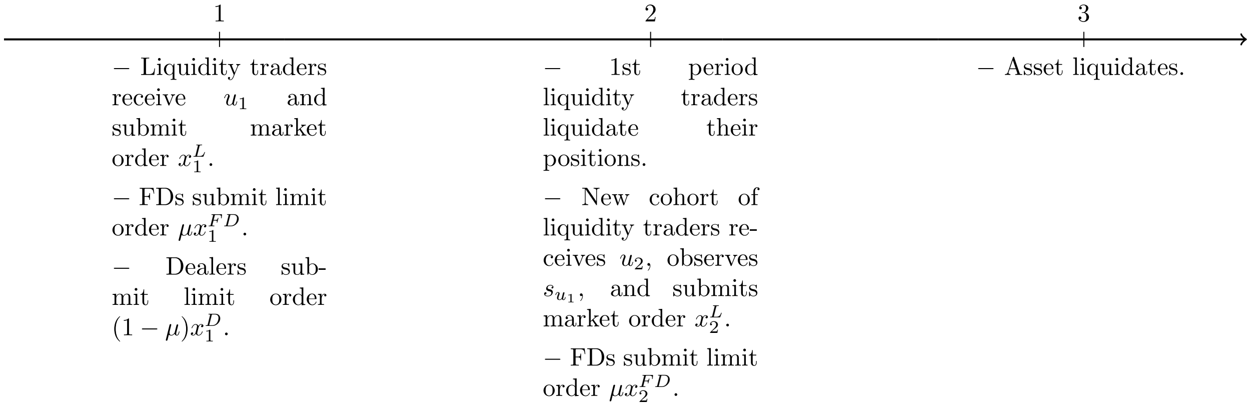


Figure 2: The timeline.

2.4 Limited market participation and the propagation of endowment shocks

Due to limited market participation, the first period endowment shock propagates to the second trading round, thereby affecting p_2. To see this, consider the first period market clearing equation

$$\mu x_1^{FD} + (1-\mu)x_1^D + x_1^L = 0. \tag{1}$$

At equilibrium the orders of first period traders are absorbed by both FDs and Ds. Thus, when $\mu < 1$, FDs' aggregate position falls short of x_1^L: $\mu x_1^{FD} + x_1^L \neq 0$. As a consequence, the inventory FDs carry over from the first period is insufficient to absorb the reverting orders that first period traders post in period 2. This creates an order imbalance driven by the first period endowment shock u_1 that adds to the one originating from second period trades, and affects the second period price. Formally, from the second period market clearing equation we have

$$\mu(x_2^{FD} - x_1^{FD}) + (x_2^L - x_1^L) = 0.$$

Substituting (1) in the latter and rearranging yields:

$$\mu x_2^{FD} + x_2^L + (1-\mu)x_1^D = 0. \tag{2}$$

According to Lemma 1, at equilibrium x_1^D depends on u_1. Thus, when $\mu < 1$, p_2 also reflects the first period endowment shock.

2.5 Strategies

We now discuss the strategies of the different market participants. In the second period, FDs act like in a static market:

$$X_2^{FD}(p_1, p_2) = -\gamma\tau_v p_2.$$

Therefore, they speculate on the asset payoff (recall that $E[v] = 0$), and supply liquidity, demanding a compensation that is inversely related to the risk they bear. In the first period, as we show in the appendix, we have

$$X_1^D(p_1) = -\frac{\gamma}{\mathrm{Var}[v]}p_1 \tag{3a}$$

$$X_1^{FD}(p_1) = \underbrace{\gamma\frac{E[p_2 - p_1|u_1]}{\mathrm{Var}[p_2|u_1]}}_{\text{Speculation}} - \underbrace{\frac{\gamma}{\mathrm{Var}[v]}p_1}_{\text{Market making}} \quad . \tag{3b}$$

The above expressions imply that standard dealers accommodate the residual imbalance, while FDs also speculate on short term returns. The speculative component in FDs first period strategy has two implications. First, it makes the price adjustment FDs require to accommodate an increase in the aggregate demand for liquidity smaller compared to that required by Ds:

$$\left(\frac{\partial X_1^{FD}(p_1;u_1)}{\partial p_1}\right)^{-1} = \frac{1}{\gamma}\left(\frac{1}{\mathrm{Var}[p_2|u_1]} + \frac{1}{\mathrm{Var}[v]}\right)^{-1} < \left(\frac{\partial X_1^{D}(p_1;u_1)}{\partial p_1}\right)^{-1} = \frac{\mathrm{Var}[v]}{\gamma}. \tag{4}$$

Second, it reduces the imbalance that liquidity providers (both Ds and FDs) have to clear at equilibrium. In particular, the larger is FDs speculative position, the smaller is the residual imbalance.

Consider now short-term traders. In the appendix we show that a second period trader trades according to

$$\begin{aligned} X_2^L(u_2, s_{u_1}) &= \underbrace{\gamma_2^L\frac{E[v - p_2|\Omega_2^L]}{\mathrm{Var}[v - p_2|\Omega_2^L]}}_{\text{Speculation}} - \underbrace{\frac{\mathrm{Cov}[v - p_2, v|\Omega_2^L]}{\mathrm{Var}[v - p_2|\Omega_2^L]}u_2}_{\text{Hedging}} \\ &= \underbrace{\frac{\gamma_2^L\mathrm{Cov}[v - p_2, u_2]}{\mathrm{Var}[v - p_2|\Omega_2^L]\mathrm{Var}[u_2]}u_2}_{\text{Speculation on } u_2} + \underbrace{\frac{\gamma_2^L\mathrm{Cov}[v - p_2, s_{u_1}]}{\mathrm{Var}[v - p_2|\Omega_2^L]\mathrm{Var}[s_{u_1}]}s_{u_1}}_{\text{Speculation on } u_1} - \underbrace{\frac{\mathrm{Cov}[v - p_2, v|\Omega_2^L]}{\mathrm{Var}[v - p_2|\Omega_2^L]}u_2}_{\text{Hedging}}. \end{aligned} \tag{5}$$

Thus, a trader's strategy has a speculative and a hedging component. According to the first line in (5), a trader speculates on value change the more, the less liquid is the market (see the first term on the r.h.s. in (5)), while lowering his exposure to the asset risk the more, the higher is the covariance between the return on his position (i.e., $v - p_2$) and the final liquidation value (v), given his information. In this way he reduces the risk that his speculative strategy goes sour precisely when the value of his endowment collapses. Expanding the expectation operator at the numerator of (5) shows that there are two sources of speculation. Other things equal, given u_2 a trader retains part of his asset exposure to the extent that this is positively correlated with the capital gain $v - p_2$, to profit from the latter. Additionally, he uses his information on u_1 to speculate on the reverting orders of first period traders.

First period traders' strategies are similar to (5):

$$X_1^L(u_1) = \underbrace{\gamma_1^L \frac{E[p_2 - p_1|u_1]}{\text{Var}[p_2|u_1]}}_{\text{Speculation on } u_1} - \underbrace{u_1}_{\text{Hedging}} . \tag{6}$$

First period traders can partially anticipate the second period price, and thus speculate on it, for example by holding part of their endowment when $u_1 > 0$. Substituting (6) in (3b) yields the following expression:

$$X_1^{FD}(p_1) = \frac{\gamma}{\gamma_1^L}\left(X_1^L(u_1) + u_1\right) - \frac{\gamma}{\text{Var}[v]}p_1. \tag{7}$$

According to (7), for given u_1, a contraction of first period traders' holdings (i.e., an increase in their demand for liquidity), leads to a corresponding contraction in FDs speculative activity, and thus to an increase in the residual imbalance that liquidity suppliers have to clear in equilibrium.

3 Market transparency and the rationing effect of illiquidity

In this section, we assume that second period traders have a perfect signal on the first period endowment shock: $\tau_\eta \to \infty$. This captures a scenario in which information on the first period imbalance is public, as is the case in a low frequency trade environment (e.g., intradaily). Alternatively, it represents an ideal setup in which second period traders have access to the same information as FDs. In this case, we obtain the following result:

Proposition 1. *When the market is transparent there exists a unique equilibrium in linear strategies, where* $x_1^L = b_1^L u_1$, $x_2^L = b_{21}^L u_2 + b_{22}^L s_{u_1}$,

$$p_2 = \lambda_2(b_{21}^L u_2 + b_{22}^L s_{u_1}) + \lambda_2(1-\mu)\gamma\tau_v\Lambda_1^* u_1 \tag{8a}$$

$$p_1 = -\Lambda_1^* u_1, \tag{8b}$$

$\lambda_2 = 1/(\mu\gamma\tau_v) > 0$, *and*

$$\Lambda_1^* = \frac{1}{\gamma\tau_v}\left(1 - \frac{(\mu\gamma + \gamma_1^L)(1 + b_1^L)}{\gamma_1^L}\right) \tag{9a}$$

$$b_1^L = \gamma_1^L \frac{\text{Cov}[p_2, u_1]\tau_{u_1} + \Lambda_1^*}{\text{Var}[p_2|u_1]} - 1 \in \left(-1, -\frac{\mu\gamma}{\mu\gamma + \gamma_1^L}\right) \tag{9b}$$

$$b_{21}^L = -\frac{\mu\gamma}{\mu\gamma + \gamma_2^L} \tag{9c}$$

$$b_{22}^L = \frac{\gamma_2^L b_{21}^L (1-\mu)\Lambda_1^*\tau_v}{\mu}. \tag{9d}$$

The coefficient Λ_1^*, i.e. the first period endowment shock's negative price impact, is our measure of liquidity:

$$\Lambda_1^* = -\frac{\partial p_1}{\partial u_1}. \tag{10}$$

As is standard in economies with noise traders and risk-averse liquidity suppliers, Λ_1^* reflects dealers' compensation to absorb the outstanding imbalance in their inventory: the cost of supplying liquidity. However, differently from a noise trader economy, in this model dealers' inventory depends on the equilibrium trading decisions of FDs *and* first period traders. To see this, consider (9a). In view of (6) and (9b), at equilibrium first period traders hold a fraction

$$1 + b_1^L = \gamma_1^L \frac{\text{Cov}[p_2, u_1]\tau_{u_1} + \Lambda_1^*}{\text{Var}[p_2|u_1]}, \tag{11}$$

of their endowment shock. At the same time, comparing (3b) with (6), FDs aggregate speculative position per unit of endowment shock is given by

$$\mu\gamma \frac{E[p_2 - p_1|u_1]}{\text{Var}[p_2|p_1]u_1} = \mu\gamma \frac{1 + b_1^L}{\gamma_1^L}. \tag{12}$$

Thus, summing (11) and (12) yields the total speculative exposure of FDs and first period traders per unit of u_1 (i.e., the fraction of the endowment shock that is not absorbed by liquidity suppliers):

$$1 + b_1^L + \mu\gamma\frac{1 + b_1^L}{\gamma_1^L} = \frac{(\mu\gamma + \gamma_1^L)(1 + b_1^L)}{\gamma_1^L}, \tag{13}$$

and the complement to one of (13) captures dealers' inventory (per unit of endowment shock):

$$\text{Dealer's inventory per unit of endowment shock } = 1 - \frac{(\mu\gamma + \gamma_1^L)(1 + b_1^L)}{\gamma_1^L}. \tag{14}$$

At date 1 FDs know that they will be able to unwind their inventory in the second trading round, when x_1^L reverts. However, at that point in time, a new generation of traders enters the market. These traders hedge a new endowment shock, exposing FDs to the risk of holding their initial inventory until the liquidation date. Thus, for given inventory (14), the riskier is the asset, and the more risk averse FDs are, the higher is the risk borne by liquidity suppliers, and, according to (9a), the less liquid is the market.

According to (9b) and (9c), first and second period traders demand liquidity to hedge a fraction of their endowment. In the second period, such a fraction corresponds to FDs' relative risk-bearing capacity (see (9c)); in the first period, instead, it is larger than that (see (9b)). This is because second period traders' hedging activity creates price volatility which heightens first period traders' uncertainty, and leads them to demand more liquidity and FDs to cut back on their speculative activity. Indeed, when second period traders' endowment shock is null, first period traders' hedging aggressiveness reaches its upper bound, and the market is infinitely liquid:

Corollary 1. *In a transparent market, when the second period endowment shock is null ($\tau_{u_2} \to \infty$), first period traders' liquidity demand matches FDs' relative risk-bearing capacity, and the market is infinitely liquid ($b_1^L \to -(\mu\gamma + \gamma_1^L)^{-1}\mu\gamma$, and $\Lambda_1^* \to 0$).*

According to (9d), second period traders also speculate on the propagated order imbalance by putting a negative weight on their signal ($b_{22}^L < 0$), which is increasing in Λ_1^*. This is because, for $u_1 > 0$, the reversion of first period trades creates a positive imbalance at date 2, which prompts second period traders to short the asset. A less liquid first period market makes it more profitable for Ds to absorb u_1, which strengthens the positive dependence between p_2, and u_1:

$$\mathrm{Cov}[p_2, u_1] = \frac{(1-\mu)\lambda_2\tau_v\Lambda_1^*}{\tau_{u_1}}\left(\frac{\gamma_2^L b_{21}^L}{\mu} + \gamma\right). \tag{15}$$

Thus, as Λ_1^* increases, second period traders step up their speculative aggressiveness.

Analytically, the equilibrium obtains as the unique solution to the system (9a)–(9b):

$$\Lambda_1^* = \frac{1}{\gamma\tau_v}\left(1 - \frac{(\mu\gamma + \gamma_1^L)(1+b_1^L)}{\gamma_1^L}\right) \tag{16a}$$

$$b_1^L = \gamma_1^L(\gamma + \gamma_2^L)(\mu\gamma + \gamma_2^L)\tau_v^2\tau_{u_2}\Lambda_1^* - 1, \tag{16b}$$

which can be understood as the intersection between the inverse supply and demand of liquidity (respectively, (16a) and (16b)). This is so because b_1^L measures the fraction of the endowment shock that first period traders hedge in the market, while Λ_1^* captures the price adjustment dealers require to accommodate the order imbalance. A less liquid first period market increases the cost of scaling down traders' exposure, and leads the latter to hedge less of their endowment. Thus, in this case a drop in liquidity has a "rationing" effect on liquidity consumption, and the demand for liquidity is a *decreasing* function of Λ_1^*.[11] Conversely, a lower hedging aggressiveness implies a larger speculative position for FDs, which shrinks the imbalance that liquidity suppliers have to clear in the first period, and leads to a more liquid market. Hence, the (inverse) supply of liquidity is *decreasing* in b_1^L. In Figure 3 we provide a graphical illustration of the equilibrium determination.

