Hysteresis in price efficiency and the economics of slow moving capital

June 4, 2020

Abstract

Will arbitrage capital flow into markets experiencing shocks, mitigating adverse effects on price efficiency? Not necessarily. In a dynamic model with privately informed capital-constrained arbitrageurs, price efficiency plays a dual role, determining both the profitability of new arbitrage and the ability to close existing positions profitably. An adverse shock to efficiency lengthens arbitrage duration, effectively reducing the amount of arbitrage capital available for new positions. If this falls below a critical mass, arbitrage capital flows out, amplifying the impact on price efficiency. This creates endogenous regimes: temporary shocks can trigger "hysteresis," a persistent shift in price efficiency.

Keywords: rational expectations, price efficiency, history dependence, slow-moving capital, regime shift, informed trading

JEL Classification: G12, G14, D82, D83, D84

1 Introduction

Traditional finance theory derives rational prices for assets based on traders' incentive to profit from any mispricing. This arbitrage pricing mechanism may break down when traders are capital-constrained, which could arise for various reasons.[1] The shortage of arbitrage capital could be temporary, since there may be surplus capital in other markets that could flow to exploit arbitrage opportunities in one market, but this reallocation might happen 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. So the economics of slow moving capital is a priority for research, and we seek to address in this paper by endogenizing the rate of flow of arbitrage capital. With enough capital deployed, privately-informed arbitrageurs (as in Grossman and Stiglitz (1980) and Kyle (1985)) will eventually push market prices towards fundamental value; in other words, information will be revealed. But if capital is limited and arbitrageurs need to decide where to deploy it, what is the underlying economics of how quickly they will flow into markets experiencing shocks, mitigating adverse effects on price efficiency? Or, are there reasons why this might fail to happen?

To answer these questions, we study an infinite horizon model with two classes of assets: long term and short term. Some traders ("arbitrageurs") have private information about assets, but limited capital. The trading model is a competitive noisy rational expectations equilibrium (REE) model: arbitrageurs interact with noise traders and with liquidity providers who have unlimited capital but lack this information ("market makers").

The presence of market makers makes prices semi-strong efficient, however, in equilibrium prices may or may not reflect the arbitrageurs' information. In the discussion that follows, it should be understood that prices are semi-strong efficient. We say "efficient" (or equivalently "informative") to refer to efficiency with respect to the arbitrageurs' private information.

Each arbitrageur must choose whether to trade a long-term asset or a short-term asset (they can buy or short sell the asset depending on their information). If they trade a short-

[1] By arbitrage, we mean a trade which exploits price inefficiencies for profits whether based on public or private information (e.g., Dow and Gorton (1994), Dow and Han (2018)). In this paper, we use the terms "informed investors" and "arbitrageurs" interchangeably.

term asset, they receive a liquidating dividend next period, so for an arbitrageur, who knows the dividend, the value gained from the trade just depends on the price. In contrast, long-term assets liquidate with a fixed probability each period, so an arbitrageur with a position in a long-term asset also cares about how long they need to hold the position—in line with a common saying among professional traders: "buy when the market is inefficient, sell when the market is efficient." If they need to hold a position for longer, they cannot redeploy their capital. A position pays off if either the long-term asset liquidates, or the price reveals the liquidation value (as a result of the trading process).

We call the mass of arbitrageurs taking a new position in both the long and the short-term market the "active capital." For a capital-constrained arbitrageur to invest in a new position, they must close out the existing position. Unless the new position is more profitable, the arbitrageur will be "locked-in" to the existing position in another asset until the price of that asset reverts closer to fundamental value. Reversion to fundamental value can happen when subsequent trades by other privately informed arbitrageurs reveal the information. Once price reveals the information, the new, correct price is supported by the valuation of uninformed traders. Therefore arbitrage capital is no longer required to support the price.

We study how financial markets respond to shocks. Our model is fully dynamic in the sense that the possibility of shocks is anticipated by agents. For tractability, shocks are represented as changes to the market wide intensity of noise trading (realizations of individual asset values and noise trader demand in individual assets are not shocks to the economy because they average out in aggregate). An increase in noise trading intensity has a similar effect on price efficiency as a reduction in arbitrage capital: the immediate impact of both is to make prices less informative.

The amount of active capital is the endogenous state variable, and, in equilibrium, price efficiency of long-term assets is increasing in active capital. Given an initial level of active capital, we show that the model follows a unique equilibrium path which is a function of the realization of the shocks. However, the model may have two risky steady states.[2] In

[2]A risky steady state is defined to be a state to which the state variables would converge in the absence of further shocks.

one risky steady state, active capital is low, so long-term assets have less informative prices. An arbitrageur who enters this market can likely trade mispriced assets, but, offsetting this, the position will probably need to be held for a long time before the arbitrageur obtains the asset value. Because arbitrage positions are held for longer, active capital is indeed lower. In the other risky steady state, active capital is higher so prices of long-term assets are more informative: an arbitrageur in this market is less likely to be able to trade mispriced assets, but, offsetting this, a position will quickly deliver the underlying asset value. Because positions are not held for long, active capital is indeed higher.

So, price efficiency in our model has a dual role. It affects both the profitability of new positions that arbitrageurs can take ("buy when the market is inefficient") and their ability to close their positions profitably ("sell when the market is efficient"). The effect on the profitability of new positions, which we call the "entry effect," is well known from previous literature. The effect on the ability to unlock capital by closing existing positions, which we call the "exit effect," is more complex and explains why arbitrageurs might amplify the impact of a shock.

To see this, consider an adverse shock to efficiency in the long-term market. This lengthens arbitrage duration, reducing the amount of arbitrage capital that remains available for new positions, and making the market less attractive to arbitrageurs. So long as the amount of active arbitrage capital stays above a critical mass, the entry effect dominates and arbitrage capital flows in, gradually restoring the original level of price efficiency. But if the amount of active arbitrage capital falls below the critical mass, the exit effect dominates and arbitrage capital flows out: the reduction in efficiency becomes self-reinforcing, setting the market on a path towards the less efficient risky steady state. It may remain trapped there for a long time. Thus, temporary shocks can trigger "hysteresis," a persistent shift to a different price efficiency regime.[3]

Active arbitrage capital can fall below the critical mass in two ways. Either a given adverse shock persists for a long enough time, or, for a given shock, the initial amount of active arbitrage

[3] A regime is the domain of attraction of a risky steady state.

capital is low enough. Either way, once it moves to a new regime, the market will stay there until a long enough sequence of favorable shocks.

In our model, trading volume is monotonic in price efficiency. Reduction in efficiency goes hand in hand with low trading volume. This reduction in trading volume can be viewed as a kind of illiquidity. While long-term assets are stuck in the less efficient regime, there is an increase in the amount of arbitrage capital allocated to short-term assets. This is reminiscent of "flight to liquidity" following financial crises.

Turning to the empirical implications of our analysis, we propose a model-based measure of active capital. Using this measure, we provide testable implications of our theory.

Methodologically, our model combines features from the literature on asset pricing with limits to arbitrage and the literature on asset pricing with imperfect information (noisy REE). This combination yields our hysteresis result.

On the one hand, in models with public information (as in most of the limits to arbitrage literature) there is no role for active capital because the marginal buyers are arbitrageurs. In noisy REE models, however, uninformed investors are marginal buyers. Therefore, the revelation of private information effectively multiplies informed capital by allowing uninformed investors to learn from informed arbitrageurs. These informed arbitrageurs are thereby released to be active elsewhere.

On the other hand, the question of unlocking arbitrage capital does not arise in the noisy REE literature, as it mostly features either static models or dynamic models with unconstrained informed arbitrageurs. Extensions of the standard REE framework have been limited because linearity, that enables tractability, is difficult to maintain. Imposing position limits for arbitrageurs makes demands inherently non-linear. Nevertheless, our framework allows the characterization of the dynamics of a stationary noisy REE with capital constraints.

Unlike static models, in which sensitivity to shocks is modelled using comparative statics or multiple equilibria, in our model it is not the case that all agents could somehow decide to flip to another regime just by coordinating their beliefs. Transition between regimes is a feature of the equilibrium.

Our results help to understand the long term response of the financial system to shocks. In a crisis, the capital and the risk-bearing capacity of financial intermediaries is reduced. This means that their ability to take positions falls, relative to the noise in the system. This can cause a period of prolonged low volume and prices that are dislocated from fundamentals.

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

2 Literature Review

There is a significant literature on the limits to arbitrage, studying how limited capital affects prices. Specifically, in Allen and Gale (1994) agents prefer to keep only a limited amount of capital available in liquid assets, in Shleifer and Vishny (1997) arbitrageurs' capital is a function of past performance, and in Gromb and Vayanos (2002) arbitrageur capital needs to cover their maximum losses. Thus, in each case there is a limit on the total amount of arbitrage capital that can be deployed to correct mispricing. In a setting with multiple assets, 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), Kondor and Vayanos (2018)) or collateral constraints (e.g., Brunnermeier and Pedersen (2009), Gromb and Vayanos (2018)).

A common theme in this literature is that arbitrageurs' ability to eliminate mispricing depends on their total amount of capital, which in turn is influenced by price movements, giving rise to amplification. In this literature all agents can see that assets are mispriced but it takes time to deploy capital in response. In Kondor (2009), the mispricing wedge varies across time because arbitrageurs allocate their limited capital across uncertain future arbitrage opportunities. In Gromb and Vayanos (2018), 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.

The above papers rely on limited capital as the main driver of price dislocation. A number of papers focus on other frictions (alone or in combination with limited capital). In Duffie and Strulovici (2012), the speed of capital flow is governed by the imbalance of capital as well as the level of intermediary competition across markets. In Buss and Dumas (2018) and Rostek and Weretka (2015), trading frictions, in the form of trading fees in the former paper and traders' price impact in the latter, make investors reluctant to trade, leading to slow recovery of prices from shocks.

In contrast to the limits to arbitrage literature, our model takes a noisy REE approach. An important difference with most of the literature is that REE is based on private, not public, information. Our model highlights the role of active arbitrage capital (as opposed to total arbitrage capital), and features different regimes with endogenous thresholds for the state variable.

Dow and Gorton (1994) study a dynamic noisy REE model of a single market where the cost of carry interacts with short trading horizons of arbitrageurs to break down price efficiency. As in our model, trade on private information relies on informed traders in the future to push price closer to fundamental value. In our model, however, future availability of informed arbitrageurs is endogenous because price efficiency decides the speed at which arbitrage capital is released. This effect plays a key role in generating our hysteresis result, and allows us to study how capital moves between markets.

In previous literature, several papers using a noisy REE approach show that feedback between price efficiency and capital availability can lead to multiple equilibria or highly sensitive comparative statics (e.g., Yuan (2005), Goldstein, Li, and Yang (2014), Cespa and Foucault (2014), Dow and Han (2018)). For example, in Dow and Han (2018), a small increase of informed capital facilitates movement of uninformed capital and induces a large increase in liquidity. Our paper differs from all these papers in two major ways. First, the feedback channel between active capital and price efficiency in our paper is intertemporal. Second, our dynamic model has a unique equilibrium but multiple efficiency regimes.[4]

[4]The literature studying dynamic information regimes such as Fajgelbaum, Schaal, and Taschereau-Dumouchel (2017) and Kurlat (2018) also features multiple information regimes which arise endogenously through intertemporal links between economic activity and information externalities.

3 Setup

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. The risk-free rate of the economy is exogenously given by $r_f = 1/\beta - 1$.

We now set out our assumptions on financial assets, participants, and timing of events. We discuss the role of our assumptions and modeling choices in Appendix A.

3.1 Financial Assets

There are two classes of risky securities: long-term assets, traded in market L, and short-term assets, traded in market S. We assume that within each market there is a continuum of securities, each of which is a claim to a single random liquidation value. An asset in market L 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. On the other hand, an asset in market S has a one period maturity; it simply pays its liquidation value in the period after issuance. At maturity, any asset i in market $h \in \{L, S\}$ 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 $h \in \{L, S\}$. We further assume that

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

where the LHS and the RHS are the present value of payoff of an asset with quality k in market L and S, respectively. We also assume that asset payoffs are independent across assets and over time.

As discussed later (in Section 4.2), equilibrium asset prices either reveal their fundamental value completely, or not at all. For tractability, we assume that the mass of unrevealed assets is fixed to one unit in each market at any point of time. That is, each period new assets are issued to replace those which have either just realized payoffs, or become fully-revealed.

3.2 Participants

There is a unit mass of capital-constrained "arbitrageurs" who trade to generate speculative profits. We denote the set of arbitrageurs by $\mathcal{A}$, and index each arbitrageur by $a \in \mathcal{A}$. Each arbitrageur can generate a private signal about the payoff of one asset of their own choosing in each period at no cost. The private signal perfectly predicts the liquidation value of the asset. 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 one unit (either long or short) of unrevealed assets. We denote by $x_i^a(t) \in \{-1, 0, 1\}$ the market order of arbitrageur a for asset i in period t. Once an arbitrageur has acquired a position in a risky asset, they can hold it until it liquidates or its value is revealed, and can then open a new position. They also have the option to close out a position early (before it has realized profits), and open another position next period.

There is a continuum of competitive 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 $\zeta_i(t)$ for each asset i in period t. We denote by $X_i(t)$ the aggregate order flow for each asset i in period t:

$$X_i(t) \equiv \int_{a \in \mathcal{A}} x_i^a(t) da + \zeta_i(t). \tag{2}$$

We assume that $\zeta_i(t)$ follows an independent uniform distribution on $[-z_h(t), z_h(t)]$. The magnitude of $z_h(t)$ captures the intensity of noise trading in market h in period t. We assume that $z_S(t)$ is a constant z_S whereas $z_L(t)$ follows a Markov process with N states $z_L^1, z_L^2, ..., z_L^N$ whose transition matrix between states is given by

$$\Omega = \begin{bmatrix} \omega_{11} & \cdots & \omega_{1N} \\ \vdots & \ddots & \vdots \\ \omega_{N1} & \cdots & \omega_{NN} \end{bmatrix}. \tag{3}$$

All agents in the model understand the Markov process governing changes in z_L.

The realization of $z_L(t)$ is publicly observable to all the agents in the economy. Note that it is the only exogenous shock to the economy in our model at the aggregate level. We further assume that $z_L^n + z_S > 1$ for any n; this assumption prevents the price for every asset from being fully-revealing. Finally, we assume that all the realizations of noise trading intensity and asset payoffs are jointly independent.

3.3 Timing of Events

The timing of events in each period is as follows. At the beginning of each period t, asset payoffs realize and they are distributed among claim holders. Next, new assets are issued as described above, and noise trading intensity $z_L(t)$ realizes. After these events, arbitrageurs with capital to invest (i.e., those who do not already hold positions) select one of the two markets, collect private information about an asset in that market, and then trade on their information, either buying or selling short depending on their information. Arbitrageurs who already hold positions decide whether to close out their positions. Arbitrageurs' orders (either to acquire new positions or liquidate their existing positions) are submitted to market makers together with noise traders' orders $\zeta_i(t)$ for each asset i. 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 t, an arbitrageur is in either of two situations: "active" or "locked-in." An active arbitrageur does not have an existing position in an unrevealed asset, thus has capital available for new investment. A locked-in arbitrageur already has a position in an unrevealed asset, thus, does not have capital available for new investment until this position is closed out (we call it locked-in because although they have the option to close out early, arbitrageurs choose not to in equilibrium, as will be shown in Lemma 2). We denote the mass of active arbitrageurs by $\xi(t)$, and the mass of locked-in arbitrageurs by $1 - \xi(t)$.

Each active arbitrageur chooses to hold a new position in either market L or S. We denote the proportion of those choosing to trade assets in market L by $\delta(t)$ (so $1-\delta(t)$ is the proportion of those choosing to trade assets in market S). Each locked-in arbitrageur chooses whether to hold on to the position one more period or to close it out in the current period. We denote the proportion of those choosing to close out early by $\eta(t)$ (so $1 - \eta(t)$ is the proportion of those choosing to hold on to the position).

We define the vector of state variables to be

$$\theta(t) \equiv (\xi(t), z_L(t)), \tag{4}$$

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 Kyle (1985), risk neutral market makers set the price equal to the expected discounted liquidation value conditional on available information:

$$P_i(t) = \mathrm{E}\left[\beta^{\tau_i} v_i \middle| \mathcal{F}(t)\right], \tag{5}$$

where τ_i is the (random) maturity of asset i, and $\mathcal{F}(t)$ is the information set of market makers in period t, which includes the history of aggregate order flows of all assets and the state variables up to period t.

Given this price formation process, an asset's price becomes more informative over time, reflecting market makers' learning from trading:

Lemma 1 *Suppose that in period t asset i is unrevealed prior to trading. If there are $\mu_i(t)$ arbitrageurs who submit an order to buy (sell) one unit if asset i is of good (bad) quality, and*

$z_i(t)$ is the noise trading intensity for asset i then,

$$P_i(t) = \begin{cases} P^B & \textit{if } -\mu_i(t) - z_i(t) \leq X_i(t) < \mu_i(t) - z_i(t) \\ P^0 & \textit{if } \mu_i(t) - z_i(t) \leq X_i(t) \leq -\mu_i(t) + z_i(t) \\ P^G & \textit{if } -\mu_i(t) + z_i(t) < X_i(t) \leq \mu_i(t) + z_i(t), \end{cases} \tag{6}$$

where

$$P^k \equiv \frac{\beta q}{1 - \beta(1-q)} V_L^k = \beta V_S^k, \quad \textit{for all } k \in \{G, B\} \tag{7}$$

is the fully-revealing price,

$$P^0 \equiv \frac{P^G + P^B}{2} \tag{8}$$

is the non-revealing price, and

$$\lambda_i(t) = \frac{\mu_i(t)}{z_i(t)} \tag{9}$$

is the probability of information revelation for asset i in period t.

