Why is capital slow moving? Liquidity hysteresis and the dynamics of limited arbitrage

James Dow[*], Jungsuk Han[†] and Francesco Sangiorgi[‡§]

May 23, 2018

Abstract

Will arbitrage capital flow into a market experiencing a liquidity shock, mitigating the adverse effect of the shock on liquidity? Using a stochastic dynamic model of equilibrium pricing with privately informed capital-constrained arbitrageurs, we show that arbitrage capital may actually flow out of the illiquid market. When some arbitrage capital flows out, the remaining capital in the market becomes trapped because it becomes too illiquid for arbitrageurs to want to close out their positions. This mechanism creates endogenous liquidity regimes under which temporary shocks can trigger flight-to-liquidity resulting in "liquidity hysteresis" which is a persistent shift in market liquidity.

[*]London Business School, Regent's Park, London, United Kingdom, NW14SA, E-mail: jdow@london.edu.

[†]Stockholm School of Economics, Drottninggatan 98, Stockholm, Sweden, SE 11160, E-mail: jungsuk.han@hhs.se

[‡]Frankfurt School of Finance and Management, Adickesallee 32-34, 60322, Frankfurt am Main, Germany, E-mail: F.Sangiorgi@fs.de

[§]We thank Laurent Bach, Jonathan Berk, Denis Gromb, Christopher Hennessy and Anna Pavlova for helpful comments, and also thank seminar participants at Aalto Business School, Collegio Carlo Alberto, Frankfurt School of Finance and Management, INSEAD, London Business School, Seoul National University, and Stockholm School of Economics.

1 Introduction

Traditional finance theory derives rational prices for assets based on the mechanism of arbitrage . Arbitrage (including trading on private information) pushes market prices towards fundamental value. This pricing mechanism may break down when arbitrageurs are capital-constrained for various reasons. An extensive literature studies the limits to arbitrage and hence, how prices may diverge from fundamental value due to constrained arbitrage capital (e.g., Allen and Gale (1994), Shleifer and Vishny (1997), Gromb and Vayanos (2002), and Brunnermeier and Pedersen (2009)). In principle, any surplus capital should flow to exploit those arbitrage opportunities, but it sometimes happens too slowly (e.g., Duffie (2010)). In other words, mispricing may still persist even with plenty of capital around because capital does not flow to the right markets. But what stops capital from flowing to correct mispricing? It seems paradoxical that capital does not quickly flow to correct prices. So endogenizing the rate of flow of arbitrage capital is a priority for research. This paper seeks to address this need.

First, we distinguish between informed and uninformed capital (much of the existing literature studies models with full information). We assume that uninformed capital is plentiful but informed capital is limited. Even so, isn't there enough informed capital in the economy to flow to where it is needed?

We find that the resolution of this paradox lies in the difference between a stock and a flow of arbitrage capital; a stock of arbitrage capital does not necessarily translate into a capital flow to arbitrage opportunities. For a capital-constrained trader to invest in a new position, he or she must close out some existing positions. Unless the new position is more profitable, the trader would stick to the existing positions until those asset prices revert closer to fundamental value because of public information or because of trades by subsequent privately informed traders. This means that arbitrage capital plays a dual role; the wedge of mispricing not only decides the profitability of new investment but also decides the speed at which engaged arbitrage capital is released (thus deciding the availability of arbitrage capital).

The dual role of arbitrage capital has several important implications. First, markets may be inefficient because arbitrage capital is "trapped" and efficiency may change over time as trapped capital is released. Second, there can be "liquidity hysteresis" in the form of a long-lasting shift in efficiency as a response to temporary changes in market liquidity. Arbitrage capital indeed does not immediately flow to seemingly profitable arbitrage opportunities, but only slowly with a delay. Third, flight-to-liquidity arises when a market suffers liquidity hysteresis. Arbitrage capital will flow to more liquid investment opportunities as illiquidity reinforces itself.

To formalize these ideas, we study a dynamic model of arbitrage with two markets where arbitrageurs freely move between the two, but are capital-constrained. One market is populated with short maturity assets (henceforth the "liquid market"), and the other market is populated with long maturity assets (henceforth, the "illiquid market").[1] Arbitrageurs collect private information on assets, and then trade those assets for speculative gains. In equilibrium, the two markets should offer the same expected speculative profits – otherwise arbitrageurs will move across to the one with higher profits. This means that the illiquid market should have a higher mispricing wedge than that of the liquid market (to compensate for the opportunity cost of longer maturity of investment); price efficiency of the illiquid market should be lower than that of the liquid market. Lower price efficiency in the illiquid market in turn implies that more capital is trapped because it becomes too illiquid for arbitrageurs to want to close out their positions.

The overall efficiency of the markets is determined by how much arbitrage capital is active as opposed to trapped. In other words, efficiency depends on the pool of active capital as a state variable. This matters because while the total stock of arbitrage capital may be large, the stock of active capital may be much smaller. The efficiency of a market may change over time as trapped capital is released from other markets. Furthermore, there is a delayed response in efficiency to changes such as shocks to liquidity trading.

The active (as opposed to trapped) capital, being the state variable of the economy, creates a feedback channel between liquidity and active capital. As more active arbitrage capital flows to the illiquid market, those who are trapped in the market become active again more quickly, and this in turn creates a larger capital flow to the illiquid market by increasing the overall size of active capital in the economy. This virtuous cycle leads to a high information regime where arbitrage capital is redeployed at a faster rate (thus giving rise to higher liquidity). On the other hand, a vicious cycle may arise, thereby leading to a low information regime in which arbitrage capital flows to the liquid market leaving locked-in investment in the illiquid market being trapped for a long time.

The feedback channel between active capital and liquidity leads to multiple regimes in our model; there is a threshold of active capital that separates domains of attraction for liquidity. We illustrate our model's implications for liquidity and capital flow dynamics by studying shock responses to a Markov stationary system where (either good or bad) liquidity shocks randomly hit the illiquid market. With a small adverse shock to the illiquid market, market liquidity recovers on its own thanks to a virtuous cycle of liquidity. As more trapped arbitrageurs become active again, they quickly replenish market liquidity. On the other hand, a large adverse shock can trigger a vicious cycle of illiquidity with flight-to-liquidity where arbitrage capital flows to the liquid market; more and more arbitrageurs choose to invest in the liquid market over time

[1] "Liquidity" is a rather broad concept in financial economics, covering a variety of different effects. In this paper we use "liquid" to describe a market where investments can easily be cashed in. An arbitrageur in a liquid market does not have to wait long before the arbitrageur's private information is incorporated in the price, hence can cash in quite soon at full value. In an illiquid market the arbitrageur could either cash in soon at a poor price, or wait longer to obtain full value.

because they expect further deterioration of low future liquidity in the illiquid market. This leads to an illiquidity regime where there is a persistent overall lack of liquidity in the market. We call this "liquidity hysteresis" because a shock to the system moves the equilibrium to a different path even after the shock is removed.

We illustrate how the market can move in and out of this illiquidity regime with numerical simulations of the stochastic equilibrium of the model. A sequence of temporary bad shocks to the illiquid market can trigger a flight-to-liquidity resulting in the illiquidity regime, from which the market can recover only after a sequence of shocks in the opposite direction. Thus, the market features persistent (endogenous) liquidity regimes even when (exogenous) liquidity trading is at its normal level most of the time. These results provide a theoretical explanation of slow-moving capital regarding why capital moves slowly, how fast (or slowly) it moves, and to which directions it moves. They further provide interesting policy implications.

The paper is organized as follows. In Section 2, we discuss related literature. In Section 3, we describe the basic model. In Section 4, we solve for the equilibrium of the model. In Section 5, we study the model's implications for liquidity and capital flow dynamics. In Section 6, we discuss empirical and regulatory implications of our model. In Section 7, we conclude.

2 Literature Review

There is an extensive literature on asset pricing with informed trading, and there is also an extensive literature on asset pricing with limits to arbitrage. However, with few exceptions, the literature on noisy rational expectation equilibrium (REE) does not study limits to arbitrage, while the literature on limits to arbitrage typically uses a full information setting. In this paper, we combine limits to arbitrage with a noisy REE model of asset prices.

Our paper contributes to the literature of noisy REE in two ways. That literature is based around a linear framework (Grossman and Stiglitz (1980), Hellwig (1980), Diamond and Verrecchia (1981), Kyle (1985), Admati (1985), Wang (1993), Wang (1994), Vives (1995)), which stems from a CARA-normal setting. Although the setup delivers mathematical tractability, further extension of the standard framework is often limited because the linear structure is difficult to maintain. Thus, it can be hard to use this framework to study a variety of interesting applications. For example, Dow and Rahi (2003) study investment response of firms to stock prices (the feedback effect) by making special assumptions on hedgers' endowment risk and on the firm's production function (linking investment to output) to maintain linearity. Some papers on the other hand are able to use a noisy REE framework without linearity, but this imposes a cost in reduced tractability (e.g., Fulghieri and Lukin (2001)). Imposing constraints on how much arbitrageurs can trade makes their demands inherently non-linear.

As well as limiting applicability, the linear framework of standard noisy REE is too simple to deliver rich dynamics. Consequently, informed trading under multiple asset (or market) environments mostly concerns correlations across fundamentals without creating dynamics due to flow of capital from one place to another. In contrast, our model focuses on nonlinear dynamic shock responses by deviating from the standard setup featuring constrained informed traders. In addition, the non-linearity in our paper can be further amplified by history-dependent regime shifts with intertemporal dependence of available capital. This allows us to study slow moving capital and flight to liquidity using the dynamics of liquidity responses to temporary liquidity shocks. Finally, we make a technical contribution to the literature by characterizing stationary rational expectations equilibrium under capital constraint.

In contrast, the limits to arbitrage literature uses constraints but with full (symmetric) information. The common framework among models in this literature is that if assets are mispriced, only a limited number of agents, with limits on how much they can trade, have the ability to arbitrage away the mispricing. Hence, the mispricing will not be fully eliminated. Traditional literature in limits to arbitrage shows that the level of arbitrage capital is the state variable that determines market liquidity: cash-in-the-market-pricing (e.g., Allen and Gale (1994)), investors' fund flow (e.g., Shleifer and Vishny (1997)), leverage constraints (e.g., Gromb and Vayanos (2002)), margin constraints (e.g., Brunnermeier and Pedersen (2009)).

A related literature explains fluctuations in market liquidity using the idea of slow-moving capital. Those papers extend the limits-to-arbitrage argument by suggesting that slow-moving capital could be the source of prolonged illiquidity even when there is enough capital in the economy. For example, Mitchell, Pedersen, and Pulvino (2007) show convertible bonds traded at prices well below the arbitrage price (relative to the stock and a straight bond) during an extended period when the convertible bond hedge funds (that normally arbitrage these assets) were short of capital, and multi-strategy hedge funds (that opportunistically redeploy capital to wherever returns are high) were slow to enter the market. Duffie (2010) suggests that institutional impediments such as search frictions, taxes, regulations, and market segmentation can slow down the flow of capital.

Our paper contributes to the literature on limits to arbitrage and slow-moving capital in two ways. First, we use a private information setting. This is a central paradigm in finance, so it seems important to integrate it into models of limited arbitrage. A key difference is that all agents observe prices and try to infer what the arbitrageurs are doing, while in models without private information the non-arbitrageurs's demands are exogenous. Some papers in the limits to arbitrage and slow-moving capital literature appeal to private information as a reason why arbitrageurs cannot raise more capital. But they do not implement the key feature of the noisy REE models of private information, that all agents are able to learn from prices. Second, we try to model the frictions that limit the movement of arbitrage capital and to derive an economic model that explains the speed of movement of arbitrage capital.

Some earlier papers take a related approach, but our paper differs in that we study capital movement across multiple markets in a dynamic setup. Dow and Gorton (1994) study a multiperiod model of limits to arbitrage where the cost of carry combined with arbitrageurs' short horizon break down the chain of arbitrage in a single asset market. Dow and Han (2018) study a static noisy rational expectations model with endogenous adverse selection in asset supply where the presence of informed capital facilitates movement of uninformed capital.

Among previous literature, a number of papers focus specifically on limited arbitrage in a dynamic setting and with multiple markets. In such models, (marginal) expected returns of new investment are equalized across markets in equilibrium. Consequently, the mispricing wedge for long duration assets is higher than for short duration assets, as discussed in Shleifer and Vishny (1990). Furthermore, a shock in one market tends to create a spillover effect where shocks are transferred to other markets through the channel of wealth effects (e.g., Kyle and Xiong (2001)) or collateral constraints (e.g., Gromb and Vayanos (2017) and Gromb and Vayanos (Forthcoming)). The literature also studies the adjustment process of capital after the arrival of a shock. Duffie, Garleanu, and Pedersen (2005) and Duffie, Garleanu, and Pedersen (2007) find a gradual process of recovery after a shock to investors' preference in search-based models. In Duffie and Strulovici (2012) where financial intermediaries trade off the cost against the benefit of intermediation, the speed of capital flow is governed by the imbalance of capital as well as the level of intermediation competition across markets. In Gromb and Vayanos (Forthcoming), there is a phase with an immediate increase in the spread where arbitrageurs decrease their positions (thus, causing a contagion effect), followed by a recovery phase.

Our paper has several important differences from this line of literature. As previously discussed, we use a private information setting where all investors learn from prices. The results of the model are also different. Because informed capital is often already engaged in investment, the level of active informed capital becomes the state variable which governs market liquidity. We find that, as a result of intertemporal informational externality, a temporary shock can generate hysteresis: there is a long-lasting (or even permanent) impact on market liquidity if active informed capital is impaired to the point at which it cannot recover on its own.

3 Setup

3.1 Financial Assets

We consider an infinite horizon discrete time economy with a continuum of long-lived agents. All agents have risk neutral preferences with a discount factor of β. There exists a risk-free asset in the economy whose return is equal to $r_f = 1/\beta - 1$.

There is a continuum of financial securities, each of which is a claim to a single random liquidation value. There are two classes of securities that differ in their maturity: (i) "liquid assets" which are short-lived, and (ii) "illiquid assets" which are long-lived. At this point in the paper, calling the assets "liquid" and "illiquid" is just convenient terminology, since liquidity is an equilibrium property of an asset and we have not yet characterized the equilibrium. We will show later that this terminology is justified.

Illiquid assets are traded in market I, and liquid assets are traded in market L. It is important to note that they are not segmented markets because capital can freely move between two markets without any friction. An asset in market L has a one period maturity; it pays its liquidation value in the period after issuance.[2] On the other hand, an asset in market I has a random maturity; if it has not liquidated in a previous period, it pays its liquidation value with probability $q > 0$ in each period. At maturity, any asset i in market $h \in \{I, L\}$ pays v_i which is either good ($v_i = V_h^G$) or bad ($v_i = V_h^B$) with equal probability where $V_h^G > V_h^B$ for all $h \in \{I, L\}$.[3] We further assume that the present value of assets in both markets are identical.[4] Asset payoffs are independent across assets and time.

As discussed later, asset prices either reveal their fundamental value completely, or not at all. The assets are called "fully-revealed" if prices reflect true fundamental value, and "unrevealed" if not. Payoffs can become known if the liquidation value is fully revealed by the trading process and asset prices. For simplicity, we assume that the mass of unrevealed assets is fixed to one unit in each market at any point of time. That is, new assets are issued to replace those which either realized payoffs, or become fully-revealed.[5]

3.2 Participants

There is a unit mass of capital-constrained "arbitrageurs" who trade to generate speculative profits. We denote $\mathcal{A}$ to be the set of arbitrageurs, and index each arbitrageur by $a \in \mathcal{A}$. Each arbitrageur can produce private information about the payoff of one asset in each period. All arbitrageurs who investigate an asset can perfectly predict its liquidation value. For tractability, we assume a simple form of capital constraint under which at any point in time, each arbitrageur can hold only one risky position of at most one unit (long or short) of unrevealed

[2]For simplicity, we assume that one of the class of assets pays every period (liquid assets). It would be possible, but considerably more complex, to analyze the case with a payoff probability less than one for all assets.

[3]The assumption that payoffs are high or low with equal probabilities simplifies the analysis by making profits from long and short positions symmetric.

[4]This is simply for mathematical convenience. Under this assumption, we have

$$\frac{\beta q}{1 - \beta(1 - q)} V_I^s = \beta V_L^s, \quad \text{for all } s \in \{G, B\}$$

where the left- and right- hand side is the present value of payoff of an asset with quality s in market I and L, respectively.

[5]One can consider that there exists a unit mass of firms which issue new securities to invest in new projects whenever their existing projects pay liquidation value or become fully-revealed. If we assume, instead, that unrevealed assets that become fully-revealed are not immediately replaced, the model would require an additional state variable and would be considerably more complex to analyze.

assets.[6] We denote $x_i^a \in \{-1, 1\}$ to be the market order of arbitrageur a for asset i.

There is a continuum of competitive risk-neutral market makers who set prices to clear the market. There are also noise traders who trade for exogenous reasons such as liquidity needs. In each period, arbitrageurs and noise traders submit market orders to the market makers. Noise traders submit an aggregate order flow of ζ_i for each asset i which follows an independent uniform distribution on $[-z_h, z_h]$. The magnitude of z_h captures the intensity of noise trading in market h in the current period. We assume that z_L is a constant whereas z_I follows a Markov process with N states $z_I^1, z_I^2, ..., z_I^N$ whose transition matrix between states is given by

$$\Omega = \begin{bmatrix} \omega_{11} & \dots & \omega_{1N} \\ \vdots & \ddots & \vdots \\ \omega_{N1} & \dots & \omega_{NN} \end{bmatrix}$$

Note that z_I is the only exogenous shock to the economy in our model at the aggregate level, and its realization is publicly observable to all the agents in the economy. We further assume that there are enough noise trading activities in the market to prevent the price for every asset from being fully-revealing; the support of aggregate noise trading is strictly greater than that of arbitrageurs' aggregate order flow: $z_I^n + z_L \geq 1$ for any n. Finally, we assume that all the realizations of noise trading intensity and asset payoff are jointly independent.

3.3 Timing of Events

The timing of events in each period is as follows. At the beginning of the period, asset payoffs realize and they are distributed among claim holders. Next, new assets are issued, and noise trading intensity z_I realizes. After these events, arbitrageurs collect private information on unrevealed assets, then submit orders to market makers together with noise traders. At the end of the period, market makers post asset prices and trades are finalized.

4 Equilibrium

4.1 Active Arbitrage Capital

In each period, an arbitrageur is in either of two situations: "active" or "locked-in". An active arbitrageur does not have an existing position in unrevealed assets, thus has capital available for new investment whereas a locked-in arbitrageur already has a position in unrevealed assets, thus, does not have capital available for new investment until this position is liquidated.

[6] Once they have acquired a position in an asset, they hold it until it liquidates or its value is revealed, and can then open a new position. They also have the option to decide to close out a position early (before learning it has realized profits), and opening another position next period.