In Figure 3 we also graphically analyze the effect of an increase in the mass of FDs on Λ_1^*.

[11] As $b_1^L < 0$, and positively sloped in Λ_1^*, a higher illiquidity implies that traders shed a lower fraction of their endowment, or that their liquidity demand subsides.

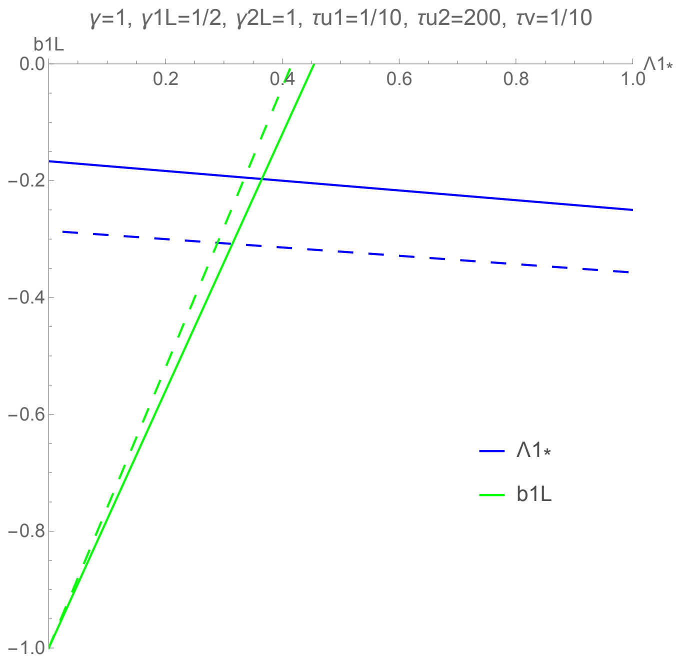


Figure 3: Transparency and equilibrium uniqueness. The solid (dashed) curves are drawn assuming $\mu = 1/10$ ($\mu = 1/5$). When $\mu = 1/10$, $\{\Lambda_1^*, b_1^L\} = \{.4, -.2\}$, while when $\mu = 1/5$, $\{\Lambda_1^*, b_1^L\} = \{.3, -.3\}$.

The solid (dashed) curves in the figure are drawn for $\mu = 1/10$ ($\mu = 1/5$). A larger μ has a positive effect on the cost of trading for all levels of b_1^L, since, according to (12), the aggregate speculative position of FDs increases, lowering dealers' inventory. As a result, when μ increases, the new function Λ_1^* shifts downwards. Consider now b_1^L. Based on (11), a larger μ has two contrasting effects on first period traders' hedging aggressiveness: on the one hand, as one can compute using (8a) and (8b), first period return uncertainty is given by:

$$\mathrm{Var}[p_2|u_1] = \frac{(\lambda_2 b_{21}^L)^2}{\tau_{u_2}} = \frac{1}{(\mu\gamma + \gamma_2^L)^2 \tau_v^2 \tau_{u_2}}, \tag{17}$$

which is decreasing in μ. Therefore, a larger μ lowers first period traders' uncertainty about p_2, and makes them consume less liquidity. However, according to (15),

$$\frac{\partial \mathrm{Cov}[p_2, u_1]}{\partial \mu} < 0 \tag{18}$$

and a higher μ lowers the positive association between the second period price and the first period endowment shock, making speculation less profitable. This pushes first period traders to shed a larger fraction of their endowment, increasing dealers' inventory, and consuming more liquidity. When the market is transparent, this latter effect is never strong enough to offset the former two and we obtain:

Corollary 2. *In a transparent market, liquidity increases in the proportion of fast dealers ($\partial\Lambda_1^*/\partial\mu < 0$).*

We concentrate our analysis on the liquidity of the first period market. However, note that as the volatility of the first period price is given by

$$\mathrm{Var}[p_1] = (\Lambda_1^*)^2 \tau_{u_1}^{-1},$$

our liquidity results can also be interpreted in terms of price volatility.

4 Opaqueness and the feedback effect of illiquidity

Suppose now that second period traders' signal on u_1 has a bounded precision ($\tau_\eta < \infty$). This setup characterizes a scenario where some traders (FDs, in our setup) have access to better market information (for example on order imbalances) compared to others (the second cohort of traders), and given our previous discussion, is likely to hold at a high trading frequency. In this case, we obtain the following result:

Proposition 2. *When $0 < \tau_\eta < \infty$, at equilibrium: $x_1^L = b_1^L u_1$, $x_2^L = b_{21}^L u_2 + b_{22}^L s_{u_1}$,*

$$b_{21}^L = -\frac{\mu\gamma}{\gamma_2^L + \mu\gamma\kappa} \tag{19a}$$

$$b_{22}^L = \gamma_2^L b_{21}^L \tau_v \tau_\eta \mathrm{Cov}[p_2, u_1|\Omega_2^L], \tag{19b}$$

where

$$\kappa \equiv \tau_v \mathrm{Var}[v - p_2|\Omega_2^L] > 1, \tag{19c}$$

and the first and second period return uncertainty are respectively given by $\mathrm{Var}[p_2|u_1] = \lambda_2^2((b_{21}^L)^2/\tau_{u_2} + (b_{22}^L)^2/\tau_\eta)$*, and* $\mathrm{Var}[v - p_2|\Omega_2^L] = \mathrm{Var}[v] + (\lambda_2(1-\mu)\gamma\tau_v\Lambda_1^*)^2\mathrm{Var}[u_1|s_{u_1}]$.

Differently from the transparent market benchmark, second period traders now face uncertainty on the price at which their order is executed, besides that on the liquidation value. This additional source of uncertainty is captured by the coefficient κ (see (19c)). As a consequence, they hedge a lower fraction of their endowment shock (see (19a)). Other things equal, as μ increases, u_1 propagates less to period 2, κ tends to 1, and second period traders (i) hedge more of their endowment shock, and (ii) speculate less aggressively on the propagated shock:

$$\lim_{\mu\to 1} b_{21}^L = -\frac{\gamma}{\gamma_2^L + \gamma}, \ \lim_{\mu\to 1} b_{22}^L = 0. \tag{20}$$

We are now ready to analyze the effect of a shock to liquidity on the equilibrium coefficients:

Corollary 3. *At equilibrium, the impact of the first period endowment shock on the second period price, second period traders' return uncertainty and hedging aggressiveness are increasing in illiquidity:*

$$\frac{\partial \mathrm{Cov}[p_2, u_1]}{\partial \Lambda_1^*} > 0, \ \frac{\partial \mathrm{Var}[v - p_2|\Omega_2^L]}{\partial \Lambda_1^*} > 0, \ \frac{\partial b_{21}^L}{\partial \Lambda_1^*} > 0. \tag{21}$$

An increase in Λ_1^* has an ambiguous effect on first period traders' hedging responsiveness and return uncertainty, and on second period traders' speculative aggressiveness (b_1^L, $\mathrm{Var}[p_2|u_1]$, and b_{22}^L).

According to (21) as in the transparent market case, a less liquid first period market increases the positive association between p_2 and u_1. Furthermore, as second period traders do not perfectly observe u_1, this also augments these traders' uncertainty and, according to (21), lowers their hedging responsiveness (recall that $b_{21}^L < 0$).

Importantly, an increase in Λ_1^* has two contrasting effects on the speculative aggressiveness of second period traders (b_{22}^L). Direct computation yields: $\mathrm{Cov}[p_2, u_1|\Omega_2^L] = \lambda_2(1-\mu)\gamma\tau_v\Lambda_1^*\mathrm{Var}[u_1|s_{u_1}]$. Thus, differentiating b_{22}^L we obtain:

$$\frac{\partial b_{22}^L}{\partial \Lambda_1^*} = \gamma_2^L \tau_v \tau_\eta \left(\underbrace{\mathrm{Cov}[p_2, u_1|\Omega_2^L] \frac{\partial b_{21}^L}{\partial \Lambda_1^*}}_{\text{Uncertainty effect } (+)} + \underbrace{b_{21}^L \frac{\partial \mathrm{Cov}[p_2, u_1|\Omega_2^L]}{\partial \Lambda_1^*}}_{\text{Speculation effect } (-)} \right). \tag{22}$$

On the one hand, like in the transparent market benchmark, an increase in Λ_1^* augments second period traders' speculative opportunities, and drives them to trade more against the u_1-led imbalance (the second term in the parenthesis in (22)). On the other hand, a higher Λ_1^* augments second period traders return uncertainty, and makes them speculate less (the first term in the parenthesis). Consider now the effect of an increase in Λ_1^* on $\mathrm{Var}[p_2|u_1]$:

$$\frac{\partial \mathrm{Var}[p_2|u_1]}{\partial \Lambda_1^*} = 2\lambda_2^2 \left(\underbrace{\frac{b_{21}^L}{\tau_{u_2}} \frac{\partial b_{21}^L}{\partial \Lambda_1^*}}_{(-)} + \underbrace{\frac{b_{22}^L}{\tau_\eta} \frac{\partial b_{22}^L}{\partial \Lambda_1^*}}_{(\pm)} \right). \tag{23}$$

In the transparent market benchmark, an increase in Λ_1^* has *no impact* on first period traders' uncertainty over p_2 (see (17)). In contrast, according to (21), opaqueness introduces two channels through which a shock to liquidity feeds back to first period traders' uncertainty. First, an increase in Λ_1^* lowers second period traders' hedging activity, lowering $\mathrm{Var}[p_2|u_1]$. However, as we argued above, a less liquid first period market can spur more speculation by second period traders. As traders' information is imprecise, this yields a second feedback channel that can instead magnify first period traders' uncertainty. Thus, according to (23), the ultimate impact of a shock to Λ_1^* on first period traders' uncertainty depends on the strength of the speculation effect. Finally, because of opaqueness, an increase in Λ_1^* introduces an additional effect on b_1^L:

$$\frac{\partial b_1^L}{\partial \Lambda_1^*} = \frac{\gamma_1^L}{\mathrm{Var}[p_2|u_1]^2} \times \left(\underbrace{\left(\frac{\partial \mathrm{Cov}[p_2, u_1]}{\partial \Lambda_1^*}\tau_{u_1} + 1\right)\mathrm{Var}[p_2|u_1]}_{\text{Direct effect } (+)} - \underbrace{\frac{\partial \mathrm{Var}[p_2|u_1]}{\partial \Lambda_1^*}(\mathrm{Cov}[p_2, u_1]\tau_{u_1} + \Lambda_1)}_{\text{Feedback effect } (\pm)} \right). \tag{24}$$

For given $\mathrm{Var}[p_2|u_1]$, as p_2 is more positively associated with u_1, a larger Λ_1^* leads first period traders to speculate more (and hedge less), as per the *direct* liquidity consumption "rationing" effect of the transparent market benchmark. However, when the speculation effect leads $\mathrm{Var}[p_2|u_1]$ to increase in Λ_1^*, a less liquid market now also has a *feedback* liquidity consumption "expanding" effect on b_1^L. As a higher Λ_1^* increases the risk to which first period traders are exposed, a less liquid market can lead them to hedge more. As a result, first period traders' demand for liquidity can become *increasing* in Λ_1^*, as shown in Figure 4: an increase in the cost of liquidity provision incites more liquidity consumption.

The expanding effect of illiquidity can be responsible for a destabilizing dynamic whereby to a sizeable evaporation of liquidity, first period traders respond with an even more aggressive liquidity consumption. In the figure we use the same parameter values of Figure 3, but assume that $\tau_\eta = 10$ (instead of $\tau_\eta \to \infty$). As a result, at equilibrium we obtain

$$b_1^L|_{\tau_\eta=10} = -0.5, \;\; \Lambda_1^*|_{\tau_\eta=10} = 3.8.$$

Compared to the values of the example of Figure 3, these results correspond to a more than two- and an almost ten-fold increase in liquidity consumption and illiquidity.

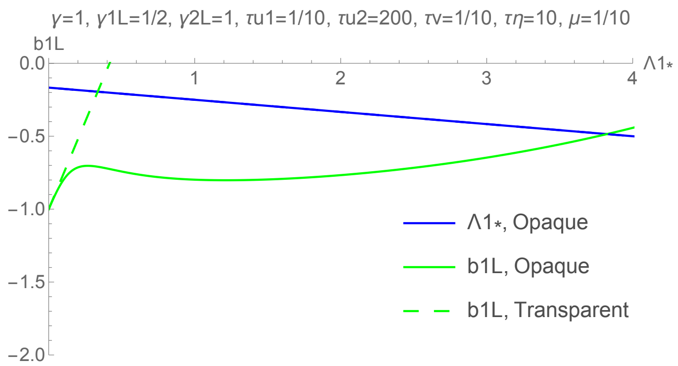


Figure 4: When the market is opaque, first period traders' demand for liquidity can turn increasing in Λ_1^*. The dashed curve corresponds to b_1^L in the transparent market case.