Proof. See Appendix B. ■

Lemma 1 implies that in our model prices start out uninformative, and may later jump straight to being fully revealing (but can never be partially revealing).[5]

The largest possible realization for the order flow $X_i(t)$ when arbitrageurs are selling the asset is $-\mu_i(t) + z_i(t)$, which obtains when noise traders buy $z_i(t)$. So if $X_i(t)$ is larger than $-\mu_i(t) + z_i(t)$, then it can only result from arbitrageurs buying the asset, so it reveals that the asset is good. Similarly, if the order flow is smaller than $\mu_i(t) - z_i(t)$, then it can only result from arbitrageurs selling the asset, so it reveals that the asset is bad. But if the order flow takes an intermediate value, then it could have resulted from either arbitrageurs buying and noise traders selling, or vice versa. Because noise trading is uniformly distributed, any level of

[5]As previously noted, the presence of market makers makes prices semi-strong efficient (Eq. (5)), so statements concerning "efficient" (or equivalently "informative") prices refer to efficiency with respect to the arbitrageurs' private information.

the order flow is equally likely regardless of whether arbitrageurs are buying or selling, so it is non-revealing.

As we explain in the next subsection, we focus on equilibria in which knowledge of the current state $\theta(t)$ (as defined in Eq. (4)) is sufficient to infer the mass of informed arbitrageurs active in each asset. As a result, the pair $(\theta(t), X_i(t))$ is a sufficient statistic for $\mathcal{F}(t)$ in Eq. (5) for any unrevealed asset.

For convenience, we call the probability of information revelation $\lambda_i(t)$ the "price efficiency" of asset i henceforth. As shown later in the paper, $\lambda_i(t)$ plays a dual role of capturing the degree of mispricing as well as the expected investment duration of asset i.[6]

4.3 Law of Motion

We focus on stationary market-wise symmetric rational expectations equilibria. Market-wise symmetry means that price efficiency $\lambda_i(t)$ is equal to $\lambda_L(t)$ for any asset i in market L, and is equal to $\lambda_S(t)$ for any asset i in market S, in each period t. Notice that, with a continuum of assets, $\lambda_h(t)$ is the fraction of assets in market $h \in \{L, S\}$ whose value is revealed by the trading process. Stationarity means that $\delta(t)$, $\eta(t)$, $\lambda_L(t)$ and $\lambda_S(t)$ are time-invariant functions of the state variables $\theta(t)$. In the rest of the paper we omit time index on state variables for notational clarity and use the prime notation to denote the value in the subsequent period. For example, ξ' denotes the value of ξ in the subsequent period.

The law of motion of the mass of active arbitrageurs is given by

$$\xi' = (1 - \delta(\theta))\xi + (\delta(\theta)\xi + 1 - \xi)(q + (1 - q)\lambda_L(\theta)) + (1 - \xi)\eta(\theta)(1 - q)(1 - \lambda_L(\theta)). \tag{10}$$

The RHS is the sum of three terms. The first term is the mass of arbitrageurs invested in market S in the current period; this mass becomes entirely active in the subsequent period as the assets are short-lived. The second and third terms are the mass of arbitrageurs invested in market L in the current period (i.e., $\delta(\theta)\xi$ new arbitrageurs from the current period and $1-\xi$ arbitrageurs

[6] Note that $\lambda_i(t)$ is not the same as price impact (or "Kyle's Lambda" in the literature). In our model individual arbitrageurs do not have a price impact because they are infinitesimal. Noise traders or arbitrageurs in aggregate have a price impact.

locked-in from the previous period) that become available for new investment in the subsequent period. This happens if the asset pays off, or if the market price fully reveals the asset value, or if locked-in arbitrageurs close out the position early (it turns out they choose not to do this in equilibrium). Overall, a fraction $q + (1-q)\lambda_L(\theta)$ of the arbitrageurs invested in market L 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 $\eta(\theta)(1-q)(1-\lambda_L(\theta))$ of locked-in arbitrageurs from the previous period becomes active next period because of the decision to close out early.

4.4 Dynamic Arbitrage

Given the current state θ, we denote by $J_L(\theta)$ and $J_S(\theta)$ the value of investing in a new position in market L and market S, respectively. Because any active arbitrageur can choose between the two markets, the value of being active given θ equals

$$J_a(\theta) = \max\left(J_L(\theta), J_S(\theta)\right). \tag{11}$$

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

$$\delta(\theta) \in \begin{cases} \{0\}, & \text{if } J_L(\theta) < J_S(\theta); \\ \{1\}, & \text{if } J_L(\theta) > J_S(\theta); \\ [0,1], & \text{otherwise,} \end{cases} \tag{12}$$

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

Some arbitrageurs are currently locked-in in market L; we denote by $J_l(\theta)$ the associated value function given θ. We can obtain $J_L(\theta)$ and $J_S(\theta)$, whose detailed derivations are relegated to Appendix B, as follows:

$$J_S(\theta) = -(\lambda_S(\theta)P^G + (1-\lambda_S(\theta))P^0) + \beta\Big[V_S^G + \mathrm{E}[J_a(\theta')|\theta]\Big]; \tag{13}$$

$$J_L(\theta) = -(\lambda_L(\theta)P^G + (1-\lambda_L(\theta))P^0) + \beta U(\theta), \tag{14}$$

where

$$\begin{aligned} U(\theta) \equiv\ & qV_L^G + (1-q)\lambda_L(\theta)P^G + (1-(1-\lambda_L(\theta))(1-q))\mathrm{E}[J_a(\theta')|\theta] \\ & + (1-\lambda_L(\theta))(1-q)\mathrm{E}[J_l(\theta')|\theta]. \end{aligned} \tag{15}$$

Eqs. (13)-(15) show expected profits for an arbitrageur with good news who takes a long position; in a market-wise symmetric equilibrium, this is the same as the expected profits for an arbitrageur with bad news taking a short position because good and bad qualities are equiprobable (i.e., $P^G - P_0 = P_0 - P^B$).

Eq. (13) shows the expression for expected profits from buying a short-term asset. With probability $\lambda_S(\theta)$ the orders reveal the information and the price is fully informative. With probability $1-\lambda_S(\theta)$ the arbitrageur is able to buy the asset at the unrevealed price. Next period the arbitrageur obtains the liquidation value of the asset plus the continuation value $\mathrm{E}[J_a(\theta')|\theta]$ of becoming active again in the following period.

Similarly, Eq. (14) shows that the expression for expected profits from buying a long-term asset consists of the expected discounted value next period, minus the expected price paid for the asset. The value next period $U(\theta)$ in Eq. (15) reflects the possibility that asset will liquidate and yield V_L^G (probability q), or the position pays off at price P^G because it is already fully revealed in the current period (probability $(1-q)\lambda_L(\theta)$). In both cases the arbitrageur will then become active again in the following period, obtaining payoff $\mathrm{E}[J_a(\theta')|\theta]$. Alternatively, the position will continue to be held for at least another period yielding value $\mathrm{E}[J_l(\theta')|\theta]$.

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

$$J_l(\theta) = \max\left(J_E(\theta), J_H(\theta)\right), \tag{16}$$

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

$$J_E(\theta) = \lambda_L(\theta)P^G + (1-\lambda_L(\theta))P^0 + \beta \mathrm{E}[J_a(\theta')|\theta]; \tag{17}$$

$$J_H(\theta) = \beta U(\theta). \tag{18}$$

Eq. (17) shows that an early-liquidating arbitrageur is able to sell the asset at the revealed price with probability $\lambda_L(\theta)$, or sells the asset at the unrevealed price with probability $1-\lambda_L(\theta)$. The arbitrageur will then become active again in the following period, obtaining payoff $\mathrm{E}[J_a(\theta')|\theta]$. Eq. (18) shows that an arbitrageur holding the position for one more period obtains the same value as an arbitrageur entering a new position in an unrevealed long-term asset but, in contrast to Eq. (14), without incurring the initial cost.

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

$$\eta(\theta) \in \begin{cases} \{0\}, & \text{if } J_E(\theta) < J_H(\theta); \\ \{1\}, & \text{if } J_E(\theta) > J_H(\theta); \\ [0,1], & \text{otherwise.} \end{cases} \tag{19}$$

4.5 Stationary Equilibrium

We define stationary equilibrium in a standard manner.[7]

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

[7] As is standard, the equilibrium is called "stationary" because all equilibrium functions 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 5.1.

1. *J_a, J_l, J_L, J_S, J_E, J_H, δ, η satisfy Eqs. (11)-(19).*

2. *λ_L and λ_S correspond to the probability that prices, which are determined by Eq. (5), reveal true asset values in market L and S, respectively.*

3. *The law of motion for ξ satisfies Eq. (10).*

We focus on interior equilibria in the sense that $J_L(\theta) = J_S(\theta)$, so that active arbitrageurs are indifferent between investing in market L and S.

Lemma 2 *In an interior equilibrium, $\eta(\theta) = 0$; it is never optimal to close out the existing position early, i.e., $J_H(\theta) > J_E(\theta)$.*

Proof. See Appendix B. ■

The intuition for Lemma 2 is as follows: at any point in time an asset's price either reflects full information, or no information. If information is unrevealed, trading on that information has the same profitability as trading on information on any other asset. An arbitrageur could close a position before the information is revealed, and redeploy the capital next trading round into another position in another asset, but would end up holding a new position with at best the same profitability, one period later. Consequently, locked-in arbitrageurs stay inactive until either the asset pays off, which occurs with exogenous probability q, or the price fully reveals the asset value, which occurs with endogenous probability $\lambda_L(\theta)$.[8]

Therefore price efficiency in our model has a dual role. On the one hand, it affects the profitability of new positions that arbitrageurs open, since it is more profitable to enter a position in an inefficient market ("entry effect"). On the other hand, it also affects their ability to close their positions profitably in the long-term market ("exit effect"). As will be shown later in the paper, the exit effect means that increased price efficiency is reinforced in two ways, first by releasing locked-in capital, and second by inducing arbitrageurs to allocate more capital to the long-term market.

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

[8] The role of the assumptions in deriving this result is discussed in Appendix A.

Lemma 3 *In an interior equilibrium, the probability of information revelation in market L equals*

$$\lambda_L(\theta) = \frac{\delta(\theta)\xi}{z_L}, \tag{20}$$

and the probability of information revelation in market S equals

$$\lambda_S(\theta) = \frac{(1-\delta(\theta))\xi}{z_S}. \tag{21}$$

Proof. See Appendix B. ■

As Lemma 3 shows, for a fixed capital allocation δ, equilibrium price efficiency in market $h \in \{L, S\}$ increases in the amount of informed arbitrage capital ξ (more active capital makes order flow more informative) and decreases in the intensity of noise trading z_L (more noise makes order flow less informative). Of course, capital allocation $\delta(\theta)$ is a function of the state θ, so the overall dependency of price efficiency on active capital ξ and noise trading intensity z_L can only be determined in equilibrium.

Proposition 1 *Under the conditions stated in Appendix C, there exists a unique interior stationary equilibrium. In equilibrium, price efficiency in the long-term market $\lambda_L(\theta)$ is monotone increasing in active capital ξ and monotone decreasing in noise trading intensity z_L.*

Proof. See Appendix C. ■

The proof of Proposition 1 shows that the technical conditions provided in the appendix ensure the contraction property of the equilibrium mapping, implying unique existence of interior stationary equilibrium. Consequently, there exists a unique path of active capital given the initial level of active capital and the sequence of shocks to noise trading intensity.

Proposition 1 also shows that equilibrium price efficiency in market L increases in ξ and decreases in z_L even when δ is endogenized. Active arbitrageurs deploy their capital across markets in response to changes in ξ and z_L. This reaction by the arbitrageurs may either exacerbate or mitigate the initial direct change in price efficiency in market L, but even in the latter case does not completely offset it.

5 Main Results

Our model can display efficiency hysteresis, in other words a transitory shock can move the system to a different level of efficiency. Due to the stochastic nature of our model, the level of price efficiency is constantly changing in response to the random arrival of shocks that change noise trading intensity. Along a sample path where noise intensity happens to be unchanged, the level of efficiency will converge to a "risky steady state" value.[9] Hysteresis is a consequence of having multiple risky steady states. To aid intuition, we initially describe the steady states for a non-stochastic version of the model without shocks to noise trading intensity; we then revert to studying the stochastic model later in this section.

5.1 Steady State Analysis

We start by considering the special case of the model under the assumption that noise trading intensity z_L is fixed at a constant level, i.e., the Markov process is degenerate ($N = 1$). Then, the state variable is just the current level of active capital (Eq. (4)). An equilibrium maps the current period's state variable ξ to the next period's state variable ξ', and a steady state is a fixed point of that mapping.

For notational convenience, we use an asterisk notation to denote the values in steady-state. For example, ξ^* denotes the steady-state-level mass of active arbitrageurs. At the steady state, the values of investing in new positions in market S and market L in Eqs. (13) and (14) admit simple closed form solutions. Using Eq. (7) together with the definitions of J_S and J_L in

[9] See Coeurdacier, Rey, and Winant (2011) for the definition of risky steady state and a discussion.

Section 4.4 and Lemma 2, we can solve for J_S^* and J_L^* as follows:

$$J_S^* = \frac{(P^G - P^0)(1 - \lambda_S^*)}{1 - \beta}. \tag{22}$$

$$J_L^* = \frac{(P^G - P^0)(1 - \lambda_L^*)[1 - \beta(1 - q)(1 - \lambda_L^*)]}{1 - \beta}. \tag{23}$$

The RHS in Eqs. (22) and (23) is the present value (discounted at the discount factor β) of future per-period expected profits of investing in market S and market L, respectively. These expressions illustrate the dual role of price efficiency in our model. On the one hand, price efficiency determines the profitability of investment opportunities: higher λ_L^* (and also λ_S^*) decreases the probability of acquiring a new position at non-revealing prices, thereby reducing an arbitrageur's per-period expected profit. This is the standard effect of price efficiency on speculative profits (the entry effect).

On the other hand, price efficiency determines the maturity of investment opportunities in long-term assets: higher λ_L^* increases the likelihood of profitably closing out a position earlier, thereby increasing arbitrageur's per-period expected profit. This effect is captured in Eq. (23) by the term $1 - \beta(1 - q)(1 - \lambda_L^*)$ which reflects the per-period probability $(1 - q)(1 - \lambda_L^*)$ of remaining locked-in a trade in market L, weighted by the discount factor β. This is the exit effect of price efficiency on speculative profits; with higher λ_L^* an arbitrageur waits less before the arbitrageur's private information is incorporated in the price, hence can cash in sooner at full value.

Equating the value functions in Eqs. (22) and (23) yields the following relationship between λ_S and λ_L if the arbitrageurs are indifferent between entering market S and L:

$$\lambda_S^* - \lambda_L^* = \beta(1 - q)(1 - \lambda_L^*)^2. \tag{24}$$

The LHS of Eq. (24) is the difference in probabilities of trading at a fully-revealing price in market S over market L. By trading in market L, an arbitrageur 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 L must compensate this opportunity cost with a higher probability of trading at a non-revealing price in the current period.[10]

We can now derive the arbitrageurs' indifference condition in terms of δ and ξ by substituting Eqs. (20) and (21) into Eq. (24):

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

For a fixed δ^*, a decrease in active arbitrage capital ξ^* decreases price efficiency in both markets. This increases speculative profits in both markets through the entry effect but the exit effect makes market L less attractive. Hence, δ^* must adjust to restore arbitrageurs' indifference condition across markets.[11]

As well as the indifference condition (IC) curve described above, we can obtain the following capital movement (CM) curve from the law of motion for active arbitrage capital in Eq. (10) together with Eq. (20) for λ_L^*:

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

Note from (CM) that an increase in the fraction of active arbitrageurs that invest in market L has two opposing effects. Clearly, as δ^* increases, more arbitrageurs enter market L where they will become locked-in. This tends to reduce the steady state value for active capital ξ^*. But the exit effect works in the other direction; an increase in δ^* improves price efficiency in market L, which increases the rate at which arbitrage capital is released from this market. This tends to increase ξ^*. The exit effect is dominated for δ^* small, in which case ξ^* decreases in δ^*. However, the exit effect may dominate for δ^* large.[12] Intuitively, when there are more

[10] This trade-off between immediate profits and holding period durations is a standard feature in dynamic trading models with multiple assets. The implication that long-term assets have a bigger mispricing wedge is noted in Shleifer and Vishny (1990), Eq. (6). Their model has fixed investment durations. Also, in Dow and Gorton (1988) long-term assets are more mispriced than short-term assets, for similar reasons.

[11] δ must increase (decrease) if the relative profitability of trading in market L is decreasing (increasing) in δ. Differentiating Eq. (IC) yields that δ must increase if

$$\lambda_L^* > 1 - \frac{1}{2\beta(1-q)}\left(1 + \frac{z_L}{z_S}\right). \tag{25}$$

A sufficient condition for Eq. (25) to hold is $q \geq 1 - \frac{1}{2\beta}\left(1 + \frac{z_L}{z_S}\right)$, which is satisfied in our numerical examples.

[12] For example, see Panels (a)-(c) of Figure 1. The bottom part of the (CM) curve is where the exit effect is

arbitrageurs in market L (and getting locked-in), increasing the rate at which locked-in capital is released has a bigger effect.