We denote ξ to be the the mass of active arbitrageurs, and π to be the mass of locked-in arbitrageurs; thus $\xi + \pi = 1$.

Each active arbitrageur chooses to hold a new position in either market I or L. δ denotes the portion of those choosing to trade assets in market I (so $1 - \delta$ is the portion of those choosing to trade assets in market L). Each locked-in arbitrageur chooses whether to hold on to the position one more period or to close it out in the current period. η denotes the portion of those choosing to close out early (so $1 - \eta$ is the portion of those choosing to hold on to the position). Note that η is included for completeness, but we show that in equilibrium in our model early liquidation is never optimal so that η is zero.

Throughout the paper we use the dot notation to denote the value of any variable in the subsequent period. For example, $\dot{\xi}$ and $\dot{\pi}$ denote the value of ξ and π in the subsequent period, respectively. We define the vector of state variables to be $\theta \equiv (\xi, z_I)$, which is a pair of the current level of active capital and the realization of noise trading intensity.

4.2 Asset Prices

Asset prices are set by the market makers given the aggregate order flows from informed arbitrageurs and noise traders as in the standard Kyle (1985) model. Because market makers are risk neutral, they set the price equal to the expected discounted liquidation value conditional on the aggregate order flow:

$$P_i = \mathrm{E}\left[\beta^{\tau_i} v_i \middle| \theta, X_i\right], \tag{1}$$

where τ_i is the (random) maturity of asset i, and $X_i = \int_{a \in \mathcal{A}} x_i^a da + \zeta_i$ is the aggregate order flow for asset i. Knowledge of θ allows market makers to make inference about informed trading activity from the order flow.

Prices are either fully-revealing or non-revealing due to the uniformly-distributed noise trading.[7] If the order flow is large (in absolute value, buy or sell) then it can only result from both informed arbitrageurs and noise traders trading in the same direction, so it is fully revealing. But if the order flow is smaller than this in absolute value, then it could have resulted from either informed traders buying and uninformed arbitrageurs selling, or vice versa. Because arbitrageurs are equally likely to buy and sell, noise trading is uniformly distributed, and the value of the asset is equally likely to be high or low, these two possibilities are equally likely and therefore the trading volume is uninformative. We denote P^G and P^B to be the fully-revealing price for good and bad quality, respectively. We also denote P^0 to be the non-revealing price.[8] We denote the probability of information revelation for asset i to be λ_i. As it is shown later in the paper, λ_i plays a dual role of capturing both "price efficiency" (which is inversely related to the degree of mispricing) and "liquidity" (which is inversely related to the expectation of

[7]For technical details, see the proof of Lemma 1 in Appendix A.

[8]Because the present values are identical across market I and L, prices given the same type of public

investment duration) of asset i. For notational convenience, we call λ_i as "price efficiency" of asset i henceforth.

4.3 Laws of Motion

We focus on market-wise symmetric rational expectations equilibria under which price efficiency is symmetric across all the assets in each market, i.e., the measure of price efficiency λ_i is equal to λ_I for any asset i in market I, and it is equal to λ_L for any asset i in market I.

The laws of motion of the mass of each group of arbitrageurs are given by

$$\begin{aligned} \dot{\xi} &= (1-\delta)\xi + (\delta\xi + \pi)(q + (1-q)\lambda_I) + \pi\eta(1-q)(1-\lambda_I); \quad (2) \\ \dot{\pi} &= (\delta\xi + \pi\,(1-\eta))(1-q)(1-\lambda_I). \quad (3) \end{aligned}$$

The first equation describes the evolution of active capital ξ. The right hand side of Eq. (2) is the sum of three terms. The first term is the mass of arbitrageurs invested in market L in the current period; this mass becomes entirely active in the subsequent period as the L assets are short-lived. The second and third terms are the mass of arbitrageurs invested in market I in the current period (i.e., $\delta\xi$ new arbitrageurs from the current period and π arbitrageurs locked-in from the previous period) that become available for new investment in the subsequent period. This happens either if the asset pays off or if the market price fully reveals the asset value (in which case the position becomes risk-free, thus relaxing the portfolio constraint), or if locked-in arbitrageurs close out the position early (it turns out they choose not to do this in equilibrium). Overall, a fraction $q + (1-q)\lambda_I$ of the arbitrageurs invested in the I market in the current period becomes free for new investment in the subsequent period because of asset paying off or information revelation through prices. A fraction $\eta(1-q)(1-\lambda_I)$ of locked-in arbitrageurs from the previous period becomes active next period because of the decision to close out early. Note that Eq. (3) is redundant given that $\xi + \pi = 1$.

4.4 Dynamic Arbitrage

Given the current state θ, we denote $J_I(\theta)$ and $J_L(\theta)$ to be the value of investing in a new position in market I and market L, respectively. Because any active arbitrageur can choose

information are also identical across the two markets:

$$P^s \equiv \frac{\beta q}{1-\beta(1-q)} V_I^s = \beta V_L^s, \quad \text{for all } s \in \{G, B\};$$
$$P^0 \equiv \frac{\beta q}{1-\beta(1-q)} \left(\frac{V_I^G + V_I^B}{2}\right) = \beta \left(\frac{V_L^G + V_L^B}{2}\right).$$

between the two markets, the value of being active given θ equals

$$J_f(\theta) = \max\left(J_I(\theta), J_L(\theta)\right). \tag{4}$$

Associated with these value functions is a capital allocation function $\delta(\theta)$ for active arbitrageurs such that

$$\delta(\theta) \in \begin{cases} \{0\}, & \text{if } J_I(\theta) < J_L(\theta); \\ \{1\}, & \text{if } J_I(\theta) > J_L(\theta); \\ [0,1], & \text{otherwise,} \end{cases} \tag{5}$$

where capital allocation $\delta(\theta)$ strikes the balance between the value of investing in market I and L if $J_I(\theta) = J_L(\theta)$.

In case an arbitrageur chooses market I, he becomes locked in until it liquidates or its value is revealed (we call it locked in because although he has the option to close out early, arbitrageurs choose not to in equilibrium). We denote $J_l(\theta)$ to be the associated value function given θ. Using the symmetry of trading profits between long and short positions, we can obtain $J_I(\theta)$ and $J_L(\theta)$, whose detailed derivations are relegated to Appendix A, as follows:

$$J_I(\theta) = -(\lambda_I P^G + (1-\lambda_I)P^0) + \beta U(\theta), \tag{6}$$

$$J_L(\theta) = -(\lambda_L P^G + (1-\lambda_L)P^0) + \beta\Big[V_L^G + \mathrm{E}[J_f(\dot{\theta})|\theta]\Big]. \tag{7}$$

where

$$\begin{aligned} U(\theta) \equiv & q V_I^G + (1-q)\lambda_I P^G + (1-(1-\lambda_I)(1-q))\mathrm{E}[J_f(\dot{\theta})|\theta] \\ & + (1-\lambda_I)(1-q)\mathrm{E}[J_l(\dot{\theta})|\theta]. \end{aligned}$$

Because any locked-in arbitrageur can choose between exiting the position or staying with it, the value function of a locked-in arbitrageur given θ equals

$$J_l(\theta) = \max\left(J_E(\theta), J_S(\theta)\right), \tag{8}$$

where $J_E(\theta)$ is the value of exiting the position and becoming active in the next period, and $J_S(\theta)$ is the value of holding the position one more period:

$$J_E(\theta) = \lambda_I P^G + (1-\lambda_I)P^0 + \beta \mathrm{E}[J_f(\dot{\theta})|\theta], \tag{9}$$

$$J_S(\theta) = \beta U(\theta). \tag{10}$$

Similarly as in $\delta(\theta)$, associated with $J_l(\theta)$ is an exit function $\eta(\theta)$ for locked-in arbitrageurs such that

$$\eta(\theta) \in \begin{cases} \{0\}, & \text{if } J_E(\theta) < J_S(\theta); \\ \{1\}, & \text{if } J_E(\theta) > J_S(\theta); \\ [0,1], & \text{otherwise.} \end{cases} \tag{11}$$

4.5 Stationary Equilibrium

We define equilibrium in a standard manner for stationary equilibrium with stochastic shocks:[9]

Definition 1 *A stationary equilibrium is a collection of value functions J_f, J_l, J_I, J_L, J_E, J_S, capital allocation function δ, exit function η, price efficiency measures λ_I, λ_L, law of motion for the mass of active arbitrageurs ξ such that*

1. *J_f, J_l, J_I, J_L, J_E, J_S, δ, η satisfy Eqs. (4)-(11).*
2. *λ_I and λ_L correspond to the probability that prices, which are determined by Eq. (1), reveal true asset values in market I and L, respectively.*
3. *The law of motion for ξ satisfies Eq. (2).*

An equilibrium is said to be interior if $J_I(\theta) = J_L(\theta)$, so that active arbitrageurs are indifferent between investing in market I and L. In an interior equilibrium, Eqs. (4),(6),(9) and (10) imply that early liquidation is strictly dominated by holding the position, i.e., $J_E(\theta) < J_S(\theta)$ (hence $J_l(\theta) = J_S(\theta)$), of which the proof is relegated to Appendix A. With early liquidation, the position is closed out and then a new position is opened. But the expected proceeds are offset by the expected cost of acquiring the new position afterwards. Early liquidation just leads to holding another locked-in position or liquid position, which in the case of an interior equilibrium, must have the same value. Consequently, locked-in arbitrageurs stay inactive until either the price fully reveals the asset value or the asset pays off (i.e., η is equal to zero in an interior equilibrium).[10]

We can now characterize price efficiency in financial markets as follows:

Lemma 1 *In an interior equilibrium, the probability of information revelation in market I equals*

$$\lambda_I = \frac{\delta\xi}{z_I}, \tag{12}$$

[9] As is standard, the equilibrium is called "stationary" because the value functions and the capital flow function are time invariant; however in general the endogenous variables will change over time. Note that "stationary" is not the same as the "steady states" which we describe in Section 5.2. While an equilibrium of our model describes the evolution of the entire system over time as a function of the state variables, the system may over time end up with the endogenous variables converging close to certain values - these are called steady states.

[10] Even if arbitrageurs are allowed to open a new risky position simultaneously with closing another one, early liquidation does not dominate staying with the existing position. Furthermore, introducing an arbitrarily small transaction or information acquisition cost would make early liquidation suboptimal.

and the probability of information revelation in market L equals

$$\lambda_L = \frac{(1-\delta)\xi}{z_L}. \tag{13}$$

Proof. See Appendix. ■

As Lemma 1 shows, for a fixed capital allocation δ, equilibrium price efficiency in market $h \in \{I, L\}$ increases in the amount of informed capital ξ and decreases in the intensity of noise trading z_h. This property is intuitive because ξ and z_h have opposite effects on the informativeness of order flows in market h. Of course, capital allocation δ is a function of the state θ, so the overall impact of a shock to z_I on market price efficiency can only be determined in equilibrium, to which we turn next.

We can find conditions which are sufficient to ensure existence and uniqueness of equilibrium as well as monotonicity of equilibrium price efficiency in the amount of active capital as well as in the intensity of noise trading. The conditions provided in the appendix are rather lengthy but straightforward. The conditions provided are satisfied by a wide range of parameter values, including all of our numerical simulations.[11]

Proposition 1 *Under the conditions stated in Appendix B, there exists a unique stationary interior equilibrium in which price efficiency in the illiquid market λ_I is monotone increasing in active capital ξ. Furthermore, λ_I is monotone decreasing in noise trading intensity z_I.*

Proof. See Appendix. ■

Proposition 1 shows that price efficiency λ_I decreases in noise trading intensity z_I when capital allocation δ is determined endogenously. Such reduction in λ_I slows down the rate at which the current mass of locked-in capital is released, which reduces the mass of active capital in subsequent periods. This effect of a current shock to z_I further impairs future price efficiency and therefore increases the expected duration of an investment in market I. This is in contrast to the contemporaneous effect of a reduction in λ_I which makes market I more attractive to informed arbitrageurs who can benefit from price inefficiency.

Note that the dynamics of price efficiency in response to a stochastic shock to z_I is solely determined by the current level of state variables ξ and z_I due to its stationary nature. In the next section, we characterize liquidity hysteresis (or regime shifts) where the dynamics of price

[11]Note that in general it is difficult to show existence and uniqueness of stationary recursive competitive equilibrium, and much of the original macroeconomics literature does not do so (e.g. Bewley (1986), Huggett (1993) and Aiyagari (1994)). However, it is preferable to be able to prove existence and uniqueness in cases where simulation or analytical results are given. More recently, Acemoglu and Jensen (2015) claims that the Bewley-Huggett-Aiyagari models do have a unique equilibrium, while Acikgoz (2018) argues that for some specifications they do not.

efficiency and its long-run evolutionary path changes as a result of crossing certain endogenous thresholds of ξ.

5 Implications

5.1 Equilibrium Price Efficiency across Two Markets

We start with a preliminary result for the cross-sectional and dynamic properties of equilibrium price efficiency and liquidity.

Lemma 2 *In an interior equilibrium price efficiency satisfies*

$$\lambda_L - \lambda_I = \beta\,(1-q)\,(1-\lambda_I)\left(1 - \mathrm{E}[\dot{\lambda}_I|\theta]\right). \tag{14}$$

Proof. See Appendix. ∎

The left-hand side of Eq. (14) is the difference in probabilities of trading at fully-revealing price in market L over market I in the current period. That is, this difference reflects how more likely an arbitrageur is to make a speculative profit when trading in market I compared to trading in market L in the current period.

The right-hand side of Eq. (14) is the probability $(1-q)(1-\lambda_I)$ of remaining locked in a trade in market I, weighted by the discount factor β, and multiplied by expected *future* illiquidity in market I, captured by the term $1-\mathrm{E}[\dot{\lambda}_I|\theta]$. By trading in market I, a speculator gives up the certainty of being able to re-trade in the next period; for arbitrageurs to be indifferent between the two markets, assets in market I must compensate this opportunity cost with a higher probability of trading at non-revealing price in the current period.

5.2 Liquidity Hysteresis

Our model can display liquidity hysteresis, in other words a transitory shock can move the system to a different level of liquidity. To study this, we start by considering the special case of the model under the assumption that noise trading intensity is fixed at a constant level, i.e., z_I is a constant in every period. We can then define the steady state equilibria of the model. An equilibrium maps the current period's state variable ξ to next period's state variable $\dot{\xi}$, and a steady state is a fixed point of that mapping. Also (as we show below) the equilibrium law of motion has the property that the endogenous state variable tends to converge to the steady state value. The argument for this special case can be carried over to the general case with stochastic noise trading intensity, but in the case without shocks we can study the steady state analytically. In the next subsection we will add small shocks and show that, when the noise trading intensity is at the "normal" level (no shocks for a while) the state variable will converge close to a stable point.[12]

We denote ξ^* to be the steady-state-level mass of active arbitrageurs, and also denote λ_L^* and λ_I^* to be the steady-state-level price efficiency in market L and I, respectively. In steady state, the indifference condition in Eq. (14) can be expressed in terms of λ_L and λ_I as follows:

$$\lambda_L^* - \lambda_I^* = \beta(1-q)(1-\lambda_I^*)^2. \tag{15}$$

Eq. (15) reveals that price efficiency plays a dual role. On the one hand, price efficiency determines the profitability of investment opportunities: higher λ_I^* (and also λ_L^*) decreases the probability of acquiring a new position at non-revealing prices. We term this the "first lambda" effect of price efficiency on speculative profits. On the other hand, price efficiency determines the maturity of investment opportunities in long lived assets: higher λ_I^* increases the likelihood of closing out a position with profits earlier. We term this the "second lambda" effect of price efficiency on speculative profits.

Substituting Eqs. (12) and (13) into Eq. (15) yields the following steady state relationship between δ and ξ implied by arbitrageurs' indifference condition:

$$\frac{z_L-(1-\delta^*)\xi^*}{z_L} = \left(\frac{z_I-\delta^*\xi^*}{z_I}\right)\left[1-\beta(1-q)\left(\frac{z_I-\delta^*\xi^*}{z_I}\right)\right]. \tag{IC}$$

For a fixed δ^*, a decrease in active arbitrage capital ξ^* decreases price efficiency in both markets. This has a (positive) first lambda effect on speculative profits in both markets but a (negative) second lambda effect in market I, which becomes relatively less attractive. Hence, δ^* must decrease to restore arbitrageurs' indifference condition across markets.[13]

An interior steady state equilibrium is found at the intersection of the (IC) curve and the following capital movement (CM) curve obtained from the law of motion for active arbitrage capital in Eq. (2) together with Eq. (12) for λ_I^*:

$$\xi^* = (1-\delta^*)\xi^* + (\delta^*\xi^* + 1 - \xi^*)\left(q+(1-q)\frac{\delta^*\xi^*}{z_I}\right). \tag{CM}$$

Note that an increase in the fraction of active arbitrageurs that invest in market I has

[12] In the literatures on dynamic macroeconomics, and dynamic systems in science, the "steady state equilibrium" of the version of the model without shocks is also called "steady state", "deterministic steady state," "stable point," "asymptotically stable point", and simply "fixed point". In the version of the model with shocks it is called "stochastic steady state", "risky steady state", "stable point" "asymptotically stable point"and "fixed point". We will call them "steady state equilibrium" or simply "steady state" for the deterministic case and "stable point" in either case. We use "regime" to refer to the stable point the economy will converge to (in the absence of any further shocks).

[13] Lemma 4 in Appendix C provides the sufficient condition for the net benefit of trading in market I to decrease as δ increases, for a fixed value of ξ.

two opposing effects. On the one hand, as δ^* increases, more arbitrageurs remain trapped in market I. This tends to reduce steady state value for active capital ξ^*. On the other hand, an increase in δ^* improves price efficiency in market I, which increases the rate at which arbitrage capital is released form this market. This feedback effect tends to increase ξ^*. Which effect dominates depends on the model parameters. The first effect dominates in panel (c) of of Figure 1 for δ^* is small, while the second effect dominates for δ^* large. Intuitively, increasing the rate at which trapped capital is released has a bigger effect when the mass of arbitrageurs that are invested in market I is larger.

We can show existence of the steady state equilibrium:

Proposition 2 *(i) A steady state equilibrium exists, and there is either one or two stable (saddle point) equilibria. (ii) There exist constants* $0 < \underline{q} < \bar{q} < 1$ *and* $0 < \underline{\beta} < \bar{\beta} < 1$ *such that the steady state equilibrium is unique if* $q > \bar{q}$ *and/or if* $\beta < \underline{\beta}$ *and there are multiple steady state equilibria if* $q < \underline{q}$ *and* $\beta > \bar{\beta}$ *and* $1 > \frac{3}{4}z_L + z_I$.