4.1 Equilibrium multiplicity

A second effect of opaqueness is the possibility of multiple, self-fulfilling equilibria. According to Corollary 3, a less liquid first period market heightens the time-propagation of the first period shock. This, in turn, can lead second period traders to speculate more aggressively on the u_1-led imbalance (see (22)), which can increase the uncertainty faced by first period traders on p_2 (see (23)). As a consequence, first period traders can decide to hedge more, and FDs to speculate less (see (24)).[12] This chain of effects turns out to be particularly strong when the risk bearing capacity of FDs is not too low, first period traders are sufficiently risk averse, second period traders have a sufficiently informative signal, and face low endowment risk, and the risk of the asset payoff is large. In these conditions, an initial dearth of liquidity escalates into a loop that sustains three equilibrium levels of liquidity:

Proposition 3. *There exists a set of parameter values* $\{\underline{\tau}_{u_2}, \bar{\tau}_v, \underline{\tau}_\eta, \bar{\mu}, \underline{\gamma}, \underline{\gamma}_1^L\}$, *such that for* $\tau_{u_2} > \underline{\tau}_{u_2}$, $\tau_v < \bar{\tau}_v$, $\tau_\eta > \underline{\tau}_\eta$, $\mu < \bar{\mu}$, $\gamma > \underline{\gamma}$, *and* $\gamma_1^L < \bar{\gamma}_1^L$, *three equilibrium levels of liquidity* $(\Lambda_1^*)^H, (\Lambda_1^*)^I, (\Lambda_1^*)^L$ *arise, where*

$$0 < (\Lambda_1^*)^H < \frac{\mu}{1-\mu} < (\Lambda_1^*)^I < \frac{1}{1-\mu} < (\Lambda_1^*)^L < \frac{1}{\gamma\tau_v}. \tag{25}$$

We will refer to the equilibrium where Λ_1^* is low (resp., intermediate, and high) as the High, (resp., Intermediate, and Low) liquidity equilibrium (HLE, ILE, and LLE). Note that since the function $\Lambda_1^*(b_1^L)$ is decreasing in b_1^L (see (16a)), the hedging activity of first period traders is respectively high, intermediate, and low along $(\Lambda_1^*)^L$, $(\Lambda_1^*)^I$, and $(\Lambda_1^*)^H$. This is a further manifestation of the fact that the feedback effect of liquidity jams the stabilizing impact of an increase in illiquidity on traders' hedging demand.

We can interpret the ratios

$$\frac{\mu}{1-\mu}, \ \frac{1}{\gamma\tau_v}, \tag{26}$$

in (25), respectively, as the likelihood that FDs second period liquidity supply is enough to absorb the demand coming from first period traders' reverting orders, and as liquidity suppliers' perceived uncertainty about the asset payoff. Then, condition (25) states that when multiplicity arises, the likelihood that FDs' inventory is sufficient to stave off a liquidity shortage is smaller than FDs' perceived asset payoff risk. Figure 5 provides a numerical example of the proposition.

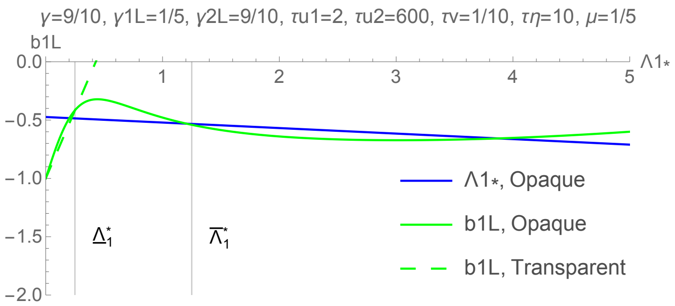


Figure 5: Market opaqueness and equilibrium multiplicity. At equilibrium $\{\Lambda_1^*, b_1^L\} \in \{\{0.4, -0.5\}, \{1, -0.5\}, \{3.9, -0.7\}\}$.

[12]Because of (12), whenever first period traders consume more liquidity, FDs speculate less, increasing the inventory held by liquidity suppliers.

The following corollary follows from Proposition 3:

Corollary 4. *When the volatility of the second period endowment shock grows unboundedly ($\tau_{u_2} \to \infty$) and the following parameter restriction applies: $\tau_v < \bar{\tau}_v$, $\tau_\eta > \underline{\tau}_\eta$, $\mu < \bar{\mu}$, $\gamma_1^L < \bar{\gamma}_1^L$, we can rank liquidity at the different equilibria as follows:*

$$0 = (\Lambda_1^*)^H < (\Lambda_1^*)^I < \frac{1}{1-\mu} < (\Lambda_1^*)^L < \frac{1}{\gamma\tau_v}. \tag{27}$$

When $\tau_{u_2} \to \infty$, second period traders have no endowment to hedge, and only trade to speculate on the u_1-induced imbalance. In the equilibrium where $\Lambda_1^* = 0$, $x_1^D = 0$, so that first period traders' orders are absorbed by FDs' speculative trades, no imbalance arises in the second period, and $b_{22}^L = 0$ (see (19b)). When second period traders' signal on u_1 is fully revealing, this equilibrium is unique (Corollary 1). For τ_η finite, however, first period traders cannot rule out the possibility that second period traders speculate on a certain realization of s_{u_1} that gives an incorrect signal about u_1 (e.g., $s_{u_1} > 0$, while $u_1 < 0$). This increases the uncertainty they face, and trigger the loop that can lead to the appearance of two further equilibria. Figure 6 provides a graphical illustration of the equilibrium determination when the conditions in Corollary 4 are satisfied.

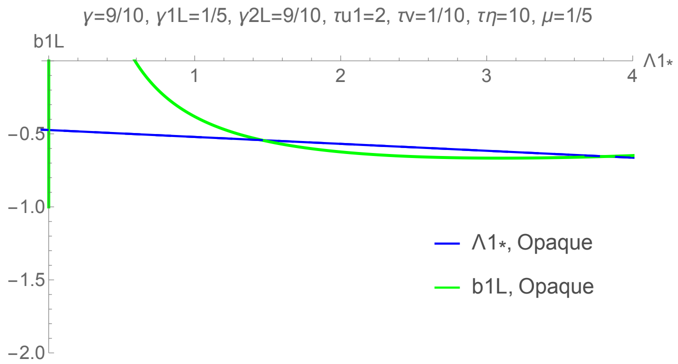


Figure 6: Equilibrium multiplicity with no second period endowment risk. At equilibrium $\{\Lambda_1^*, b_1^L\} \in \{\{0, -0.5\}, \{1.1, -0.5\}, \{4.2, -0.7\}\}$ (the function b_1^L is very steep at $\Lambda_1^* = 0$).

4.2 Uniqueness and comparative statics

Equilibrium uniqueness can occur at both an equilibrium with a high level of liquidity and one with a low level of liquidity (corresponding respectively to the HLE and the LLE when there are multiple equilibria), depending on parameters' values. A high FDs' risk bearing capacity (large γ and/or μ), low payoff risk (high τ_v), low volatility or high risk tolerance of first period traders' endowments (respectively, high τ_{u_1} and γ_1^L), high signal precision (high τ_η), and low risk tolerance of second period traders (low γ_2^L) can lead to uniqueness at a high liquidity equilibrium. Indeed, for such parameters' values, first period traders have a lower need for immediacy, face a lower risk from holding the asset, and benefit from a larger FDs presence at interim. Furthermore, second period traders' orders create less first period return uncertainty. All of these effects contribute to weaken the strategic complementarity loop, facilitating a high liquidity equilibrium. Conversely, a high volatility of second period traders' endowments works to strengthen the loop, as it heightens second period traders' demand for immediacy, which in turn increases the return uncertainty faced by first period traders. As a consequence, a lower τ_{u_2} can instead lead to uniqueness at an equilibrium with low liquidity. Figures 10 and 11 (in the Appendix) illustrate how uniqueness at an equilibrium with high and low liquidity arises.

Along the equilibrium with high liquidity, shocks to parameters have a monotone effect on market liquidity. More in detail, a higher dealers' risk bearing capacity (higher γ or μ), lower payoff risk (higher τ_v), lower volatility of first period endowment and higher first period risk-tolerance (higher τ_{u_1} or γ_1^L), make the market more liquid; a higher signal precision (τ_η), instead, makes the market less liquid. An example of these effects is presented in Figure 7, Panel (a) (for μ), and Panel (b). Along the equilibrium with low liquidity, shocks to parameters can have non-monotone effects. We illustrate this feature for the proportion of FDs in Figure 7, Panel (c). The intuition for the non-monotonicity is as follows. An increase in μ triggers two potentially contrasting effects on liquidity:

$$\frac{\partial \Lambda_1^*}{\partial \mu} = -\frac{1}{\gamma \tau_v} \times \left(\underbrace{\frac{\gamma(1+b_1^L)}{\gamma_1^L}}_{\text{Direct effect } (+)} + \underbrace{\frac{\partial b_1^L}{\partial \mu}\left(1+\frac{\mu\gamma}{\gamma_1^L}\right)}_{\text{Indirect effect } (\pm)} \right). \tag{28}$$

For given b_1^L, the direct effect captures the increase in FDs' aggregate speculative position, which works to lower dealers' inventory, and make the market more liquid. The indirect effect reflects the impact of the change in μ on first period traders' demand for liquidity ($\partial b_1^L/\partial\mu$), and on each FD's speculative position ($(\mu\gamma/\gamma_1^L)(\partial b_1^L/\partial\mu)$). The sign of this effect is, instead, ambiguous. Indeed, an increase in the mass of FDs can lower the impact of second period traders' orders on p_2, thereby lowering $\mathrm{Var}[p_2|u_1]$ and leading first period traders to hold more of their endowment, and each FD to speculate more on the short term capital gain;[13] at the same time, however, it can also lower the propagation of u_1 to the second period, impairing the predictability of p_2, and inducing traders to shed more of their endowment, and each FD to speculate less.[14] When the market is opaque, second period traders face execution risk, which tames their hedging aggressiveness (see (19a)), and lowers first period traders' return uncertainty. In this situation, the uncertainty reduction effect of μ on b_1^L can be dwarfed by the one due to reduced predictability. As a consequence, when μ increases, first period traders' demand for immediacy can increase (Figure 7, Panel (c)) and the individual speculative activity of each FD can abate, offsetting the direct positive effect of FDs' aggregate speculative trades. Hence, a wider FDs' participation can impair liquidity.

[13] The volatility reduction can happen for two different reasons. As μ increases, (i) less of the first period endowment shock propagates to the second period, and (ii) more FDs absorb second period liquidity traders' orders, enhancing risk sharing.

[14] In a transparent market, this latter effect is never strong enough to overcome the previous two, and liquidity increases in the mass of FDs (see Corollary 2).

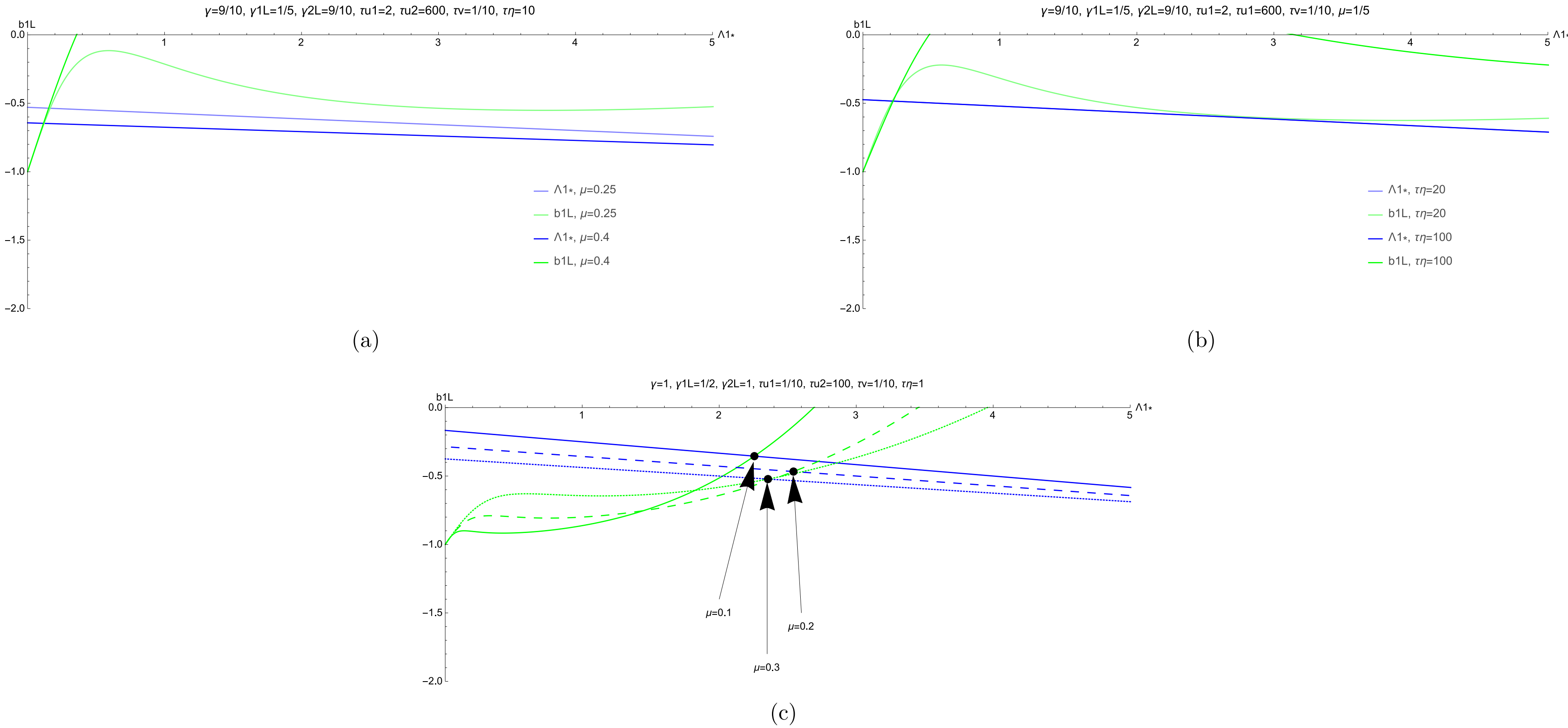


Figure 7: Comparative statics along the High and Low Liquidity Equilibrium. A higher proportion of FDs, increases liquidity (Panel (a)), while a higher signal precision, decreases it (Panel (b)). Along the low liquidity equilibrium, an increase in μ can have a non-monotone effect on market liquidity (Panel (c)).

4.3 Fragility

The feedback loop induced by market opaqueness implies that liquidity can be "fragile," in the sense that a relatively small shock to one of the model's parameters can lead to a disproportionately large change in liquidity. We show this with two examples. In the first one, we consider the effect of a parameter shock yielding a switch from the HLE to an equilibrium with low liquidity. In the second example, we show that along an equilibrium with low liquidity, Λ_1^* can be hump-shaped in μ. This implies that a sudden reduction in the mass of FDs can lead to a large drop in liquidity.