An interior steady state is found at the intersection of the IC curve and the CM curve. We can show the following results about steady states under the same conditions for Proposition 1:

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

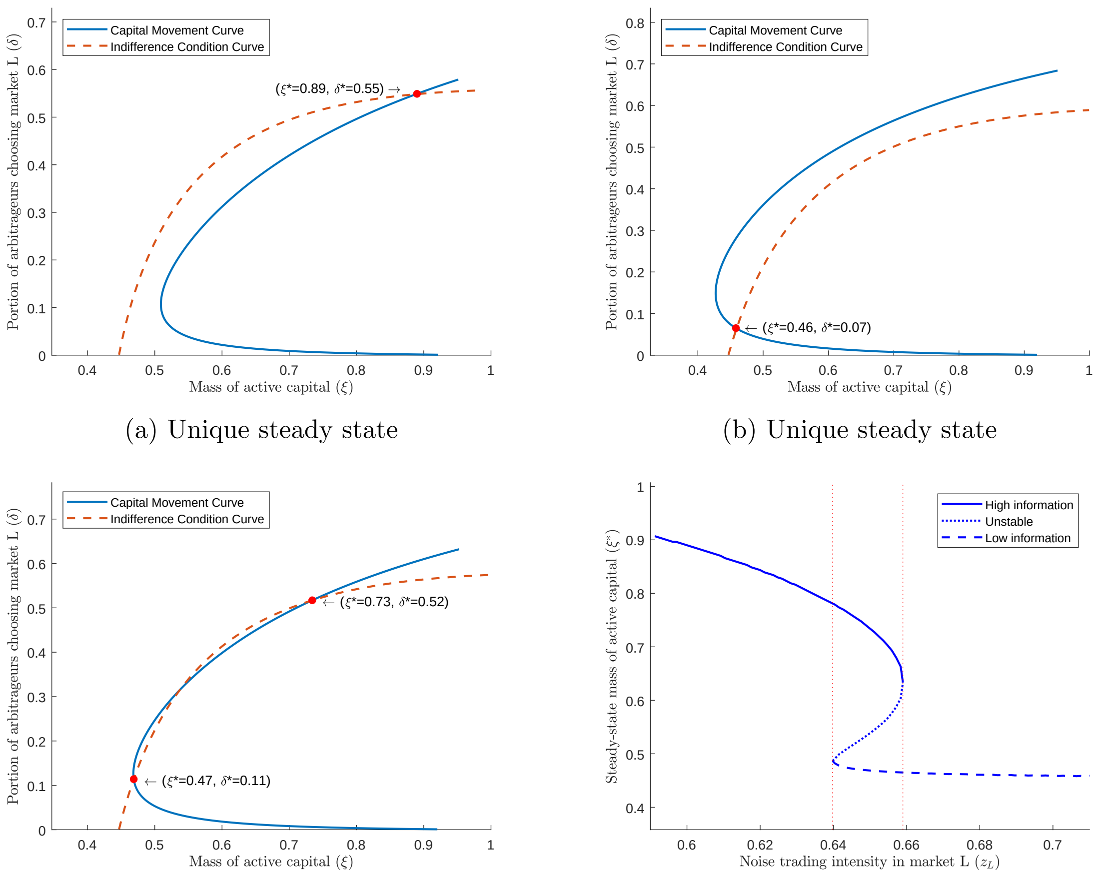
Proof. See Appendix D. ■

Figure 1 illustrates the steady state values for ξ^* and δ^* determined by the intersection of the IC and CM curves for different values of the exogenous parameter z_L (the noise trading intensity in the long-term market). 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 regions of values of z_L where there is uniqueness or multiplicity.

The multiplicity of steady states depends on the relative positions and curvatures of the IC and CM curves which are linked to the exit effect. To see this, consider the limit as $q \to 1$, which corresponds to turning the long-term market L into a short-term market. The IC curve approaches a horizontal line, and the CM curve approaches a vertical line.[13] Therefore, the multiplicity disappears in the limit. On the other hand, the multiplicity may return as the exit effect is more pronounced. When the long-term assets liquidate infrequently (q low), arbitrageurs rely mainly on information revelation to exit their positions. This creates a feedback effect between price efficiency and active capital: an increase in price efficiency makes the long-term asset more attractive to arbitrageurs through shorter investment duration and releases more active capital from locked-in arbitrageurs. This self-reinforcing nature of price efficiency leads to multiple steady states.

dominated (when δ^* is small), and the top part is where the exit effect dominates (when δ^* is large).

[13]The limit of the IC curve is $\delta^* = z_L/(z_S + z_L)$, and the limit of the CM curve is $\xi^* = 1$.



(a) Unique steady state

(b) Unique steady state

(c) Multiple steady states

(d) Steady states under various z_L.

Figure 1: **Steady States. (non-stochastic model)** Parameter values common across all panels: $q = .01, z_S = .475, \beta = .95$. Values for z_L in the unique steady state in panel (a) is $z_L = .6$ and in panel (b) is $z_L = .7$; in the multiple steady states in panel (c): $z_L = .65$.

Propositions 1 and 2 imply the following comparison across steady states:

Corollary 1 *With multiple steady states, a steady state with a larger amount of active capital features higher levels of price efficiency and trading volume in market L (and lower levels of price efficiency and trading volume in market S) compared to a steady state with a smaller amount of active capital.*

Proof. See Appendix D. ■

Corollary 1 shows that, since trading volume is monotone increasing in price efficiency in our model, a larger amount of active capital which increases price efficiency is also associated with higher trading volume. The corollary also shows opposite comparison across steady states for market S. Intuitively, with a smaller amount of active capital market L is less efficient and holding periods of arbitrage positions are longer in this market, which leads more arbitrageurs to invest in market S. This "flight-to-liquidity" pushes up price efficiency and trading volume in market S compared to a steady state with larger amount of active capital.

5.2 Analysis of the Stochastic Model

In this subsection, we return to analysis of the system in the general case with stochastic shocks. In our numerical illustrations we will consider the case where one level of noise trading intensity is more likely than the others and we refer to this as the "normal" level, and any deviation to another level of noise trading intensity as a "shock." A price efficiency regime is the region associated with the risky steady state to which the state variables would converge in the absence of further shocks. The system displays hysteresis because a regime shift can occur in which shocks cause the economy to enter the region of attraction to a different risky steady state.

To shed light on the response to a shock to noise trading intensity in market L, we obtain arbitrageurs' indifference condition which is the dynamic version of Eq. (24):

Lemma 4 *In an interior equilibrium price efficiency satisfies*

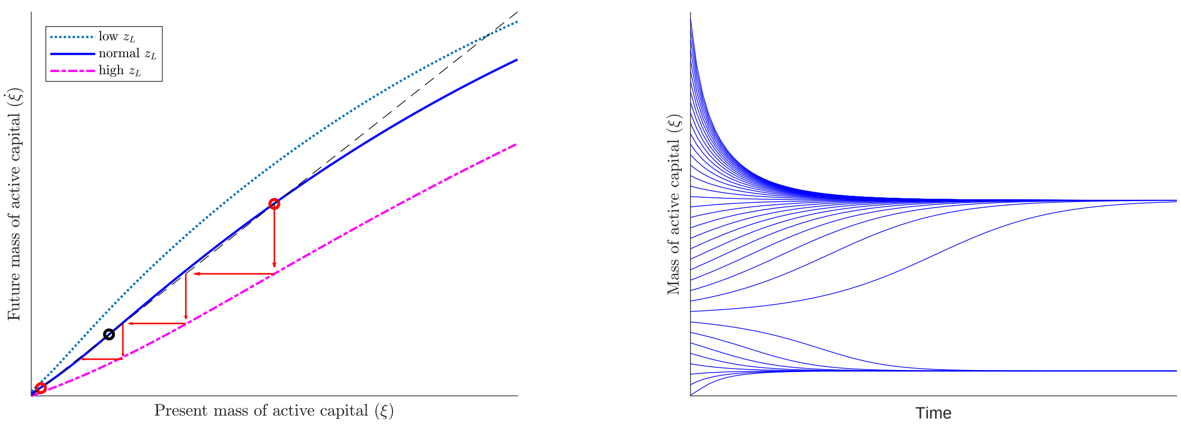
$$1 - \lambda_S(\theta) = (1 - \lambda_L(\theta))(1 - \beta(1 - q)(1 - \mathrm{E}[\lambda_L(\theta')|\theta])). \tag{26}$$

Proof. See Appendix B. ■

Suppose that z_L increases. To aid intuition, assume $\delta(\theta)$ does not react to this shock, and consider the effect on both sides of Eq. (26). By Eq. (21), the LHS of Eq. (26) is unaffected, while the RHS is affected via two channels. By Eq. (20), $\lambda_L(\theta)$ drops, making investment in market L more attractive (the entry effect). But, by Eq. (10), lower $\lambda_L(\theta)$ implies that ξ' also drops because current locked-in capital is released at a lower rate, thereby decreasing $\lambda_L(\theta')$

and implying that market L is more inefficient in the subsequent period. This leads to a longer investment duration, making investment in market L less attractive (the exit effect).

The reaction of capital allocation $\delta(\theta)$ to the shock depends on the relative strength of the entry and exit effects on the attractiveness of market L. $\delta(\theta)$ increases if the entry effect is stronger, thereby mitigating the initial effect of the shock on price efficiency in market L. On the other hand, $\delta(\theta)$ decreases if the exit effect is stronger. In this case a larger fraction of active arbitrageurs chooses market S, thus exacerbating the direct effect of the shock.



(a) Transition curves for ξ

(b) Evolutionary paths of ξ

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

Figure 2 illustrates regimes and dynamics of the mass of active capital ξ. Panel (a) plots the transition curves for ξ which map the current state $\theta = (\xi, z_L)$ into next period's active capital ξ', that is,

$$\xi' = (1 - \delta(\theta))\,\xi + (\delta(\theta)\xi + 1 - \xi)\left(q + (1-q)\frac{\delta(\theta)\xi}{z_L}\right). \tag{27}$$

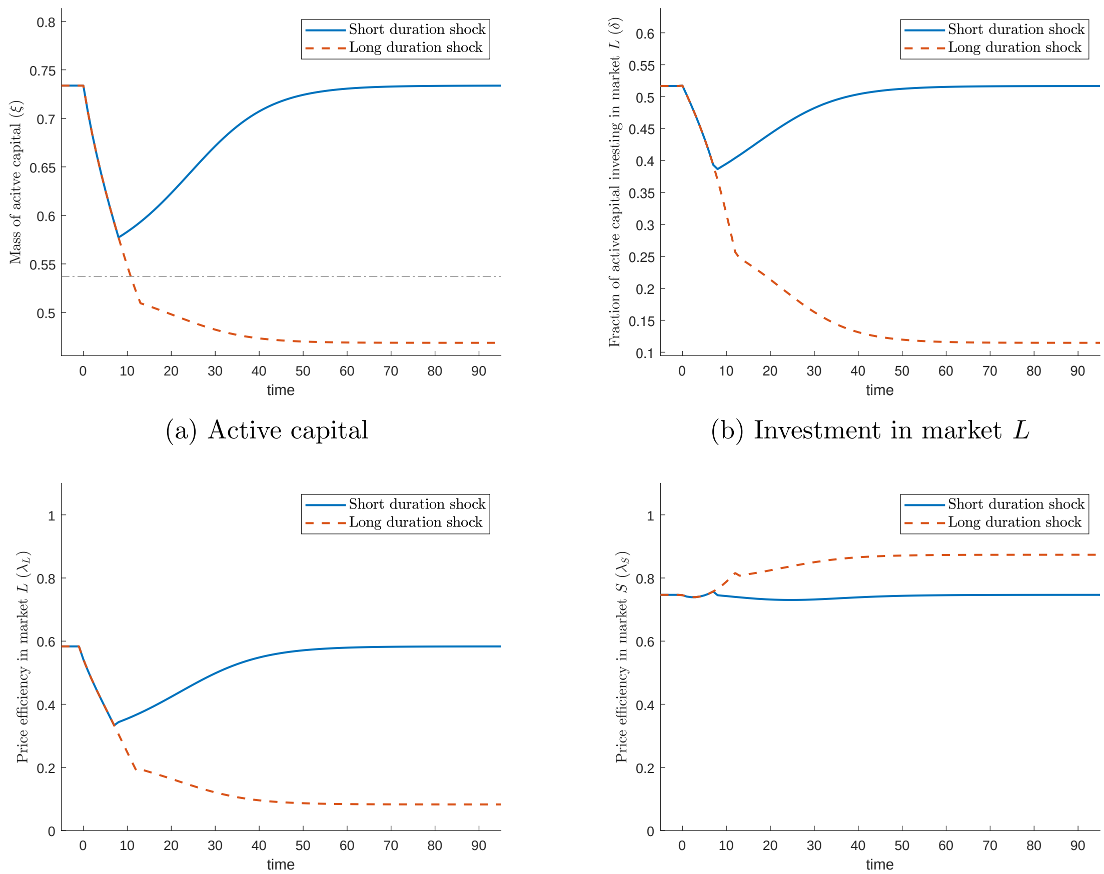
An intersection with the 45-degree line is a fixed point of the transition curve such that $\xi' = \xi$. The intermediate (solid) curve displays three fixed points corresponding to risky steady states, with the lowest and the highest being stable and the middle one unstable. The middle fixed point corresponds to the threshold value of ξ which separates the two regimes for ξ.

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 risky steady state for initial values of ξ above the threshold value (and similarly, to the lowest risky steady state for initial values below the threshold value.) Hence, if the value of active capital ξ is close to a risky steady state, it converges back to it after experiencing a small deviation due to shocks. However, the threshold value represents a critical mass for ξ. Once ξ crosses the threshold value, then it triggers a regime shift: ξ is set on a trajectory towards a different risky steady state value.

Such a regime shift is further illustrated in panel (a) of Figure 2. The arrows show the effects of shocks to z_L from its normal level starting from the risky steady state with high efficiency (and a large mass of active capital). A temporary increase in z_L pushes ξ downwards, but ξ reverts in subsequent periods if z_L 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 risky steady state characterized by low efficiency even if noise trading intensity z_L reverts to its normal state. The mass of active capital ξ cannot go back to the original high level until a sequence of favorable shocks to z_L arrives that pushes ξ back above the critical threshold; this puts the dynamics of active capital back on the upward trajectory. Our numerical simulation in the next subsection shows an example of such transitions across price efficiency regimes.

5.3 Response to Shocks

In this subsection, we provide some numerical examples of dynamic responses to stochastic shocks to noise trading intensity to illustrate the interaction between price efficiency and active capital. In the numerical examples, we consider a temporary deviation of z_L from its normal level to a higher level, after which z_L reverts to its normal level. Our examples illustrate how temporary shocks can create long-lasting effects in both markets by causing hysteresis in price efficiency.



(a) Active capital (b) Investment in market L

(c) Price efficiency in market L (d) Price efficiency in market S

Figure 3: **Transitional Dynamics for a Temporary Shock under Different Shock Durations.** A short-duration shock (duration of 8 periods, solid line) and a long-duration shock (duration of 13 periods, dashed line) is given at $t = 0$. Parameter values: $q = .01, z_L \in \{.6, .65, .7\}, z_S = .475, \beta = .95$. Transition probabilities are given by $\omega_{11} = .46, \omega_{12} = .54, \omega_{13} = 0, \omega_{21} = .12, \omega_{22} = .76, \omega_{23} = .12, \omega_{31} = 0, \omega_{32} = .5, \omega_{33} = .5$ where states 1,2 and 3 correspond to low, normal and high level of z_L, respectively.