Proof. See Appendix. ■

The sufficient conditions for multiple steady states in the proposition are intuitive. Because the feedback between price efficiency and active capital is across periods, its effect is stronger when investors care more about the future and when the duration of an investment in market I is mainly determined by future informed trading.

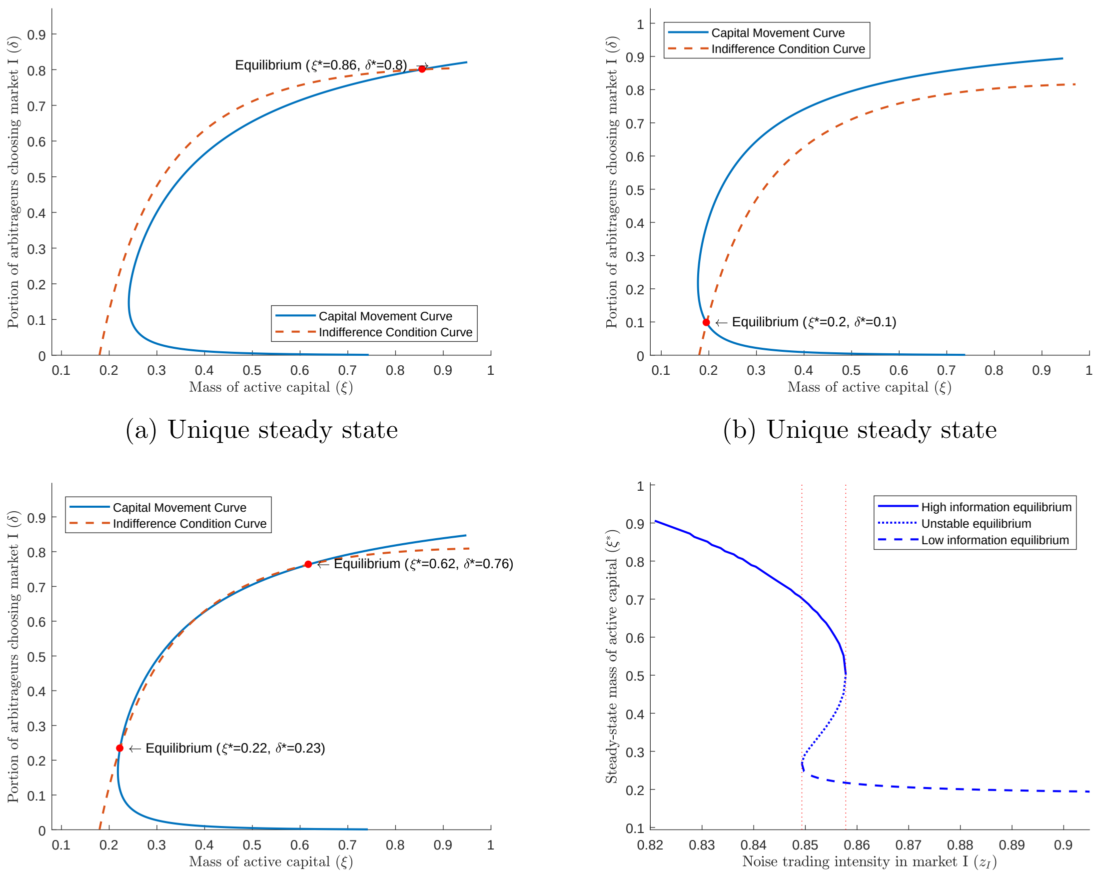
Figure 1 illustrates the steady state equilibrium values for ξ^* and δ^* determined by the intersection of the IC and CM curves. The steady state is unique in panels (a) and (b), whereas there are three steady states in panel (c), of which two are stable and one (for intermediate values of ξ^* and δ^*) is unstable. Panel (d) illustrates the region of noise trading intensity in the illiquid market where there is uniqueness or multiplicity.

5.3 Shock Response and Liquidity Regimes

Now, we return to analysis of the system in the general case of stochastic shocks, and consider the response to a stochastic shock to noise trading intensity in market I, whereby z_I deviates from its normal level to a higher value for a short period of times. In contrast to the analysis of the steady state equilibrium (in the model without shocks), arbitrageurs anticipate the possibility that a change in noise trading in market I might occur on the equilibrium path.

To shed light on the response to this temporary liquidity shock, we rearrange the indifference condition in Eq. (14) as follows:

$$1 - \lambda_L = (1 - \lambda_I)(1 - \beta(1 - q)(1 - \mathrm{E}[\dot{\lambda}_I])). \tag{16}$$



(a) Unique steady state

(b) Unique steady state

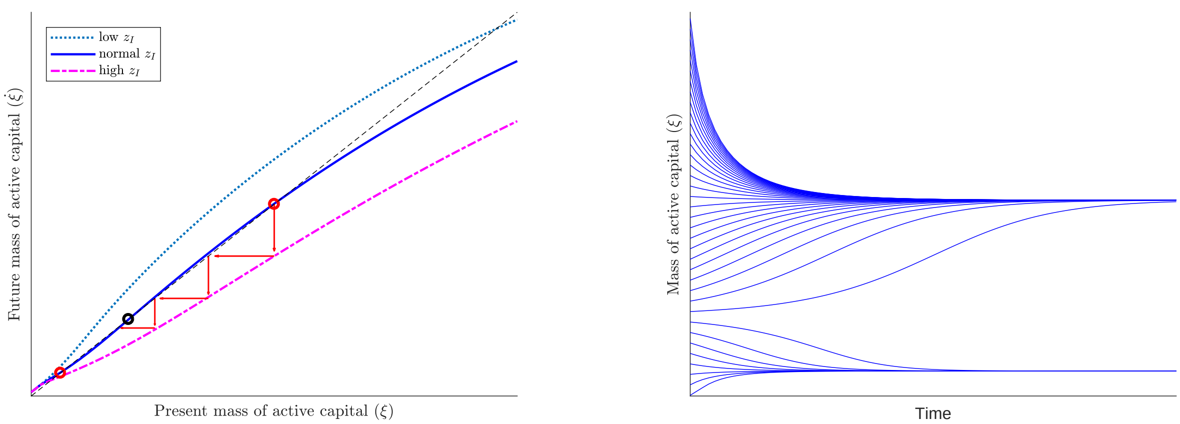
(c) Multiple steady states

(d) Steady states under various z_I.

Figure 1: **Steady State Equilibrium.** Parameter values common across all panels: $q = .002, z^L = .2, \beta = .9$. Values for z_I in the unique steady state equilibrium in panel (a) is $z_I = 0.83$ and in panel (b) is $z_I = 0.9$; in the multiple steady state equilibria in panel (c): $z_I = 0.855$.

Let us first assume δ does not react to the shock, and consider the effect of the shock to both sides of Eq. (16). By Eq. (13), the left hand side of Eq. (16) is unaffected, while the right hand side is affected via two channels. First, λ_I would drop (see Eq. (12)), making investment in market L more attractive (first lambda effect). But, lower λ_I implies that $\dot{\xi}$ would also drop because current locked-in capital is released at a lower rate (see Eq. (2)). This decreases $\dot{\lambda}_I$ and implies that market I is more illiquid in the subsequent period (second lambda effect). Arbitrageurs that consider investing in market I in the current period must trade off the larger probability of trading at a non-revealing price in the current period with the longer expected duration of the investment and therefore the larger opportunity cost of being inactive in future periods. When this second effect is sufficiently strong, a larger fraction of active arbitrageurs flows into market L and away from market I. Anticipating such trade-off arising from the first and the second lambda effect, capital allocation δ endogenously readjusts given the arrival of a shock in equilibrium. We present numerical examples of shock responses in the next subsection.

Figure 2: **Regimes and Evolution of the Mass of Active Capital (ξ).** Panel (a): transition curves for ξ for each of three possible contemporaneous values of $z_I \in \{z_I^{low}, z_I^{normal}, z_I^{high}\}$; circles correspond to stable points in the transition curve for $z_I = z_I^{normal}$. Panel (b): evolutionary paths of ξ under various initial values and fixing $z_I = z_I^{normal}$.



(a) Transition curves for ξ

(b) Evolutionary paths of ξ

Figure 2 illustrates regimes and dynamics of the mass of active capital ξ. Panel (a) plots the transition curve for ξ which maps the current state $\theta = (\xi, z_I)$ into next period active capital $\dot{\xi}$, that is,

$$\dot{\xi} = [1 - \delta(\theta)]\,\xi + (\delta(\theta)\,\xi + 1 - \xi)\left(q + (1-q)\frac{\delta(\theta)\xi}{z_I}\right).$$

An intersection with the 45 degree line is a fixed point of the transition curve such that $\dot{\xi} = \xi$. The intermediate (solid) curve displays three fixed points, with the lowest and the highest being stable and the middle one unstable. The middle fixed point corresponds to the threshold value of ξ which separates the two regions of attraction for ξ.[14]

Panel (b) of Figure 2 illustrates these regions of attraction by plotting the evolution, as implied by the intermediate curve in panel (a), of the mass of active capital for different initial values. The value of ξ converges to the highest (lowest) stable point for initial values of ξ above (below) the threshold value. Hence, if the value of active capital ξ is close to a stable

[14]In case of a deterministic model described in Section 5.2, this would be the value of ξ corresponding to the the unstable equilibrium in panel (c) of Figure 1 (the middle intersection between the IC and CM curves).

point, it converges back to it after experiencing a small deviation due to shocks. However, once ξ crosses the threshold value, then ξ is set on a different trajectory and converges to a different stable point. This situation describes a regime shift in active capital and liquidity in the economy.

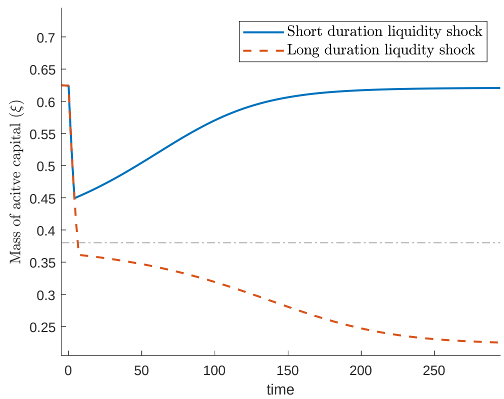
Such a regime shift is further illustrated in panel (a) of Figure 2. The arrows show the effects of shocks to z_I from its normal level starting from the stable point with high liquidity (or large mass of active capital). A temporary increase in z_I pushes ξ downward in the next periods, but such shock is absorbed in subsequent periods if z_I goes back to its normal state. However, the figure illustrates that if the higher level of noise trading persists for more periods (three periods for the parameters in the figure), then ξ crosses the threshold value. After this happens, ξ is set on a downward trajectory toward the stable point characterized by low liquidity even if noise trading intensity z_I reverts back to its normal state. The mass of active capital ξ can go back from this low level to the original high level only after a sequence of favorable shocks to z_I that push ξ above the critical threshold which would put the dynamics of active capital on the upward trajectory. Our numerical simulation in the next subsection shows an example of such transitions across liquidity regimes.

5.4 Numerical Examples

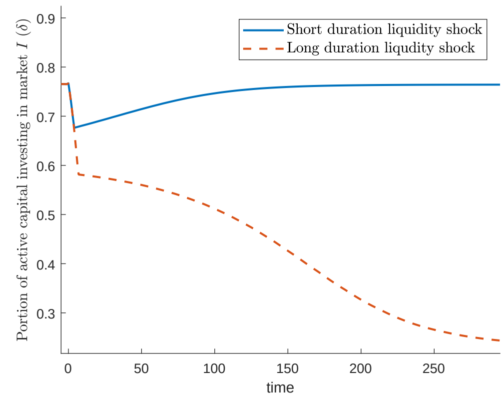
In this subsection, we provide some numerical examples of dynamic responses to stochastic shocks to noise trading intensity in light of the feedback channel between price efficiency and active capital. A shock is a temporary deviation of z_I from its normal level to a higher level, after which z_I reverts back to its normal level. Our examples illustrate how temporary shocks can trigger flight-to liquidity and cause liquidity hysteresis, and have near permanent effects on both markets.

Figures 3 shows responses to shocks to z_I with two different durations. The shock in the current period ($t = 0$) leads to a drop in price efficiency in market I and therefore a decrease in ξ from its initial value starting from the subsequent period. In both cases of short and long duration shocks, arbitrageurs react to the shock by flowing out of market I (i.e., δ decreases) as they anticipate lower liquidity and larger opportunity cost of being locked in this market going forward. Therefore, market I suffers further decreases in price efficiency, thereby triggering further decreases in ξ until the shock is removed.[15] In case of a short duration shock, such market illiquidity is gradually resolved as soon as the shock is removed. This replenishes active capital and the economy converges back to the initial stable point. By contrast, the response to a longer duration shock, which drags the level of active capital below the critical

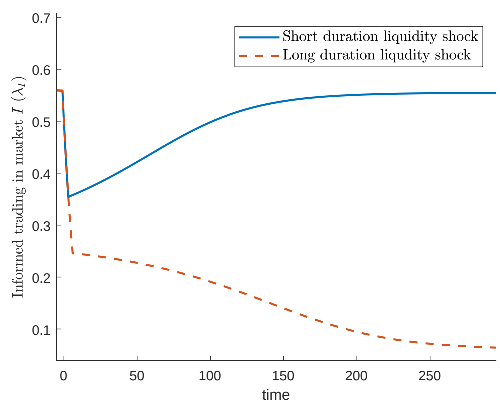
[15] Price efficiency in market L, however, may increase or decrease depending on the relative magnitudes of two confounding effects; price efficiency in market L tends to increase as arbitrageurs flow into this market, but the overall reduction in active capital has the opposite effect. The latter effect dominates after the short duration shock, while the former effect dominates after the long duration shock.



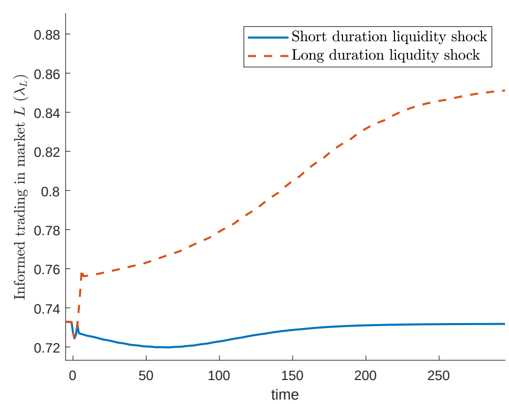
(a) Active capital



(b) Investment in market I



(c) Price efficiency in market I

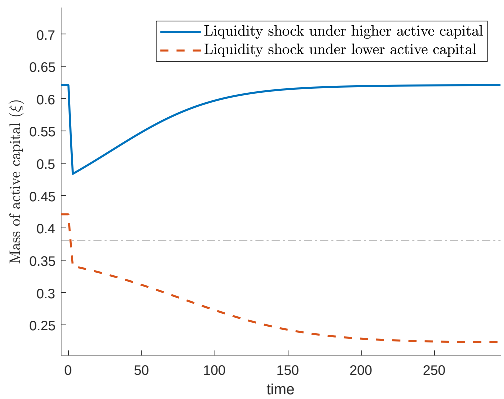


(d) Price efficiency in market L

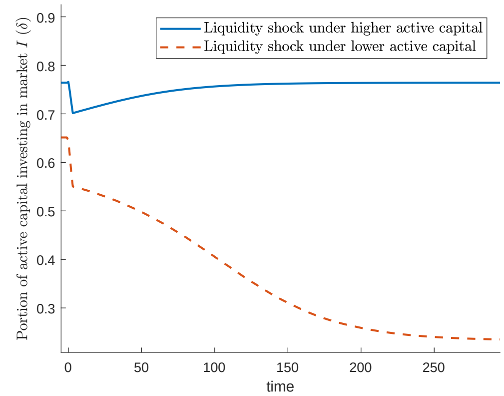
Figure 3: **Transitional Dynamics for a Temporary Shock under Different Shock Durations.** A short duration shock (duration of four periods, solid line) and a long duration shock (duration of seven periods, dashed line) is given at $t = 0$. Parameter values: $q = .002, z_I \in \{.8, .86, .96\}, z_L = .2, \beta = .9$. Transition probabilities are given by $\omega_{11} = .2, \omega_{12} = .8, \omega_{12} = 0, \omega_{21} = .1, \omega_{22} = .82, \omega_{23} = .08, \omega_{31} = 0, \omega_{32} = .85, \omega_{33} = .15$ where states 1,2 and 3 correspond to low, normal and high level of z_I, respectively.

threshold, has different dynamics due to liquidity hysteresis. Instead of reverting back, the flow of arbitrageurs out of market I and into market L persists after the shock is removed. This "flight-to-liquidity" continues as the economy transitions to the low information regime that features low values for δ and ξ.