4.3.1 Equilibrium switch

Consider Panel (a) in Figure 8, and suppose that initially the market is at the high liquidity equilibrium, where $\Lambda_1^* = 0.22$. Suppose that a technical "glitch" disconnects 6% of the FDs. In this new situation, as argued in Section 4.1, the plot for $\Lambda_1^*(b_1^L)$ shifts upwards, while the one for $b_1^L(\Lambda_1^*)$ moves downwards, as illustrated in panel (b) of the figure. As a result, a new, unique equilibrium obtains with $\Lambda_1^* = 3.5$, which corresponds to a 16-fold liquidity decrease.

A similar effect also arises if we shock the volatility of first or second period traders' endowment. To see this, suppose now that starting from Panel (a) in the figure, we introduce a 5% negative shock to τ_{u_1} (i.e., we move τ_{u_1} from 2 to 1.9), which increases the likelihood that an order of an unusual magnitude hits the first period market. As argued in Section 4.2 this leads to a downward shift in the plot for b_1^L which, as we show in Panel (c) of Figure 8, is large enough to eliminate the HLE and move the market towards a new equilibrium with low liquidity in which $\Lambda_1^* = 3.5$ and $b_1^L = -0.6$. Finally, suppose that we increase the volatility of the second period endowment shock, introducing a 7% negative shock to τ_{u_2} (lowering it to 620). In this new situation, the plot for the function $b_1^L(\Lambda_1^*)$ moves downwards, while the one for $\Lambda_1^*(b_1^L)$ is unchanged (see Panel (d) in Figure 8). A unique equilibrium obtains, where $\Lambda_1^* = 3.5$, implying liquidity dry-up comparable to the one of the previous examples.

Table 1 summarizes the results of these exercises and compares them with the effects that obtain in the transparent market case.

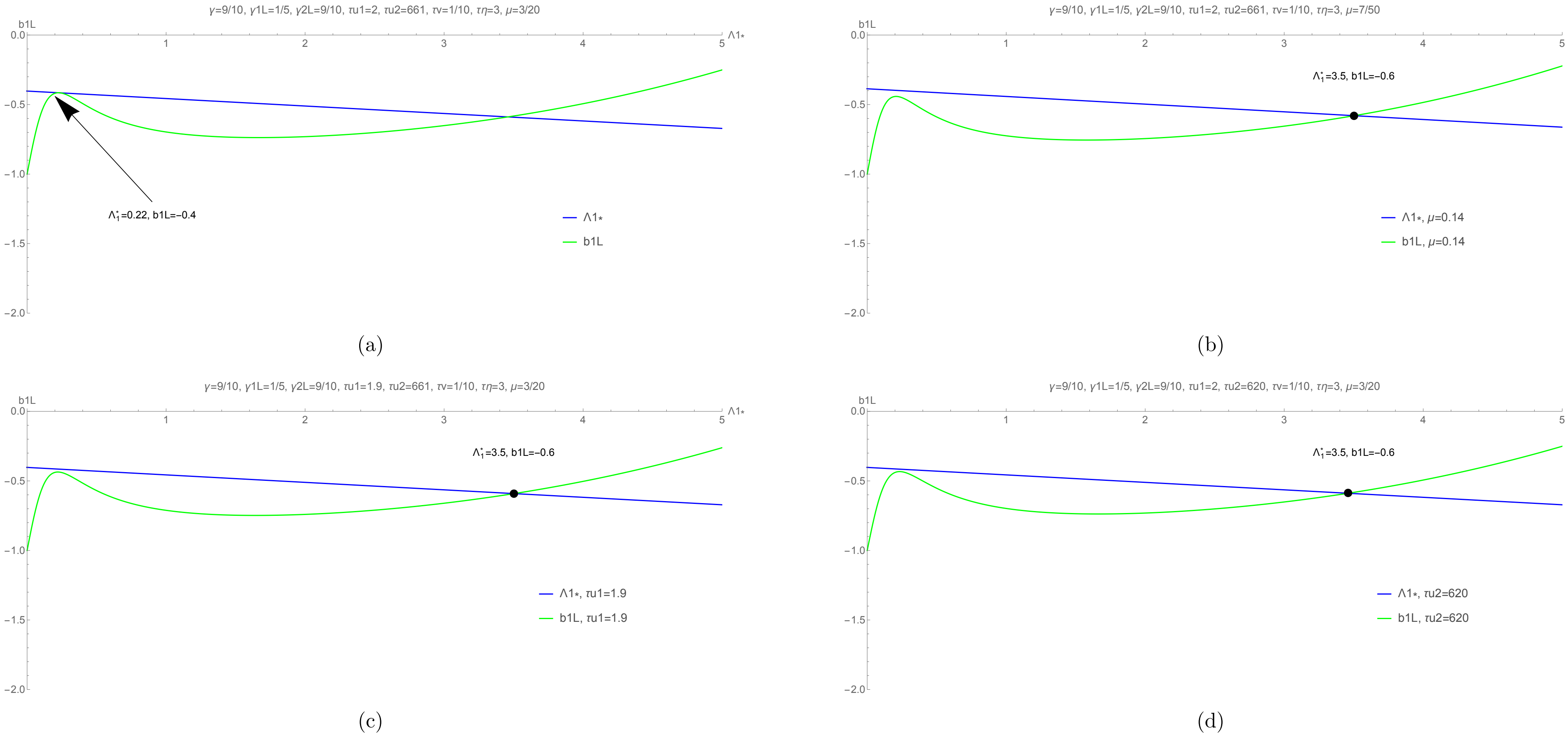


Figure 8: Liquidity fragility. Comparing panel (a) with (b) illustrates the effect of a decrease in the mass of FDs. Comparing panel (a) with (c) and (d) illustrates the effect of an increase in the volatility of first and second period liquidity traders' demand.

	Status quo Λ_1^*	Shock to parameter	New Λ_1^*	$\Delta\Lambda_1^*/\Lambda_1^*$
		$\Delta\mu/\mu = -6\%$	0.245	3.2%
Transparent market ($\tau_\eta \to \infty$)	0.237	$\Delta\tau_{u_1}/\tau_{u_1} = -5\%$	0.237	0%
		$\Delta\tau_{u_2}/\tau_{u_2} = -7\%$	0.252	6.3%
		$\Delta\mu/\mu = -6\%$	3.5	1470%
Opaque market ($\tau_\eta < \infty$)	0.22	$\Delta\tau_{u_1}/\tau_{u_1} = -5\%$	3.5	1470%
		$\Delta\tau_{u_2}/\tau_{u_2} = -7\%$	3.5	1470%

Table 1: Equilibrium switch: The impact of a shock to μ, τ_{u_1}, and τ_{u_2}, in the transparent, and opaque market case. Other parameters' values are as in Figure 8.

4.3.2 Fragility along a unique equilibrium with low liquidity

In Section 4.2 we have shown that along an equilibrium with low liquidity, an increase in the mass of FDs can have a non-monotone impact on Λ_1^*. This implies that illiquidity can be hump-shaped in the proportion of FDs, as shown in Figure 9 (Panel (a)). The non-monotone relationship between Λ_1^* and μ illustrates an additional channel through which liquidity fragility can arise. To see this, suppose that the proportion of FDs in the market is initially $\mu = 0.4$. According to the figure, for this fraction of FDs' participation, in the opaque market case we have $\Lambda_1^* = 0.5$, and $b_1^L = -0.47$ (see Panel (a) and Panel (b)). Suppose now that a glitch disconnects 10% of FDs, implying that a proportion $\mu = 0.36$ of FDs supplies liquidity. In the opaque market case, this implies a new illiquidity level $\Lambda_1^* = 1.58$, corresponding to a 216% liquidity withdrawal. Conversely, in the case with transparent markets, when $\mu = 0.4$, $\Lambda_1^* = 0.38$, while when $\mu = 0.36$, $\Lambda_1^* = 0.41$, corresponding to an 8% liquidity decrease. This shows that along an equilibrium with low liquidity, following a reduction in FD participation, liquidity can dry up quite dramatically.

The example highlights an additional implication of our analysis. When $\mu = 0.36$, $b_1^L = -0.51$ (see Figure 9, Panel (b)), which, compared to the status quo liquidity demand, corresponds to a 9% increase in liquidity consumption by first period traders (i.e., $b_1^L = -0.47$ when $\mu = 0.4$). How can such a comparatively small increase in liquidity consumption generate an illiquidity spike of this magnitude? To understand this, note that increased liquidity consumption only accounts for *part* of the total effect, as it occurs *jointly* with a steep increase in first period return uncertainty (according to Panel (c) in the figure, $\text{Var}[p_2|u_1]$ experiences a 310% increase across the two equilibrium outcomes). This, as noted in Section 4.2, leads each FD to scale down his speculative position, thereby adding to the aggregate effect of a reduced liquidity supply. Using expression (28) to break down the different effects yields:

$$\underbrace{\frac{\Delta\Lambda_1^*}{\Delta\mu}}_{\approx -27} \approx -\underbrace{\frac{1}{\gamma\tau_v}}_{=-10} \times \tag{29}$$

$$\underbrace{\left(\frac{\gamma(1+b_1^L)}{\gamma_1^L}\right.}_{\Delta\text{ in FDs' aggregate speculative position}\approx 1} + \underbrace{\frac{\mu\gamma}{\gamma_1^L}}_{=4/5} \times \underbrace{\frac{\Delta b_1^L}{\Delta\mu}}_{\Delta\text{ in each FD speculative position}\approx 1} + \underbrace{\left.\frac{\Delta b_1^L}{\Delta\mu}\right)}_{\Delta\text{ in liquidity demand}\approx 1}$$

Thus, the increase in liquidity consumption accounts for roughly 36% of the drought, whereas the lion share of it (about 64%) is due to the combined effect of the aggregate and individual reduction in FDs' speculative activity.[15] This suggests that the empirical analysis seeking to explain sudden and large changes in liquidity has to look beyond the impact of changes in the demand for immediacy, and also account for the effect of changes in HFTs' short-term speculative activity.[16]

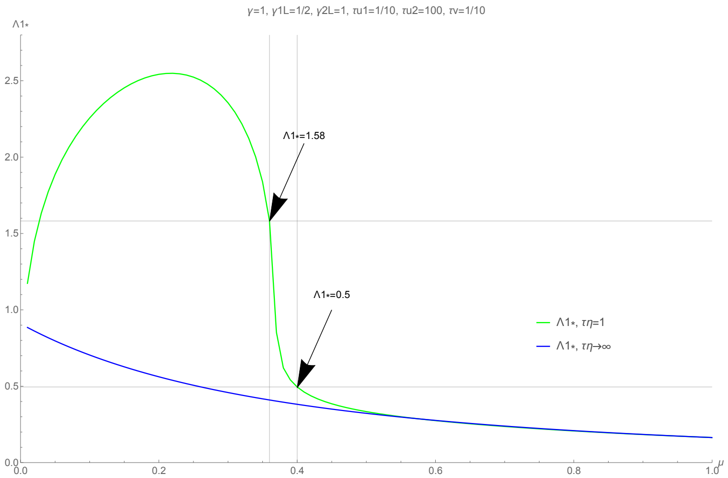
A final implication of this example is that when information on prices and/or order imbalances is opaque, an increase in the mass of HFTs (promoting full participation), can lower market liquidity. This finding is consistent with Boehmer et al. (2015) who show that greater algorithmic trading intensity is associated with more liquidity for average firm size, the same is not true for small market cap firms. For these firms, when algorithmic trading increases, liquidity declines.[17]

Summarizing: this section highlights the role of informational frictions in generating a liquidity feedback loop that can have a destabilizing effect on the market. Second period traders, endowed with a noisy signal on the first period endowment shock, speculate against the propagated order imbalance, generating additional volatility. This can feed back on first period traders' strategies, leading them to consume more liquidity and FDs to retreat from speculation, thereby magnifying the inventory held by liquidity suppliers, and further lowering market liquidity. This self-sustaining loop can induce multiple equilibria and liquidity fragility. Equilibria can be ranked in terms of liquidity and first period traders' hedging activity, with the most (least) liquid equilibrium occurring with the least (highest) liquidity consumption. Thus, with market opaqueness, the self-stabilizing mechanism whereby an illiquidity spike depresses liquidity consumption can jam, and instead be replaced by a vicious cycle that creates a liquidity rout.

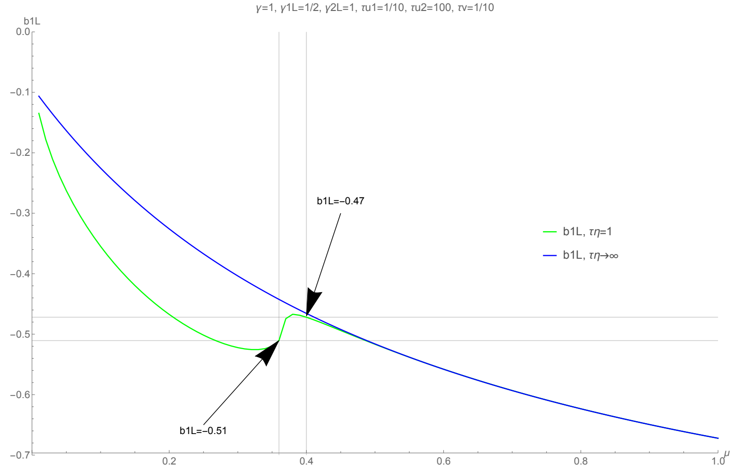
[15] According to (29), the sum of Δ in FDs' aggregate speculative position, Δ in each FD speculative position, and Δ in traders' liquidity demand amounts to 2.8, of which $\Delta b_1^L/\Delta\mu$ accounts for 1.

[16] In this example too, for μ small, the stabilizing effect of illiquidity is jammed and first period traders demand more immediacy *precisely* when the cost of liquidity supply is increasing.