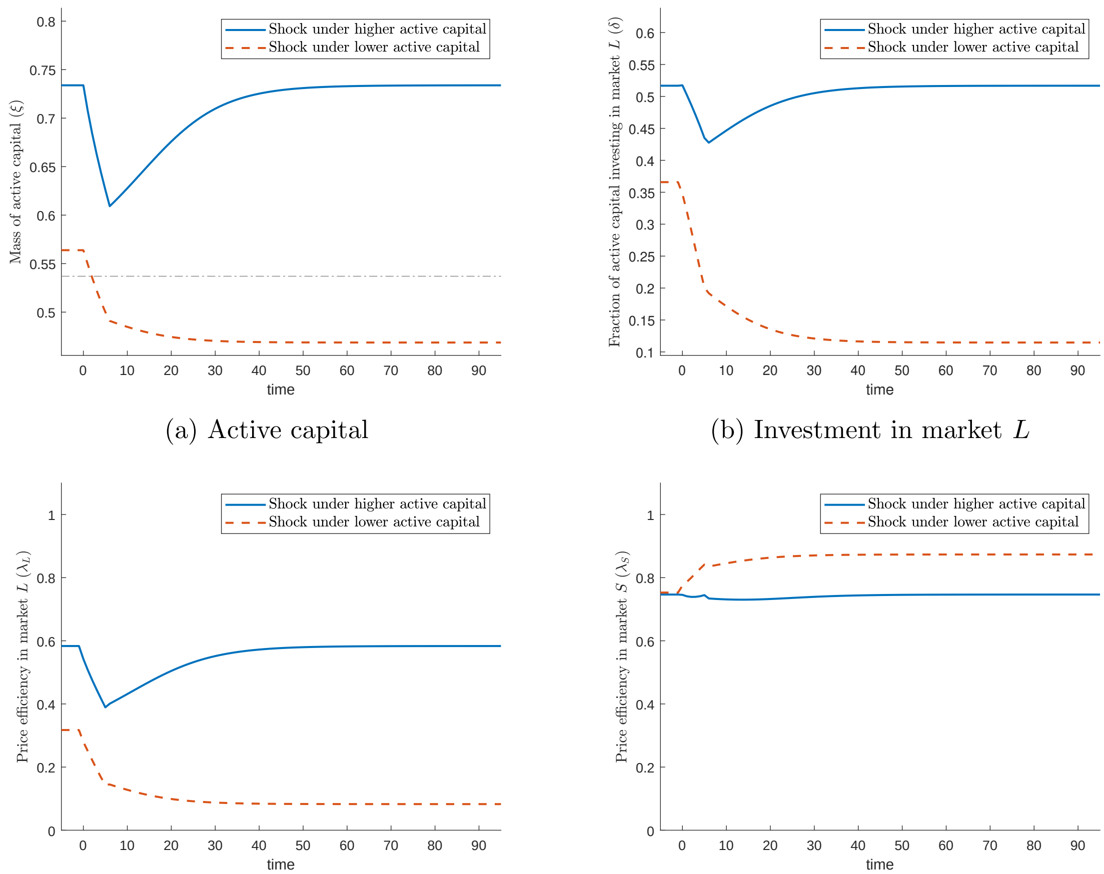
Figures 3 shows responses to shocks to z_L with two different durations. The shock staring at $t = 0$ leads to an immediate drop in price efficiency in market L and therefore a decrease in ξ starting from the subsequent period. In cases of both short and long-duration shocks, arbitrageurs react by flowing out of market L (i.e., δ decreases) as they anticipate lower efficiency and a higher opportunity cost of being locked in this market going forward. Therefore, market

L suffers further decreases in price efficiency, thereby triggering further decreases in ξ until the shock is removed. In the case of a short-duration shock, price efficiency is gradually restored once the shock is removed. This replenishes active capital and the economy converges back to the initial risky steady state. By contrast, the response to a longer duration shock, which drags the level of active capital below the critical threshold, has different dynamics due to hysteresis. Instead of reverting back, the flow of arbitrageurs out of market L and into market S persists after the shock is removed. This "flight-to-liquidity" continues as the economy transitions from the high efficiency regime to the low efficiency regime that features low values for δ and ξ.[14]

Figure 4 shows the responses to shocks to z_L at two different initial levels of active capital ξ (but with an identical duration of shocks). As in Figure 3, active capital and price efficiency in market L decrease as a result of the shocks and the ensuing capital flow out of market L. But the resilience of the market depends on the current level of active capital. Intuitively, a reduction in price efficiency in market L has a stronger effect when the mass of locked-in arbitrageurs is relatively high (equivalently, when active capital ξ is relatively low). As shown in Panel (a), when the initial level of active capital is high relative to the critical threshold, the market is resilient and the level of ξ reverts back to its level in the risky steady state with high information efficiency. On the other hand, with a relatively small initial amount of active capital, the market is fragile; the outflow of arbitrageurs from market L persists after the shock is removed as active capital crosses the critical threshold and the economy transitions to the low efficiency regime.

With the market experiencing the low efficiency regime in both examples (as illustrated in Figure 3 and Figure 4), price efficiency in market L deteriorates unambiguously in the aftermath of the shock because of two effects: (i) the reduction in ξ which is the stock of active capital, and (ii) the reduction in δ which is the flow of active capital to market L. On the other hand, these two effects move in opposite directions in the case of market S. The shock in market L leads to contagion to market S by reducing the level of active capital ξ, but also leads to flight-to-liquidity by increasing the capital flow $1 - \delta$ to market S. Therefore, the

[14]We call this flight-to-liquidity because capital is flowing to the short-term asset. We discuss this further in Section 6.



(a) Active capital

(b) Investment in market L

(c) Price efficiency in market L

(d) Price efficiency in market S

Figure 4: **Transitional Dynamics for a Temporary Shock under Different Levels of the Initial Active Capital.** A short-duration shock (duration of 6 periods) is given at $t = 0$ with a higher active capital level (initial value of $\xi = .73$, solid line) and a lower active capital level (initial value of $\xi = .56$, dashed line). Parameter values: $q = .01, z_L \in \{.6, .65, .7\}, z_S = .475, \beta = .95$. Transition probabilities are given by $\omega_{11} = .46, \omega_{12} = .54, \omega_{13} = 0, \omega_{21} = .12, \omega_{22} = .76, \omega_{23} = .12, \omega_{31} = 0, \omega_{32} = .5, \omega_{33} = .5$ where states 1,2 and 3 correspond to low, normal and high level of z_L, respectively.

direction of change in λ_S depends on the magnitude of the two effects. In these numerical examples (dashed lines in panel (d) in Figure 3 and Figure 4), the flow effect (increase in $1-\delta$) dominates the level effect (decrease in ξ), thus, increasing price efficiency in market S instead of decreasing it. This illustrates Corollary 1.

Along the equilibrium path, the market can move in and out of the low efficiency regime. We illustrate this in Figure 5, which shows a simulation of the stochastic model. The occurrence

(a) Active capital

(b) Investment in market L

(c) Price efficiency

(d) noise trading intensity in market L

Figure 5: **Simulation.** Parameter values: $q = .01, z_L \in \{.6, .65, .7\}, z_S = .475, \beta = .95$. Transition probabilities are given by $\omega_{11} = .46, \omega_{12} = .54, \omega_{13} = 0, \omega_{21} = .12, \omega_{22} = .76, \omega_{23} = .12, \omega_{31} = 0, \omega_{32} = .5, \omega_{33} = .5$ where states 1,2 and 3 correspond to low, normal and high level of z_L, respectively.

of temporary shocks does not have persistent effects in the first portion of the simulation. It is only when 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 efficiency regime for market L (white area) is followed by a low efficiency regime (shaded area) after the occurrence of a sequence of shocks in which noise trading intensity in market L increases. The economy is therefore trapped in this regime for many periods even though noise trading intensity has long since reverted to its normal level. It takes a sequence of opposite shocks where noise trading intensity falls below its normal level for the economy to revert to the high efficiency regime for market L. Along the transition, capital flows to market L, thereby improving price efficiency in this market. As a result locked-in capital is released at a faster rate, further increasing price efficiency.

6 Discussion

6.1 Empirical Proxy for Active Capital

In our model, the mass of active capital is the key state variable that determines price efficiency as well as market resilience to shocks (Figure 4). The measurement of active capital itself may be empirically challenging, but our model suggests that the cross-sectional difference in price efficiency can be used as a proxy for the level of active capital.[15]

In equilibrium, the cross-sectional difference in efficiency today is related to the expectation of efficiency in the future. To see this, we can use recursive substitution and Eq. (26) to obtain the following equation:

$$\lambda_{S,t} - \lambda_{L,t} = \mathrm{E}\left[\sum_{\tau=1}^{\infty} \beta^{\tau} \left(1-q\right)^{\tau} \prod_{j=0}^{\tau-1} \left(1-\lambda_{L,t+j}\right)\left(1-\lambda_{S,t+\tau}\right)\right], \tag{28}$$

where $\lambda_{h,t}$ is price efficiency in market h at time t. The LHS of Eq. (28) captures the difference in mispricing across markets in the current period, and the RHS captures the opportunity cost arising from future inefficiency in market L, which is the expected loss in future speculative profits because of locked-in capital. Eq. (28) has the following cross-sectional and dynamic implications. First, there should be larger mispricing for long-term assets than for short-term assets. Second, a large difference in mispricing between long-term and short-term assets

[15] Building a direct measure of active capital, which is the difference between total capital and locked-in capital, is challenging because one has to find measures for both total capital and locked-in capital. One possibility is to use the available capital of financial institutions (such as primary dealers' market equity capital ratio, as in He, Kelly, and Manela (2017)) as a proxy for active capital. Another possibility is to measure portfolio turnover (active capital) and identify long-term portfolios (locked-in capital). For example, one could use transaction data (von Beschwitz, Schmidt, and Lunghi (2019)) or regulatory disclosure such as 13F filings (Koijen and Yogo (2019)) to estimate long-term holdings as well as portfolio turnover. A further challenge of using these alternative measures for our purposes is that they only measure changes in the availability of capital and not changes in investment opportunities (i.e., the degree of mispricing). Therefore, our suggested measure based on cross-sectional differences in price efficiency seems the most practical choice for testing predictions of our model.

predicts slower convergence of price to fundamental in the future in the long-term market.

As an empirically-implementable measure, we propose the standard deviation of price efficiency across assets, which does not require us to identify the cash-flow duration of assets:

$$D_t \equiv Stdev(\lambda_{i,t}) = \sqrt{\sum_{i=1}^{N} \frac{1}{N}\left(\lambda_{i,t} - \frac{\sum_{i=1}^{N}\lambda_{i,t}}{N}\right)^2}, \tag{29}$$

where $\lambda_{i,t}$ is the price efficiency measure for asset i at time t and N is the number assets. The idea behind this measure is that price efficiency is more dispersed across assets when there is less active capital, since in our model D_t is proportional to the difference in price efficiency between short-term and long-term assets:

$$D_t = \sqrt{\int_i \frac{1}{m}\left(\lambda_{i,t} - \frac{\int_i \lambda_{i,t} di}{m}\right)^2 di} = \frac{\lambda_{S,t} - \lambda_{L,t}}{2}, \tag{30}$$

where $m \equiv \int_i di$ is the total mass of traded assets. By Corollary 1, the difference in price efficiency $\lambda_{S,t} - \lambda_{L,t}$ is inversely related to active capital. Therefore, the dispersion in efficiency D_t can serve as an empirical proxy for active capital. Note that it is an inverse measure: dispersion in price efficiency is higher with less active capital.

There are various empirical price efficiency/mispricing measures in the literature. For example, there are measures based on anomalies in terms of standard factor models (e.g., Stambaugh and Yuan (2017), von Beschwitz, Schmidt, and Lunghi (2019)), index future basis (e.g., Roll, Schwartz, and Subrahmanyam (2007)), non-random walk component in price (e.g., Hasbrouck (1993)), price delay (e.g., Hou and Moskowitz (2005), Saffi and Sigurdsson (2011)), return predictability from order imbalances (e.g., Chordia, Roll, and Subrahmanyam (2005)), price deviation from valuation models (e.g., Doukas, Kim, and Pantzalis (2010)), and violations of parities (e.g., Rosenthal and Young (1990), Lee, Shleifer, and Thaler (1991)).

These tests refer to mispricing with respect to public information. Our model is a model of private information, but the boundary between public and private information is blurred. Implementing a textbook arbitrage often requires specialized knowledge. For example, a standard formula relates convertible bond values to the underlying stock and bond, but implementing it requires checking the covenants, having specialized knowledge about volatility, and executing the trade cheaply. It is likely that only a few traders have this knowledge. The same logic applies to any option.

In the following subsections, we offer our interpretations of empirical phenomena based on our results. We also provide some testable empirical predictions based on our theory using these proxies for active capital.

6.2 Slow-Moving Capital

There are several well-known episodes of crises 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 price efficiency in the aftermath (e.g., Mitchell, Pedersen, and Pulvino (2007), Coval and Stafford (2007)). Existing literature often explains those crises as a result of shock amplifications which impair capital itself.[16] In our model, a 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 crisis is merely a transient shock which causes arbitrage capital to get redeployed more slowly; this can trigger a change in regime and have a long-lasting impact. At the core of this argument lies the multiplicity of risky steady states; a sufficiently large shock can disturb the system enough to put the state variable (active capital) on another path. This mechanism allows us to give a distinctive prediction that the market may be shifted toward low liquidity as a result of shocks. This prediction matches empirical observations of long periods of inefficiency in the market.

Our results in Section 5.2 show that the level of active capital endogenously follows a two-state regime-switching process in which the regime is determined by the level of active capital (see Panel (b) of Figure 2). Furthermore, in our model the probability of switching to a different regime depends on the level of active capital. For example, Figure 4 shows that a switch from high to low price efficiency regime as a result of shocks is less likely to occur

[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)).

for higher levels of active capital. (Similarly, a switch from low to high efficiency is less likely for lower levels of active capital.) Because the dispersion in price efficiency D_t is an empirical proxy for the mass of active capital ξ_t, it should follow similar dynamics as ξ_t. This hypothesis can be tested with regime-switching analysis (e.g., Hamilton (1989), Ang and Bekaert (2002), Watanabe and Watanabe (2008)) and in particular with regime-switching models featuring time-varying transition probabilities as in Diebold, Lee, and Weinbach (1994).

Empirical prediction 1. (Price efficiency regimes) The dispersion of price efficiency D_t follows a two-state regime-switching process with time varying transition probabilities which depend on the level of D_t.

Our results suggest that price efficiency is generally low for long-term assets compared to short-term assets. Given this result, we can consider using average price efficiency of an asset as a proxy for its payoff duration. For example, we can sort assets into different groups (such as deciles) based on the average of an empirical measure of price efficiency. In a regime with a smaller amount of active capital, long-term assets (or assets with low average efficiency) would suffer a longer period of price inefficiency than short-term assets.

Empirical prediction 2. (Slow-moving capital) A higher level of dispersion in price efficiency D_t predicts slower convergence to fundamentals for long-term assets (or low-average-efficiency assets.)

Furthermore, our model suggests increased price efficiency for short-term assets in a regime with low price efficiency.

Empirical prediction 3. (Flight-to-liquidity) A higher level of dispersion in price efficiency D_t predicts faster convergence to fundamentals for short-term assets (or high-average-price-efficiency assets.)

Upon the arrival of a shock, arbitrageurs may optimally choose to invest in the short-term (more liquid) market. Therefore, capital tends to flow out of the long-term market, and this response becomes particularly persistent and magnified if it involves a regime shift. These predictions are broadly consistent with several patterns identified in the empirical literature in various markets. Acharya, Amihud, and Bharath (2013) document the existence of liquid and illiquid regimes for corporate bonds. In particular, they find that flight-to-liquidity happens in conjunction with the illiquid regime during which prices of investment-grade bonds rise while prices of speculative-grade bonds fall.[17]

6.3 Welfare and Policy Implications

Our model may switch between efficient and inefficient regimes for the long-term assets. In the literature, price inefficiency and associated notions of illiquidity are typically considered (formally or informally) to be welfare reducing.

Prices that more accurately reflect information can be used to guide investment or other long-term decisions such as investment decisions, so long term information is socially valuable.[18] On the other hand short-term information is likely determined by decisions that have already been made, and is therefore not socially valuable. A classic example would be predicting quarterly earnings shortly before their announcement.

This is in line with Hirshleifer (1971)'s discussion of "foreknowledge" versus "discovery." Foreknowledge is finding out something that will become known anyway. He argues that foreknowledge has low social value. Discovery means finding out something that will not otherwise become known. If we fix a given time horizon, discovery relates to information that will not otherwise become known within that time horizon, such as the values of long-term assets in our model. Foreknowledge relates to information about short-term assets, because their values will become known anyway next period.

[17]Beber, Brandt, and Kavajecz (2007) and Ben-Rephael (2017) provide further evidence of flight-to-liquidity in the Euro-area bond market and in the U.S. stock market, respectively.

[18]There is a literature modelling asset price formation when the prices of the assets are themselves affected by those decisions (e.g., Dow and Gorton (1994), Dow and Rahi (2003), Edmans, Goldstein, and Jiang (2015), Dow, Goldstein, and Guembel (2017)). However, asset prices may of course also guide other decisions even when the asset prices are not affected by the decisions. Price efficiency has implications for both allocative efficiency and for risk sharing. The implications for risk sharing can go either way. Risk sharing is not the focus of our paper (for a comprehensive analysis see Dow and Rahi (2000)). Also, we do not consider the distribution of wealth between informed and uninformed traders since this is outside the scope of standard welfare economics.

Hence, we can consider the efficiency of the long-term market as our welfare criterion. We can compare the two regimes, for parameter values where the model displays hysteresis, as illustrated in Figure 1. In one regime, the long-term asset has plentiful arbitrage capital and is efficient. Hence this regime is welfare superior to the regime with low capital in the long-term market, judged by the efficiency of the market for the long-term asset. Sometimes the drop in efficiency in the long-term market is accompanied by a rise in efficiency in the short-term market (see our discussion of Figure 5), but it still seems reasonable to use efficiency in the long-term market to judge welfare. As noted previously, information is likely to be more useful for guiding long-term decisions. In addition, our simulations all show that a regime with much larger efficiency in the long-term market has only slightly lower efficiency in the short-term market. Thus, even if information had equal value in both markets, the switch to the "better" regime in the long-term market has only small costs in the short-term market because of diminishing returns: the short-term market is already more efficient than the long-term market (Lemma 4).

7 Conclusion

In this paper, we have presented a dynamic stationary model of capital-constrained informed trading with short-term and long-term assets. Price efficiency plays a dual role in our model; the mispricing wedge not only determines the profitability of new investment but also determines the speed at which engaged arbitrage capital is released. This creates a feedback channel between active capital and price inefficiency. There is a critical mass of active capital separating different regimes. Active arbitrage capital falls below the critical mass either if an adverse shock persists for long enough, or if, for a given shock, there is a sufficiently low initial level of active arbitrage capital. It may take a long time to revert to an efficient regime from an inefficient regime; it requires a sequence of beneficial shocks strong enough to push the mass of active capital back above the critical mass. Our results shed light on why capital moves slowly, how fast it moves, and to which directions it moves. The results provide empirical implications on cross-sectional as well as dynamic patterns of price efficiency.