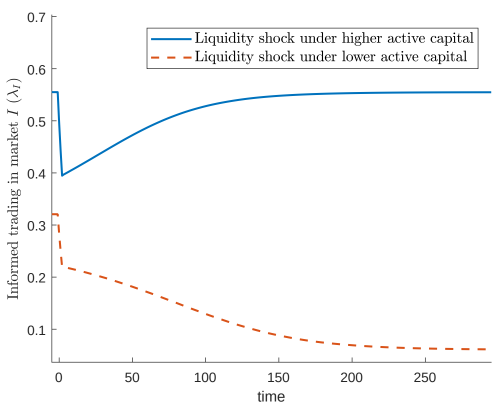
Figure 4 shows the responses to shocks to z_I at two different initial levels of active capital ξ (but with an identical duration). As in Figures 3, active capital and price efficiency in market I decrease as a result of shocks and the ensuing capital flow out of market I. In addition, Figure 4 shows how the resiliency or fragility of the economy depends on the current level of active capital. Intuitively, a reduction in price efficiency in market I has a stronger feedback effect on



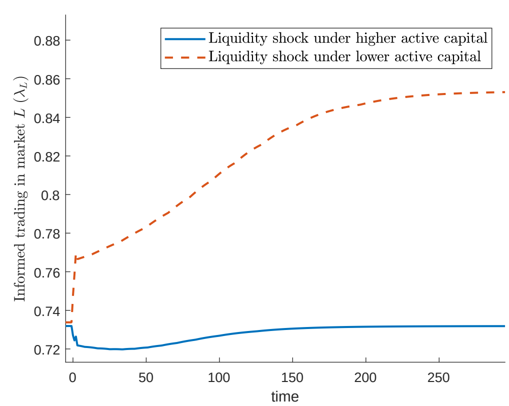
(a) Active capital



(b) Investment in market I



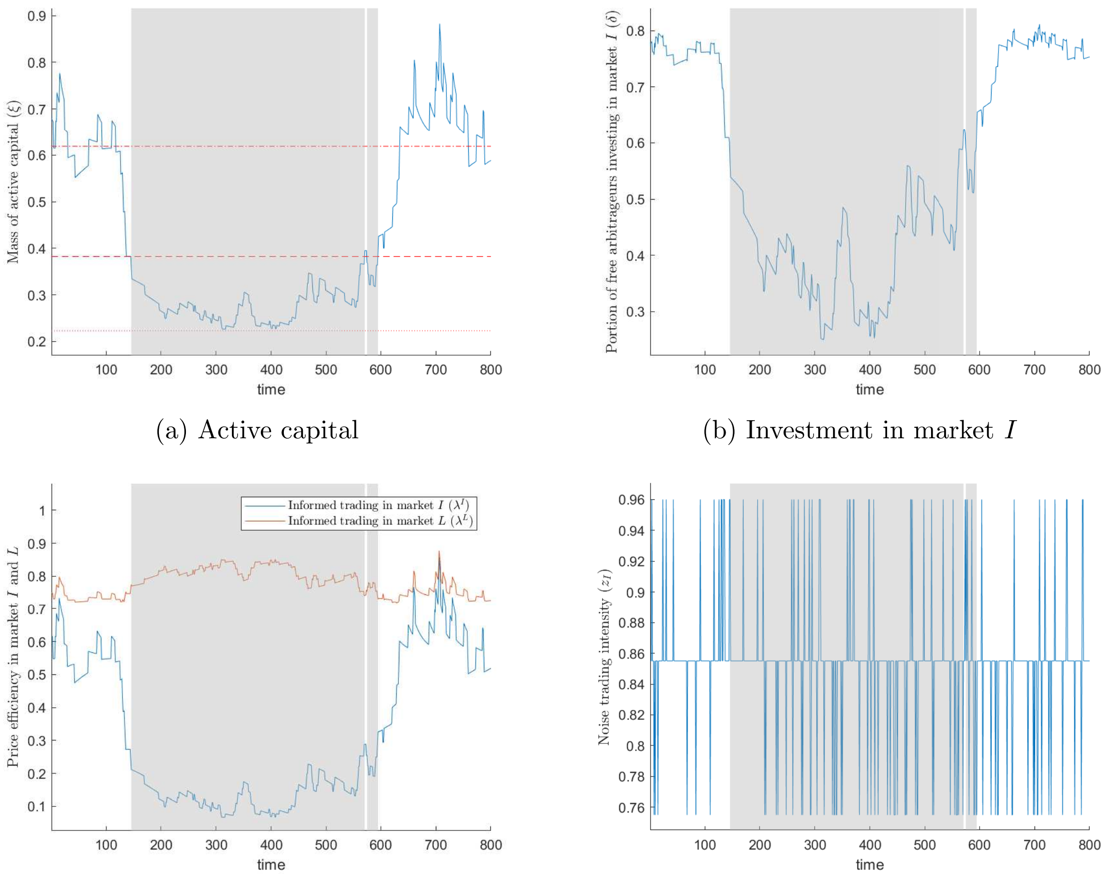
(c) Price efficiency in market I



(d) Price efficiency in market L

Figure 4: **Transitional Dynamics for a Temporary Shock under Different Levels of the Initial Active Capital.** A short duration shock (duration of three periods) is given at $t = 0$ with a higher active capital level (initial value of $\xi = .62$, solid line) and a lower active capital level (initial value of $\xi = .42$, dashed line). Parameter values: $q = .002, z_I \in \{.8, .86, .96\}, z_L = .2, \beta = .9$. Transition probabilities are given by $\omega_{11} = .2, \omega_{12} = .8, \omega_{12} = 0, \omega_{21} = .1, \omega_{22} = .82, \omega_{23} = .08, \omega_{31} = 0, \omega_{32} = .85, \omega_{33} = .15$ where states 1,2 and 3 correspond to low, normal and high level of z_I, respectively.

future active capital when the mass of locked in arbitrageurs is relatively high (equivalently, when active capital is relatively low). As shown in Panel (a), when the initial level of active capital is relatively higher than the critical threshold, the economy is resilient and the level of ξ reverts back to the stable point with high information. On the other hand, with relatively smaller initial amount active capital, the economy is fragile and the outflow of arbitrageurs out of market I persists after the shock is removed as active capital crosses the critical threshold and the economy transitions to the low information regime that features low values for δ and ξ.



(a) Active capital

(b) Investment in market I

(c) Price efficiency

(d) noise trading intensity in market I

Figure 5: **Simulation.** Parameter values: $q = .002, z_I \in \{.8, .86, .96\}, z_L = .2, \beta = .9$. Transition probabilities are given by $\omega_{11} = .2, \omega_{12} = .8, \omega_{12} = 0, \omega_{21} = .1, \omega_{22} = .82, \omega_{23} = .08, \omega_{31} = 0, \omega_{32} = .85, \omega_{33} = .15$ where states 1,2 and 3 correspond to low, normal and high level of z_I, respectively. .

In equilibrium, the market can move in and out the illiquidity regime. We illustrate this in Figure 5, which shows a simulation of the stochastic model when there is a "normal" and persistent level for noise trading intensity in market I but with small probability noise trading intensity can jump up or down from its normal level, and these shocks are not persistent. The occurrence of temporary shocks does not have persistent effects in the first portion of the simulation. It is only when bad shocks occur for several consecutive periods (around $t = 150$ in the figure) that there is a sustained flight-to-liquidity and the economy enters a different regime. In the figure, the initial high liquidity regime for market I (white area) is followed by a low liquidity regime for market I (shaded area) after the occurrence of a sequence of bad shocks in this market. The economy is therefore trapped in this regime for many periods even though noise trading is in its normal state most of the time during these periods. It takes a sequence of good shocks in market I for the economy to exit this illiquidity regime and revert back to the high liquidity regime for market I. Along the transition, capital flows to market I and improves liquidity in this market; as a result locked in capital is released at a faster rate, further increasing liquidity and price efficiency.

6 Discussion

In this section, we discuss empirical implications of our model.

6.1 First Lambda vs. Second Lambda Effects

In our model, price efficiency (lambda) plays a dual role in long-lived assets. It determines mispricing wedge which determines the profitability of investment opportunities but also determines the maturity of new investment which determines its liquidity. Upon the arrival of a liquidity shock, two opposing effects arise together. We call the effect of increased profitability the first lambda effect, and the effect of decreased liquidity the second lambda effect. The first lambda effect is closer to the traditional interpretation of Kyle's lambda. In that context, lambda measures price impact of trade and is often interpreted as a measure of illiquidity from the point of view of uninformed traders. On the other hand, the second lambda determines the speed at which arbitrageurs can close out their positions at a profit because subsequent trades make prices efficient. Hence, second lambda is a measure of liquidity from the point of view of informed arbitrageurs in our model.

Recursively substituting Eq. (14) into itself yields

$$\lambda_{L,t} - \lambda_{I,t} = \mathrm{E}\left[\sum_{\tau=1}^{\infty} \beta^{\tau} (1-q)^{\tau} \prod_{j=0}^{\tau-1} (1-\lambda_{I,t+j}) (1-\lambda_{L,t+\tau})\right], \tag{17}$$

where $\lambda_{h,t}$ is price efficiency in market h at time t.

Note that $(1-q)^{\tau} (1-\lambda_{I,t}) \ldots (1-\lambda_{I,t+\tau-1})$ is the probability that by investing in market I in the current period, an arbitrageur's capital is not available for a new trade in period $t+\tau$. Because $(1-\lambda_{L,t+\tau})$ is the probability of realizing a speculative profit in $t+\tau$ in market L, the right hand side of Eq. (17) measures the expected loss in future speculative profits arising from capital being trapped in future periods. This is the opportunity cost of trading in market I.

Our model implies that the mass of active capital is the key state variable that determines the market liquidity. However, active capital is difficult to observe. On the other hand, price efficiency is often measured using various empirical measures. Our theory predicts that those empirical measures of price efficiency can be used as a proxy for the actual state variable–the mass of active capital. In the following subsections, we discuss how this idea can be applied to some empirical implications.

6.2 Liquidity Crises

There are several well-known episodes of liquidity crisis such as the 1987 stock market crash, the 1998 Long-Term Capital Management crisis, and the subprime mortgage crisis of 2007-2009. These episodes are often characterized by a delayed recovery of liquidity in the aftermath (e.g., Mitchell, Pedersen, and Pulvino (2007); Coval and Stafford (2007)). Existing literature often explains those liquidity crises as a result of shock amplifications which impair capital itself.[16] In our model, a liquidity crisis can happen even in the absence of any reduction in arbitrage capital itself – what matters is a reduction in *active* arbitrage capital.

Our simulations illustrate that all it takes to create a full-blown liquidity crisis is merely a transient shock which causes engaged capital to get redeployed more slowly. While market liquidity recovers rather quickly after a small shock, a sizable shock (or a sequence of small shocks) can trigger a change in regime and have long-lasting impact. At the core of this argument lies the multiplicity of steady state equilibria (or stable fixed points); a sufficiently large shock can disturb the system enough to put the state variable (active capital) in another path.[17] This mechanism allows us to give a distinct prediction that equilibrium may be shifted toward low liquidity as a result of shocks. In the case of stochastic shocks, it takes a long time to have a series of good shocks that push active capital to the upward trajectory. This prediction matches empirical observations of long periods of illiquidity in the market.

6.3 Flight-to-Liquidity

Using our model, we show that active arbitrageurs may optimally choose to invest in the liquid market upon the arrival of liquidity shocks. There is indeed ample evidence about flight-to-liquidity in various markets: investors tend to prefer liquid assets during bad times. For example, Beber, Brandt, and Kavajecz (2007) find that capital flow in case of the Euro-area government bonds is mostly determined by liquidity rather than credit quality. Acharya, Amihud, and Bharath (2013) also document that there are two liquidity regimes for corporate bonds. In one regime, liquidity shocks have mostly insignificant effects on bond prices whereas in another regime, liquidity shocks produce significant effects. In particular, they find empirical evidence of flight-to-liquidity: prices of investment-grade bonds rise while prices of speculative-

[16]For example, capital becomes increasingly less available through the channel of tightened collateral (e.g., Gromb and Vayanos (2002)) or margin constraints (e.g., Brunnermeier and Pedersen (2009)).

[17]Even in cases where the system has only one regime (or one steady state), a shock that initially reduces liquidity may lead to further drops in liquidity and long delays before liquidity is re-established. But, it can take arbitrarily long time to recover in case of multiple regimes.

grade bonds fall. Ben-Rephael (2017) also find that mutual funds reduce their holdings of illiquid stocks during bad times.

In our model, capital tends to flow out of a market if this market's future liquidity is expected to deteriorate. Because lower future liquidity means longer maturity of new investment, mispricing wedge should become larger to compensate arbitrageurs with lower price efficiency in return for longer maturity. Furthermore, we also show the conditions under which capital flows in or out of a market hit by a liquidity shock.

One of the key observations in our model is well summarized by Eq. (17). It states that the cross-sectional difference in liquidity across markets predict future illiquidity. That is, when there is a larger cross-sectional difference in liquidity, we expect a period of low liquidity in the subsequent periods. This prediction is consistent with flight-to-liquidity episodes in which there is a divergence of liquidity across markets, and this is followed by a period of overall low liquidity. There is some empirical support on this hypothesis. Cao, Chen, Liang, and Lo (2013) find that hedge fund managers can time market liquidity based on their forecasts of future market liquidity conditions. Furthermore, Cao, Liang, Lo, and Petrasek (2017) find that hedge funds contribute to price efficiency by investing in relatively more mispriced stocks, but those stocks tend to experience large decline in price efficiency during liquidity crises.

6.4 Reaching-for-Yield

As an opposite situation to flight-to-liquidity, traders sometimes seek more risk by investing in illiquid assets. That is, traders tend to reach for yield during good times ample with liquidity (e.g., Becker and Ivashina (2015)). Our model can also contribute to the discussion by suggesting an alternative mechanism of reaching for yield. Our theory suggests that as the market starts having more capital, there would a reinforcement effect in which more locked-in capital is further released. This will raise price efficiency and shorten maturities of investment in illiquid assets. Consequently, capital starts flowing into more illiquid asset classes as active capital expands. This can reduce mispricing wedge greatly by transferring to high liquidity equilibrium. While lower mispricing wedge is good for price efficiency, it puts pressure on financial institutions to reach for higher yields. We interpret this situation as reaching-for-yield because arbitrageurs invest more in riskier assets in that situation.

7 Conclusion

We study a dynamic stationary model of informed trading with two markets. The model features endogenous liquidity regimes where temporary shocks to noise trading can trigger a shift of the regime. We show that upon the arrival of a shock arbitrage capital may actually flow out of the illiquid market. With some arbitrage capital flowing out, the remaining capital in the market becomes trapped because it is too illiquid for arbitrageurs to want to close out their positions. This in turn deepens illiquidity in a self-reinforcing manner, thereby creating liquidity hysteresis where illiquidity persists even when the initial cause is removed.

In our model, arbitrage capital plays a dual role; the wedge of mispricing not only decides the profitability of new investment but also decides the speed at which engaged arbitrage capital is released (thus deciding the availability of arbitrage capital). The dual role of arbitrage capital implies that efficiency depends on the pool of active capital as a state variable. Furthermore, it creates a feedback channel between active capital and liquidity which leads to multiple regimes where there is a threshold of active capital that separates domains of attraction for liquidity. Therefore, a large adverse shock can trigger a vicious cycle of illiquidity with flight-to-liquidity where arbitrage capital flows to the market with short-lived assets.

Although the market can move in and out of different regimes, it may take quite a long time to come back to a normal liquidity regime from an illiquidity regime; it requires a sequence of good shocks strong enough to push the mass of active capital toward the path of a normal liquidity regime. This result provides a mechanism for slow moving capital under which a seemingly temporary liquidity shock creates long lasting illiquidity in the market. Our results shed light on why capital moves slowly, how fast (or slowly) it moves, and to which directions it moves. The results further provide interesting implications on liquidity crises, flight-to-liquidity, and cross-sectional patterns of liquidity.

Appendix A: proofs for Section 4

The derivation of the value functions in Section 4.4:
We first derive J_L. Because asset qualities are equally likely, the continuation value of active arbitrageurs making new investment in market L is

$$J_L(\theta) = \frac{1}{2} J_L(\theta; G) + \frac{1}{2} J_L(\theta; B), \tag{18}$$

where $J_L(\theta; s)$ conditions on the quality of the chosen asset being $s \in \{G, B\}$. We have:

$$\begin{aligned} J_L(\theta; G) &= -(\lambda_L P^G + (1 - \lambda_L)P^0) + \beta\Big[V_L^G + \mathrm{E}[J_f(\dot{\theta})|\theta]\Big]; \\ J_L(\theta; B) &= (\lambda_L P^B + (1 - \lambda_L)P^0) + \beta\Big[-V_L^B + \mathrm{E}[J_f(\dot{\theta})|\theta]\Big]. \end{aligned}$$

Because $-(\lambda_L P^G + (1 - \lambda_L)P^0) + \beta V_L^G = (\lambda_L P^B + (1 - \lambda_L)P^0) - \beta V_L^B$, it is immediate that $J_L(\theta)$ in Eq. (18) is equivalent to the one in Eq. (7).

We turn to the derivation of J_I. In a similar fashion, the continuation value of an active arbitrageur making a new investment in market I is given by

$$J_I(\theta) = \frac{1}{2} J_I(\theta; G) + \frac{1}{2} J_I(\theta; B), \tag{19}$$

where

$$\begin{aligned} J_I(\theta; G) &= -(\lambda_I P^G + (1 - \lambda_I)P^0) + \beta U(\theta; G); \\ J_I(\theta; B) &= (\lambda_I P^B + (1 - \lambda_I)P^0) + \beta U(\theta; B), \end{aligned} \tag{20}$$

and

$$\begin{aligned} U(\theta; G) \equiv& qV_I^G + (1 - q)\lambda_I P^G + (1 - (1 - \lambda_I)(1 - q))\mathrm{E}[J_f(\dot{\theta})|\theta] \\ &+ (1 - \lambda_I)(1 - q)\mathrm{E}[J_l(\dot{\theta}; G)|\theta]; \\ U(\theta; B) \equiv& - qV_I^B - (1 - q)\lambda_I P^B + (1 - (1 - \lambda_I)(1 - q))\mathrm{E}[J_f(\dot{\theta})|\theta] \\ &+ (1 - \lambda_I)(1 - q)\mathrm{E}[J_l(\dot{\theta}; B)|\theta]. \end{aligned} \tag{21}$$

We define $J_l(\theta; s)$ to be the continuation value of a locked-in arbitrageur holding an asset with quality s in market I. Because locked-in arbitrageurs can either liquidate or keep holding onto their existing positions, we have

$$J_l(\theta; s) = \max\left(J_E(\theta; s), J_S(\theta; s)\right), \tag{22}$$

where

$$
\begin{aligned}
J_E(\theta; G) &= \lambda_I P^G + (1-\lambda_I)P^0 + \beta \mathrm{E}[J_f(\dot{\theta})|\theta]; \\
J_E(\theta; B) &= -\lambda_I P^B - (1-\lambda_I)P^0 + \beta \mathrm{E}[J_f(\dot{\theta})|\theta]; \\
J_S(\theta; s) &= \beta U(\theta; s).
\end{aligned}
$$

Now, we conjecture that

$$
U(\theta; G) = U(\theta; B) + \frac{2P^0}{\beta}. \tag{23}
$$

Then, Eq. (22) implies that $J_E(\theta; G) = J_E(\theta; B) + 2P^0$ and $J_S(\theta; G) = J_S(\theta; B) + 2P^0$, therefore, we have

$$
J_l(\theta; G) = J_l(\theta; B) + 2P^0. \tag{24}
$$

Eqs. (23) and (24) imply that

$$
U(\theta; G) - U(\theta; B) = q(V_I^G + V_I^B) + 2(1-q)P^0. \tag{25}
$$

Because $P^0 = \frac{\beta q}{1-\beta(1-q)}\left(\frac{V_I^G + V_I^B}{2}\right)$, Eq. (25) implies that $U(\theta; G) = U(\theta; B) + \frac{2P^0}{\beta}$, which proves that the initial conjecture in Eq. (23) is indeed true.

Finally, Eq. (20) implies that $J_I(\theta; G) - J_I(\theta; B) = -2P^0 + \beta[U(\theta; G) - U(\theta; B)]$, which in turn implies that $J_I(\theta; G) = J_I(\theta; B)$ due to Eq. (23). Therefore, we conclude that $J_I(\theta)$ in Eq. (19) is equivalent to the one in Eq. (6).