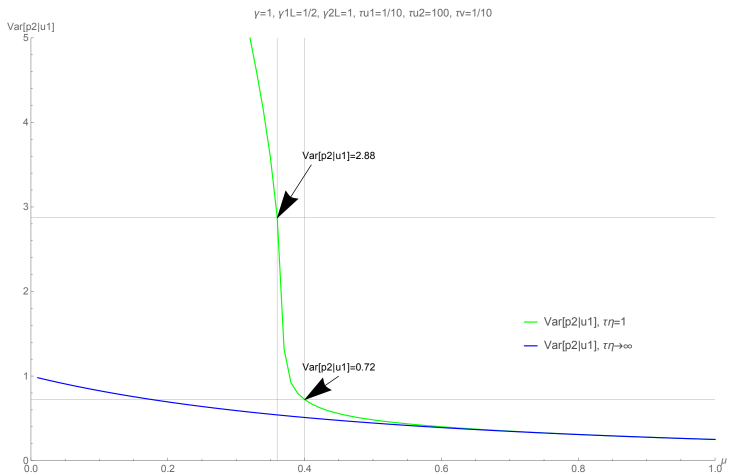
[17] See also Breckenfelder (2014) for other evidence on the negative impact of an increase in HFT competition on market liquidity for a sample of stocks traded on the Stockholm Stock Exchange.



(a)



(b)



(c)

Figure 9: Fragility at an equilibrium with low liquidity. In panel (a), (b), and (c) we plot Λ_1^*, b_1^L, and $\mathrm{Var}[p_2|u_1]$, for $\tau_v = \tau_{u_1} = 0.1$, $\tau_{u_2} = 100$, $\tau_\eta = \gamma = \gamma_2^L = 1$, $\gamma_1^L = 0.5$, and $\mu \in \{0.01, .02, \ldots, 1\}$; the blue (green) plot relates to the transparent (opaque) case. A glitch disconnecting 10% of FDs yields a 9% increase in the demand for immediacy, a 310% increase in first period uncertainty and a 216% illiquidity spike. The gridlines are drawn at $\mu = \hat{\mu} \in \{0.36, 0.4\}$ and at the corresponding values for Λ_1^*, b_1^L, and $\mathrm{Var}[p_2|u_1]$.

	Transparent ($\tau_\eta \to \infty$)	Opaque ($0 < \tau_\eta < \infty$)
Equilibrium	Unique	Possible ME
Liquidity	Increasing in μ	Can be 'fragile' and hump-shaped in μ (Numerical)
First period traders' demand for immediacy	Decreasing in Λ_1^*	Can be increasing in Λ_1^*

Table 2: Summary of results.

5 Concluding remarks

We study a 2-period model in which two classes of dealers—full and standard—intermediate the orders of two successive cohorts of short-term traders, in a context where markets are fragmented due to both an informational and a participation friction. We show that dealers' limited market participation favors the propagation of first period traders' endowment shock across time, inducing a predictable price pressure. This, in turn, leads second period traders to speculate against the propagated endowment shock. The effect of speculation crucially depends on the transparency regime governing the market. More in detail, our main findings can be summarized as follows (see Table 2):

1. When second period traders have perfect information about the first period endowment shock, speculation exerts a stabilizing effect. In this context, a unique equilibrium obtains, and a dearth of liquidity increases first period traders' cost of hedging, reducing their liquidity consumption. Furthermore, higher FDs participation always has a beneficial impact on liquidity.

2. When the market is opaque—in that second period traders' information is imprecise—speculation can *augment first period traders' uncertainty*, leading them to demand more immediacy when the market is less liquid. This can offset the rationing impact of illiquidity, and trigger a liquidity feedback loop in which a liquidity dry-up breeds a further, larger liquidity withdrawal. We show that in this scenario,

   (a) Multiple equilibria—that can be ranked in terms of liquidity, price volatility, and demand for immediacy—can arise.

   (b) Uniqueness can obtain with either a high or a low level of liquidity, and comparative statics is not necessarily monotone in the latter case. For example, along an equilibrium with low liquidity, an increase in the mass of FDs can impair market liquidity.

   (c) Liquidity can be fragile, either because a shock to parameter values can prompt a switch from the high liquidity equilibrium to an equilibrium with low liquidity;

alternatively, because, along an equilibrium with low liquidity, a reduction in FDs participation can generate a large spike in illiquidity.

From a methodological point of view, our work shows that fragility can arise in a context where prices are driven by a non-payoff related shock. We view this as a realistic feature of trading at high frequencies since in those conditions, the chances that payoff fundamentals drive prices are negligible. This also allows us to offer an alternative explanation for how the buildup of a large imbalance can precipitate the market into a crash, which does not rely on the effect of order flow toxicity (Easley et al. (2011, 2012)).

From a policy perspective, our paper has two important implications. First, our analysis of the opaque market model shows that favoring FDs' entry (i.e., reducing the participation friction) doesn't necessarily enhance liquidity. Indeed, illiquidity can be hump-shaped in the proportion of FDs. This can also serve as a guide to empirical analysis, as it suggests that the liquidity impact of HFT entry should be assessed taking into account the effect of frictions in the access to market information. Second, with noisy market information, the presence of liquidity providers acting with different trading frequencies may make liquidity fragile, either because a shock to parameters can prompt a switch across equilibria; or because, along an equilibrium with low liquidity, due to the hump-shaped relationship between illiquidity and the proportion of FDs, a sudden reduction of these dealers' participation can lead to a large liquidity withdrawal. This supports regulatory concerns about the potential drawbacks of automated trading due to operational and transmission risks.[18] This also implies that fragility can arise in the absence of order flow toxicity, suggesting two possible lines of intervention to reduce the likelihood of flash episodes. First, allowing access to transparent market information to *all market participants at the same time*, would limit the uncertainty-increasing effect of speculation on predictable price pressures, besides reducing the toxicity of order flows. Furthermore, making liquidity provision by different agents more in sync, would reduce fragmentation and help avoiding predictable, short-lived, price pressures that, as we have argued, with opaque information can make liquidity fragile.

Finally, our analysis of fragility along an equilibrium with low liquidity highlights the role that changes in HFTs' strategies in the wake of crashes have to explain huge illiquidity spikes. As we argued in our numerical example, the total reduction in HFTs' speculative activity can act as a multiplier of the initial increase in traders' liquidity consumption. This can be of help in empirical analyses of these events, in that an exclusive focus on changes in liquidity consumption can miss a potentially important explanatory factor.

[18]See Joint Staff Report: The U.S. Treasury Market on October 15, 2014.

References

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

Boehmer, E., K. Fong, and J. Wu (2015). International evidence on algorithmic trading. *Working paper*.

Bogousslavsky, V. (2014). Infrequent rebalancing and short-term return predictability. *Working Paper*.

Bongaerts, D. and M. Van Achter (2015). High frequency trading and market stability. *Working Paper, Erasmus University*.

Breckenfelder, H. J. (2014). Competition between high-frequency traders, and market quality. *Working Paper*.

Budish, E., P. Cramton, and J. Shim (2015). The high-frequency trading arms race: Frequent batch auctions as a market design response. *Quarterly Journal of Economics 139*, 1547–1621.

Cespa, G. and T. Foucault (2014). Illiquidity contagion and liquidity crashes. *Review of Financial Studies 27*(6), 1615–1660.

Cespa, G. and X. Vives (2012). Dynamic trading and asset prices: Keynes vs. Hayek. *Review of Economic Studies 79*, 539–580.

Cespa, G. and X. Vives (2015). The Beauty Contest and short-term trading. *Journal of Finance 70*(5), 2099–2154.

Chien, Y., H. Cole, and H. Lustig (2012). Is the volatility of the market price of risk due to intermittent portfolio rebalancing? *American Economic Review 102*(6), 2859–96.

Collin-Dufresne, P. and V. Vos (2015). Do prices reveal the presence of informed trading? *Journal of Finance 70*, 1555–1582.

Ding, S., J. Hanna, and T. Hendershott (2014). How slow is the NBBO? a comparison with direct exchange feeds. *Financial Review* (49), 313–332.

Du, S. and H. Zhu (2014). Welfare and optimal trading frequency in dynamic double auctions. *Working paper*.

Duffie, D. (2010). Presidential address: Asset price dynamics with slow-moving capital. *Journal of Finance 65*(4), 1237–1267.

Easley, D., M. O'Hara, and M. López de Prado (2011). The microstructure of the Flash-Crash: flow toxicity, liquidity crashes, and the probability of informed trading. *Journal of Portfolio Management 37*, 118–128.

Easley, D., M. O'Hara, and M. López de Prado (2012). Flow toxicity and liquidity in a high-frequency world. *Review of Financial Studies 25*, 1457–1493.

Foucault, T., J. Hombert, and I. Rosu (2015). News trading and speed. *Journal of Finance (Forthcoming)*.

Goldstein, I., Y. Li, and L. Yang (2014). Speculation and hedging in segmented markets. *Review of Financial Studies 27*(3), 881–922.

Goldstein, I. and L. Yang (2015). Information diversity and complementarities in trading and information acquisition. *Journal of Finance 70*, 1723–1765.

Hendershott, T., S. Li, A. Menkveld, and M. Seasholes (2014). Asset price dynamics with limited attention. *Working Paper*.

Heston, S. L., R. A. Korajczyk, and R. Sadka (2010). Intraday patterns in the cross-section of stock returns. *The Journal of Finance 65*(4), 1369–1407.

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

Huang, J. and J. Wang (2009). Liquidity and market crashes. *Review of Financial Studies 22*(7), 2607–2643.

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

Menkveld, A. J. and B. Yueshen (2012). Middlemen interaction and its effect on market quality. *Working Paper*.

Menkveld, A. J. and M. A. Zoican (2015). Need for speed? exchange latency and liquidity. *Working Paper, Timbergen Institute*.

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

Peress, J. and D. Schmidt (2015). Noise traders incarnate: Describing a realistic noise trading process. *Working paper*.

SEC (2010). Concept release on equity market structure. *Federal Register*.

Spiegel, M. (1998). Stock price volatility in a multiple security overlapping generations model. *Review of Financial Studies 11*(2), 419–447.

Vayanos, D. and J. Wang (2012). Liquidity and asset returns under asymmetric information and imperfect competition. *Review of Financial Studies 25*(5), 1339–1365.

Vives, X. (2008). *Information and learning in markets: the impact of market microstructure.* Princeton University Press.

A Appendix

The following is a standard results (see, e.g. Vives (2008), Technical Appendix, pp. 382–383) that allows us to compute the unconditional expected utility of market participants.

Lemma 2. *Let the n-dimensional random vector $z \sim N(0, \Sigma)$, and $w = c + b'z + z'Az$, where $c \in \mathbb{R}$, $b \in \mathbb{R}^n$, and A is a $n \times n$ matrix. If the matrix $\Sigma^{-1} + 2\rho A$ is positive definite, and $\rho > 0$, then*

$$E[-\exp\{-\rho w\}] = -|I + 2\rho\Sigma A|^{-1/2} \exp\{-\rho(c - \rho b'(\Sigma + 2\rho A)^{-1} b)\}.$$

Proof of Lemma 1

Denote by $\mu x_1^{FD} = \int_0^\mu x_1^{FD} di$, $(1-\mu)x_1^D = \int_\mu^1 x_1^D di$, and by x_1^L respectively the aggregate position of FDs, dealers and liquidity traders in the first period. Imposing market clearing yields:

$$\mu x_1^{FD} + (1-\mu)x_1^D + x_1^L = 0 \iff \mu\varphi_1^{FD}(p_1) + (1-\mu)\varphi^D(p_1) + b_1^L u_1 = 0. \tag{A.1}$$

At equilibrium the coefficients of traders' strategies are known, which implies that p_1 is observationally equivalent to u_1 and that both FDs and dealers can retrieve u_1 from the price. Therefore, the information set of a FD and a dealer in the first period coincide and are given by $\Omega_1^{FD} = \Omega_1^D = \{u_1\}$. In the second period, denote by $\mu x_2^{FD} = \int_0^\mu x_2^{FD} di$ and by x_2^L, respectively the aggregate position of FDs and second period liquidity traders. Impose market clearing:

$$\mu(x_2^{FD} - x_1^{FD}) + (x_2^L - x_1^L) = 0,$$

and rearrange the first period market clearing condition as follows

$$(1-\mu)x_1^D = -\left(\mu x_1^{FD} + x_1^L\right).$$

Substitute the latter in the second period clearing equation to obtain

$$\mu x_2^{FD} + x_2^L + (1-\mu)x_1^D = 0. \tag{A.2}$$

Once again, at a linear equilibrium the coefficient of traders' strategies are known, which implies that the price sequence $\{p_1, p_2\}$ is observationally equivalent to $\{u_1, x_2^L\}$. Thus, the second period information set of a FD is given by $\Omega_2^{FD} = \{s_v, u_1, u_2\}$. □

Proof of Proposition 1

When $\tau_\eta \to \infty$, explicit computation of b_1^L yields

$$b_1^L = \gamma_1^L(\gamma_2^L + \gamma)(\gamma_2^L + \mu\gamma)\Lambda_1^* \tau_{u_2} \tau_v^2 - 1.$$

The above, together with (9a) yield a system of two equations in (b_1^L, Λ_1^*) which can be solved to obtain the following explicit expressions:

$$b_1^L = -\frac{\gamma(1+(\gamma+\gamma_2^L)(\mu\gamma+\gamma_2^L)\mu\tau_{u_2}\tau_v)}{\gamma+(\gamma+\gamma_2^L)(\mu\gamma+\gamma_2^L)(\mu\gamma+\gamma_1^L)\tau_{u_2}\tau_v} \tag{A.3a}$$

$$\Lambda_1^* = \frac{1}{\tau_v(\gamma+(\mu\gamma+\gamma_1^L)(\mu\gamma+\gamma_2^L)(\gamma_2^L+\gamma)\tau_{u_2}\tau_v)}. \tag{A.3b}$$