Our analysis has policy implications for government intervention following crises such as the 2009 financial crisis or the Covid-19 crisis. A key implication is that a market's ability to recover from a shock is determined by the level of active capital (rather than the total stock of capital) in the market. That is, it is difficult for the market to recover once it transitions to the inefficient regime even when the stock of capital itself is plentiful. This could be helped by policies such as easing capital requirements in banks, or providing cheap financing to financial intermediaries (such as the ECB's "liquidity support" to the banking system). Furthermore, the observable measures we have suggested such as the cross sectional dispersion in price efficiency could serve as indicators for decision making regarding market interventions. A possible extension of our approach could be to study optimal policies, using a policy rule that is within the model.

Appendix A: Discussion of Assumptions and Modeling Choices

In this section, we provide further clarity on the role of our main assumptions.

Financial Assets

The assumption of a continuum of assets simplifies the analysis by reducing the number of state variables. This is because the law of large numbers eliminates the randomness in payoffs as well as the randomness in the revelation of fundamentals in the aggregate level (e.g., Guerrieri and Shimer (2014)).

The assumption that good and bad qualities are equiprobable simplifies the analysis by making profits from long and short positions symmetric (see Section 4.4 where we present arbitrageurs' value functions).

Eq. (1) has the intuitive interpretation that the present values are equalized as mentioned in the text. Eq. (1) is simply a normalization of payoffs. In case we drop this assumption, we just have an extra parameter in the model, the ratio $(V_L^G - V_L^B)/(V_S^G - V_S^B)$, which does not affect our results qualitatively. For expositional convenience, we keep Eq. (1).

In our model, the mass of unrevealed assets in the long-term market is assumed equal to one in each period. If we assume, instead, that unrevealed assets that become fully-revealed are not immediately replaced, the model requires an additional state variable, which is the mass of unrevealed assets, and is considerably more complex to analyze.

Shocks

We want to study how our model responds to shocks. In practical terms, one of the most important shocks to study is a shock to arbitrage capital. However, in our model, this locus for the shock would increase modeling complexity because of two reasons. First, total capital would become an additional state variable. Second, a shock to active capital can influence inferences and create asymmetries between the long and short-term markets. Intuitively, a shock that reduces the amount of capital is similar to a shock that increases the amount of noise. This is because in each market efficiency equals the ratio of active arbitrage capital that is deployed in that market to noise trading intensity in that market (Lemma 3).

Capital constraint

For tractability, our model features exogenous position limits for arbitrageurs ($x_i^a(t) \in \{-1, 0, 1\}$). This assumption has been used in the literature (for example, Gârleanu and Pedersen (2003), Chamley (2007), Goldstein and Guembel (2008), Goldstein, Ozdenoren, and Yuan (2013), and Edmans, Goldstein, and Jiang (2015)).

Endogenizing position limits would not change our results qualitatively. For example, we can consider an endogenous financial constraint similar to that of Dow and Han (2018) by assuming that arbitrageurs have one unit of their own capital, receive a noisy signal on the payoff (instead of a perfect signal), and collateralize their positions with riskless loans from outside financiers. In that case, the position limit becomes $x_i^a(t) \in \{-m, 0, m\}$ where m is the present value of the maximum loss from the position. This means that the mass of arbitrageurs' total and active capital in our model simply needs to be changed by a proportion of m.

Price formation and noise trading

The assumption that noise trading is uniformly distributed with finite support simplifies the price history to be either fully revealing or non-revealing. This feature of the price process simplifies arbitrageurs' equilibrium strategies (Lemma 2) and preserves analytical tractability. Several papers in the literature make similar assumptions on noise trade to yield equilibrium prices that are either fully revealing or non-revealing (for example, Dow and Gorton (1994) and Makinen and Ohl (2015)).

We remark that whereas each asset price can only jump from non-revealing to fully revealing, price efficiency at the market level is a continuous variable. This is because, with a continuum of assets, $\lambda_h(\theta)$ is the fraction of assets in market h whose value is revealed by the trading process. This is a measure of price efficiency in market h, and $\lambda_h(\theta)$ is a continuous function of the state θ.[19] The hysteresis result in our paper depends on arbitrageurs'

[19] More technically, $\lambda_h(\theta)$ is Lipschitz continuous–see the proof of Proposition 1 in the Appendix.

response to changes in price efficiency at the market level, and not on the discrete nature of the equilibrium price process for individual assets.

We have assumed that noise trading and asset payoffs are independent across assets. This assumption implies that there is no cross-asset learning in this model. Similarly, because noise trading is independent over time and there is no partial revelation, an unrevealed asset's past order flows are uninformative. Therefore, market makers do not learn anything about the payoff of an unrevealed asset using other assets' order flows or past order flows.

Early liquidation

In our model, arbitragers never close out an existing position early (Lemma 2). In reality, arbitrageurs may redeploy their capital by closing positions before prices have completely converged (e.g., von Beschwitz, Schmidt, and Lunghi (2019)). They face a complex dynamic optimization problem reflecting the facts that prices may converge gradually and that different trades have different expected profits per unit of capital. This channel is not open in our model, which allows us to maintain tractability.

We have assumed that arbitragers who close out a position early can open a new position in the next period. 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. To see this, notice that with this alternative assumption the payoff $J_a(\theta)$ would replace $\beta \mathrm{E}[J_a(\theta')|\theta]$ in Eq. (17), and therefore the RHS of Eq. (B.11) (see the proof of Lemma 2) would be equal to zero. Furthermore, introducing an arbitrarily small transaction or information acquisition cost would make early liquidation suboptimal.

Appendix B: Proofs for Section 4

The derivation of the value functions in Section 4.4:

We first derive J_S. Because good and bad qualities are equally likely, the continuation value of an active arbitrageur making new investment in market S is given by

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

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

$$\begin{aligned} J_S(\theta; G) &= -(\lambda_S P^G + (1 - \lambda_S) P^0) + \beta \Big[V_S^G + \mathrm{E}[J_a(\dot{\theta})|\theta] \Big]; \\ J_S(\theta; B) &= (\lambda_S P^B + (1 - \lambda_S) P^0) + \beta \Big[-V_S^B + \mathrm{E}[J_a(\dot{\theta})|\theta] \Big]. \end{aligned}$$

Because $-(\lambda_S P^G + (1-\lambda_S)P^0) + \beta V_S^G = (\lambda_S P^B + (1-\lambda_S)P^0) - \beta V_S^B$, it is immediate that $J_S(\theta; G) = J_S(\theta; B)$, thus, we find that $J_S(\theta)$ in Eq. (B.1) is equivalent to the one in Eq. (13).

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

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

where

$$\begin{aligned} J_L(\theta; G) &= -(\lambda_L P^G + (1 - \lambda_L) P^0) + \beta U(\theta; G); \\ J_L(\theta; B) &= (\lambda_L P^B + (1 - \lambda_L) P^0) + \beta U(\theta; B), \end{aligned} \tag{B.3}$$

and

$$\begin{aligned} U(\theta; G) \equiv\, & q V_L^G + (1-q)\lambda_L P^G + (1 - (1-\lambda_L)(1-q)) \mathrm{E}[J_a(\dot{\theta})|\theta] \\ & + (1-\lambda_L)(1-q)\mathrm{E}[J_l(\dot{\theta}; G)|\theta]; \\ U(\theta; B) \equiv & - q V_L^B - (1-q)\lambda_L P^B + (1 - (1-\lambda_L)(1-q)) \mathrm{E}[J_a(\dot{\theta})|\theta] \\ & + (1-\lambda_L)(1-q)\mathrm{E}[J_l(\dot{\theta}; B)|\theta]. \end{aligned} \tag{B.4}$$

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

$$J_l(\theta;k) = \max\left(J_E(\theta;k), J_H(\theta;k)\right), \tag{B.5}$$

where

$$\begin{aligned}
J_E(\theta;G) &= \lambda_L P^G + (1-\lambda_L)P^0 + \beta \mathrm{E}[J_a(\dot{\theta})|\theta];\\
J_E(\theta;B) &= -\lambda_L P^B - (1-\lambda_L)P^0 + \beta \mathrm{E}[J_a(\dot{\theta})|\theta];\\
J_H(\theta;k) &= \beta U(\theta;k).
\end{aligned}$$

It is immediate that $J_E(\theta;G) = J_E(\theta;B) + 2P^0$. Now, we conjecture that

$$U(\theta;G) = U(\theta;B) + \frac{2P^0}{\beta}. \tag{B.6}$$

Then, Eq. (B.6) implies that $J_H(\theta;G) = J_H(\theta;B) + 2P^0$, therefore, using Eq. (B.5) we have

$$J_l(\theta;G) = J_l(\theta;B) + 2P^0. \tag{B.7}$$

Eqs. (B.4) and (B.7) imply that

$$U(\theta;G) - U(\theta;B) = q(V_L^G + V_L^B) + 2(1-q)P^0. \tag{B.8}$$

Because $P^0 = \frac{\beta q}{1-\beta(1-q)}\left(\frac{V_L^G+V_L^B}{2}\right)$, Eq. (B.8) implies that $U(\theta;G) = U(\theta;B) + \frac{2P^0}{\beta}$, which proves that the initial conjecture in Eq. (B.6) is indeed true.

Finally, Eq. (B.3) implies that $J_L(\theta;G) - J_L(\theta;B) = -2P^0 + \beta[U(\theta;G) - U(\theta;B)]$, which in turn implies that $J_L(\theta;G) = J_L(\theta;B)$ due to Eq. (B.6). Therefore, we conclude that $J_L(\theta)$ in Eq. (B.2) is equivalent to the one in Eq. (14). ■

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 and z_i is noise trading intensity for 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. Suppose that the market makers have an initial prior belief of $Pr(v_i = V_h^G) = p_i$ in the beginning of the period, and also 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

$$\hat{p}_i(X_i) = \frac{p_i f_X^i(X_i|G)}{p_i f_X^i(X_i|G) + (1-p_i) f_X^i(X_i|B)}, \tag{B.9}$$

where $f_X^i(\cdot|G)$ 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 $[-z_i, z_i]$ and is independent across periods, the market makers know that 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. (B.9) implies

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

This shows that either the order flow fully reveals the asset's liquidation value, i.e., $\hat{p}_i(X_i) \in \{0, 1\}$, or the order flow realization is uninformative, i.e., $\hat{p}_i(X_i) = p_i$. Hence, for an unrevealed asset the history of past order flows is irrelevant; because good and bad realizations are equally likely, $p_i = \frac{1}{2}$ in every period as long as the asset quality has not been revealed.

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

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

■

Proof of Lemma 2:

We can rewrite the value of holding the existing position $J_H(\theta)$ in Eq. (18) using Eqs. (11)

and (14) as

$$J_H(\theta) = \lambda_L(\theta)P^G + (1 - \lambda_L(\theta))P^0 + J_a(\theta). \tag{B.10}$$

Subtracting Eq. (17) from Eq. (B.10) yields

$$J_H(\theta) - J_E(\theta) = J_a(\theta) - \beta\mathrm{E}[J_a(\theta')|\theta], \tag{B.11}$$

where the RHS represents the opportunity cost from being out of the market for one period as a result of premature liquidation of the position. By Eqs. (11) and (13), this opportunity cost equals $-(\lambda_S(\theta)P^G + (1 - \lambda_S(\theta))P^0) + \beta V_S^G$, which is strictly positive, where $\lambda_S(\theta)P^G + (1 - \lambda_S(\theta))P^0$ is the expected cost paid to invest in a new position in market S, and βV_S^G is the present value of its payoff. ■

Proof of Lemma 3:

In a market-wise symmetric equilibrium, all the future λ_i's are equalized across assets in each market. Then, Eqs. (14) and (13) imply that the continuation value of arbitrageurs making new investment is identical across all assets except for the expected cost of acquiring the position in the current period, which is determined by current period λ_i's. If current period λ_i's are equalized across assets in each market for all asset i in market h, then arbitrageurs are indifferent among all the unrevealed assets in market h. Therefore, it is optimal to randomize across all the unrevealed assets with uniform probability. This equilibrium strategy gives $\mu_i = \delta\xi$ for market L, and $\mu_i = (1 - \delta)\xi$ for market S. Therefore, we have

$$\lambda_L = \min\left(\frac{\delta\xi}{z_L}, 1\right), \quad \lambda_S = \min\left(\frac{(1-\delta)\xi}{z_S}, 1\right). \tag{B.12}$$

Eq. (26) implies that $\lambda_L = 1$ if and only if $\lambda_S = 1$. However, it is impossible to have $\lambda_L = \lambda_S = 1$ because $z_L + z_S > 1$ by assumption. Therefore, we have $\lambda_L < 1$ and $\lambda_S < 1$, which together with Eq. (B.12) implies the desired results in Eqs. (20) and (21). ■

Proof of Lemma 4: Using $P^G = \beta q V_L^G/(1-\beta(1-q))$, we can represent Eq. (14) as

$$J_L(\theta) = \left(P^G - P^0\right)(1-\lambda_L) + \beta(1-\lambda_L)(1-q)\left(\mathrm{E}[J_l(\theta')|\theta] - P^G - \mathrm{E}[J_a(\theta')|\theta]\right) + \beta\mathrm{E}[J_a(\theta')|\theta].$$

Because $J_l(\theta) = J_H(\theta)$ and $J_a(\theta) = J_L(\theta)$ in an interior equilibrium, the above equation together with Eq. (B.10) implies

$$J_L(\theta) = \left(P^G - P^0\right)(1-\lambda_L)\left[1 - \beta(1-q)\left(1 - \mathrm{E}[\lambda_L(\theta')|\theta]\right)\right] + \beta\mathrm{E}[J_a(\theta')|\theta]. \tag{B.13}$$

Similarly, using $P^G = \beta V_S^G$, we can represent Eq. (13) as

$$J_S(\theta) = \left(P^G - P^0\right)(1-\lambda_S) + \beta\mathrm{E}[J_a(\theta')|\theta]. \tag{B.14}$$

Because $J_S(\theta) = J_L(\theta)$ in an interior equilibrium, equating Eqs. (B.13) and (B.14) yields

$$(1-\lambda_S) = (1-\lambda_L)\left[1 - \beta(1-q)\left(1 - \mathrm{E}[\lambda_L(\theta')|\theta]\right)\right],$$

which in turn implies the desired result in Eq. (26). ■

Appendix C: 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.[20]

Notations

Here we introduce some notations used in the appendices. We let $Z \equiv \left\{z_L^1, z_L^2, ..., z_L^N\right\}$ be the set of possible values for noise trading intensity in market L, and let $\bar{z}_L \equiv \max\left\{z_L^1, z_L^2, ..., z_L^N\right\}$

[20] See Fajgelbaum, Schaal, and Taschereau-Dumouchel (2017) for a similar method of proof of existence and uniqueness of stationary equilibrium of a different class of models. There are also other papers (e.g., Follmer, Horst, and Kirman (2005), Acharya and Viswanathan (2011)) that use a related approach.

and $\underline{z}_L \equiv \min\left\{z_L^1, z_L^2, ..., z_L^N\right\}$. Let M be a constant such that $M\left|z_L^n - z_L^m\right| \geq \bar{z}_L - \underline{z}_L$ for all n, m. We denote $\omega\left(z_L^n \middle| z_L^m\right) \equiv \omega_{mn}$ for the transition probability from state m to n. We also let

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

Note that $\alpha = 0$ in case noise trading intensity process z_L is independently and identically distributed, in which case $\omega\left(z_L' \middle| z_L^m\right) - \omega\left(z_L' \middle| z_L^n\right) = 0$ for all $z_L^n, z_L^m, z_L' \in Z$.

We let $\Xi \equiv \left[\underline{\xi}, 1\right]$ be the interval of possible values for ξ where the lower bound $\underline{\xi} = \max\left\{1 - \bar{z}_L \frac{1-\sqrt{q}}{1+\sqrt{q}}, 0\right\}$ is derived in Lemma E.6. We also let $B\left(\Xi \times Z\right)$ be the set of bounded, continuous functions $\lambda : (\xi, z_L) \in \Xi \times Z \to \mathbb{R}$ with the sup-norm $\|\lambda\| = \sup_{\xi \in \Xi, z_L \in Z}\left|\lambda\left(\xi, z_L\right)\right|$.

We can reformulate the indifference condition $J_L\left(\theta\right) = J_S\left(\theta\right)$ in terms of λ_L. Let $\lambda : (\xi, z_L) \in \Xi \times Z \to \mathbb{R}$ be the level of price efficiency in market L given the state variable $\theta = (\xi, z_L)$. Using Eqs. (20) and (21) to substitute out λ_S in Eq. (26) and rearranging, we obtain the following functional equation which should be satisfied in an interior equilibrium:

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

where

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

Note that $\xi' = C(\xi, z_L)$ follows from the law of motion Eq. (10) and the definition for λ_L in Eq. (20).

Definitions and Assumptions

Here we introduce some assumptions needed to ensure existence and uniqueness of stationary equilibrium of our model, and also introduce other definitions related to those assumptions.