Proof of Lemma 1: Let X_i^a be the aggregate order flow of arbitrageurs for asset i. Suppose that there are μ_i mass of arbitrageurs investing in asset i. Because arbitrageurs are risk-neutral and informed, their aggregate order flow is given by $X_i^a = \mu_i$ if $v_i = V_h^G$, and $X_i^a = -\mu_i$ otherwise. Then, the market makers observe the aggregate order flow $X_i = X_i^a + \zeta_i$. Bayes' theorem implies that the market makers' posterior belief that $v_i = V_h^G$ is given by

$$
\hat{p}_i(X_i, \mu_i) = \frac{p f_X^i(X_i|G)}{p f_X^i(X_i|G) + (1-p) f_X^i(X_i|B)}, \tag{26}
$$

where $p = \frac{1}{2}$ is the prior belief and $f_X^i(\cdot|G)$ and $f_X^i(\cdot|B)$ are the distribution of X_i given $v_i = V_h^G$ and $v_i = V_h^B$, respectively.

Because ζ_i follows a uniform distribution on the interval $[-z_i, z_i]$ in each period, X_i follows a uniform distribution either on the interval $[\mu_i - z_i, \mu_i + z_i]$ if $v_i = V_h^G$, or on the interval

$[-\mu_i - z_i, -\mu_i + z_i]$ otherwise. Therefore, Eq. (26) implies

$$\hat{p}_i = \begin{cases} 0 & \text{if } -\mu_i - z_i \leq X_i < \mu_i - z_i \\ p & \text{if } \mu_i - z_i \leq X_i \leq -\mu_i + z_i \\ 1 & \text{if } -\mu_i + z_i < X_i \leq \mu_i + z_i \end{cases}$$

Therefore, the probability of revealing the true value of v_i is given by

$$\begin{aligned} \lambda_i &= Pr(\hat{p}_i = 0 \text{ or } \hat{p}_i = 1) \\ &= Pr(-\mu_i - z_i \leq X_i < \mu_i - z_i) + Pr(-\mu_i + z_i < X_i \leq \mu_i + z_i) = \frac{\mu_i}{2z_i} + \frac{\mu_i}{2z_i} = \frac{\mu_i}{z_i}. \end{aligned}$$

In a market-wise symmetric equilibrium, all the future lambdas are equalized across assets in each market. Then, Eqs. (6) and (7) imply that the continuation value of arbitrageurs making new investment is identical across all assets except for the cost of acquiring the position in the current period. Therefore, all arbitrageurs would want to invest in an asset with the lowest λ_i in the current period until λ_i are equalized across assets in each market, i.e., $\lambda_i = \lambda_h$ for all asset i in market h. It is immediate that $\mu_i = \delta\xi$ for market I, and $\mu_i = (1-\delta)\xi$ for market L. Therefore, we obtain the desired results in Eqs. (12) and (13). ■

Lemma 3 *In an interior equilibrium, it is never optimal to close out the existing position early, i.e.,* $J_l(\theta) = J_S(\theta)$.

Proof. We can write $J_S(\theta)$ as

$$J_S(\theta) = J_I(\theta) + \lambda_I P^G + (1-\lambda_I)P^0.$$

In an interior equilibrium, $J_I(\theta) = J_L(\theta)$ and therefore

$$\begin{aligned} J_S(\theta) &= J_L(\theta) + \lambda_I P^G + (1-\lambda_I)P^0 \\ &= -(\lambda_L P^G + (1-\lambda_L)P^0) + \beta V_L^G + \beta \mathrm{E}[J_f(\dot{\theta})] + \lambda_I P^G + (1-\lambda_I)P^0 \\ &= -(\lambda_L P^G + (1-\lambda_L)P^0) + \beta V_L^G + J_E(\theta) \end{aligned}$$

Because $(\lambda_L P^G + (1-\lambda_L)P^0) < \beta V_L^G$, we have $J_S(\theta) > J_E(\theta)$. ■

Proof of Lemma 2: We start rewriting $J_I(\theta)$ as

$$J_I(\theta) \overset{(i)}{=} \left(P^G - P^0\right)(1-\lambda_I) + \beta[(1-\lambda_I)(1-q)\left(\mathrm{E}[J_l(\dot{\theta})|\theta] - P^G - \mathrm{E}[J_f(\dot{\theta})|\theta]\right)]$$
$$+\beta\mathrm{E}[J_f(\dot{\theta})|\theta];$$
$$J_I(\theta) \overset{(ii)}{=} \left(P^G - P^0\right)(1-\lambda_I) + \beta[(1-\lambda_I)(1-q)\mathrm{E}[J_I(\dot{\theta}) - \left(P^G - P^0\right)\left(1-\dot{\lambda}^I\right) - J_f(\dot{\theta})|\theta]$$
$$+\beta\mathrm{E}[J_f(\dot{\theta})|\theta];$$
$$J_I(\theta) \overset{(iii)}{=} \left(P^G - P^0\right)(1-\lambda_I)\left[1-\beta(1-q)\left(1-\mathrm{E}[\dot{\lambda}^I|\theta]\right)\right] + \beta\mathrm{E}[J_f(\dot{\theta})|\theta],$$

where (i) uses $qV_I^G = P^G\left(R_f - (1-q)\right)$ and (ii) uses Lemma 3 and therefore $J_l(\dot{\theta}) = J_I(\dot{\theta}) + \dot{\lambda}_I P^G + \left(1-\dot{\lambda}_I\right)P^0$ and (iii) uses $J_f(\dot{\theta}) = J_I(\dot{\theta})$. Similarly, we can use $\beta V_L^G = P^G$ to write

$$J_L(\theta) = \left(P^G - P^0\right)(1-\lambda_L) + \beta\mathrm{E}[J_f(\dot{\theta})|\theta]$$

Combining the above expressions for $J_L(\theta), J_I(\theta)$ we obtain that in an interior equilibrium,

$$J_L(\theta) = J_I(\theta) \Leftrightarrow (1-\lambda_L) = (1-\lambda_I)\left[1-\beta(1-q)\left(1-\mathrm{E}\left[\dot{\lambda}_I|\theta\right]\right)\right].$$

Rearranging the above equation gives Eq. (14). ■

Appendix B: proof of Proposition 1

In this section, we prove existence and uniqueness of stationary equilibrium of our model by using the contraction property of equilibrium mapping for price efficiency in the class of Lipschitz continuous functions. See, for example, Follmer, Horst, and Kirman (2005), Acharya and Viswanathan (2011), and Fajgelbaum, Schaal, and Taschereau-Dumouchel (2017) for similar methods of proof of existence and uniqueness of stationary equilibrium of different classes of models.

Notation

Denote $Z = \left\{z_I^1, z_I^2, ..., z_I^N\right\}$ the set of possible values for noise trading intensity in market I, and let $\bar{z}_I = \max\left\{z_I^1, z_I^2, ..., z_I^N\right\}$ and $\underline{z}_I = \min\left\{z_I^1, z_I^2, ..., z_I^N\right\}$. Define the constant M to be such that $M\left|z_I^n - z_I^m\right| \geq \bar{z}_I - \underline{z}_I$ for all n, m. We further use the notation $\omega\left(z_I^n\middle| z_I^m\right) = \omega_{nm}$ and we define

$$\alpha \equiv \max_{z_I', z_I^n, z_I^m \in Z}\left|\omega\left(z_I'\middle| z_I^m\right) - \omega\left(z_I'\middle| z_I^n\right)\right|.$$

We remark that $\alpha = 0$ in case noise trading intensity process z_I is i.i.d. in which case $\omega\left(z_I'\middle| z_I^m\right) - \omega\left(z_I'\middle| z_I^n\right) = 0$ for all $z_I^n, z_I^m, z_I' \in Z$.

Denote $\Xi = \left[\underline{\xi}, 1\right]$ the interval of possible values for ξ; the lower bound $\underline{\xi} = \max\left\{1 - \bar{z}_I \frac{1-\sqrt{q}}{1+\sqrt{q}}, 0\right\}$ is derived in Lemma 5.

We denote $B(\Xi \times Z)$ the set of bounded, continuous functions $\lambda : (\xi, z_I) \in \Xi \times Z \to \mathbb{R}$ with the sup-norm $\|\lambda\| = \sup_{\xi \in \Xi, z_I \in Z} |\lambda(\xi, z_I)|$. λ is decreasing in z_I if $\lambda(\xi, z_I^n) - \lambda(\xi, z_I^m) \leq 0$ for all $\xi \in \Xi$ and $z_I^n, z_I^m \in Z$ such that $z_I^n > z_I^m$. The rate of change in z_I is bounded by some constant κ if for all $z_I^n, z_I^m \in Z$ we have

$$\sup_{\xi \in \Xi} |\lambda(\xi, z_I^n) - \lambda(\xi, z_I^m)| \leq \kappa |z_I^n - z_I^m|.$$

Reformulation

We can reformulate the indifference condition $J_I(\theta) = J_L(\theta)$ in terms of λ_I. Using Eqs. (12)-(13) to substitute out λ_L in Eq. (14) and rearranging, we obtain that in an interior equilibrium there exists a function $\lambda : (\xi, z_I) \in \Xi \times Z \to \mathbb{R}$ for price efficiency in market I that satisfies the following functional equation:

$$\lambda(\xi, z_I) = A(z_I)\xi - B(z_I)(1 - \lambda(\xi, z_I)) \left(1 - \sum_{z_I' \in Z} \omega\left(z_I' \middle| z_I\right) \lambda(C(\xi, z_I), z_I')\right), \tag{27}$$

where

$$\begin{aligned} A(z_I) &= \frac{1}{z_L + z_I}; \\ B(z_I) &= \frac{\beta(1-q)z_L}{z_L + z_I}; \\ C(\xi, z_I) &= q + (1-q)\xi + (1-q)(1-z_I)\lambda(\xi, z_I) - (1-q)\xi\lambda(\xi, z_I) + (1-q)z_I[\lambda(\xi, z_I)]^2 \\ &= 1 - (1-q)(1 - \lambda(\xi, z_I))(1 - \xi + z_I\lambda(\xi, z_I)), \end{aligned} \tag{28}$$

and where $\dot{\xi} = C(\xi, z_I)$ follows from the law of motion Eq. (2) and the definition for λ_I in Eq. (12).

Definitions and assumptions

Definition 2 *Define the functions $f, g : (u, z_I) \in [0,1] \times Z \to \mathbb{R}$ as*

$$f(u, z_I) = \max\left\{1 - q, (1-q)(z_I u - 1), \left(1 - \underline{\xi}\right) u, \left(1 - \underline{\xi}\right) u + (1-q)(1 - z_I u)\right\} \tag{29}$$

and

$$g(u, z_I) = \max\left\{f(u, z_I), (1-q) z_I u\right\} \tag{30}$$

and the function $\Gamma : z_I \in Z \to \mathbb{R}$ *as*

$$\Gamma\left(z_I\right) = \begin{cases} \frac{z_I}{4}\left(1 + \frac{\bar{z}_I}{z_I}\frac{1-\sqrt{q}}{1+\sqrt{q}}\right)^2, & \text{if } \bar{z}_I \frac{1-\sqrt{q}}{1+\sqrt{q}} \leq z_I; \\ \bar{z}_I \frac{1-\sqrt{q}}{1+\sqrt{q}}, & \text{otherwise.} \end{cases} \tag{31}$$

Definition 3 *Let* $\hat{\lambda}_\xi, \lambda_\xi^*$ *be the constant values*

$$\hat{\lambda}_\xi = \frac{1 + \sqrt{1 + \frac{4\bar{z}_I}{\beta(1-q)^2 z_L}}}{2\bar{z}_I}, \tag{32}$$

and

$$\lambda_\xi^* = \min_{z_I \in Z} \frac{1/B(z_I) - 2}{(1-q)\Gamma\left(z_I\right)}. \tag{33}$$

Let Λ_ξ *be the set*

$$\Lambda_\xi = \left\{ \lambda_\xi \in \mathbb{R}^+ \middle| \, \lambda_\xi \geq \max_{z_I \in Z} A(z_I) + B(z_I)\left[1 + f\left(\lambda_\xi, z_I\right)\right]\lambda_\xi, \lambda_\xi \leq \hat{\lambda}_\xi, \lambda_\xi < \lambda_\xi^* \right\}, \tag{34}$$

and define

$$\overline{\lambda}_\xi = \inf \Lambda_\xi. \tag{35}$$

Assumption 1 *Parameters are chosen such that* Λ_ξ *is non-empty.*

Assumption 2 *Parameters are chosen such that* $1 > \beta\left(1-q\right)z_L + \bar{z}_I \frac{1-\sqrt{q}}{1+\sqrt{q}}$.

Definition 4 *Let* $\overline{\lambda}_\gamma$ *be the constant value*

$$\overline{\lambda}_\gamma = \frac{1 - \bar{z}_I \frac{1-\sqrt{q}}{1+\sqrt{q}} - \beta\left(1-q\right)z_L}{\left(z_L + \bar{z}_I\right)\beta\left(1-q\right)z_L \alpha M} \tag{36}$$

Let Λ_z *be the set*

$$\Lambda_z = \left\{ \lambda_z \in \mathbb{R}^+ \middle| \, \lambda_z \geq \max_{z_I^n, z_I^m \in Z, n \neq m} A(z_I^n)A(z_I^m) + B(z_I^n)\left[\overline{\lambda}_z\left(1 + \alpha M\right) + g\left(\overline{\lambda}_z, z_I^n\right)\overline{\lambda}_\xi\right], \lambda_z \leq \overline{\lambda}_\gamma \right\} \tag{37}$$

and define

$$\overline{\lambda}_z = \inf \Lambda_z. \tag{38}$$

Assumption 3 *Parameters are chosen such that* Λ_z *is non-empty and*

$$\overline{\lambda}_\xi \leq \min\left\{ \frac{1-\alpha M}{(1-q)\bar{z}_I \frac{2\sqrt{q}}{1+\sqrt{q}}}, \frac{1}{\left(1-q\right)\bar{z}_I \frac{2\sqrt{q}}{1+\sqrt{q}}}\left(\frac{1}{\left(z_L + \bar{z}_I\right)\overline{\lambda}_z} - \alpha M\right), \frac{1}{\beta\left(1-q\right)^2 z_L\left(\bar{z}_I + z_L\right)\overline{\lambda}_z} \right\}. \tag{39}$$

It is easy to verify that Assumptions 1, 2 and 3 are jointly satisfied, for example, when q is large enough, or z_L is small enough, or β is small enough, and when α is small enough.

From Eq. (27) we define the following mapping:

Definition 5 *Let $\mathcal{T} : \lambda \in B(\Xi \times Z) \to B(\Xi \times Z)$ be the mapping*

$$\mathcal{T}\lambda(\xi, z_I) = \max\left\{0, A(z_I)\xi - B(z_I)(1 - \lambda(\xi, z_I))\left(1 - \sum_{z_I' \in Z} \omega\left(z_I' \middle| z_I\right)\lambda(C(\xi, z_I), z_I')\right)\right\}, \tag{40}$$

Definition 6 *Let $\mathcal{F}_0 \subset B(\Xi \times Z)$ be the set of bounded continuous functions $\lambda : (\xi, z_I) \in \Xi \times Z \to \mathbb{R}$ which are bounded below by zero and above by one, monotone increasing in ξ, and Lipschitz continuous of modulus $\bar{\lambda}_\xi$ in ξ.*

We remark that if $\mathcal{T}$ has a strictly positive fixed point in $\mathcal{F}_0$, such a fixed point satisfies Eq. (27) by construction.

Definition 7 *Let $\mathcal{F}_1 \subset \mathcal{F}_0$ be the subset of functions in $\mathcal{F}_0$ that are decreasing in z_I with rate of change bounded by $\bar{\lambda}_z$.*

Proposition 1

We restate here the text Proposition 1 in full: *Under Assumptions 1 and 2, there exists a unique stationary interior equilibrium in which price efficiency in the illiquid market λ_I is monotone increasing in active capital ξ. Furthermore, under Assumption 3, λ_I is monotone decreasing in noise trading intensity z_I.*

Proof. The proof is divided in six steps. As a first step, we show that $\mathcal{T}$ maps $\mathcal{F}_0$ into $\mathcal{F}_0$. As a second step, we prove that $\mathcal{F}_0$ is a complete metric space. As a third step, we prove that $\mathcal{T}$ is a contraction on $\mathcal{F}_0$. Contraction Mapping Theorem (See, for example, Theorem 3.2 in Stokey and Lucas (1996)) then implies that $\mathcal{T}$ has a unique fixed point λ_I in $\mathcal{F}_0$. As a fourth step we show that under Assumption 2, such a fixed point λ_I is strictly positive and therefore satisfies Eq. (27); since λ_I is in $\mathcal{F}_0$, then it is increasing in ξ. As a fifth step, we show that under Assumption 3, λ_I is decreasing in z_I. As a sixth and last step, we show that all equilibrium functions in Definition 1 can be uniquely recovered given λ_I.

Step 1: $\mathcal{T}$ maps $\mathcal{F}_0$ into $\mathcal{F}_0$.

Let $\lambda \in \mathcal{F}_0$. Then, λ is bounded between zero and one by assumption. Because $B(z_I) > 0$ and $A(z_I) \in (0, 1)$, then it is immediate by Eq. (40) that $\mathcal{T}\lambda$ is bounded between zero and one. Lemma 7 shows that under Assumption 1, $\mathcal{T}\lambda$ is Lipschitz continuous of modulus $\bar{\lambda}_\xi$ in

ξ for every $\lambda \in \mathcal{F}_0$. Lemma 8 shows that under Assumption 1, $\mathcal{T}\lambda$ is monotone increasing in ξ for every $\lambda \in \mathcal{F}_0$. This concludes the proof of Step 1.

Step 2: $\mathcal{F}_0$ is a complete metric space.

$\mathcal{F}_0$ with metric induced by the sup-norm is a metric space. We must show it is complete. For this, take a Cauchy sequence $\{\lambda_n\}$ of functions in $\mathcal{F}_0$. Because $\mathcal{F}_0$ is a subset of $B(\Xi \times Z)$ and $B(\Xi \times Z)$ is complete (see, for example, Theorem 3.1 in Stokey and Lucas (1996)), $\{\lambda_n\}$ converges to an element λ^* in $B(\Xi \times Z)$. We must show λ^* is in $\mathcal{F}_0$. Because each λ_n is bounded between zero and one, so is the limit. Hence, λ^* is bounded between zero and one. Next, we show λ^* is monotone increasing in ξ. Take $\xi_2 > \xi_1$ and $\varepsilon > 0$, and let n_0 be such that $|\lambda^*(\xi_1, z_I) - \lambda_n(\xi_1, z_I)|, |\lambda^*(\xi_2, z_I) - \lambda_n(\xi_2, z_I)| < \varepsilon/2$ for all $n \geq n_0$. Then,

$$\begin{aligned} \lambda_n(\xi_2, z_I) - \lambda^*(\xi_2, z_I) &\leq \varepsilon/2 \\ -(\lambda_n(\xi_1, z_I) - \lambda^*(\xi_1, z_I)) &\leq \varepsilon/2 \end{aligned}$$

and therefore

$$0 \leq \lambda_n(\xi_2, z_I) - \lambda_n(\xi_1, z_I) \leq \varepsilon + \lambda^*(\xi_2, z_I) - \lambda^*(\xi_1, z_I).$$

Because ε can be taken to be arbitrarily small, then it must be $0 \leq \lambda^*(\xi_2, z_I) - \lambda^*(\xi_1, z_I)$.