From the expression in (A.3b) it is easy to see that an increase in any one of the model's parameters increases the liquidity of the market. For (A.3a) the expressions for the relevant derivatives are as follows:

$$\frac{\partial b_1^L}{\partial \mu} = \frac{\gamma\gamma_1^L\tau_{u_2}\tau_v(\gamma-\tau_{u_2}\tau_v(\gamma+\gamma_2^L)(\mu\gamma+\gamma_2^L)^2)}{(\gamma+(\gamma+\gamma_2^L)(\mu\gamma+\gamma_2^L)(\mu\gamma+\gamma_1^L)\tau_{u_2}\tau_v)^2} \tag{A.4a}$$

$$\frac{\partial b_1^L}{\partial \gamma} = \frac{\gamma_1^L\tau_{u_2}\tau_v(\mu\gamma^2-(\gamma_2^L)^2-\mu\tau_{u_2}\tau_v(\gamma+\gamma_2^L)^2(\mu\gamma+\gamma_2^L)^2)}{(\gamma+(\gamma+\gamma_2^L)(\mu\gamma+\gamma_2^L)(\mu\gamma+\gamma_1^L)\tau_{u_2}\tau_v)^2} \tag{A.4b}$$

$$\frac{\partial b_1^L}{\partial \gamma_1^L} = \frac{\gamma(\gamma+\gamma_2^L)(\mu\gamma+\gamma_2^L)\tau_{u_2}\tau_v(1+\mu\tau_{u_2}\tau_v(\gamma+\gamma_2^L)(\mu\gamma+\gamma_2^L))}{(\gamma+(\gamma+\gamma_2^L)(\mu\gamma+\gamma_2^L)(\mu\gamma+\gamma_1^L)\tau_{u_2}\tau_v)^2} \tag{A.4c}$$

$$\frac{\partial b_1^L}{\partial \gamma_2^L} = \frac{\gamma\gamma_1^L\tau_{u_2}\tau_v(\mu\gamma+\gamma+\gamma_2^L)}{(\gamma+(\gamma+\gamma_2^L)(\mu\gamma+\gamma_2^L)(\mu\gamma+\gamma_1^L)\tau_{u_2}\tau_v)^2} \tag{A.4d}$$

$$\frac{\partial b_1^L}{\partial \tau_v} = \frac{\gamma\gamma_1^L\tau_{u_2}(\gamma+\gamma_2^L)(\mu\gamma+\gamma_2^L)}{(\gamma+(\gamma+\gamma_2^L)(\mu\gamma+\gamma_2^L)(\mu\gamma+\gamma_1^L)\tau_{u_2}\tau_v)^2} \tag{A.4e}$$

$$\frac{\partial b_1^L}{\partial \tau_{u_2}} = \frac{\gamma\gamma_1^L\tau_v(\gamma+\gamma_2^L)(\mu\gamma+\gamma_2^L)}{(\gamma+(\gamma+\gamma_2^L)(\mu\gamma+\gamma_2^L)(\mu\gamma+\gamma_1^L)\tau_{u_2}\tau_v)^2}. \tag{A.4f}$$

Expressions (A.4c)-(A.4f) are positive. For (A.4a) to be positive, we need

$$\tau_v < \frac{\gamma}{(\gamma+\gamma_2^L)(\mu\gamma+\gamma_2^L)^2\tau_{u_2}}, \tag{A.5}$$

whereas for (A.4b) to be positive the following two conditions are required:

$$\mu > \left(\frac{\gamma_2^L}{\gamma}\right)^2, \ \tau_v < \frac{\mu\gamma^2-(\gamma_2^L)^2}{\mu\tau_{u_2}(\gamma+\gamma_2^L)^2(\mu\gamma+\gamma_2^L)^2}. \tag{A.6}$$

□

Proof of Corollary 1

Taking the limit for $\tau_{u_2} \to \infty$ in (A.3a) yields the desired result.

□

Proof of Proposition 2

In the second period a new mass of liquidity traders endowed with risk-tolerance coefficient $\gamma_2^L > 0$ enter the market. A date-2 liquidity trader submits a market order

$$X_2^L(u_2, s_{u_1}) = b_{21}^L u_2 + b_{22}^L s_{u_1}, \tag{A.7}$$

with $u_2 \sim N(0, \tau_{u_2}^{-1})$, and $s_{u_1} = u_1 + \eta$, with $\eta \sim N(0, \tau_\eta^{-1})$ and u_2, η independent of all the other random variables in the model. Consider the sequence of market clearing equations

$$\mu x_1^{FD} + (1-\mu)x_1^D + x_1^L = 0 \tag{A.8a}$$

$$\mu(x_2^{FD} - x_1^{FD}) + (b_{21}^L u_2 + b_{22}^L s_{u_1} - x_1^L) = 0. \tag{A.8b}$$

Rearrange (A.8a) as follows:

$$(1-\mu)x_1^D = -\left(\mu x_1^{FD} + x_1^L\right).$$

Substitute the latter in (A.8b):

$$\mu x_2^{FD} + b_{21}^L u_2 + b_{22}^L s_{u_1} + (1-\mu)x_1^D = 0. \tag{A.9}$$

A FD maximizes the expected utility of his second period wealth:

$$\begin{aligned}
&E\Big[-\exp\Big\{-\frac{1}{\gamma}\Big((p_2-p_1)x_1^{FD} + (v-p_2)x_2^{FD}\Big)\Big\}|p_1, p_2\Big] = \\
&= E\left[\exp\Big\{-\frac{1}{\gamma}(p_2-p_1)x_1^{FD}\Big\}\Big(-\exp\Big\{-\frac{1}{\gamma}(v-p_2)x_2^{FD}\Big\}\Big)|p_1,p_2\right] \\
&= \exp\Big\{-\frac{1}{\gamma}(p_2-p_1)x_1^{FD}\Big\}E\left[-\exp\Big\{-\frac{1}{\gamma}(v-p_2)x_2^{FD}\Big\}|p_1,p_2\right] \qquad \text{(A.10)} \\
&= \exp\Big\{-\frac{1}{\gamma}(p_2-p_1)x_1^{FD}\Big\}\Big(-\exp\Big\{-\frac{1}{\gamma}\Big(E[v-p_2|p_1,p_2]x_2^{FD} - \frac{(x_2^{FD})^2}{2\gamma}\mathrm{Var}[v-p_2|p_1,p_2]\Big)\Big\}\Big),
\end{aligned}$$

where the last expression in (A.10) is due to CARA and normality. For given x_1^{FD} the above is a concave function of the second period strategy x_2^{FD}. Solving the FOC, yields that in the second period a FD's limit order is given by $X_2^{FD}(p_1, p_2) = -\gamma\tau_v p_2$. Similarly, due to CARA and normality, in the first period a traditional market maker maximizes

$$E\Big[-\exp\Big\{-\frac{1}{\gamma}(v-p_1)x_1^D\Big\}|p_1\Big] = -\exp\Big\{-\frac{1}{\gamma}\Big(E[v-p_1|p_1]x_1^D - \frac{(x_1^D)^2}{2\gamma}\mathrm{Var}[v-p_1|p_1]\Big)\Big\}. \tag{A.11}$$

Hence, his strategy is given by $X_1^D(p_1) = -\gamma\tau_v p_1$. Substituting these strategies in (A.9) and solving for p_2 yields

$$p_2 = \lambda_2\left(b_{21}^L u_2 + b_{22}^L s_{u_1}\right) - \frac{1-\mu}{\mu}p_1, \tag{A.12}$$

where $\lambda_2 = 1/\mu\gamma\tau_v$. The assumption that first period liquidity traders' strategies are linear implies that $p_1 = -\Lambda_1 u_1$ (see below). As a consequence we can rewrite (A.12) as follows:

$$p_2 = \lambda_2(b_{21}^L u_2 + b_{22}^L s_{u_1}) + \lambda_2(1-\mu)\gamma\tau_v\Lambda_1 u_1. \tag{A.13}$$

CARA and normality assumptions imply that the objective function of a second period liquidity trader is given by

$$E[-\exp\{-\pi_2^L/\gamma_2^L\}|\Omega_2^L] = -\exp\left\{-\frac{1}{\gamma}\left(E[\pi_2^L|\Omega_2^L] - \frac{1}{2\gamma}\text{Var}[\pi_2^L|\Omega_2^L]\right)\right\}, \tag{A.14}$$

where $\Omega_2^L = \{u_2, s_{u_1}\}$, and $\pi_2^L \equiv (v - p_2)x_2^L + u_2 v$. Maximizing (A.14) with respect to x_2^L, the strategy of a second period liquidity trader is given by

$$X_2^L(u_2, s_{u_1}) = \gamma_2^L \frac{E[v - p_2|\Omega_2^L]}{\text{Var}[v - p_2|\Omega_2^L]} - \frac{\text{Cov}[v - p_2, v|\Omega_2^L]}{\text{Var}[v - p_2|\Omega_2^L]}u_2. \tag{A.15}$$

Computing

$$E[v - p_2|\Omega_2^L] = -\left(\lambda_2(b_{21}^L u_2 + b_{22}^L s_{u_1}) + \frac{1-\mu}{\mu}\Lambda_1 \frac{\tau_\eta}{\tau_\eta + \tau_{u_1}} s_{u_1}\right) \tag{A.16a}$$

$$\text{Var}[v - p_2|\Omega_2^L] = \frac{\mu^2(\tau_{u_1} + \tau_\eta) + ((1-\mu)\Lambda_1)^2\tau_v}{\mu^2(\tau_{u_1} + \tau_\eta)\tau_v} \tag{A.16b}$$

$$\text{Cov}[v - p_2, v|\Omega_2^L] = \frac{1}{\tau_v}. \tag{A.16c}$$

Substituting (A.16a), (A.16b), and (A.16c) in (A.15) and identifying coefficients yields

$$X_2^L(u_2, s_{u_1}) = b_{21}^L u_2 + b_{22}^L s_{u_1},$$

where

$$b_{21}^L = -\frac{1}{\tau_v(\gamma_2^L\lambda_2 + \text{Var}[v - p_2|\Omega_2^L])} \tag{A.17a}$$

$$b_{22}^L = -\frac{\gamma_2^L\tau_\eta\lambda_2(1-\mu)\gamma\tau_v\Lambda_1}{(\tau_\eta + \tau_{u_1})(\gamma_2^L\lambda_2 + \text{Var}[v - p_2|\Omega_2^L])}. \tag{A.17b}$$

According to (A.15) second period liquidity traders' strategies react both to endowment and informational shocks. Thus, there are *two* measures of the price impact of trades in the second period (see (A.13)):

$$\lambda_{21} \equiv \frac{\partial p_2}{\partial u_2} = \lambda_2 b_{21}^L \tag{A.18a}$$

$$\lambda_{22} \equiv \frac{\partial p_2}{\partial s_{u_1}} = \lambda_2 b_{22}^L. \tag{A.18b}$$

Expressions (A.18a) and (A.18b) respectively correspond to the price impact of a marginal increase in the endowment shock and in the realization of the signal about u_1 observed by second period liquidity traders.

Consider now the first period. We start by characterizing the strategy of a FD. Substituting the optimal strategy in (A.10), rearranging and applying Lemma 2 yields the following expression for the first period objective function of a FD:

$$E[U((p_2 - p_1)x_1^{FD} + (v - p_2)x_2^{FD})|u_1] = -\left(1 + \frac{\mathrm{Var}[p_2|u_1]}{\mathrm{Var}[v]}\right)^{-1/2} \times \tag{A.19}$$
$$\exp\left\{-\frac{1}{\gamma}\left(\frac{\gamma\tau_v}{2}\nu^2 + (\nu - p_1)x_1^{FD} - \frac{(x_1^{FD} + \gamma\tau_v\nu)^2}{2\gamma}\left(\frac{1}{\mathrm{Var}[p_2|u_1]} + \frac{1}{\mathrm{Var}[v]}\right)^{-1}\right)\right\},$$

where

$$\nu \equiv E[p_2|u_1] = \left(\lambda_2 b_{22}^L + \frac{1-\mu}{\mu}\Lambda_1\right) u_1 \tag{A.20a}$$
$$\mathrm{Var}[p_2|u_1] = \lambda_2^2\left(\frac{(b_{21}^L)^2}{\tau_{u_2}} + \frac{(b_{22}^L)^2}{\tau_\eta}\right). \tag{A.20b}$$

Maximizing (A.19) with respect to x_1^{FD} and solving for the first period strategy yields

$$X_1^{FD}(p_1) = \frac{\gamma}{\mathrm{Var}[p_2|u_1]}\nu - \gamma\left(\frac{1}{\mathrm{Var}[p_2|u_1]} + \frac{1}{\mathrm{Var}[v]}\right)p_1. \tag{A.21}$$

As we argued above, due to CARA and normality, for traditional market makers at date 1 we have $X_1^D(p_1) = -\gamma\tau_v p_1$. At equilibrium we then have

$$\mu\left(\frac{\gamma}{\mathrm{Var}[p_2|u_1]}\nu - \gamma\left(\frac{1}{\mathrm{Var}[p_2|u_1]} + \frac{1}{\mathrm{Var}[v]}\right)p_1\right) + (-(1-\mu)\gamma\tau_v p_1) + b_1^L u_1 = 0,$$

implying that p_1 is linear in u_1: $p_1 = -\Lambda_1 u_1$, with Λ_1 to be determined.