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

$$f(u, z_L) \equiv \max\left\{1-q, (1-q)(z_L u - 1), \left(1-\underline{\xi}\right)u, \left(1-\underline{\xi}\right)u + (1-q)(1-z_L u)\right\},$$

and

$$g(u, z_L) \equiv \max\left\{f(u, z_L), (1-q) z_L u\right\},$$

and

$$\Gamma(z_L) \equiv \begin{cases} \frac{z_L}{4}\left(1+\frac{\bar{z}_L}{z_L}\frac{1-\sqrt{q}}{1+\sqrt{q}}\right)^2, & \textit{if } \bar{z}_L\frac{1-\sqrt{q}}{1+\sqrt{q}} \leq z_L; \\ \bar{z}_L\frac{1-\sqrt{q}}{1+\sqrt{q}}, & \textit{otherwise.} \end{cases}$$

Definition C.3 *Define the constant values*

$$\begin{aligned}
\bar{\lambda}_a &\equiv \frac{1}{\beta(1-q)^2 z_S}, \\
\bar{\lambda}_b(z_L) &\equiv \frac{1-\beta(1-q)+\frac{z_L}{z_S}+\sqrt{\left(1-\beta(1-q)+\frac{z_L}{z_S}\right)^2+\frac{4\beta(1-q)\left(1-\underline{\xi}\right)}{z_S}}}{2\beta(1-q)\left(1-\underline{\xi}\right)}, \\
\bar{\lambda}_c(z_L) &\equiv \frac{1-\beta(1-q)(2-q)+\frac{z_L}{z_S}+\sqrt{\left(1-\beta(1-q)(2-q)+\frac{z_L}{z_S}\right)^2+4\beta(1-q)^2\frac{z_L}{z_S}}}{2\beta(1-q)^2 z_L}.
\end{aligned}$$

Definition C.4 *Let* $\tilde{\lambda}_\xi, \hat{\lambda}_\xi, \lambda^*_\xi$ *be the constant values*

$$\tilde{\lambda}_\xi \equiv \min\left\{\bar{\lambda}_a, \min_{z_L\in Z}\bar{\lambda}_b(z_L), \min_{z_L\in Z}\bar{\lambda}_c(z_L)\right\}, \tag{C.3}$$

$$\hat{\lambda}_\xi \equiv \frac{1+\sqrt{1+\frac{4\bar{z}_L}{\beta(1-q)^2 z_S}}}{2\bar{z}_L}, \tag{C.4}$$

and

$$\lambda_\xi^* \equiv \min_{z_L \in Z} \frac{1/B(z_L)-2}{(1-q)\Gamma(z_L)}. \tag{C.5}$$

Let Λ_ξ *be the set*

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

and also let $\bar{\lambda}_\xi$ *be its infimum*

$$\bar{\lambda}_\xi \equiv \inf \Lambda_\xi.$$

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

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

Definition C.5 *Let* $\bar{\lambda}_\gamma$ *be the constant value*

$$\bar{\lambda}_\gamma \equiv \frac{1 - \bar{z}_L \frac{1-\sqrt{q}}{1+\sqrt{q}} - \beta\left(1-q\right) z_S}{\left(z_S + \bar{z}_L\right)\beta\left(1-q\right) z_S \alpha M}.$$

Let Λ_z *be the set*

$$\Lambda_z \equiv \left\{ \lambda_z \in \mathbb{R}^+ \middle| \, \lambda_z \geq \max_{z_L^n, z_L^m \in Z, n \neq m} A(z_L^n)A(z_L^m) + B(z_L^n)\left[\lambda_z\left(1+\alpha M\right) + g\left(\lambda_z, z_L^n\right)\bar{\lambda}_\xi\right], \lambda_z \leq \bar{\lambda}_\gamma \right\}.$$

and let $\bar{\lambda}_z$ *be its infimum*

$$\bar{\lambda}_z \equiv \inf \Lambda_z.$$

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

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

It is easy to verify that Assumptions C.1, C.2 and C.3 can be jointly satisfied. For example, consider the case when q is large enough, or z_S is small enough, or β is small enough, and when α is small enough.

From Eq. (C.1) we define the following mapping:

Definition C.6 *Let $\mathcal{T} : \lambda \in B(\Xi \times Z) \to B(\Xi \times Z)$ be the mapping*

$$\mathcal{T}\lambda(\xi, z_L) \equiv \max\left\{0, A(z_L)\xi - B(z_L)(1-\lambda(\xi, z_L))\left(1 - \sum_{z'_L \in Z} \omega\left(z'_L \middle| z_L\right)\lambda(C(\xi, z_L), z'_L)\right)\right\}.$$

Definition C.7 *Let $\mathcal{S} \subset B(\Xi \times Z)$ be the set of functions $\lambda : (\xi, z_L) \in \Xi \times Z \to \mathbb{R}$ which are bounded below by zero and above by one, and Lipschitz continuous of modulus $\bar{\lambda}_\xi$ in ξ.*

Note that if $\mathcal{T}$ has a strictly positive fixed point in $\mathcal{S}$, such a fixed point satisfies Eq. (C.1) by construction.

Definition C.8 *Let $\mathcal{F}_0 \subset \mathcal{S}$ be the subset of functions in $\mathcal{S}$ that are monotone increasing in ξ.*

We define that λ is decreasing in z_L if $\lambda(\xi, z_L^n) - \lambda(\xi, z_L^m) \leq 0$ for all $\xi \in \Xi$ and $z_L^n, z_L^m \in Z$ such that $z_L^n > z_L^m$, and also define that the rate of change in z_L is bounded by some constant κ if for all $z_L^n, z_L^m \in Z$ we have

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

Definition C.9 *Let $\mathcal{F}_1 \subset \mathcal{F}_0$ be the subset of functions in $\mathcal{F}_0$ that are decreasing in z_L with the rate of change bounded by $\bar{\lambda}_z$.*

Proof of Proposition 1

Recall that λ denotes an element of the set of bounded continuous functions $B(\Xi \times Z)$ whereas λ_L denotes the equilibrium price efficiency function in market L which is a fixed point of the mapping $\mathcal{T}$ defined in Definition C.6. Now, we restate Proposition 1 with the full details:

Proposition 1. *Under Assumptions C.1 and C.2, there exists a unique interior stationary equilibrium. In equilibrium, price efficiency in the long-term market λ_L is monotone increasing* *in active capital ξ. Under the additional Assumption C.3, λ_L is monotone decreasing in noise trading intensity z_L.*

Proof. The proof is divided in seven steps. First, we prove that in an interior equilibrium, price efficiency in the long-term market λ_L must be monotone increasing in active capital ξ. Second, we show that $\mathcal{T}$ maps $\mathcal{F}_0$ into $\mathcal{F}_0$. Third, we prove that $\mathcal{F}_0$ with metric induced by the sup-norm is a complete metric space. Fourth, we prove that $\mathcal{T}$ is a contraction on $\mathcal{F}_0$. By the contraction mapping theorem (see, for example, Theorem 3.2 in Stokey and Lucas (1996)), $\mathcal{T}$ has a unique fixed point in $\mathcal{F}_0$. We denote this fixed point λ_L. Fifth, we show that under Assumption C.2, λ_L is strictly positive and therefore satisfies Eq. (C.1); since λ_L is in $\mathcal{F}_0$, then it is increasing in ξ. Sixth, we show that under Assumption C.3, λ_L is decreasing in z_L. Seventh, we show that all equilibrium functions in Definition 1 can be uniquely recovered given λ_L.

Step 1: Equilibrium λ_L in increasing in ξ.

Lemma E.8 shows that in an interior stationary equilibrium, price efficiency in the long-term market λ_L must be increasing in active capital ξ. Therefore, we can confine the search for an equilibrium among functions in $\mathcal{F}_0$.

Step 2: $\mathcal{T}$ 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 $B(z_L) > 0$ and $A(z_L) \in (0,1)$, then it is immediate from Definition C.6 that $\mathcal{T}\lambda$ is bounded between zero and one. Lemma E.9 shows that under Assumption C.1, $\mathcal{T}\lambda$ is Lipschitz continuous of modulus $\bar{\lambda}_\xi$ in ξ for every $\lambda \in \mathcal{F}_0$. Lemma E.10 shows that under Assumption C.1, $\mathcal{T}\lambda$ is monotone increasing in ξ for every $\lambda \in \mathcal{F}_0$.

Step 3: $\mathcal{F}_0$ with metric induced by the sup-norm 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 $|\lambda^*(\xi_1, z_L) - \lambda_n(\xi_1, z_L)|, |\lambda^*(\xi_2, z_L) - \lambda_n(\xi_2, z_L)| < \varepsilon/2$ for all $n \geq n_0$. Then,

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

and therefore

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

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

Finally, we have

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

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

Step 4: $\mathcal{T}$ is a contraction mapping on $\mathcal{F}_0$.

Lemma E.11 shows that under Assumption C.1, the mapping $\mathcal{T}$ is a contraction on $\mathcal{F}_0$. Then, Steps 2-4 and the Contraction Mapping Theorem imply that $\mathcal{T}$ has a unique fixed point in $\mathcal{F}_0$.

Step 5: λ_L is strictly positive.

It is immediate to verify that Assumption C.2 together with Lemma E.6 implies that $A(z_L)\underline{\xi} - B(z_L) > 0$, and therefore $\mathcal{T}\lambda > 0$ for all $\lambda \in \mathcal{F}_0$ and all $(\xi, z_L) \in \Xi \times Z$. Because λ_L is a fixed point of the $\mathcal{T}$ mapping, λ_L must be a strictly positive function and it satisfies Eq. (C.1) by construction.

Step 6: λ_L is decreasing in z_L.

By Steps 2-4, $\mathcal{F}_0$ with metric induced by the sup-norm is a complete metric space and $\mathcal{T}$:

$\mathcal{F}_0 \to \mathcal{F}_0$ is a contraction mapping with fixed point λ_L. Since $\mathcal{F}_1$ is a closed subset of $\mathcal{F}_0$ and Lemmas E.12 and E.13 imply that, under Assumptions C.1 to C.3, $\mathcal{T}$ maps $\mathcal{F}_1$ into $\mathcal{F}_1$, then $\lambda_L \in \mathcal{F}_1$ (see Corollary 1 in Stokey and Lucas (1996)). By construction, this is decreasing in noise trading intensity z_L.

Step 7: There exists a unique interior stationary equilibrium.

The previous steps prove that in an interior equilibrium there exists a unique function λ_L that satisfies Eq. (C.1). By Lemma 3, given λ_L we can uniquely recover the capital allocation function δ as well as market S price efficiency λ_S. In an in interior equilibrium $J_a(\theta) = J_S(\theta)$, so Eq. (13) gives a functional equation for J_S. Consider the mapping $\mathcal{T}_S : J \in B(\Xi \times Z) \to B(\Xi \times Z)$ given by

$$\mathcal{T}_S J(\xi, z_L) = -\left(\lambda_S(\xi, z_L) P^G + (1 - \lambda_S(\xi, z_L)) P^0\right) + \beta \left[V_S^G + \sum_{z_L' \in Z} \omega\left(z_L' \middle| z_L\right) J(C(\xi, z_L), z_L') \right].$$

It is immediate that $\mathcal{T}_S$ satisfies Blackwell's sufficient conditions for a contraction on $B(\Xi \times Z)$. Hence, given λ_S, $\mathcal{T}_S$ has a unique fixed point $J_S \in B(\Xi \times Z)$ satisfying Eq. (13). Furthermore, in an interior equilibrium $J_a(\theta) = J_L(\theta)$ and $J_l(\theta) = J_H(\theta)$ and therefore Eqs. (14) and (18) give two functional equations for J_L and J_H. Given λ_L and J_a, the same argument as above shows that Eqs. (14) and (18) have a unique solution. This uniquely pins down J_a, J_l, J_L, J_S, J_E, J_H in an interior equilibrium and concludes the proof. ■

Appendix D: Proof of Proposition 2

Lemma D.5 *When* $\frac{z_L}{z_S} + 1 \geq 2\beta(1-q)$, *the IC curve implicitly defines* δ *as an increasing function of* ξ.

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

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

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

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

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

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

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

Proof of Proposition 2: Part (i) Since z_L is constant we denote $J_j(\theta) = J_j(\xi)$ for $j = L, S, f$. Using Eqs. (11)-(18) we can write, in steady state,

$$J_L(\xi) \leq (P^G - P^0)(1-\lambda_L)\left[1 - \beta(1-q)(1-\lambda_L)\right] + \beta J_a(\xi) \tag{D.1}$$

$$J_L(\xi) \geq (P^G - P^0)(1-\lambda_L)\left[1 - \beta(1-q)(1-\lambda_L)\right] + \beta J_L(\xi). \tag{D.2}$$

Suppose that Eq. (IC) is not satisfied. Then, it is one of the two cases: either everyone chooses market L or everyone chooses market S. In the former case, $\delta = 1$ and therefore $\lambda_S = 0$ and $\lambda_L \in (0,1]$. By Eq. (12), $\delta = 1$ is an equilibrium only if $(J_L(\xi) - J_S(\xi))\big|_{\lambda_S=0} \geq 0$. Using Eq. (D.1) we find

$$(J_L(\xi) - J_S(\xi))\big|_{\lambda_S=0} \leq -\left(P^G - P^0\right)\left[\lambda_L + (1-\lambda_L)^2\beta(1-q)\right] < 0,$$

a contradiction.

In the latter case, we have $\delta = 0$ and therefore $\xi = 1$, $\lambda_L = 0$ and $\lambda_S \in (0,1]$. By Eq. (12), $\delta = 0$ is an equilibrium only if $(J_S(1) - J_L(1))\big|_{\lambda_L=0} \geq 0$. Using Eq. (D.2) and $J_S(1) = (P^G - P^0)(1-\lambda_S)/(1-\beta)$, we find

$$(J_S(1) - J_L(1))\big|_{\lambda_L=0} \leq \frac{P^G - P^0}{1-\beta}\left[\beta(1-q) - \min\left\{1, \frac{1}{z_S}\right\}\right] < 0,$$

where the last inequality follows by $\beta(1-q)z_S < 1$, which is implied by Assumption C.2. We conclude that there is no steady state in which $\delta \in \{0,1\}$.

Next, we proceed to show that there exist either one or three interior steady states. We define $\hat{\xi} \equiv \delta\xi$ as the net mass of arbitrageurs who are investing in the long-term market at time t. Likewise, we define $\hat{\delta} \equiv \delta\xi + 1 - \xi$ as the total mass of investors who are investing in the long-term market at time t. Instead of the original problem stated in terms 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

$$\xi = \hat{\xi} + 1 - \hat{\delta}, \quad \delta = \frac{\hat{\xi}}{\hat{\xi} + 1 - \hat{\delta}}, \quad \lambda_L = \frac{\hat{\xi}}{z_L}, \quad \lambda_S = \frac{1 - \hat{\delta}}{z_S}. \tag{D.3}$$

Using Eq. (D.3), Eq. (CM) can be represented as

$$\hat{\delta} = \frac{\hat{\xi}}{q + (1-q)\frac{\hat{\xi}}{z_L}}. \tag{D.4}$$

Likewise, Eq. (IC) can be represented as

$$\frac{1 - \hat{\delta}}{z_S} - \frac{\hat{\xi}}{z_L} = \beta(1-q)\left(1 - \frac{\hat{\xi}}{z_L}\right)^2. \tag{D.5}$$

By substituting Eq. (D.4) into Eq. (D.5), 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

$$\begin{aligned}
a_0 &\equiv q\,(z_L)^3\,(1 - (1-q)\,z_S\beta)\,; \\
a_1 &\equiv -\,(z_L)^2\,(z_L + qz_S - (1-q)\,(1 + (3q-1)z_S\beta)); \\
a_2 &\equiv -z_L z_S(1-q)(1 + (3q-2)\beta); \\
a_3 &\equiv -(1-q)^2 z_S\beta.
\end{aligned}$$

Since $\beta\,(1-q)\,z_S < 1$, then $a_0 > 0$ for all $q > 0$ and therefore $Q\,(0) > 0$. Using the fact that $z_L + z_S > 1$, we can verify that $Q(\min\{1, z_L\}) < 0$, which implies that Q has either one or three real roots in the open interval of $(0, \min\{1, z_L\})$. Each of these roots is an interior steady state equilibrium in which $\delta \in (0,1)$.

Next, we turn to the proof of stability. Proposition 1 implies as a special case that there exists a unique equilibrium price efficiency function $\lambda_L : [\underline{\xi}, 1] \to \mathbb{R}$ satisfying Definition C.6 at the given level of z_L. For notational convenience, we define $\hat{C}(\xi) \equiv C(\xi, z_L)$ as the transition equation Eq. (C.2) over the interval $[\underline{\xi}, 1]$. A solution for the equation $\xi = \hat{C}(\xi)$ is a steady state, and the previous result shows that there can be at most three such solutions on the interval $[\underline{\xi}, 1]$. We call them ξ^s, ξ^m and ξ^l in the order of size. Lemma E.6 implies that $\hat{C}(\underline{\xi}) \geq \underline{\xi}$. Because $B(z_L) > 0$ and $A(z_L) \in (0,1)$, then it is immediate from Definition C.6 that λ_L is strictly less than one, and therefore $\hat{C}(1) < 1$. Because $\hat{C}$ is continuous and $\hat{C}(\underline{\xi}) \geq \underline{\xi}$ and $\hat{C}(1) < 1$, $\hat{C}$ crosses the 45-degree line from above in $[\underline{\xi}, 1]$ at the largest steady state ξ^l, so ξ^l is a risky steady state. Likewise, $\hat{C}$ crosses the 45-degree from above in $[\underline{\xi}, 1]$ at the smallest steady state ξ^s, implying that ξ^s is also a risky steady state. Therefore, if $\hat{C}$ crosses the 45-degree three times in $[\underline{\xi}, 1]$, then it must cross from below at ξ^m, implying ξ^m is an unstable point.