Finally, we have

$$|\lambda^*(\xi_1, z_I) - \lambda^*(\xi_2, z_I)| = \lim_{n\to\infty} |\lambda_n(\xi_1, z_I) - \lambda_n(\xi_2, z_I)|.$$

Because each term in the r.h.s. is bounded by $\bar{\lambda}_\xi |\xi_1 - \xi_2|$ by assumption, so is the limit. Hence, λ^* is Lipschitz continuous with modulus $\bar{\lambda}_\xi$. This concludes Step 2.

Step 3: $\mathcal{T}$ is a contraction mapping on $\mathcal{F}_0$.

Lemma 9 shows that under Assumption 1, the mapping $\mathcal{T}$ is a contraction on $\mathcal{F}_0$. Then, steps 1-3 and the Contraction Mapping Theorem imply that $\mathcal{T}$ has a unique fixed point in $\mathcal{F}_0$. We denote such a fixed point with λ_I.

Step 4: λ_I is strictly positive

It is immediate to verify that Assumption 2 together with Lemma 5 imply that $A(z_I)\underline{\xi} - B(z_I) > 0$, and therefore $\mathcal{T}\lambda > 0$ for all $(\xi, z_I) \in \Xi \times Z$. Because λ_I is a fixed point of the $\mathcal{T}$ mapping, λ_I must be a strictly positive function and it satisfies Eq. (27) by construction.

Step 5: λ_I is decreasing in z_I

Lemmas 10 and 11 imply that under Assumptions 1 to 3, $\mathcal{T}$ maps $\mathcal{F}_1$ into $\mathcal{F}_1$. By arguments analogous to Step 2, $\mathcal{F}_1$ is a complete metric space. By Step 3, $\mathcal{T}$ is a contraction mapping on

$\mathcal{F}_1$. By Contraction Mapping Theorem, $\mathcal{T}$ has a unique fixed point in $\mathcal{F}_1$. By construction, this is decreasing in noise trading intensity z_I. This concludes Step 5.

Step 6: unique interior equilibrium

The previous steps prove that in an interior equilibrium there exists a unique function λ_I that satisfies Eq. (27). By Lemma 1, given λ_I we can uniquely recover the capital allocation function δ as well as market L price efficiency λ_L. In an in interior equilibrium equilibrium $J_f(\theta) = J_L(\theta)$, so Eq. (7) gives a functional equation for J_L. Consider the mapping $\mathcal{T}_L : J \in B(\Xi \times Z) \to B(\Xi \times Z)$ given by

$$\mathcal{T}_L J(\xi, z_I) = -\left(\lambda_L(\xi, z_I) P^G + (1 - \lambda_L(\xi, z_I)) P^0\right) + \beta \left[V_L^G + \sum_{z_I' \in Z} \omega\left(z_I' \middle| z_I\right) J(C(\xi, z_I), z_I') \right]. \quad (41)$$

It is immediate that $\mathcal{T}_L$ satisfies Blackwell's sufficient conditions for a contraction on $B(\Xi \times Z)$. Hence, given λ_L, $\mathcal{T}_L$ has a unique fixed point $J_L \in B(\Xi \times Z)$ satisfying Eq. (7). Furthermore, in an in interior equilibrium $J_f(\theta) = J_I(\theta)$ by definition and $J_l(\theta) = J_S(\theta)$ as explained in the text and therefore Eqs. (6) and (10) give two functional equations for J_I and J_S. Given λ_I and J_f, the same argument as above shows that Eqs. (6) and (10) have a unique solution. This uniquely pins down J_f, J_l, J_I, J_L, J_E, J_S in an interior equilibrium and concludes the proof. ■

Appendix C: proofs for Section 5

Lemma 4 *When $\frac{z_I}{z_L} + 1 \geq 2\beta(1-q)$, the IC curve implicitly defines δ as an increasing function of ξ.*

Proof. Write the IC curve as $F(\delta, \xi) = 0$, where

$$F(\delta, \xi) = \frac{z_L - (1-\delta)\xi}{z_L} - \left(\frac{z_I - \delta\xi}{z_I}\right)\left[1 - \beta(1-q)\left(\frac{z_I - \delta\xi}{z_I}\right)\right].$$

We wish to show that $\frac{\partial F(\delta,\xi)}{\partial \delta} > 0$ and $\frac{\partial F(\delta,\xi)}{\partial \xi} < 0$. We have:

$$\frac{\partial F(\delta, \xi)}{\partial \xi} = \frac{(1-\delta)}{z_L} + \frac{\delta}{z_I} - 2\frac{\delta}{z_I}\beta(1-q)\left(\frac{z_I - \delta\xi}{z_I}\right) = \frac{1}{\xi}\left(\lambda_I - \lambda^L - 2\lambda_I\beta(1-q)(1-\lambda_I)\right).$$

Because $F(\delta, \xi) = 0$ requires $\lambda_I < \lambda_L$, then $\frac{\partial F(\delta,\xi)}{\partial \xi} < 0$. Furthermore,

$$\frac{\partial F(\delta, \xi)}{\partial \delta} = \frac{\xi}{z_L} + \frac{\xi}{z_I} - 2\frac{\xi}{z_I}\beta(1-q)\left(\frac{z_I - \delta\xi}{z_I}\right) = \frac{\xi}{z_I}\left(\frac{z_I}{z_L} + 1 - 2\beta(1-q)(1-\lambda_I)\right).$$

Clearly, $\frac{\partial F(\delta,\xi)}{\partial \delta} > 0$ if $\frac{z_I}{z_L} + 1 - 2\beta(1-q) \geq 0$. ■

Proof of Proposition 2: Suppose that Eq. (IC) is not satisfied. Then, it is one of the two cases: either everyone chooses market I or everyone chooses market L. In the former case, $\delta = 1$ and therefore $\lambda_L = 0$ and $\lambda_I \in (0, 1]$. However, we can show that there is no such equilibrium that satisfies Eq. (IC) because, for all $\lambda_I \in (0, 1]$

$$1 > (1-\lambda_I)\left(1-\beta\left(1-\lambda_I\right)(1-q)\right),$$

which implies $J_L(\xi) > J_I(\xi)$. In the latter case, we have $\delta = 0$ and therefore $\xi = 1$, $\lambda_I = 0$ and $\lambda_L = \min\{1, \frac{1}{z_L}\}$. Hence, $\delta = 0$ is an equilibrium if $J_L(1)|_{\lambda_L=\min\{1,\frac{1}{z_L}\}} \geq J_I(1)|_{\lambda_I=0}$ which is equivalent to

$$1-\min\{1, \frac{1}{z_L}\} \geq 1-\beta\left(1-q\right) \Leftrightarrow \beta\left(1-q\right) z_L \geq 1. \tag{42}$$

Next, we let $\beta\left(1-q\right) z_L < 1$, for which there is no corner equilibrium, and proceed to show that there exist either one or three interior equilibria. We define $\hat{\xi} = \delta\xi$ as the net mass of arbitrageurs who are investing in the illiquid market at time t. Likewise, we define $\hat{\delta} = \delta\xi + \pi$ as the total mass of investors who are investing in the illiquid market at time t. Instead of the original problem stated in terms of of δ and ξ, we can solve an equivalent problem in terms of $\hat{\delta}$ and $\hat{\xi}$. Using the definition of $\hat{\xi}$ and $\hat{\delta}$, we find

$$\xi = \hat{\xi} + 1 - \hat{\delta}, \quad \delta = \frac{\hat{\xi}}{\hat{\xi} + 1 - \hat{\delta}}, \quad \lambda_I = \frac{\hat{\xi}}{z_I}, \quad \lambda_L = \frac{1-\hat{\delta}}{z_L}. \tag{43}$$

Using Eq. (43), Eq. (CM) can be represented as

$$\hat{\delta} = \frac{\hat{\xi}}{q + (1-q)\frac{\hat{\xi}}{z_I}}. \tag{44}$$

Likewise, Eq. (IC) can be represented as

$$\frac{1-\hat{\delta}}{\bar{z}^L} - \frac{\hat{\xi}}{\bar{z}^I} = \beta(1-q)\left(1 - \frac{\hat{\xi}}{\bar{z}^I}\right)^2. \tag{45}$$

By substituting Eq. (44) into Eq. (45), we obtain

$$Q(\hat{\xi}) \equiv a_0 + a_1\hat{\xi} + a_2\hat{\xi}^2 + a_3\hat{\xi}^3 = 0,$$

where Q is a third degree polynomial with coefficients

$$\begin{aligned} a_0 &= q\,(z_I)^3\,(1-(1-q)\,z_L\beta) \\ a_1 &= -\,(z_I)^2\,(z_I+qz_L-(1-q)\,(1+(3q-1)z_L\beta)) \\ a_2 &= -z_I z_L(1-q)(1+(3q-2)\beta) \\ a_3 &= -(1-q)^2 z_L\beta \end{aligned}$$

Assumption 2 implies $\beta\,(1-q)\,z_L < 1$, which in turn implies $a_0 > 0$ and therefore $Q\,(0) > 0$. Using the fact that $z_I + z_L > 1$, we can verify that $Q\,(1) < 0$, which implies that Q has either one or three real roots in the $(0,1)$ interval. Each of these roots is an interior steady state equilibrium in which $\delta \in (0,1)$.

Now, we turn to the proof stability. We can rewrite Eqs. (44)-(45), and define functions $\phi, \psi : [0,1] \to \mathbb{R}$ such that

$$\phi(\hat{\xi},\hat{\delta}) = \frac{\hat{\xi}}{q+(1-q)\frac{\hat{\xi}}{\bar{z}^I}} \tag{46}$$

$$\psi(\hat{\xi},\hat{\delta}) = 1-\bar{z}^L\frac{\hat{\xi}}{\bar{z}^I}-\bar{z}^L\beta(1-q)\left(1-\frac{\hat{\xi}}{\bar{z}^I}\right)^2. \tag{47}$$

Because $\beta\,(1-q)\,z_L < 1$, we have $g(0) = \frac{1}{\bar{z}^L} - \beta(1-q) > f(0) = 0$. Thus, it has to be the case that $\phi'(\xi^s) > \psi'(\xi^s)$ at the smallest steady state $\xi^s \in (0,1)$. Then, this implies $|D+1| < |T|$ which implies that ξ^s is a saddle point.

It can be shown that D an T are always positive, and we can also show that $T > 3$ if $\bar{z}^I > \bar{z}^L$. In case $|D+1| > |T|$, this implies $D > 1$. Thus, the middle steady state is unstable (source) if $z_I > z_L$.

Finally, we prove part (ii) in the proposition. For $q = 1$ we have that $a_2 = a_3 = 0$, so Q has a unique root equal to $x^* = z_I/(z_I+z_L)$. For $\beta = 0$, we have that $a_0 > 0, a_2 < 0, a_3 = 0$ which implies that Q at most one root in the $[0,1]$ interval. For $q = 0$ we have that $a_0 = 0$ and Q has three roots x_1, x_2, x_3 equal to

$$\begin{aligned} x_1 &= 0 \\ x_2 &= \frac{z_I}{2\beta}\left(2\beta-1-\sqrt{1+\frac{4\beta}{z_L}\,(1-z_I-z_L)}\right) \\ x_3 &= \frac{z_I}{2\beta}\left(2\beta-1+\sqrt{1+\frac{4\beta}{z_L}\,(1-z_I-z_L)}\right) \end{aligned}$$

If $1 > \frac{3}{4}z_L + z_I$, then x_2, x_3 are real. It is immediate to see that $0 < x_2 < x_3 < 1$ for β sufficiently close to one. The claim in the proposition follows by continuity of the coefficients

a_0, a_1, a_2, a_3 in q and β and by continuous dependence of the roots of a polynomial on its coefficients. ■

Appendix D: Auxiliary lemmas

Lemma 5 *ξ is bounded from below by*

$$\underline{\xi} = \max \left\{ 1 - \bar{z}_I \frac{1-\sqrt{q}}{1+\sqrt{q}}, 0 \right\}.$$

Proof. Let $\underline{\xi} = q + \varepsilon$ be such that $C(\xi, z_I) \geq \underline{\xi}$ for all $\xi \geq \underline{\xi}$ and $z_I \in Z$. It is sufficient that, for all $\lambda \in [0,1]$ and $z_I \in Z$,

$$1 - (1-q)(1-\lambda)(1-(q+\varepsilon)+z_I\lambda) \geq q + \varepsilon,$$

or equivalently

$$\varepsilon \leq (1-q) \frac{(1-(1-\lambda)(1-q+z_I\lambda))}{1-(1-q)(1-\lambda)}.$$

Notice that the r.h.s. is convex in λ and minimized at $\lambda = \frac{\sqrt{q}}{1+\sqrt{q}}$, so

$$\min_{\lambda, z_I \in Z} q + (1-q) \frac{(1-(1-\lambda)(1-q+z_I\lambda))}{1-(1-q)(1-\lambda)} = \min_{z_I \in Z} 1 - z_I \frac{1-\sqrt{q}}{1+\sqrt{q}}. \tag{48}$$

■

Lemma 6 *We have*

$$|(1-\lambda(\xi_1, z_I))(C(\xi_2, z_I) - C(\xi_1, z_I))| \leq f\left(\bar{\lambda}_\xi, z_I\right) |\xi_2 - \xi_1|,$$

where the function f is as in Definition 2, Eq. (29).

Proof. Start from:

$$\begin{aligned} C(\xi_2, z_I) - C(\xi_1, z_I) &= (1-q) \left[\begin{array}{c} (\xi_2 - \xi_1) + (1-q)(1-z_I)(\lambda(\xi_2, z_I) - \lambda(\xi_1, z_I)) - (\xi_2 - \xi_1)\lambda(\xi_2, z_I) \\ -\xi_1(\lambda(\xi_2, z_I) - \lambda(\xi_1, z_I)) + z_I \left(\lambda(\xi_2, z_I)^2 - \lambda(\xi_1, z_I)^2\right) \end{array} \right] \\ &= (1-q) \left[\begin{array}{c} (1-\lambda(\xi_2, z_I))(\xi_2 - \xi_1) + \\ (\lambda(\xi_2, z_I) - \lambda(\xi_1, z_I))(1 - \xi_1 - z_I(1 - \lambda(\xi_1, z_I) - \lambda(\xi_2, z_I))) \end{array} \right] \end{aligned}$$

Lipschitz continuity and monotonicity of λ in ξ imply that there exists a value $\lambda_\xi \in [0, \overline{\lambda}_\xi]$ such that

$$\lambda(\xi_2, z_I) - \lambda(\xi_1, z_I) = \lambda_\xi (\xi_2 - \xi_1), \tag{49}$$

and therefore

$$\begin{aligned} C(\xi_2, z_I) - C(\xi_1, z_I) &= (1-q)\left[(1-\lambda(\xi_2, z_I)) + \lambda_\xi\left(1-\xi_1 - z_I\left(1-\lambda(\xi_1, z_I) - \lambda(\xi_2, z_I)\right)\right)\right](\xi_2-\xi_1) \\ &= (1-q)\left[(1-\lambda(\xi_2, z_I))(1-\lambda_\xi z_I) + \lambda_\xi\left(1-\xi_1 + z_I\lambda(\xi_1, z_I)\right)\right](\xi_2-\xi_1). \end{aligned}$$

Hence, we can write

$$\begin{aligned} & (1-\lambda(\xi_1, z_I))\left(C(\xi_2, z_I) - C(\xi_1, z_I)\right) \\ = \ & \left[(1-q)(1-\lambda(\xi_1, z_I))(1-\lambda(\xi_2, z_I))(1-\lambda_\xi z_I) + \lambda_\xi(1-q)(1-\lambda(\xi_1, z_I))(1-\xi_1 + z_I\lambda(\xi_1, z_I))\right](\xi_2-\xi_1) \\ = \ & \left[(1-q)(1-\lambda(\xi_1, z_I))(1-\lambda(\xi_2, z_I))(1-\lambda_\xi z_I) + \lambda_\xi\left(1-C(\xi_1, z_I)\right)\right](\xi_2-\xi_1), \end{aligned} \tag{50}$$

where in the second line we make use of Eq. (28). Using the fact that $\lambda_\xi \in [0, \bar{\lambda}_\xi]$ and $\lambda(\xi, z_I) \in [0,1]$ and $C(\xi_1, z_I) \in [\underline{\xi}, 1]$, it is easy to verify that

$$\begin{aligned} & \left|(1-q)(1-\lambda(\xi_1, z_I))(1-\lambda(\xi_2, z_I))(1-\lambda_\xi z_I) + \lambda_\xi\left(1-C(\xi_1, z_I)\right)\right| \\ \leq \ & \max\left\{(1-q), (1-q)\left(z_I\bar{\lambda}_\xi - 1\right), \left(1-\underline{\xi}\right)\bar{\lambda}_\xi, \left(1-\underline{\xi}\right)\bar{\lambda}_\xi + (1-q)\left(1-z_I\bar{\lambda}_\xi\right)\right\} = f\left(\bar{\lambda}_\xi, z_I\right). \end{aligned}$$

■

Lemma 7 *Under Assumption 1, $\mathcal{T}\lambda$ is Lipschitz continuous of modulus $\bar{\lambda}_\xi$ in ξ for every $\lambda \in \mathcal{F}_0$.*

Proof. Take $\lambda \in \mathcal{F}_0$. We decompose

$$\mathcal{T}\lambda(\xi_2, z_I) - \mathcal{T}\lambda(\xi_1, z_I) = T_1 + T_2 + T_3,$$

where

$$\begin{aligned} T_1 &= A(z_I)(\xi_2-\xi_1); \\ T_2 &= B(z_I)\left[\lambda(\xi_2, z_I) - \lambda(\xi_1, z_I)\right]\left(1 - \sum_{z_I' \in Z}\omega\left(z_I' \middle| z_I\right)\lambda(C(\xi_2, z_I), z_I')\right); \\ T_3 &= B(z_I)(1-\lambda(\xi_1, z_I))\sum_{z_I' \in Z}\omega\left(z_I' \middle| z_I\right)\left(\lambda(C(\xi_2, z_I), z_I') - \lambda(C(\xi_1, z_I), z_I')\right). \end{aligned}$$