We now turn to the characterization of first period liquidity traders' strategies. CARA and normality imply

$$E[-\exp\{-\pi_1^L/\gamma_1^L\}] = -\exp\left\{-\frac{1}{\gamma}\left(E[\pi_1^L|u_1] - \frac{1}{2\gamma_1^L}\mathrm{Var}[\pi_1^L|u_1]\right)\right\}, \tag{A.22}$$

where $\pi_1^L \equiv (p_2 - p_1)x_1^L + u_1 p_2$. Maximizing (A.22) with respect to x_1^L, and solving for the optimal strategy, yields

$$X_1^L(u_1) = \gamma_1^L\frac{E[p_2 - p_1|u_1]}{\mathrm{Var}[p_2 - p_1|u_1]} - \frac{\mathrm{Cov}[p_2 - p_1, p_2|u_1]}{\mathrm{Var}[p_2 - p_1|u_1]}u_1. \tag{A.23}$$

Computing

$$p_2 - p_1 = \left(\lambda_2 b_{22}^L + \frac{\Lambda_1}{\mu}\right)u_1 + \lambda_2\left(b_{21}^L u_2 + b_{22}^L \eta\right),$$

and

$$E[p_2 - p_1|u_1] = \left(\lambda_2 b_{22}^L + \frac{\Lambda_1}{\mu}\right)u_1 \tag{A.24a}$$
$$\mathrm{Cov}[p_2 - p_1, p_2|u_1] = \mathrm{Var}[p_2|u_1]. \tag{A.24b}$$

Substituting the above in the strategy of a first period liquidity trader and identifying yields

$$X_1^L(u_1) = b_1^L u_1, \tag{A.25}$$

where

$$b_1^L = \gamma_1^L \frac{\mu\lambda_2 b_{22}^L + \Lambda_1}{\mu \text{Var}[p_2|u_1]} - 1. \tag{A.26}$$

Substituting (A.21), x_1^D, and (A.25) in the first period market clearing condition and solving for the price yields $p_1 = -\Lambda_1 u_1$, where

$$\begin{aligned}\Lambda_1 = \psi(\Lambda_1) &\equiv -\left(\mu\gamma\left(\frac{1}{\text{Var}[p_2|u_1]} + \frac{1}{\text{Var}[v]}\right) + (1-\mu)\gamma\frac{1}{\text{Var}[v]}\right)^{-1}\left(\mu\frac{\gamma\text{Cov}[p_2,u_1]}{\text{Var}[p_2|u_1]\text{Var}[u_1]} + b_1^L\right)\\ &= -\frac{\mu\gamma\text{Cov}[p_2,u_1]\tau_{u_1} + b_1^L\text{Var}[p_2|u_1]}{\gamma(\mu + \tau_v\text{Var}[p_2|u_1])}.\end{aligned} \tag{A.27}$$

According to (A.26), the equilibrium coefficient of a first period liquidity trader depends on b_{21}^L, and b_{22}^L. Therefore, recursive substitution of the equilibrium strategies' coefficients in (A.27) shows that Λ_1 is pinned down by the solution of the following equation in Λ_1:

$$\psi(\Lambda_1) - \Lambda_1 = \frac{(\mu\gamma + \gamma_1^L)(\text{Cov}[p_2,u_1]\tau_{u_1} + \Lambda_1) + \text{Var}[p_2|u_1](\gamma\tau_v\Lambda_1 - 1)}{\gamma(\mu + \tau_v\text{Var}[p_2|u_1])} = 0. \tag{A.28}$$

For $\mu \in (0,1]$ the denominator in the above expression is positive, which implies that equilibria are pinned down by solutions to the quintic at the numerator of (A.28):

$$f(\Lambda_1) \equiv g_1(\Lambda_1) + g_2(\Lambda_1) + g_3(\Lambda_1) = 0, \tag{A.29}$$

where

$$g_1(\Lambda_1) \equiv -\mu^3\tau_{u_1}^2(1-\gamma\tau_v\Lambda_1) + \Lambda_1\tau_{u_2}\tau_v^2(\gamma_1^L + \mu\gamma)(\gamma\tau_v\Lambda_1^2(1-\mu)^2 + \mu\tau_{u_1}(\gamma_2^L + \mu\gamma))^2 \tag{A.30a}$$

$$g_2(\Lambda_1) \equiv \mu^3\tau_\eta^2(\Lambda_1\tau_v(\gamma + (\gamma_1^L + \gamma)(\gamma_1^L + \mu\gamma)(\gamma_2^L + \mu\gamma)\tau_{u_2}\tau_v) - 1), \tag{A.30b}$$

and

$$\begin{aligned}g_3(\Lambda_1) \equiv \tau_\eta\big(&-2\mu^3\tau_{u_1} - (\gamma_2^L)^2\Lambda_1^2(1-\mu)^2\mu\tau_{u_2}\tau_v^2 + \gamma\Lambda_1^3(1-\mu)^2\mu\tau_{u_2}\tau_v^3\times\\ &(\gamma_2^L(\gamma_1^L + \gamma_2^L) + (2\gamma\gamma_1^L + (\gamma + \gamma_1^L)\gamma_2^L)\mu + \mu^2\gamma(\gamma_2^L + 2\gamma))+\\ &+\mu^2\Lambda_1\tau_{u_1}\tau_v(2\mu\gamma + (\gamma_1^L + \mu\gamma)(\gamma_2^L + \mu\gamma)(\gamma_2^L + 2\mu\gamma + \mu\gamma_2^L)\tau_{u_2}\tau_v)\big).\end{aligned} \tag{A.30c}$$

Collecting the terms in Λ_1 in the quintic equation (A.29) yields

$$
\begin{aligned}
f(\Lambda_1) &= \gamma^2(1-\mu)^4\Lambda_1^5(\mu\gamma+\gamma_1^L)\tau_{u_2}\tau_v^4 \\
&+ \mu\gamma(1-\mu)^2\tau_{u_2}\tau_v^3\Lambda_1^3\big(((\mu\gamma+\gamma_1^L)(2\mu\gamma+\gamma_2^L(1+\mu))+(\gamma_2^L)^2)\tau_\eta+2\tau_{u_1}(\mu\gamma+\gamma_1^L)(\mu\gamma+\gamma_2^L)\big) \\
&- (\gamma_2^L)^2(1-\mu)^2\mu\tau_\eta\tau_{u_2}\tau_v^2\Lambda_1^2 \qquad\qquad (A.31)\\
&+ \mu^2(\tau_\eta+\tau_{u_1})\tau_v\Lambda_1\big(\mu\gamma(\tau_\eta+\tau_{u_1})+(\mu\gamma+\gamma_1^L)(\mu\gamma+\gamma_2^L)(\mu\gamma(\tau_\eta+\tau_{u_1})+\gamma_2^L(\mu\tau_\eta+\tau_{u_1}))\tau_{u_2}\tau_v\big) \\
&- \mu^3(\tau_\eta+\tau_{u_1})^2 = 0.
\end{aligned}
$$

The above expression shows that there are three sign changes in the sequence formed by the quintic's coefficients. Therefore, by Descartes' rule of sign, there are up to three positive roots of the equation $f(\Lambda_1)=0$.

Compute $\text{Cov}[p_2,u_1]$:

$$
\text{Cov}[p_2,u_1] = \frac{(1-\mu)\Lambda_1(\tau_{u_1}\gamma_2^L\lambda_2+(\tau_{u_1}+\tau_\eta)\text{Var}[v-p_2|\Omega_2^L])}{\mu\tau_{u_1}(\tau_{u_1}+\tau_\eta)(\gamma_2^L\lambda_2+\text{Var}[v-p_2|\Omega_2^L])}, \qquad (A.32)
$$

which is positive if and only if $\Lambda_1>0$. Consider (A.28) and suppose that at equilibrium $\Lambda_1^*<0$. From (A.32), $\text{Cov}[p_2,u_1]<0$. Due to (A.28) this implies $f(\Lambda_1^*)<0$, which is impossible. Thus, at equilibrium, $\Lambda_1^*>0$, and $\text{Cov}[p_2,u_1]\geq 0$. Similarly,

$$
\text{Cov}[p_2,u_1|\Omega_2^L] = \lambda_2(1-\mu)\gamma\tau_v\Lambda_1\text{Var}[u_1|s_{u_1}] \geq 0.
$$

To sign the strategy coefficient of a first period liquidity trader, we use (A.26):

$$
b_1^L = \gamma_1^L\frac{\text{Cov}[p_2,u_1]\tau_{u_1}+\Lambda_1}{\text{Var}[p_2|u_1]} - 1. \qquad (A.33)
$$

From (A.33) we obtain

$$
\frac{\text{Var}[p_2|u_1]}{\gamma_1^L}(1+b_1^L) = \text{Cov}[p_2,u_1]\tau_{u_1}+\Lambda_1,
$$

which substituted in (A.28) yields

$$
f(\Lambda_1) = \frac{\text{Var}[p_2|u_1]}{\gamma_1^L}\big((\mu\gamma+\gamma_1^L)(1+b_1^L)+\gamma_1^L(\gamma\tau_v\Lambda_1-1)\big) = 0. \qquad (A.34)
$$

Solving the above for Λ_1 yields:

$$
\begin{aligned}
\Lambda_1^* &= \frac{1}{\gamma\tau_v}\left(1-\frac{(1+b_1^L)(\mu\gamma+\gamma_1^L)}{\gamma_1^L}\right) \qquad (A.35a)\\
&= \frac{1}{\gamma_1^L\gamma\tau_v}(-\mu\gamma-b_1^L(\mu\gamma+\gamma_1^L)). \qquad (A.35b)
\end{aligned}
$$

Since $\Lambda_1^*>0$, the last expression in (A.35a) implies that at equilibrium $b_1^L<0$. Furthermore, using (A.33), $1+b_1^L>0$, which proves our result.

Taking the limit for $\tau_\eta \to \infty$ in $\psi(\Lambda_1)$ yields:

$$\lim_{\tau_\eta \to \infty} \psi(\Lambda_1) = \frac{1 - \Lambda_1(\gamma_2^L + \mu\gamma)(\gamma_1^L(\gamma + \gamma_2^L) + \mu\gamma^2(1-\mu))\tau_{u_2}\tau_v^2}{\gamma\tau_v(1 + \mu(\gamma_2^L + \mu\gamma)^2\tau_{u_2}\tau_v)}. \tag{A.36}$$

Identifying Λ_1:

$$f(\Lambda_1) = \Lambda_1(\tau_v(\gamma + (\gamma + \gamma_2^L)(\gamma_1^L + \mu\gamma)(\gamma_2^L + \gamma\mu)\tau_{u_2}\tau_v) - 1 = 0,$$

and a unique solution with

$$\Lambda_1^*|_{\tau_\eta \to \infty} \equiv \frac{1}{\tau_v(\gamma + (\mu\gamma + \gamma_1^L)(\mu\gamma + \gamma_2^L)(\gamma_2^L + \gamma)\tau_{u_2}\tau_v)}, \tag{A.37}$$

obtains. Note that liquidity is in this case increasing in μ. Also, according to (A.36) we have $\psi'(\Lambda_1^*) < 0$.

□

Proof of Corollary 3

From Proposition 2 it is immediate that $\text{Var}[v - p_2|\Omega_2^L]$ is increasing in Λ_1. Differentiating $\text{Cov}[p_2, u_1]$ yields

$$\frac{\partial \text{Cov}[p_2, u_1]}{\partial \Lambda_1} = \frac{1-\mu}{\mu(\tau_{u_1} + \tau_\eta)} + \frac{(1-\mu)\tau_\eta(\text{Var}[v - p_2|\Omega_2^L](\gamma_2^L\lambda_2 + \text{Var}[v - p_2|\Omega_2^L]) + \gamma_2^L\lambda_2\Lambda_1\text{Var}[v - p_2|\Omega_2^L])}{\mu\tau_{u_1}(\tau_{u_1} + \tau_\eta)(\gamma_2^L\lambda_2 + \text{Var}[v - p_2|\Omega_2^L])^2} \geq 0,$$

for $\mu \leq 1$. Finally,

$$\text{Cov}[p_2 - p_1, p_1] = -\frac{\Lambda_1^2}{\mu\tau_{u_1}}\left(\frac{\gamma_2^L\lambda_2(\mu\tau_\eta + \tau_{u_1}) + \text{Var}[v - p_2|\Omega_2^L](\tau_\eta + \tau_{u_1})}{(\tau_\eta + \tau_{u_1})(\gamma_2^L\lambda_2 + \text{Var}[v - p_2|\Omega_2^L])}\right) < 0.$$

□

Proof of Corollary 3

The equilibrium quintic (A.31) can be expressed as the sum of two polynomials: a quintic in Λ_1^* that multiplies τ_{u_2}, and a first degree polynomial in Λ_1^* that does not depend on τ_{u_2}, as shown in the expression below:

$$\begin{aligned} f(\Lambda_1) = \Big[&\Lambda_1(\tau_{u_1}(\gamma_2^L\lambda_2 + \text{Var}[v - p_2|\Omega_2^L]) + \tau_\eta(\mu\gamma_2^L\lambda_2 + \text{Var}[v - p_2|\Omega_2^L]))(\mu\gamma + \gamma_1^L)\times \\ &(\tau_{u_1} + \tau_\eta)(\gamma_2^L\lambda_2 + \text{Var}[v - p_2|\Omega_2^L])\tau_\eta + \lambda_2^2(\gamma_2^L\tau_\eta\lambda_2(1-\mu)\gamma\tau_v\Lambda_1)^2(\gamma\tau_v\Lambda_1 - 1)\mu\Big]\tau_{u_2}\tau_v^2 \\ &+ \lambda_2^2(\tau_{u_1} + \tau_\eta)^2\mu\tau_\eta(\gamma\tau_v\Lambda_1 - 1). \end{aligned} \tag{A.38}$$

Inspection of the equilibrium mapping $\psi(\Lambda_1)$ shows that if we let $\tau_{u_2} \to \infty$, the corresponding equilibrium quintic is proportional to the term in square brackets in (A.38) (i.e., the one that multiplies τ_{u_2}). We first concentrate on the analysis of this quintic:

$$\hat{f}(\Lambda_1) = \Lambda_1(\tau_{u_1}(\gamma_2^L\lambda_2 + \mathrm{Var}[v - p_2|\Omega_2^L]) + \tau_\eta(\mu\gamma_2^L\lambda_2 + \mathrm{Var}[v - p_2|\Omega_2^L]))(\mu\gamma + \gamma_1^L)\times \\ (\tau_{u_1} + \tau_\eta)(\gamma_2^L\lambda_2 + \mathrm{Var}[v - p_2|\Omega_2^L])\tau_\eta + \lambda_2^2(\gamma_2^L\tau_\eta\lambda_2(1-\mu)\gamma\tau_v\Lambda_1)^2(\gamma\tau_v\Lambda_1 - 1)\mu. \tag{A.39}$$

First, note that $\hat{f}(0) = 0$, implying that when $\tau_{u_2} \to \infty$, $\Lambda_1^* = 0$ is an equilibrium of the model. Additionally, considering $h(\Lambda_1) \equiv \hat{f}(\Lambda_1)/\Lambda_1$, a quartic in Λ_1, we can pin down parameter restrictions that ensure the existence of two additional equilibria. To see this, we start by evaluating $h(\cdot)$ at $\Lambda_1 = 0$ obtaining:

$$h(0) = \frac{\tau_\eta(\mu\gamma + \gamma_1^L)(\mu\gamma + \gamma_2^L)(\tau_{u_1} + \tau_\eta)(\mu\gamma(\tau_{u_1} + \tau_\eta) + \gamma_2^L(\tau_{u_1} + \mu\tau_\eta))}{\gamma^2\mu^2} > 0. \tag{A.40}$$