Part (ii) For $q = 1$ we have that $a_2 = a_3 = 0$, so Q has a unique root equal to $x^* = z_L/(z_L + z_S)$. 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

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

If $1 > \frac{3}{4}z_S + z_L$, 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. ■

Proof of Corollary 1: Denote with ξ^s, ξ^l two steady state values of active capital such that $\xi^l > \xi^s$. Proposition 1 implies as a special case that the price efficiency function $\lambda_L\left(\xi^l\right) > \lambda_L\left(\xi^s\right)$. Using Eq. (10) together with Eqs. (20) and (21) for λ_L^* and λ_S^* we can express Eq. (CM) as follows:

$$z_L\lambda_L^* = \left(1 - z_S\lambda_S^*\right)\left(q + (1-q)\lambda_L^*\right). \tag{D.6}$$

Implicit differentiation of Eq. (D.6) shows that λ_S^* is decreasing in λ_L^* along the CM curve. Because in any steady state equilibrium the pair $(\lambda_L^*, \lambda_S^*)$ must satisfy Eq. (D.6) and $\lambda_L\left(\xi^l\right) > \lambda_L\left(\xi^s\right)$, the reverse inequality must hold for λ_S^* across steady states.

Trading volume for asset i is defined in the standard way as the expectation of the absolute value of the order flow, $E\left[|X_i|\right]$. Letting μ_i and z_i denote the mass of arbitrageurs and noise trading intensity in asset i, we have:

$$\begin{aligned}
2E\left[|X_i|\right] &= E\left[|\mu_i + \zeta_i|\right] + E\left[|-\mu_i + \zeta_i|\right] \\
&= \int_{-z}^{z_i} |\mu_i + \zeta_i| \frac{1}{2z_i} d\zeta_i + \int_{-z}^{z_i} |-\mu_i + \zeta_i| \frac{1}{2z_i} d\zeta_i \\
&= \int_{-\mu_i}^{z_i} (\mu_i + \zeta_i) \frac{1}{2z_i} d\zeta_i - \int_{-z_i}^{-\mu_i} (\mu_i + \zeta_i) \frac{1}{2z_i} d\zeta_i + \int_{\mu_i}^{z_i} (-\mu_i + \zeta_i) \frac{1}{2z_i} d\zeta_i - \int_{-z_i}^{\mu_i} (-\mu_i + \zeta_i) \frac{1}{2z_i} d\zeta_i \\
&= \frac{z_i}{2}\left(1 + \lambda_i^2\right)
\end{aligned}$$

We conclude that trading volume in market h is monotonically increasing in λ_h. ■

Appendix E: Auxiliary Lemmas

Lemma E.6 *ξ is bounded from below by*

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

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

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

or equivalently,

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

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

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

■

Lemma E.7 *We have*

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

where the function f is from Definition C.2.

Proof. We first obtain

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

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

$$\lambda(\xi_2,z_L)-\lambda(\xi_1,z_L) = \lambda_\xi\,(\xi_2-\xi_1)\,, \tag{E.1}$$

and therefore

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

Hence, we can write

$$
\begin{aligned}
&(1-\lambda(\xi_1, z_L))\,(C(\xi_2, z_L) - C(\xi_1, z_L)) \\
&= \begin{bmatrix} (1-q)\,(1-\lambda(\xi_1, z_L))\,(1-\lambda(\xi_2, z_L))\,(1-\lambda_\xi z_L) \\ +\lambda_\xi(1-q)\,(1-\lambda(\xi_1, z_L))\,(1-\xi_1 + z_L\lambda(\xi_1, z_L)) \end{bmatrix}(\xi_2 - \xi_1) \qquad \text{(E.2)} \\
&= \left[(1-q)\,(1-\lambda(\xi_1, z_L))\,(1-\lambda(\xi_2, z_L))\,(1-\lambda_\xi z_L) + \lambda_\xi\,(1 - C(\xi_1, z_L))\right](\xi_2 - \xi_1),
\end{aligned}
$$

where in the second line we make use of Eq. (C.2). Using the fact that $\lambda_\xi \in [0, \bar{\lambda}_\xi]$ and $\lambda(\xi, z_L) \in [0,1]$ and $C(\xi_1, z_L) \in [\underline{\xi}, 1]$, it is easy to verify that

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

■

Lemma E.8 *If $\lambda \in \mathcal{S}$ is an interior fixed point of the mapping $\mathcal{T}$ in Definition C.6, then λ is monotone increasing in ξ.*

Proof. The proof is by contradiction. Assume there exist $\xi_2, \xi_1 \in \Xi$ such that $\xi_2 > \xi_1$ and $\lambda(\xi_2, z_L) - \lambda(\xi_1, z_L) < 0$ for some $z_L \in Z$. Since $\lambda(\xi_2, z_L) - \lambda(\xi_1, z_L) < 0$, by Lipschitz continuity of λ there exists a value $x \in [-\bar{\lambda}_\xi, 0)$ such that

$$
\lambda(\xi_2, z_L) - \lambda(\xi_1, z_L) = x\,(\xi_2 - \xi_1) \qquad \text{(E.3)}
$$

We decompose

$$\mathcal{T}\lambda(\xi_2, z_L) - \mathcal{T}\lambda(\xi_1, z_L) = T_1 + T_2 + T_3, \tag{E.4}$$

where

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

For some $\lambda_3 \in [0,1]$ we can write T_2 as

$$T_2 = B(z_L)\left(1 - \lambda_3\right) x \left(\xi_2 - \xi_1\right).$$

By Lipschitz continuity of λ, there is some value $y \in [-\bar{\lambda}_\xi, \bar{\lambda}_\xi]$ such that we can write T_3 as

$$\begin{aligned}
T_3 &= B(z_L)(1 - \lambda(\xi_1, z_L))\left(C(\xi_2, z_L) - C(\xi_1, z_L)\right) y\\
&= B(z_L)\left[(1-q)\left(1 - \lambda(\xi_1, z_L)\right)\left(1 - \lambda(\xi_2, z_L)\right)\left(1 - x z_L\right) + x\left(1 - C(\xi_1, z_L)\right)\right] y(\xi_2 - \xi_1)
\end{aligned}$$

where the second line we used Eq. (E.2) to simplify the term

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

Combining the expressions for T_1, T_2, T_3, we rewrite Eq. (E.4) as

$$\mathcal{T}\lambda(\xi_2, z_L) - \mathcal{T}\lambda(\xi_1, z_L) = H\left(x, \lambda_1, \lambda_2, \lambda_3, \xi_3, y\right)\left(\xi_2 - \xi_1\right) \tag{E.5}$$

where we define

$$\begin{aligned} H\left(x, \lambda_1, \lambda_2, \lambda_3, \xi, y\right) &= A(z_L) + B(z_L)\left(1-\lambda_3\right)x \\ &\quad + B(z_L)\left[\left(1-q\right)\left(1-\lambda_1\right)\left(1-\lambda_2\right)\left(1-xz_L\right) + x\left(1-\xi_3\right)\right]y. \end{aligned}$$

For $\lambda(\xi, z_L)$ to be an interior equilibrium, $\lambda(\xi, z_L)$ must be a fixed point of the $\mathcal{T}$ mapping: the R.H.S. in Eq. (E.3) must be the same as R.H.S. in Eq. (E.5). We will prove that $x < H$ for all $y \in [-\bar{\lambda}_\xi, \bar{\lambda}_\xi]$ and $(\lambda_1, \lambda_2, \lambda_3) \in [0,1]^3$ and $\xi_3 \in [\underline{\xi}, 1]$ and $x \in [-\bar{\lambda}_\xi, 0)$. This will lead to the required contradiction.

Since H is linear in y, then, for fixed values of $(x, \lambda_1, \lambda_2, \lambda_3, \xi_3)$, H achieves its minimum either at $y = -\bar{\lambda}_\xi$ or at $y = \bar{\lambda}_\xi$. Furthermore, for all $x < 0$, we have

$$H\left(x, \lambda_1, \lambda_2, \lambda_3, \xi_3, \bar{\lambda}_\xi\right) \geq A(z_L) + B(z_L)\left[1 + \bar{\lambda}_\xi\left(1 - \underline{\xi}\right)\right]x \equiv H_1(x) \tag{E.6}$$

$$H\left(x, \lambda_1, \lambda_2, \lambda_3, \xi_3, -\bar{\lambda}_\xi\right) \geq A(z_L) + B(z_L)\left[x - \bar{\lambda}_\xi(1-q)\left(1-xz_L\right)\right] \equiv H_2(x) \tag{E.7}$$

It is therefore sufficient to prove that for all $x \in [-\bar{\lambda}_\xi, 0)$,

$$x < \min\left\{H_1(x), H_2(x)\right\}. \tag{E.8}$$

Using the definitions of H_1, H_2 in Eqs. (E.6) and (E.7) it is immediate to verify that $H_1(0) > 0$ and that $H_2(0) \geq 0$ if and only if $\bar{\lambda}_\xi \leq \bar{\lambda}_a$, where $\bar{\lambda}_a$ is defined in Definition C.3 Eq. (C.3). Since H_1, H_2 are linear in x and $0 \leq \min\{H_1(0), H_2(0)\}$, then a sufficient condition for (E.8) to hold for all $x \in [-\bar{\lambda}_\xi, 0)$ is that

$$-\bar{\lambda}_\xi < \min\left\{H_1\left(-\bar{\lambda}_\xi\right), H_2\left(-\bar{\lambda}_\xi\right)\right\}. \tag{E.9}$$

Using the definitions of H_1, H_2 in Eqs. (E.6) and (E.7) yields that $-\bar{\lambda}_\xi < H_1\left(-\bar{\lambda}_\xi\right)$ if and only if $\bar{\lambda}_\xi < \bar{\lambda}_b(z_L)$, where $\bar{\lambda}_b(z_L)$ is defined in Definition C.3. Similarly, we obtain that $-\bar{\lambda}_\xi < H_2\left(-\bar{\lambda}_\xi\right)$ if and only if $\bar{\lambda}_\xi < \bar{\lambda}_c(z_L)$, where $\bar{\lambda}_c(z_L)$ is defined in Definition C.3. Since $\tilde{\lambda}_\xi$ is defined in Eq. (C.3) as the minimum of $\bar{\lambda}_a, \bar{\lambda}_b$ and $\bar{\lambda}_c$ over all $z_L \in Z$ and $\bar{\lambda}_\xi < \tilde{\lambda}_\xi$ by

Definition C.4, then (E.9) holds for all $x \in [-\bar{\lambda}_\xi, 0)$ and $z_L \in Z$. This concludes the proof. ■

Lemma E.9 *Under Assumption C.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

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

where

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

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

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

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

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

and therefore,

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

By Lemma E.7, the previous inequality can be written as

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

Summing up terms, we get

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

Taking the maximum of the RHS over z_L values yields that $\mathcal{T}\lambda$ is Lipschitz continuous of modulus $\bar{\lambda}_\mathcal{T}$ in ξ, where

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

and the function f is as in Definition C.2. Under Assumption C.1, Definition C.4 implies $\bar{\lambda}_\xi \geq \bar{\lambda}_\mathcal{T}$. This concludes the proof. ■

Lemma E.10 *Under Assumption C.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 E.9, there exist $\lambda_{\xi_0}, \lambda_{\xi_1} \in [0, \bar{\lambda}_\xi]$ such that

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

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

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

Using Eq. (E.2) in the proof of Lemma E.7, there exists some $\lambda_\xi \in [0, \bar{\lambda}_\xi]$ such that the above inequality is equivalent to

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

which is satisfied if

$$H(\bar{\lambda}_\xi, z_L) \geq 0,$$

where

$$H(\bar{\lambda}_\xi, z_L) \equiv \min_{\lambda_1, \lambda_2 \in [0,1], x \in [0,1], \lambda_\xi \in [0, \bar{\lambda}_\xi]} A(z_L) + B(z_L) \begin{bmatrix} (1-q)(1-\lambda_1)(1-\lambda_2)(1-\lambda_\xi z_L) \\ +\lambda_\xi (1-x) \end{bmatrix} \bar{\lambda}_\xi.$$

For $\bar{\lambda}_\xi \leq 1/z_L$, it is immediate that $H(\bar{\lambda}_\xi, z_L)$ is positive. For $\bar{\lambda}_\xi > 1/z_L$, $H(\bar{\lambda}_\xi, z_L)$ is positive if

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

or equivalently, if

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

Taking the minimum of the RHS over z_L values in Z yields the expression for $\hat{\lambda}_\xi$ in Definition C.4. Under Assumption C.1, Definition C.4 implies $\bar{\lambda}_\xi \leq \hat{\lambda}_\xi$. This concludes the proof.

■

Lemma E.11 *Under Assumption C.1, the mapping $\mathcal{T}$ is a contraction on $\mathcal{F}_0$.*

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

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

where

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

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

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

First, we have

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

Second, we have

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

Third, using Eq. (C.2) we have

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

and therefore

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

Let $\Gamma\left(z_L\right)$ be the value

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

It is immediate to verify that $\Gamma\left(z_L\right)$ is from Definition C.2. Therefore,

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

Summing up terms, we have

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

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

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

or equivalently, if

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

Under Assumption C.1, $\bar{\lambda}_\xi < \bar{\lambda}^*_\xi$ by Definition C.4. This concludes the proof. ■

Lemma E.12 *Under Assumptions C.1, C.2 and C.3, $T\lambda$ is decreasing in z_L for all $\lambda \in \mathcal{F}_1$.*

Proof. Let $z_L^m > z_L^n$. The difference of $\mathcal{T}\lambda(\xi, z_L)$ with respect to z_L is given by

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

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

$$\begin{aligned}
(A(z_L^n) - A(z_L^m))\xi &= A(z_L^n)A(z_L^m)\xi(z_L^m - z_L^n); \\
(A(z_L^n) - A(z_L^m))\beta(1-q) z_S &= A(z_L^n)B(z_L^m)(z_L^m - z_L^n).
\end{aligned} \tag{E.10}$$

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

$$\lambda(\xi, z_L^n) - \lambda(\xi, z_L^m) = \lambda_z \left(z_L^m - z_L^n\right). \tag{E.11}$$

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

$$\lambda(C(\xi, z_L^n), z_L') - \lambda(C(\xi, z_L^m), z_L') = \lambda_\xi \left(C(\xi, z_L^n) - C(\xi, z_L^m)\right). \tag{E.12}$$

Fourth, because λ is decreasing in z_L, for all $\xi \in \Xi$ we have[21]

$$\alpha \left(\lambda(\xi, \bar{z}_L) - \lambda(\xi, \underline{z}_L)\right) \leq \sum_{z_L' \in Z} \left(\omega\left(z_L' \middle| z_L^n\right) - \omega\left(z_L' \middle| z_L^m\right)\right) \lambda(\xi, z_L') \leq \alpha \left(\lambda(\xi, \underline{z}_L) - \lambda(\xi, \bar{z}_L)\right),$$

and furthermore, because the rate of change of λ in z_L is bounded by $\bar{\lambda}_z$ and $M \left|z_L^n - z_L^m\right| \geq \bar{z}_L - \underline{z}_L$ by definition, then

$$\lambda(C(\xi, z_L^m), \underline{z}_L) - \lambda(C(\xi, z_L^m), \bar{z}_L) \leq \bar{\lambda}_z \left(\bar{z}_L - \underline{z}_L\right) \leq \bar{\lambda}_z M \left|z_L^n - z_L^m\right|.$$

Hence, the above inequalities imply

$$\begin{aligned} &\sum_{z_L' \in Z} \left(\omega\left(z_L' \middle| z_L^n\right) - \omega\left(z_L' \middle| z_L^m\right)\right) \lambda(C(\xi, z_L^m), z_L') \\ &\qquad \in \left[-\alpha M \bar{\lambda}_z \left|z_L^n - z_L^m\right|, \alpha M \bar{\lambda}_z \left|z_L^n - z_L^m\right|\right]. \end{aligned} \tag{E.13}$$

Using Eqs. (E.10)-(E.13) we can rewrite the difference of $\mathcal{T}\lambda(\xi, z_L)$ with respect to z_L as

$$\mathcal{T}\lambda(\xi, z_L^n) - \mathcal{T}\lambda(\xi, z_L^m) = \Pi(\xi, z_L^n, z_L^m)(z_L^m - z_L^n) \tag{E.14}$$

[21] Recall the definition $\alpha \equiv \max_{z_L^n, z_L^m, z_L' \in Z} \left|\omega\left(z_L' \middle| z_L^m\right) - \omega\left(z_L' \middle| z_L^n\right)\right|$.

where

$$\Pi(\xi, z_L^n, z_L^m) = A(z_L^n)A(z_L^m)\xi - A(z_L^n)B(z_L^m)(1-\lambda(\xi, z_L^m))\left(1 - \sum_{z_L' \in Z} \omega\left(z_L' \middle| z_L^m\right)\lambda(C(\xi, z_L^m), z_L')\right)$$
$$+ B(z_L^n)\lambda_z\left(1 - \sum_{z_L' \in Z} \omega\left(z_L' \middle| z_L^m\right)\lambda(C(\xi, z_L^m), z_L')\right)$$
$$+ B(z_L^n)(1-\lambda(\xi, z_L^n))\left[\frac{C(\xi, z_L^n) - C(\xi, z_L^m)}{z_L^m - z_L^n}\lambda_\xi + \chi\right],$$

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

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

$$\begin{aligned}
&C(\xi, z_L^n) - C(\xi, z_L^m) \\
=&(1-q)\left[\begin{array}{c} (\lambda(\xi, z_L^n) - \lambda(\xi, z_L^m))[1-\xi - z_L^n(1-\lambda(\xi, z_L^n) - \lambda(\xi, z_L^m))] \\ +\lambda(\xi, z_L^m)(1-\lambda(\xi, z_L^m))\,(z_L^m - z_L^n) \end{array}\right] \\
=&(1-q)\left[\begin{array}{c} \lambda_z[1-\xi - z_L^n(1-\lambda(\xi, z_L^n) - \lambda(\xi, z_L^m))] \\ +\lambda(\xi, z_L^m)(1-\lambda(\xi, z_L^m)) \end{array}\right](z_L^m - z_L^n),
\end{aligned} \tag{E.15}$$

where the second line makes use of Eq. (E.11). Using Eq. (E.15) we can write Eq. (E.14) as