First, it is immediate that $|T_1| \leq A(z_I)|\xi_2 - \xi_1|$. Second, Lipschitz continuity and monotonicity of λ in ξ imply that there exists $\lambda_{\xi_0} \in [0, \bar{\lambda}_\xi]$ such that $\lambda(\xi_2, z_I) - \lambda(\xi_1, z_I) = \lambda_{\xi_0}(\xi_2 - \xi_1)$. Because $\sum_{z_I' \in Z}\omega\left(z_I' \middle| z_I\right)\lambda(C(\xi_2, z_I), z_I') \leq 1$, we have

$$|T_2| \leq B(z_I)\bar{\lambda}_\xi|\xi_2 - \xi_1|.$$

Again by Lipschitz continuity and monotonicity of λ in ξ, there exist $\lambda_{\xi_1} \in [0, \bar{\lambda}_\xi]$ such that

$$T_3 = B(z_I)(1 - \lambda(\xi_1, z_I))\left(C(\xi_2, z_I) - C(\xi_1, z_I)\right)\lambda_{\xi_1},$$

and therefore,

$$|T_3| \leq B(z_I)\left|(1 - \lambda(\xi_1, z_I))\left(C(\xi_2, z_I) - C(\xi_1, z_I)\right)\right|\bar{\lambda}_\xi.$$

By Lemma 6, the previous inequality can be written as

$$|T_3| \leq B(z_I) f\left(\bar{\lambda}_\xi, z_I\right)\bar{\lambda}_\xi |\xi_2 - \xi_1|.$$

Summing up terms, we get

$$|\mathcal{T}\lambda(\xi_2, z_I) - \mathcal{T}\lambda(\xi_1, z_I)| \leq \left(A(z_I) + B(z_I)\left[1 + f\left(\bar{\lambda}_\xi, z_I\right)\right]\bar{\lambda}_\xi\right)|\xi_2 - \xi_1|.$$

Taking the maximum of the r.h.s. over z_I values yields that $\mathcal{T}\lambda$ is Lipschitz continuous of modulus $\bar{\lambda}_\mathcal{T}$ in ξ, where

$$\bar{\lambda}_\mathcal{T} = \max_{z_I \in Z} A(z_I) + B(z_I)\left[1 + f\left(\bar{\lambda}_\xi, z_I\right)\right]\bar{\lambda}_\xi$$

and the function f is as in Definition 2. Under Assumption 1, Definition 3 implies $\bar{\lambda}_\xi \geq \bar{\lambda}_T$. This concludes the proof. ■

Lemma 8 *Under Assumption 1, $\mathcal{T}\lambda$ is monotone increasing in ξ for every $\lambda \in \mathcal{F}_0$.*

Proof. Take $\lambda \in \mathcal{F}_0$ and let $\xi_2 > \xi_1$. By the proof of Lemma 7, there exist $\lambda_{\xi_0}, \lambda_{\xi_1} \in [0, \bar{\lambda}_\xi]$ such that

$$\mathcal{T}\lambda(\xi_2, z_I) - \mathcal{T}\lambda(\xi_1, z_I) = \left\{ A(z_I) + B(z_I)\left[\begin{array}{c} \left(1 - \sum_{z_I' \in Z} \omega\left(z_I' \middle| z_I\right)\lambda(C(\xi_2, z_I), z_I')\right)\lambda_{\xi_0} \\ + \frac{(1-\lambda(\xi_1, z_I))(C(\xi_2, z_I) - C(\xi_1, z_I))}{(\xi_2 - \xi_1)}\lambda_{\xi_1} \end{array} \right] \right\}(\xi_2 - \xi_1).$$

Hence, $T\lambda$ is increasing in ξ if

$$A(z_I) + B(z_I)\left[\frac{(1 - \lambda(\xi_1, z_I))\left(C(\xi_2, z_I) - C(\xi_1, z_I)\right)}{(\xi_2 - \xi_1)}\bar{\lambda}_\xi\right] \geq 0.$$

Using Eq. (50) in the proof of Lemma 6, there exists some $\lambda_\xi \in [0, \bar{\lambda}_\xi]$ such that the above inequality is equivalent to

$$A(z_I) + B(z_I)\left[(1 - q)(1 - \lambda(\xi_1, z_I))(1 - \lambda(\xi_2, z_I))(1 - \lambda_\xi z_I) + \lambda_\xi\left(1 - C\left(\xi_1, z_I\right)\right)\right]\bar{\lambda}_\xi \geq 0,$$

which is satisfied if

$$\min_{\lambda_1,\lambda_2\in[0,1],x\in[0,1],\lambda_\xi\in[0,\bar{\lambda}_\xi]} A(z_I)+B(z_I)\left[(1-q)\left(1-\lambda_1\right)\left(1-\lambda_2\right)\left(1-\lambda_\xi z_I\right)+\lambda_\xi\left(1-x\right)\right]\bar{\lambda}_\xi \geq 0. \tag{51}$$

For $\bar{\lambda}_\xi \leq 1/z_I$, it is immediate that the l.h.s. of (51) is positive. For $\bar{\lambda}_\xi > 1/z_I$, the l.h.s. of (51) is positive if

$$A(z_I)+B(z_I)(1-q)\left(1-\bar{\lambda}_\xi z_I\right)\bar{\lambda}_\xi \geq 0,$$

or equivalently, if

$$\bar{\lambda}_\xi \leq \frac{1+\sqrt{1+\frac{4z_I}{\beta(1-q)^2 z_L}}}{2z_I}.$$

Taking the minimum of the r.h.s. over z_I values in Z yields the expression for $\hat{\lambda}_\xi$ in Definition 3. Under Assumption 1, Definition 3 implies $\bar{\lambda}_\xi \leq \hat{\lambda}_\xi$. This concludes the proof. ■

Lemma 9 *Under Assumption 1, the mapping $\mathcal{T}$ is a contraction on $\mathcal{F}_0$.*

Proof. Take $\lambda_1, \lambda_2 \in \mathcal{F}_0$. We decompose

$$\mathcal{T}\lambda_2(\xi,z_I)-\mathcal{T}\lambda_1(\xi,z_I)=\mathcal{T}_1+\mathcal{T}_2+\mathcal{T}_3,$$

where

$$\mathcal{T}_1 = B(z_I)\left[\lambda_2(\xi,z_I)-\lambda_1(\xi,z_I)\right]\left(1-\sum_{z_I'\in Z}\omega\left(z_I'\middle|\,z_I\right)\lambda_2(C_2(\xi,z_I),z_I')\right);$$

$$\mathcal{T}_2 = B(z_I)(1-\lambda_1(\xi,z_I))\sum_{z_I'\in Z}\omega\left(z_I'\middle|\,z_I\right)\left(\lambda_2(C_1(\xi,z_I),z_I')-\lambda_1(C_1(\xi,z_I),z_I')\right);$$

$$\mathcal{T}_3 = B(z_I)(1-\lambda_1(\xi,z_I))\sum_{z_I'\in Z}\omega\left(z_I'\middle|\,z_I\right)\left(\lambda_2(C_2(\xi,z_I),z_I')-\lambda_2(C_1(\xi,z_I),z_I')\right)$$

First, we have

$$|\mathcal{T}_1| \leq B(z_I)\left(1-\sum_{z_I'\in Z}\omega\left(z_I'\middle|\,z_I\right)\lambda_2(C(\xi,z_I),z_I')\right)||\lambda_2-\lambda_1|| \leq B(z_I)||\lambda_2-\lambda_1||.$$

Second,

$$|\mathcal{T}_2| \leq B(z_I)||\lambda_2-\lambda_1||$$

Third, using Eq. (28) we have

$$C_2(\xi, z_I) - C_1(\xi, z_I) = (1-q)\left(\lambda_2(\xi, z_I) - \lambda_1(\xi, z_I)\right)\left[1 - \xi - z_I\left(1 - \lambda_1(\xi, z_I) - \lambda_2(\xi, z_I)\right)\right],$$

and therefore

$$|\mathcal{T}_3| \leq B(z_I)\bar{\lambda}_\xi(1-q)|(1-\lambda_1(\xi, z_I))\left(1-\xi - z_I(1-\lambda_1(\xi, z_I) - \lambda_2(\xi, z_I))\right)|||\lambda_2 - \lambda_1||.$$

Let $\Gamma(z_I)$ be the value

$$\Gamma(z_I) = \max_{\lambda_1, \lambda_2 \in [0,1], \xi \in [\underline{\xi}, 1]} \left|(1-\lambda_1)\left(1 - \xi - z_I(1-\lambda_1-\lambda_2)\right)\right|.$$

It is immediate to verify that $\Gamma(z_I)$ is as in Eq. (31) in Definition 2. Therefore,

$$|\mathcal{T}_3| \leq B(z_I)\bar{\lambda}_\xi(1-q)\Gamma(z_I)\,||\lambda_2 - \lambda_1||.$$

Summing up terms, we have

$$|\mathcal{T}\lambda_2(\xi, z_I) - \mathcal{T}\lambda_1(\xi, z_I)| \leq B(z_I)(2 + \bar{\lambda}_\xi(1-q)\Gamma(z_I))||\lambda_2 - \lambda_1||$$

Therefore, $\mathcal{T}$ is a contraction mapping if for all $z_I \in Z$

$$B(z_I)(2 + \bar{\lambda}_\xi(1-q)\Gamma(z_I)) < 1,$$

or equivalently, if

$$\bar{\lambda}_\xi < \bar{\lambda}_\xi^* = \min_{z_I \in Z} \frac{1/B(z_I) - 2}{(1-q)\Gamma(z_I)}.$$

Under Assumption 1, $\bar{\lambda}_\xi < \bar{\lambda}_\xi^*$ by Definition 3. This concludes the proof. ∎

Lemma 10 *Under Assumptions 1, 2 and 3, $T\lambda$ is decreasing in z_I for all $\lambda \in \mathcal{F}_1$.*

Proof. The difference of $\mathcal{T}\lambda(\xi, z_I)$ with respect to z_I is given by

$$\begin{aligned}
&\mathcal{T}\lambda(\xi, z_I^n) - \mathcal{T}\lambda(\xi, z_I^m) \\
&= (A(z_I^n) - A(z_I^m)) \left(\xi - \beta\,(1-q)\,z_L(1-\lambda(\xi, z_I^m)) \left(1 - \sum_{z_I' \in Z} \omega\left(z_I' \middle| z_I^m\right) \lambda(C(\xi, z_I^m), z_I')\right)\right) \\
&+ B(z_I^n)(\lambda(\xi, z_I^n) - \lambda(\xi, z_I^m)) \left(1 - \sum_{z_I' \in Z} \omega\left(z_I' \middle| z_I^m\right) \lambda(C(\xi, z_I^m), z_I')\right) \\
&+ B(z_I^n)(1-\lambda(\xi, z_I^n)) \sum_{z_I' \in Z} \omega\left(z_I' \middle| z_I^n\right) \left[\lambda(C(\xi, z_I^n), z_I') - \lambda(C(\xi, z_I^m), z_I')\right] \\
&+ B(z_I^n)(1-\lambda(\xi, z_I^n)) \sum_{z_I' \in Z} \left(\omega\left(z_I' \middle| z_I^n\right) - \omega\left(z_I' \middle| z_I^m\right)\right) \lambda(C(\xi, z_I^m), z_I').
\end{aligned}$$

We can simplify each line in the expression above as follows. First, using the definitions of A and B we can write

$$(A(z_I^n) - A(z_I^m))\,(\xi - \beta\,(1-q)\,z_L) = A(z_I^n)\,(A(z_I^m)\xi - B(z_I^m))\,(z_I^m - z_I^n). \tag{52}$$

Second, since λ is decreasing in z_I, then, for any $\xi \in \Xi$, $z_I^m, z_I^n \in Z$ there exists some $\lambda_z \in \left[0, \overline{\lambda}_z\right]$, which depends on ξ, z_I^n, z_I^m, such that,

$$\lambda(\xi, z_I^n) - \lambda(\xi, z_I^m) = \lambda_z\,(z_I^m - z_I^n)\,. \tag{53}$$

Third, because λ is increasing and Lipschitz in ξ with modulus $\overline{\lambda}_\xi$, there exists some $\lambda_\xi \in \left[0, \overline{\lambda}_\xi\right]$, which depends on ξ, z_I^n, z_I^m, z_I', such that

$$\lambda(C(\xi, z_I^n), z_I') - \lambda(C(\xi, z_I^m), z_I') = \lambda_\xi\,(C(\xi, z_I^n) - C(\xi, z_I^m))\,. \tag{54}$$

Fourth, because λ is decreasing in z_I, we have[18]

$$\alpha\,(\lambda(\xi, \overline{z}_I) - \lambda(\xi, \underline{z}_I)) \leq \sum_{z_I' \in Z} \left(\omega\left(z_I' \middle| z_I^n\right) - \omega\left(z_I' \middle| z_I^m\right)\right) \lambda(\xi, z_I') \leq \alpha\,(\lambda(\xi, \underline{z}_I) - \lambda(\xi, \overline{z}_I)\,)\;\forall \xi \in \Xi \tag{55}$$

and furthermore, because the rate of change of λ in z_I is bounded by $\overline{\lambda}_z$ and $M\,|z_I^n - z_I^m| \geq \overline{z}_I - \underline{z}_I$ by definition, then

$$\lambda(C(\xi, z_I^m), \underline{z}_I) - \lambda(C(\xi, z_I^m), \overline{z}_I) \leq \overline{\lambda}_z\,(\overline{z}_I - \underline{z}_I) \leq \overline{\lambda}_z M\,|z_I^n - z_I^m|\,. \tag{56}$$

[18]We remind the definition $\alpha \equiv \max_{z_I^n, z_I^m, z_I' \in Z} \left|\omega\left(z_I' \middle| z_I^m\right) - \omega\left(z_I' \middle| z_I^n\right)\right|$.

Hence, (55) and (56) imply

$$\sum_{z_I'\in Z}\left(\omega\left(z_I'\middle|\, z_I^n\right)-\omega\left(z_I'\middle|\, z_I^m\right)\right)\lambda(C(\xi,z_I^m),z_I')\in\left[-\alpha M\overline{\lambda}_z,\alpha M\overline{\lambda}_z\right]. \tag{57}$$

Using (52)-(57) we can rewrite the difference of $\mathcal{T}\lambda(\xi,z_I)$ with respect to z_I as

$$\begin{aligned}&\mathcal{T}\lambda(\xi,z_I^n)-\mathcal{T}\lambda(\xi,z_I^m)\\&=\left\{\begin{array}{c}A(z_I^n)\left(A(z_I^m)\xi-B(z_I^m)(1-\lambda(\xi,z_I^m))\right)\left(1-\sum_{z_I'\in Z}\omega\left(z_I'\middle|\, z_I^m\right)\lambda(C(\xi,z_I^m),z_I')\right)\\+B(z_I^n)\left[\lambda_z\left(1-\sum_{z_I'\in Z}\omega\left(z_I'\middle|\, z_I^m\right)\lambda(C(\xi,z_I^m),z_I')\right)+(1-\lambda(\xi,z_I^n))\left[\frac{C(\xi,z_I^n)-C(\xi,z_I^m)}{z_I^m-z_I^n}\lambda_\xi+\chi\right]\right]\end{array}\right\}(z_I^m-z_I^n)\end{aligned} \tag{58}$$

for some $\lambda_z\in\left[0,\overline{\lambda}_z\right]$, $\lambda_\xi\in\left[0,\overline{\lambda}_\xi\right]$ and $\chi\in\left[-\alpha M\overline{\lambda}_z,\alpha M\overline{\lambda}_z\right]$.

The difference of $C(\xi,z_I)$ with respect to z_I can be written as

$$\begin{aligned}&C(\xi,z_I^n)-C(\xi,z_I^m)\\=\;&(1-q)\left[(\lambda(\xi,z_I^n)-\lambda(\xi,z_I^m))[1-\xi-z_I^n(1-\lambda(\xi,z_I^n)-\lambda(\xi,z_I^m))]+\lambda(\xi,z_I^m)(1-\lambda(\xi,z_I^m))\,(z_I^m-z_I^n)\right]\\=\;&(1-q)\left[\lambda_z[1-\xi-z_I^n(1-\lambda(\xi,z_I^n)-\lambda(\xi,z_I^m))]+\lambda(\xi,z_I^m)(1-\lambda(\xi,z_I^m))\right](z_I^m-z_I^n),\end{aligned} \tag{59}$$

where the second line makes use of Eq. (53). Using Eq. (59) we can write Eq. (58) as

$$\begin{aligned}&\mathcal{T}\lambda(\xi,z_I^n)-\mathcal{T}\lambda(\xi,z_I^m)\\=\;&A(z_I^n)A(z_I^m)\xi(z_I^m-z_I^n)\\&+\left(B(z_I^n)\lambda_z-A(z_I^n)B(z_I^m)(1-\lambda(\xi,z_I^m))\right)\left(1-\sum_{z_I'\in Z}\omega\left(z_I'\middle|\, z_I^m\right)\lambda(C(\xi,z_I^m),z_I')\right)(z_I^m-z_I^n)\\&+B(z_I^n)(1-\lambda(\xi,z_I^n))\left[(1-q)\left(\begin{array}{c}\lambda_z[1-\xi-z_I^n(1-\lambda(\xi,z_I^n)-\lambda(\xi,z_I^m))]\\+\lambda(\xi,z_I^m)(1-\lambda(\xi,z_I^m))\end{array}\right)\lambda_\xi+\chi\right](z_I^m-z_I^n).\end{aligned}$$

Hence, we obtain that $\mathcal{T}\lambda$ is decreasing in z_I if for all $\lambda_1,\lambda_2,\lambda_3\in[0,1],\xi\in[\underline{\xi},1],\lambda_z\in[0,\overline{\lambda}_z]$, $\lambda_\xi\in[0,\overline{\lambda}_\xi]$, we have

$$\begin{aligned}&A(z_I^n)A(z_I^m)\xi+\left(B(z_I^n)\lambda_z-A(z_I^n)B(z_I^m)(1-\lambda_2)\right)(1-\lambda_3)\\&\times B(z_I^n)(1-\lambda_1)\left[(1-q)\left[\lambda_z[1-\xi-z_I^n(1-\lambda_1-\lambda_2)]+\lambda_2\,(1-\lambda_2)\right]\lambda_\xi-\alpha M\overline{\lambda}_z\right]\geq 0\end{aligned} \tag{60}$$

It is easy to verify that the l.h.s. of (60) is minimized at $\lambda_1=\lambda_2=0$ for all $\lambda_3\in[0,1],\xi\in$

$[\underline{\xi}, 1], \lambda_z \in [0, \overline{\lambda}_z],\ \lambda_\xi \in [0, \overline{\lambda}_\xi]$, which leaves

$$A(z_I^n)A(z_I^m)\xi + (B(z_I^n)\lambda_z - A(z_I^n)B(z_I^m))(1-\lambda_3) + B(z_I^n)\left[(1-q)\lambda_z(1-\xi-z_I^n)\lambda_\xi - \alpha M\overline{\lambda}_z\right] \geq 0. \tag{61}$$