Next, evaluating $h(\cdot)$ at the point $\bar{\Lambda}_1^* = 1/(1-\mu)$, yields

$$h(\bar{\Lambda}_1^*) = \frac{\tau_\eta}{\gamma^2\tau_\eta^4}\times \\ \Big((\mu\gamma + \gamma_1^L)(\mu(\mu\gamma + \gamma_2^L)(\tau_{u_1} + \tau_\eta) + \gamma\tau_v)(\mu\gamma_2^L\mu(\tau_{u_1} + \mu\tau_\eta) + \gamma(\mu^2(\tau_\eta + \tau_{u_1}) + \tau_v)) \\ - \mu(\gamma_2^L)^2\tau_\eta(1 - \mu - \gamma\tau_v)\Big), \tag{A.41}$$

which is negative when the following parameter restrictions are satisfied:

$$0 < \mu < \bar{\mu} \equiv \frac{\gamma_2^L(\sqrt{5\gamma^2 + \gamma_2^L(2\gamma + \gamma_2^L)} - (\gamma + \gamma_2^L))}{2\gamma^2} \tag{A.42a}$$

$$0 < \tau_v < \bar{\tau}_v \equiv \frac{(1-\mu)(\gamma_2^L)^2 - \mu\gamma(\mu\gamma + \gamma_2^L)}{\gamma(\gamma_2^L)^2} \tag{A.42b}$$

$$\tau_\eta > \underline{\tau}_\eta \equiv \frac{\gamma(\mu(\mu\gamma + \gamma_2^L)\tau_{u_1} + \gamma\tau_v)}{(1 - \mu - \gamma\tau_v)(\gamma_2^L)^2 - \mu\gamma(\mu\gamma + \gamma_2^L)} \tag{A.42c}$$

$$0 < \gamma_1^L < \bar{\gamma}_1^L \equiv \frac{\mu((\gamma_2^L)^2\tau_\eta(1 - \mu - \gamma\tau_v) - \mu\gamma\gamma_2^L(\tau_{u_1} + \tau_\eta) - \gamma^2(\tau_v + \mu^2(\tau_{u_1} + \tau_\eta)))}{\gamma\tau_v + \mu(\mu\gamma + \gamma_2^L)(\tau_{u_1} + \tau_\eta)}. \tag{A.42d}$$

Therefore, when (A.42a)-(A.42d) hold, two additional equilibria exist $(\Lambda_1^*)^I \in (0, \bar{\Lambda}_1^*)$, and $(\Lambda_1^*)^L \in (\bar{\Lambda}_1^*, 1/\gamma\tau_v)$. This establishes that in the case $\tau_{u_2} \to \infty$, when (A.42a)-(A.42d) hold, three equilibria: $0 < (\Lambda_1^*)^I < \bar{\Lambda}_1^* < (\Lambda_1^*)^L < 1/\gamma\tau_v$, arise.

Consider now the general quintic (A.38). First, note that

$$f(0) = -\frac{\tau_\eta(\tau_\eta + \tau_{u_1})^2}{\mu(\gamma\tau_v)^2} < 0. \tag{A.43}$$

Next, evaluating $f(\cdot)$ at $\underline{\Lambda}_1^* = \mu/(1-\mu) < \bar{\Lambda}_1^*$ yields

$$f(\underline{\Lambda}_1^*) = \frac{\tau_\eta}{\mu\gamma^2}\Bigg\{\frac{(\tau_\eta+\tau_{u_1})^2(\mu\gamma\tau_v-(1-\mu))}{(1-\mu)\tau_v^2} + \frac{\tau_{u_2}}{1-\mu}\Bigg((\gamma_2^L)^2\tau_\eta(\mu\gamma\tau_v-(1-\mu)) + \tag{A.44}$$
$$+ (\mu\gamma+\gamma_1^L)(\gamma_2^L(\tau_{u_1}+\tau_\eta)+\mu\gamma(\tau_{u_1}+\tau_\eta+\tau_v))(\gamma_2^L(\tau_{u_1}+\mu\tau_\eta)+\mu\gamma(\tau_{u_1}+\tau_\eta+\tau_v))\Bigg)\Bigg\}.$$

The sign of (A.44) is determined by the sign of the expression inside the curly brackets. As $\underline{\Lambda}_1^* < 1/\gamma\tau_v$, the term $\mu\gamma\tau_v-(1-\mu) < 0$. Also, by inspection, the expression within parentheses is positive provided that

$$\gamma > \underline{\gamma} \equiv \frac{(1-\mu)\tau_\eta - \gamma_1^L(\tau_{u_1}+\tau_\eta)(\tau_{u_1}+\mu\tau_\eta)}{\mu(\tau_v\tau_\eta+(\tau_{u_1}+\tau_\eta)(\tau_{u_1}+\mu\tau_\eta))}. \tag{A.45}$$

Hence, if (A.45) holds, and

$$\begin{aligned}\tau_{u_2} > \underline{\tau}_{u_2} \equiv \big(&(\tau_v^2((\gamma_2^L)^2(\tau_\eta(\mu\gamma\tau_v-(1-\mu)) + (\mu\gamma+\gamma_1^L)(\tau_1+\tau_\eta)(\tau_{u_1}+\mu\tau_\eta)) + \mu\gamma\gamma_2^L\times\\ &(\mu\gamma+\gamma_1^L)(\tau_1+\tau_v+\tau_\eta)(2\tau_{u_1}+\tau_\eta(1+\mu)) + (\mu\gamma+\gamma_1^L)(\mu\gamma)^2(\tau_{u_1}+\tau_\eta+\tau_v)^2))\big)^{-1}\times\\ &\big((\tau_{u_1}+\tau_\eta)^2(1-\mu-\mu\gamma\tau_v)\big),\end{aligned} \tag{A.46}$$

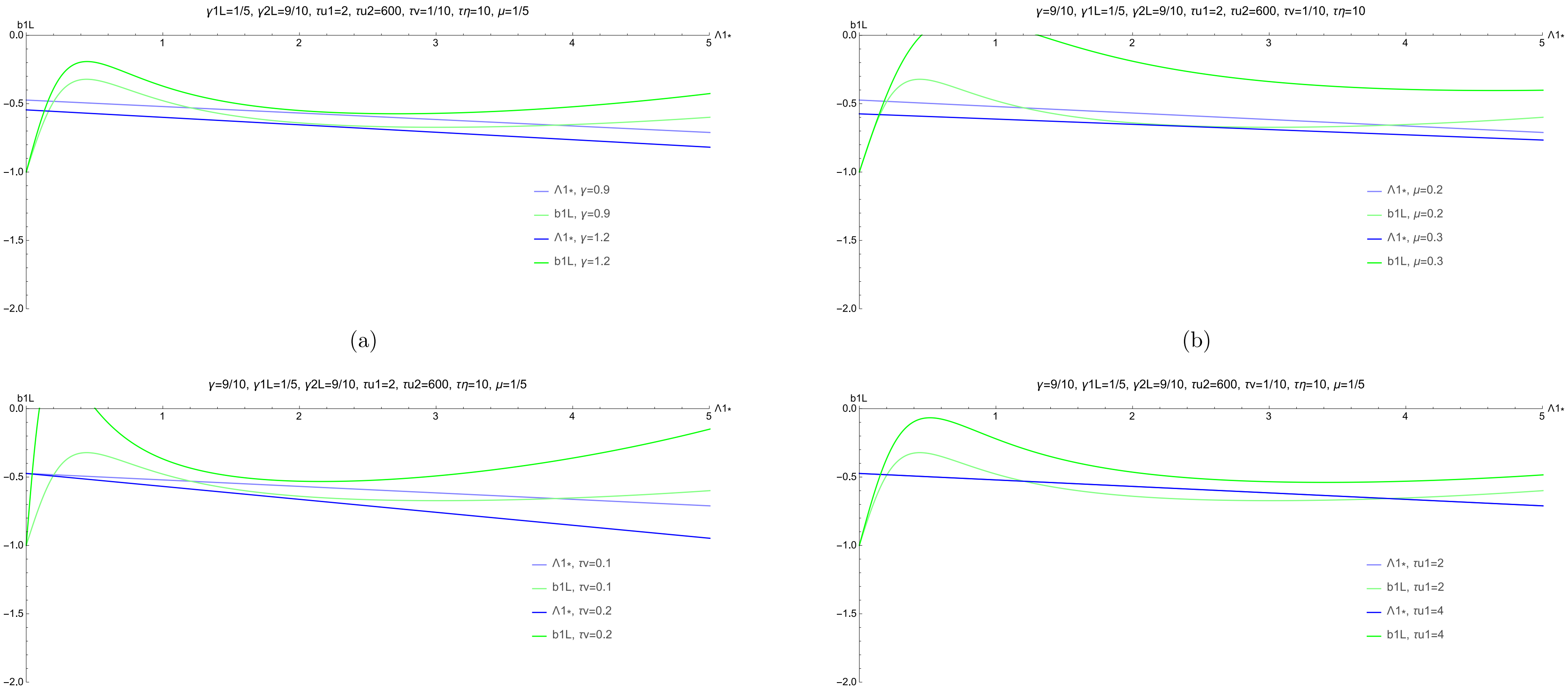
expression (A.44) is positive. This establishes the existence of an equilibrium $0 < (\Lambda_1^*)^H < \underline{\Lambda}_1^*$. Finally, provided (A.42a)-(A.42d) hold, $f(\bar{\Lambda}_1^*) < 0$, since $\hat{f}(\bar{\Lambda}_1^*) < 0$ and

$$\frac{\gamma\tau_v}{1-\mu} < 1.$$

Therefore, we can conclude that when $\tau_{u_2} < \infty$, if (A.42a)-(A.42d), and (A.45), (A.46) hold, the model displays three equilibria:

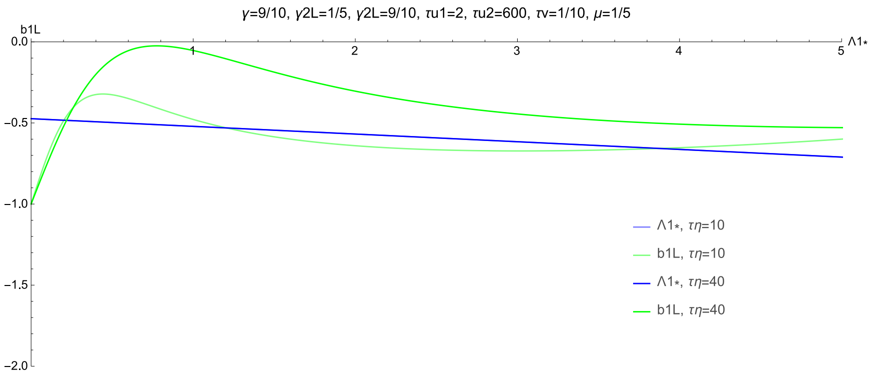
$$0 < (\Lambda_1^*)^H < \underline{\Lambda}_1^* < (\Lambda_1^*)^I < \bar{\Lambda}_1^* < (\Lambda_1^*)^L < \frac{1}{\gamma\tau_v}. \tag{A.47}$$

□

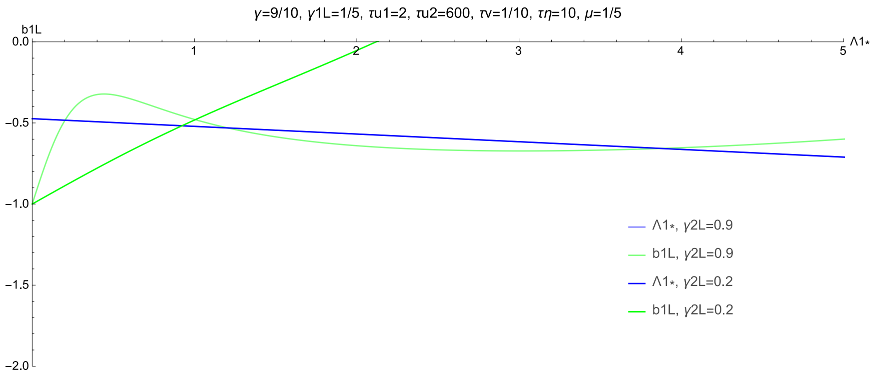


(a) (b) (c) (d)

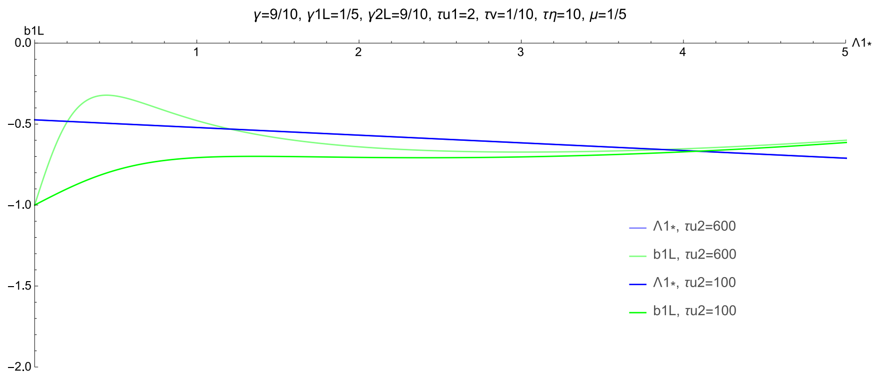
Figure 10: Uniqueness at an equilibrium with high liquidity. A higher dealers' risk tolerance (Panel (a)), proportion of FDs (Panel (b)), precision of the payoff distribution (Panel (c)), or precision of the first period endowment shock (Panel (d)) can lead to a unique equilibrium with high liquidity. Other parameters' values are as in Figure 5.



(a)



(b)



(c)

Figure 11: Uniqueness at an equilibrium with high and low liquidity. A higher signal precision (Panel (a)), or a lower second period traders' risk tolerance (Panel (b)) can lead to a unique equilibrium with high liquidity. A high volatility of the second period endowment can lead to a unique equilibrium with low liquidity (Panel (c)). Other parameters' values are as in Figure 5.