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

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

$\lambda_\xi \in [0, \bar{\lambda}_\xi]$, we have

$$\begin{aligned} & A(z_L^n)A(z_L^m)\xi + (B(z_L^n)\lambda_z - A(z_L^n)B(z_L^m)(1-\lambda_2))(1-\lambda_3) \\ & + B(z_L^n)(1-\lambda_1)\left[(1-q)\left[\lambda_z[1-\xi-z_L^n(1-\lambda_1-\lambda_2)] + \lambda_2(1-\lambda_2)\right]\lambda_\xi - \alpha M\bar{\lambda}_z\right] \geq 0. \end{aligned}$$

It is easy to verify that the LHS of the above inequality is minimized at $\lambda_1 = \lambda_2 = 0$ for all $\lambda_3 \in [0,1], \xi \in [\underline{\xi}, 1], \lambda_z \in [0, \bar{\lambda}_z]$, $\lambda_\xi \in [0, \bar{\lambda}_\xi]$, which leaves

$$\begin{array}{c} A(z_L^n)A(z_L^m)\xi + (B(z_L^n)\lambda_z - A(z_L^n)B(z_L^m))(1-\lambda_3) \\ +B(z_L^n)\left[(1-q)\lambda_z(1-\xi-z_L^n)\lambda_\xi - \alpha M\bar{\lambda}_z\right] \end{array} \geq 0. \tag{E.16}$$

Next, it is immediate to check that the LHS of Eq. (E.16) is minimized at $\xi = \underline{\xi}$ for all $\lambda_3 \in [0,1], \lambda_z \in [0, \bar{\lambda}_z]$, $\lambda_\xi \in [0, \bar{\lambda}_\xi]$ if the following condition on $\bar{\lambda}_\xi$ holds:

$$\bar{\lambda}_\xi \leq \frac{1}{(\bar{z}_L + z_S)\,\beta\,(1-q)^2\, z_S\bar{\lambda}_z}. \tag{E.17}$$

Hence, if Eq. (E.17) holds, Eq. (E.16) is satisfied if

$$\begin{array}{c} A(z_L^n)A(z_L^m)\underline{\xi} + (B(z_L^n)\lambda_z - A(z_L^n)B(z_L^m))(1-\lambda_3) \\ +B(z_L^n)\left[(1-q)\left(1-\underline{\xi}-z_L^n\right)\lambda_z\lambda_\xi - \alpha M\bar{\lambda}_z\right] \end{array} \geq 0.$$

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

$$\frac{1}{z_S + z_L^m}\left(1 - \bar{z}_L\frac{1-\sqrt{q}}{1+\sqrt{q}}\right) + \beta\,(1-q)\,z_S\left[\begin{array}{c} \left(\lambda_z - \frac{1}{z_S+z_L^m}\right)(1-\lambda_3) \\ -(1-q)\lambda_z\bar{z}_L\frac{2\sqrt{q}}{1+\sqrt{q}}\lambda_\xi - \alpha M\bar{\lambda}_z \end{array}\right] \geq 0.$$

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

$$\frac{1}{z_S + z_L^m}\left(1 - \bar{z}_L\frac{1-\sqrt{q}}{1+\sqrt{q}}\right) - \beta\,(1-q)\,z_S\left(\frac{1}{z_S + z_L^m} + \alpha M\bar{\lambda}_z\right), \tag{E.18}$$

and at $\lambda_z = \bar{\lambda}_z$ the LHS is equal to

$$\frac{1}{z_S + z_L^m}\left(1 - \bar{z}_L \frac{1-\sqrt{q}}{1+\sqrt{q}}\right) + \beta\left(1-q\right) z_S \left[\begin{array}{c} \left(\bar{\lambda}_z - \frac{1}{z_S + z_L^m}\right)\left(1-\lambda_3\right) \\ -\bar{\lambda}_z \left((1-q)\bar{z}_L \frac{2\sqrt{q}}{1+\sqrt{q}} \bar{\lambda}_\xi + \alpha M\right) \end{array}\right]. \tag{E.19}$$

It is immediate that Eq. (E.18) is positive if

$$\bar{\lambda}_z \leq \frac{1 - \bar{z}_L \frac{1-\sqrt{q}}{1+\sqrt{q}} - \beta\left(1-q\right) z_S}{\left(z_S + \bar{z}_L\right) \beta \left(1-q\right) z_S \alpha M}, \tag{E.20}$$

which is satisfied under Assumptions C.2 and C.3. For (E.19), we see that either $\bar{\lambda}_z \leq \frac{1}{z_S + z_L^m}$, in which case Eq. (E.19) is minimized at

$$\frac{1}{z_S + z_L^m}\left[\begin{array}{c} 1 - \bar{z}_L \frac{1-\sqrt{q}}{1+\sqrt{q}} - \beta\left(1-q\right) z_S \\ +\beta\left(1-q\right) z_S \left(1 - \max\left\{(1-q)\bar{z}_L \frac{2\sqrt{q}}{1+\sqrt{q}} \bar{\lambda}_\xi + \alpha M, 1\right\}\right) \end{array}\right], \tag{E.21}$$

which is positive under Assumption C.2 if

$$\bar{\lambda}_\xi \leq \frac{1 - \alpha M}{(1-q)\bar{z}_L \frac{2\sqrt{q}}{1+\sqrt{q}}}, \tag{E.22}$$

or $\bar{\lambda}_z > \frac{1}{z_S + z_L^m}$, in which case Eq. (E.19) is minimized at

$$\frac{1}{z_S + \bar{z}_L}\left[\begin{array}{c} 1 - \bar{z}_L \frac{1-\sqrt{q}}{1+\sqrt{q}} - \beta\left(1-q\right) z_S \\ +\beta\left(1-q\right) z_S \left(1 - \left(z_S + \bar{z}_L\right) \bar{\lambda}_z \left(\left(1-q\right) \bar{z}_L \frac{2\sqrt{q}}{1+\sqrt{q}} \bar{\lambda}_\xi + \alpha M\right)\right) \end{array}\right]. \tag{E.23}$$

Under Assumption C.2, Eq. (E.23) is positive if

$$\bar{\lambda}_\xi \leq \frac{1}{\left(1-q\right) \bar{z}_L \frac{2\sqrt{q}}{1+\sqrt{q}}} \left(\frac{1}{\left(z_S + \bar{z}_L\right) \bar{\lambda}_z} - \alpha M\right). \tag{E.24}$$

Putting together the bounds in Eqs. (E.17), (E.22) and (E.24) gives the inequality in Assumption C.3. ■

Lemma E.13 *Under Assumptions C.1, C.2 and C.3, the rate of change of $\mathcal{T}\lambda$ in z_L is bounded by $\bar{\lambda}_z$ for all $\lambda \in \mathcal{F}_1$.*

Proof. We bound the rate of change of $\mathcal{T}\lambda$ in z. Using Eq. (E.15) and the definition of C in Eq. (C.2) we compute

$$
\begin{aligned}
&(1-\lambda(\xi,z_L^n))(C(\xi,z_L^n)-C(\xi,z_L^m)) \\
=&(1-q)\left[\begin{array}{c} \lambda_z(1-\lambda(\xi,z_L^n))[1-\xi-z_L^n(1-\lambda(\xi,z_L^n)-\lambda(\xi,z_L^m))] \\ +\lambda(\xi,z_L^m)(1-\lambda(\xi,z_L^n))(1-\lambda(\xi,z_L^m)) \end{array}\right](z_L^m-z_L^n) \\
=&(1-q)\left[\begin{array}{c} \lambda_z(1-\lambda(\xi,z_L^n))(1-\xi+z_L^n\lambda(\xi,z_L^n)) \\ +\left(\lambda(\xi,z_L^m)-z_L^n\lambda_z\right)(1-\lambda(\xi,z_L^n))(1-\lambda(\xi,z_L^m)) \end{array}\right](z_L^m-z_L^n) \\
=&\left[\begin{array}{c} \lambda_z(1-C(\xi,z_L^n)) \\ +(1-q)\left(\lambda(\xi,z_L^m)-z_L^n\lambda_z\right)(1-\lambda(\xi,z_L^n))(1-\lambda(\xi,z_L^m)) \end{array}\right](z_L^m-z_L^n)\,.
\end{aligned}
\tag{E.25}
$$

Using Eqs. (E.14) and (E.25), we have:

$$
\begin{aligned}
&|\mathcal{T}\lambda(\xi,z_L^n)-\mathcal{T}\lambda(\xi,z_L^m)| \\
&\leq A(z_L^n)A(z_L^m)\left|z_L^m-z_L^n\right| + B(z_L^n)\lambda_z\left(1-\sum_{z'_L\in Z}\omega\left(z'_L\middle| z_L^m\right)\lambda(C(\xi,z_L^m),z'_L)\right)\left|z_L^m-z_L^n\right| \\
&\quad + B(z_L^n)G(\xi,z_L^m,z_L^n)\lambda_\xi\left|z_L^m-z_L^n\right| + B(z_L^n)(1-\lambda(\xi,z_L^n))\bar{\lambda}_z\alpha M\left|z_L^m-z_L^n\right| \\
&\leq A(z_L^n)A(z_L^m)\left|z_L^m-z_L^n\right| + B(z_L^n)\left[\bar{\lambda}_z\left(1+\alpha M\right)+G(\xi,z_L^m,z_L^n)\bar{\lambda}_\xi\right]\left|z_L^m-z_L^n\right|,
\end{aligned}
$$

where

$$
G(\xi,z_L^m,z_L^n)\equiv\left|\lambda_z(1-C(\xi,z_L^n))+(1-q)\left[\lambda(\xi,z_L^m)-z_L^n\lambda_z\right](1-\lambda(\xi,z_L^n))(1-\lambda(\xi,z_L^m))\right|.
$$

From Definition C.2, we have

$$
g\left(\bar{\lambda}_z,z_L^n\right)=\max\left\{1-q,(1-q)\,z_L^n\bar{\lambda}_z,\bar{\lambda}_z\left(1-\underline{\xi}\right),\bar{\lambda}_z\left(1-\underline{\xi}\right)+(1-q)\left(1-z_L^n\bar{\lambda}_z\right)\right\}.
$$

Then, it is immediate that

$$G(\xi, z_L^m, z_L^n) \leq g\left(\bar{\lambda}_z, z_L^n\right).$$

Therefore, we have

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

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

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

Under Assumption C.3, Definition C.5 implies $\bar{\lambda}_z \leq \bar{\lambda}_\zeta$. This concludes the proof. ■

References

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.

Allen, F., and D. Gale, 1994, "Limited market participation and volatility of asset prices," *American Economic Review*, 84(4), 933–955.

Ang, A., and G. Bekaert, 2002, "Regime switches in interest rates," *Journal of Business and Economic Statistics*, 20(2), 163–182.

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.

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

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

Buss, A., and B. Dumas, 2018, "The dynamic properties of financial-market equilibrium with trading fees," *Journal of Finance*, Forthcoming.

Cespa, G., and T. Foucault, 2014, "Illiquidity Contagion and Liquidity Crashes," *Review of Financial Studies*, 27(6), 1615–1660.

Chamley, C. P., 2007, "Complementarities in information acquisition with short-term trades," *Theoretical Economics*, 2(4), 441–467.

Chordia, T., R. Roll, and A. Subrahmanyam, 2005, "Evidence on the speed of convergence to market efficiency," *Journal of Financial Economics*, 76(2), 271–292.

Coeurdacier, N., H. Rey, and P. Winant, 2011, "The risky steady state," *American Economic Review*, 101(3), 398–401.

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

Diebold, F. X., L.-H. Lee, and G. Weinbach, 1994, "Regime Switching with Time-Varying Transition Probabilities," in *Nonstationary Time Series Analysis and Cointegration. (Advanced Texts in Econometrics, C.W.J. Granger and G. Mizon, eds.)*. Oxford: Oxford University Press, pp. 283–302.

Doukas, J. A., C. Kim, and C. Pantzalis, 2010, "Arbitrage risk and stock mispricing," *Journal of Financial and Quantitative Analysis*, 45(4), 907–934.

Dow, J., I. Goldstein, and A. Guembel, 2017, "Incentives for information production in markets where prices affect real investment," *Journal of the European Economic Association*, 15(4), 877–909.

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 liquidity," *Journal of Finance*, 73(1), 229–274.

Dow, J., and R. Rahi, 2000, "Should speculators be taxed?," *Journal of Business*, 73(1), 89–107.

——— , 2003, "Informed trading, investment and welfare," *Journal of Business*, 76(3), 439–454.

Duffie, D., and B. Strulovici, 2012, "Capital mobility and asset pricing," *Econometrica*, 80(6), 2469–2509.

Edmans, A., I. Goldstein, and W. Jiang, 2015, "Feedback Effects, Asymmetric Trading, and the Limits to Arbitrage," *The American Economic Review*, 105(12), 3766–3797.

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.

Goldstein, I., and A. Guembel, 2008, "Manipulation and the Allocational Role of Prices," *The Review of Economic Studies*, 1(75), 133–164.

Goldstein, I., Y. Li, and L. Yang, 2014, "Speculation and Hedging in Segmented Markets," *The Review of Financial Studies*, 27(3), 881–922.

Goldstein, I., E. Ozdenoren, and K. Yuan, 2013, "Trading frenzies and their impact on real investment," *Journal of Financial Economics*, 109(2), 566 – 582.

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

——— , 2018, "The dynamics of financially constrained rbitrage," *Journal of Finance*, 73(4), 1713–1750.

Grossman, S. J., and J. E. Stiglitz, 1980, "On the impossiblity of informationally efficient markets," *American Economic Review*, 70(3), 393–408.

Guerrieri, V., and R. Shimer, 2014, "Dynamic Adverse Selection: A Theory of Illiquidity, Fire Sales, and Flight to Quality," *American Economic Review*, 104(7), 1875–1908.

Gârleanu, N., and L. H. Pedersen, 2003, "Adverse Selection and the Required Return," *The Review of Financial Studies*, 17(3), 643–665.

Hamilton, J. D., 1989, "A New Approach to the Economic Analysis of Nonstationary Time Series and the Business Cycle," *Econometrica*, 57(2), 357.

Hasbrouck, J., 1993, "Assessing the quality of a security market: a new approach to transaction- cost measurement," *Review of Financial Studies*, 6(1), 191–212.

He, Z., B. Kelly, and A. Manela, 2017, "Intermediary Asset Pricing: New Evidence from Many Asset Classes," *Journal of Financial Economics*, 126(1), 1–35.

Hirshleifer, J., 1971, "The Private and Social Value of Information and the Reward to Inventive Activity," *The American Economic Review*, 61(4), 561–574.

Hou, K., and T. J. Moskowitz, 2005, "Market frictions, price delay, and the cross-section of expected returns," *Review of Financial Studies*, 18(3), 981–1020.

Koijen, R. S. J., and M. Yogo, 2019, "A Demand System Approach to Asset Pricing," *Journal of Political Economy*, 127(4), 1475–1515.

Kondor, P., 2009, "Risk in dynamic arbitrage: the price effects of convergence trading," *Journal of Finance*, 64(2), 631–655.

Kondor, P., and D. Vayanos, 2018, "Liquidity Risk and the Dynamics of Arbitrage Capital," *Journal of Finance*, Forthcoming.

Kurlat, P., 2018, “Liquidity as social expertise,” *Journal of Finance*, 73(2), 619–656.

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), 1401–1440.

Lee, C. M. C., A. Shleifer, and R. H. Thaler, 1991, “Investor sentiment and the closed-end fund puzzle,” *Journal of Finance*, 46(1), 75–109.

Makinen, T., and B. Ohl, 2015, “Information acquisition and learning from prices over the business cycle,” *Journal of Economic Theory*, 158, 585 – 633, Symposium on Information, Coordination, and Market Frictions.

Mitchell, M., L. H. Pedersen, and T. Pulvino, 2007, “Slow moving capital,” *American Economic Review*, 97(2), 215–220.

Roll, R., E. Schwartz, and A. Subrahmanyam, 2007, “Liquidity and the law of one price: the case of the futures-cash basis,” *Journal of Finance*, 62(5), 2201–2234.

Rosenthal, L., and C. Young, 1990, “The seemingly anomalous price behavior of Royal Dutch/Shell and Unilever N.V./PLC,” *Journal of Financial Economics*, 26(1), 123–141.

Rostek, M., and M. Weretka, 2015, “Dynamic Thin Markets,” *Review of Financial Studies*, 28(10), 2946–2992.

Saffi, P. A. C., and K. Sigurdsson, 2011, “Price efficiency and short selling,” *Review of Financial Studies*, 24(3), 821–852.

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.

Stambaugh, R. F., and Y. Yuan, 2017, "Mispricing factors," *Review of Financial Studies*, 30(4), 1270–1315.

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

von Beschwitz, B., D. Schmidt, and S. Lunghi, 2019, "Fundamental Trading under the Microscope: Evidence from detailed Hedge Fund Transaction Data," Working paper.

Watanabe, A., and M. Watanabe, 2008, "Time-varying liquidity risk and the cross section of stock returns," *Review of Financial Studies*, 21(6), 2449–2486.

Yuan, K., 2005, "Asymmetric price movements and borrowing constraints: A rational expectations equilibrium model of crises, contagion, and confusion," *Journal of Finance*, 60(1), 379–411.