Next, it is immediate to check that the l.h.s. of (61) is minimized at $\xi = \underline{\xi}$ for all $\lambda_3 \in [0,1], \lambda_z \in [0, \overline{\lambda}_z],\ \lambda_\xi \in [0, \overline{\lambda}_\xi]$ if the following condition on $\overline{\lambda}_\xi$ holds:

$$\overline{\lambda}_\xi \leq \frac{1}{(\bar{z}_I + z_L)\,\beta\,(1-q)^2\, z_L \overline{\lambda}_z}. \tag{62}$$

Hence, if (62) holds, (61) is satisfied if

$$A(z_I^n)A(z_I^m)\underline{\xi} + (B(z_I^n)\lambda_z - A(z_I^n)B(z_I^m))(1-\lambda_3) + B(z_I^n)\left[(1-q)\left(1-\underline{\xi}-z_I^n\right)\lambda_z\lambda_\xi - \alpha M\overline{\lambda}_z\right] \geq 0.$$

Using the definitions of A, B and $\underline{\xi} \geq 1 - \bar{z}_I \frac{1-\sqrt{q}}{1+\sqrt{q}}$, the above inequality can be rearranged as

$$\frac{1}{z_L + z_I^m}\left(1 - \bar{z}_I \frac{1-\sqrt{q}}{1+\sqrt{q}}\right) + \beta\,(1-q)\,z_L\left[\left(\lambda_z - \frac{1}{z_L + z_I^m}\right)(1-\lambda_3) - (1-q)\lambda_z \bar{z}_I \frac{2\sqrt{q}}{1+\sqrt{q}}\lambda_\xi - \alpha M\overline{\lambda}_z\right] \geq 0.$$

Because the l.h.s. is linear in λ_z, it is minimized either at $\lambda_z = 0$ or $\lambda_z = \overline{\lambda}_z$. At $\lambda_z = 0$ the l.h.s. bounded from below by the value

$$\frac{1}{z_L + z_I^m}\left(1 - \bar{z}_I \frac{1-\sqrt{q}}{1+\sqrt{q}}\right) - \beta\,(1-q)\,z_L\left(\frac{1}{z_L + z_I^m} + \alpha M\overline{\lambda}_z\right), \tag{63}$$

and at $\lambda_z = \overline{\lambda}_z$ the l.h.s. is equal to

$$\frac{1}{z_L + z_I^m}\left(1 - \bar{z}_I \frac{1-\sqrt{q}}{1+\sqrt{q}}\right) + \beta\,(1-q)\,z_L\left[\left(\overline{\lambda}_z - \frac{1}{z_L + z_I^m}\right)(1-\lambda_3) - \overline{\lambda}_z\left((1-q)\bar{z}_I \frac{2\sqrt{q}}{1+\sqrt{q}}\overline{\lambda}_\xi + \alpha M\right)\right]. \tag{64}$$

It is immediate that (63) is positive if

$$\overline{\lambda}_z \leq \frac{1 - \bar{z}_I \frac{1-\sqrt{q}}{1+\sqrt{q}} - \beta\,(1-q)\,z_L}{(z_L + \bar{z}_I)\,\beta\,(1-q)\,z_L \alpha M}, \tag{65}$$

which is satisfied under Assumptions 2 and 3. For (64), we see that either $\overline{\lambda}_z \leq \frac{1}{z_L + z_I^m}$, in which case (64) is minimized at

$$\frac{1}{z_L + z_I^m}\left[1 - \bar{z}_I \frac{1-\sqrt{q}}{1+\sqrt{q}} - \beta\,(1-q)\,z_L + \beta\,(1-q)\,z_L\left(1 - \max\left\{(1-q)\bar{z}_I \frac{2\sqrt{q}}{1+\sqrt{q}}\overline{\lambda}_\xi + \alpha M, 1\right\}\right)\right], \tag{66}$$

which is positive under Assumption 2 if

$$\overline{\lambda}_\xi \leq \frac{1-\alpha M}{(1-q)\bar{z}_I \frac{2\sqrt{q}}{1+\sqrt{q}}}, \tag{67}$$

or $\overline{\lambda}_z > \frac{1}{z_L + z_I^m}$, in which case (64) is minimized at

$$\frac{1}{z_L+\bar{z}_I}\left[1-\bar{z}_I\frac{1-\sqrt{q}}{1+\sqrt{q}}-\beta(1-q)z_L+\beta(1-q)z_L\left(1-(z_L+\bar{z}_I)\overline{\lambda}_z\left((1-q)\bar{z}_I\frac{2\sqrt{q}}{1+\sqrt{q}}\overline{\lambda}_\xi+\alpha M\right)\right)\right]. \tag{68}$$

Under Assumption 2, (68) is positive if

$$\overline{\lambda}_\xi \leq \frac{1}{(1-q)\bar{z}_I\frac{2\sqrt{q}}{1+\sqrt{q}}}\left(\frac{1}{(z_L+\bar{z}_I)\overline{\lambda}_z}-\alpha M\right). \tag{69}$$

Putting together the bounds (62), (67) and (69) gives (39). ■

Lemma 11 *Under Assumptions 1, 2 and 3, the rate of change of $\mathcal{T}\lambda$ in z_I is bounded by $\overline{\lambda}_z$ for all $\lambda \in \mathcal{F}_1$.*

Proof. We bound the rate of change of $\mathcal{T}\lambda$ in z. Using Eq. (59) and the definition of C in Eq. (28) we compute

$$\begin{aligned}
& (1-\lambda(\xi,z_I^n))\left[C(\xi,z_I^n)-C(\xi,z_I^m)\right] \\
= \ & (1-q)\left[\lambda_z(1-\lambda(\xi,z_I^n))[1-\xi-z_I^n(1-\lambda(\xi,z_I^n)-\lambda(\xi,z_I^m))]+\lambda(\xi,z_I^m)(1-\lambda(\xi,z_I^n))(1-\lambda(\xi,z_I^m))\right](z_I^m-z_I^n) \\
= \ & (1-q)\left[\lambda_z(1-\lambda(\xi,z_I^n))[1-\xi+z_I^n\lambda(\xi,z_I^n)]+[\lambda(\xi,z_I^m)-z_I^n\lambda_z](1-\lambda(\xi,z_I^n))(1-\lambda(\xi,z_I^m))\right](z_I^m-z_I^n) \\
= \ & \left[\lambda_z(1-C(\xi,z_I^n))+(1-q)[\lambda(\xi,z_I^m)-z_I^n\lambda_z](1-\lambda(\xi,z_I^n))(1-\lambda(\xi,z_I^m))\right](z_I^m-z_I^n).
\end{aligned} \tag{70}$$

Using (58) and (70), we have:

$$\begin{aligned}
& |\mathcal{T}\lambda(\xi,z_I^n)-\mathcal{T}\lambda(\xi,z_I^m)| \\
\leq \ & A(z_I^n)A(z_I^m)|z_I^m-z_I^n|+B(z_I^n)\lambda_z\left(1-\sum_{z_I'\in Z}\omega\left(z_I'\middle|z_I^m\right)\lambda(C(\xi,z_I^m),z_I')\right)|z_I^m-z_I^n| \\
& +B(z_I^n)\left|\left[\lambda_z(1-C(\xi,z_I^n))+(1-q)[\lambda(\xi,z_I^m)-z_I^n\lambda_z](1-\lambda(\xi,z_I^n))(1-\lambda(\xi,z_I^m))\right]\lambda_\xi\right||z_I^m-z_I^n| \\
& +B(z_I^n)(1-\lambda(\xi,z_I^n))\overline{\lambda}_z\alpha M|z_I^m-z_I^n| \\
\leq \ & A(z_I^n)A(z_I^m)|z_I^m-z_I^n| \\
& +B(z_I^n)\left[\overline{\lambda}_z(1+\alpha M)+\left|\begin{array}{c}\lambda_z(1-C(\xi,z_I^n)) \\ +(1-q)[\lambda(\xi,z_I^m)-z_I^n\lambda_z](1-\lambda(\xi,z_I^n))(1-\lambda(\xi,z_I^m))\end{array}\right|\overline{\lambda}_\xi\right]|z_I^m-z_I^n|
\end{aligned}$$

Notice that

$$\begin{aligned}
& \left|\lambda_z(1-C(\xi,z_I^n)) + (1-q)\left[\lambda(\xi,z_I^m) - z_I^n\lambda_z\right](1-\lambda(\xi,z_I^n))(1-\lambda(\xi,z_I^m))\right| \\
\leq \ & \max\left\{1-q, (1-q)\,\overline{\lambda}_z z_I^n, \overline{\lambda}_z\left(1-\underline{\xi}\right), \overline{\lambda}_z\left(1-\underline{\xi}\right) + (1-q)\left(1-z_I^n\overline{\lambda}_z\right)\right\} = g\left(\overline{\lambda}_z, z_I^n\right)
\end{aligned}$$

Therefore, we can write

$$|\mathcal{T}\lambda(\xi,z_I^n) - \mathcal{T}\lambda(\xi,z_I^m)| \leq \left[A(z_I^n)A(z_I^m) + B(z_I^n)\left(\overline{\lambda}_z\left(1+\alpha M\right) + g\left(\overline{\lambda}_z, z_I^n\right)\overline{\lambda}_\xi\right)\right]|z_I^m - z_I^n|\,.$$

Taking the maximum of this bound, we obtain that the rate of change of $\mathcal{T}\lambda$ in z_I is bounded by $\overline{\lambda}_\zeta$, which we define as

$$\bar{\lambda}_\zeta = \max_{z_I^n, z_I^m \in Z, n\neq m} A(z_I^n)A(z_I^m) + B(z_I^n)\left[\overline{\lambda}_z\left(1+\alpha M\right) + g\left(\overline{\lambda}_z, z_I^n\right)\overline{\lambda}_\xi\right]$$

Under Assumption 3, Definition 4 implies $\bar{\lambda}_z \leq \overline{\lambda}_\zeta$. This concludes the proof. ■

Lemma 12 *Consider the following dynamic system of x and y:*

$$\begin{aligned}
\dot{u} &= f(u,v) \\
\dot{v} &= g(u,v)
\end{aligned}$$

where f and g are continuous and differentiable. If there are three steady-state equilibria and two extreme steady state equilibria are stable, the middle steady state equilibrium is either a source (unstable) or a sink.

Proof. The steady state solution (u^*, v^*) solves

$$u = f(u,v) \tag{71}$$

$$v = g(u,v), \tag{72}$$

and it also has to satisfy

$$du = f_u(u^*,v^*)du + f_v(u^*,v^*)dv \tag{73}$$

$$dv = g_u(u^*,v^*)du + g_v(u^*,v^*)dv \tag{74}$$

This implies

$$0 = [f_u(u^*,v^*) - 1]du + f_v(u^*,v^*)dv \tag{75}$$

$$0 = g_u(u^*,v^*)du + [g_v(u^*,v^*) - 1]dv \tag{76}$$

Or equivalently,

$$\left.\frac{dv}{du}\right|_{u=f(u,v)} = -\frac{f_u(u^*,v^*)-1}{f_v(u^*,v^*)} \tag{77}$$

$$\left.\frac{dv}{du}\right|_{v=g(u,v)} = -\frac{g_u(u^*,v^*)}{g_v(u^*,v^*)-1} \tag{78}$$

Now, suppose that there are three steady state equilibria. Because f and g are continuous, it has to be the case that the middle equilibrium has an opposite inequality on the slopes from the extreme ones. That is, if $\left.\frac{dv}{du}\right|_{u=f(u,v)} > \left.\frac{dv}{du}\right|_{v=g(u,v)}$ for the extreme equilibria, the middle one should have $\left.\frac{dv}{du}\right|_{u=f(u,v)} < \left.\frac{dv}{du}\right|_{v=g(u,v)}$, and vice versa. Equivalently, if $|D+1| < |T|$ for the extreme ones, $|D+1| > |T|$ for the middle one, and vice versa where

$$T = f_u(u^*,v^*) + g_v(u^*,v^*) \tag{79}$$

$$D = f_u(u^*,v^*)g_v(u^*,v^*) - g_u(u^*,v^*)f_v(u^*,v^*). \tag{80}$$

The linearized system of the original dynamic system around (u^*,v^*) is

$$d\dot{u} = f_u(u^*,v^*)du + f_v(u^*,v^*)dv \tag{81}$$

$$d\dot{v} = g_u(u^*,v^*)du + g_v(u^*,v^*)dv \tag{82}$$

But, it happens that the stability condition for the linearized system is given by $|D+1| < |T|$. Suppose the extreme equilibria are stable (i.e., $|D+1| < |T|$). Then it has to be the case that $|D+1| > |T|$ for the middle steady state. Then, the middle steady state is either sink if $|D| < 1$, or source if $|D| > 1$. ■

References

Acemoglu, D., and M. K. Jensen, 2015, "Robust Comparative Statics in Large Dynamic Economies," *Journal of Political Economy*, 123(3), 587–640.

Acharya, V. V., Y. Amihud, and S. T. Bharath, 2013, "Liquidity risk of corporate bond returns: conditional approach," *Journal of Financial Economics*, 110(2), 358–386.

Acharya, V. V., and S. Viswanathan, 2011, "Leverage, Moral Hazard, and Liquidity," *Journal of Finance*, 66(1), 99–138.

Acikgoz, O. T., 2018, "On the existence and uniqueness of stationary equilibrium in Bewley economies with production," *Journal of Economic Theory*, 173, 18 – 55.

Admati, A. R., 1985, "A Noisy Rational Expectations Equilibrium for Multi-Asset Securities Markets," *American Economic Review*, 53(3), 629–657.

Aiyagari, S. R., 1994, "Uninsured Idiosyncratic Risk and Aggregate Saving," *Quarterly Journal of Economics*, 109(3), 659–684.

Allen, F., and D. Gale, 1994, "Limited Market Participation and Volatility of Asset Prices," *American Economic Review*, 84(4), 933–955.

Beber, A., M. W. Brandt, and K. A. Kavajecz, 2007, "Flight-to-Quality or Flight-to-Liquidity? Evidence from the Euro-Area Bond Market," *Review of Financial Studies*, 22(3), 925–957.

Becker, B., and V. Ivashina, 2015, "Reaching for Yield in the Bond Market," *Journal of Finance*, 70(5), 1863–1902.

Ben-Rephael, A., 2017, "Flight-to-liquidity, market uncertainty, and the actions of mutual fund investors," *Journal of Financial Intermediation*, 31, 30–44.

Bewley, T., 1986, "Stationary Monetary Equilibrium with a Continuum of Independently Fluctuating Consumers," in *Contributions to Mathematical Economics in Honor of Gerard Debreu*, ed. by W. Hildenbrand, and A. Mas-Colell. North-Halland, Amsterdam.

Brunnermeier, M. K., and L. H. Pedersen, 2009, "Market Liquidity and Funding Liquidity," *Review of Financial Studies*, 22(6), 2201–2199.

Cao, C., Y. Chen, B. Liang, and A. W. Lo, 2013, "Can hedge funds time market liquidity?," *Journal of Financial Economics*, 109(2), 493–516.

Cao, C., B. Liang, A. W. Lo, and L. Petrasek, 2017, "Hedge fund holdings and stock market efficiency," *Review of Asset Pricing Studies*, (October), 1–61.

Coval, J., and E. Stafford, 2007, "Asset fire sales (and purchases) in equity markets," *Journal of Financial Economics*, 86(2), 479–512.

Diamond, D. W., and R. E. Verrecchia, 1981, "Information Aggregation in a Noisy Rational Expectations Economy," *Journal of Financial Economics*, 9(3), 221–235.

Dow, J., and G. Gorton, 1994, "Arbitrage Chains," *Journal of Finance*, 49(3), 819–849.

Dow, J., and J. Han, 2018, "The Paradox of Financial Fire Sales: the Role of Arbitrage Capital in Determining Liquidityeness," *Journal of Finance*, 73(1), 229–274.

Dow, J., and R. Rahi, 2003, "Informed Trading, Investment and Welfare," *Journal of Business*, 76(3), 439–454.

Duffie, D., 2010, "Asset Price Dynamics with Slow-Moving Capital," *Journal of Finance*, 65(4), 1237–1267.

Duffie, D., N. Garleanu, and L. H. Pedersen, 2005, "Over-the-Counter Markets," *Econometrica*, 73(6), 1815–1847.

——— , 2007, "Valuation in Over-the-Counter Markets," *Review of Financial Studies*, 20(6), 1865–1900.

Duffie, D., and B. Strulovici, 2012, "Capital Mobility and Asset Pricing," *Econometrica*, 80(6), 2469–2509.

Fajgelbaum, P. D., E. Schaal, and M. Taschereau-Dumouchel, 2017, "Uncertainty Traps," *Quarterly Journal of Economics*, 132(4), 1641–1692.

Follmer, H., U. Horst, and A. Kirman, 2005, "Equilibria in financial markets with heterogeneous agents: a probabilistic perspective," *Journal of Mathematical Economics*, 41(1), 123 – 155, Special Issue on Evolutionary Finance.

Fulghieri, P., and D. Lukin, 2001, "Information production, dilution costs, and optimal security design," *Journal of Financial Economics*, 61(1), 3 – 42.

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

——— , 2017, "Financially Constrained Arbitrage and Cross-Market Contagion," Working paper.

——— , Forthcoming, "The Dynamics of Financially Constrained Arbitrage," *Journal of Finance*.

Grossman, S. J., and J. E. Stiglitz, 1980, "On the Impossiblity of Informationally Efficient Markets," *American Economic Review*, 70(3), 393–408.

Hellwig, M. F., 1980, "On the Aggregation of Information in Competitive Markets," *Journal of Economic Theory*, 22, 477–498.

Huggett, M., 1993, "The risk-free rate in heterogeneous-agent incomplete-insurance economies," *Journal of Economic Dynamics and Control*, 17(5), 953 – 969.

Kyle, A. S., 1985, "Continuous Auctions and Insider Trading," *Econometrica*, 53(6), 1315–1335.

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

Mitchell, M., L. H. Pedersen, and T. Pulvino, 2007, "Slow Moving Capital," *American Economic Review*, 97(2), 215–220.

Shleifer, A., and R. W. Vishny, 1990, "Equilibrium Short Horizons of Investors and Firms," *American Economic Review*, 80(2), 148–153.

Shleifer, A., and R. W. Vishny, 1997, "The Limits of Arbitrage," *Journal of Finance*, 52(1), 35–55.

Stokey, N., and R. Lucas, 1996, *Recursive methods in economic dynamics.* Harvard Univ. Press, Cambridge, Mass., 4. print edn.

Vives, X., 1995, "Short-Term Investment and the Informational Efficiency of the Market," *The Review of Financial Studies*, 8(1), 125–160.

Wang, J., 1993, "A Model of Intertemporal Asset Prices Under Asymmetric Information," *Review of Financial Studies*, 60, 249–282.

——— , 1994, "A Model of Competitive Stock Trading," *Journal of Political Economy*, 102(1), 127–168.