Tech-Driven Intermediation in the Originate-to-Distribute Model*

Zhiguo He Sheila Jiang Douglas Xu

February 11, 2026

Abstract

This paper develops a general equilibrium model to examine the role of information technology when intermediaries facilitate the origination and distribution of assets given information asymmetry. Information technology measures the informativeness of asset-quality signals received by intermediaries, who purchase assets produced by originators and then resell them to uninformed investors. Allowing intermediaries to operate has a mixed social welfare effect: Uninformed intermediation can be welfare reducing when adverse selection is severe in the economy, while informed intermediation always improves social welfare.

Keywords: Information Technology, Intermediation, Directed Search, Adverse Selection, Originate to Distribute, Monitoring the Monitor.

JEL codes: D52, D82, G21, G23, O33

*He: Graduate School of Business, Stanford University and NBER, hezhg@stanford.edu; Jiang: Warrington College of Business, University of Florida, sheila.jiang@warrington.ufl.edu; Xu: Warrington College of Business, University of Florida, douglas.xu@warrington.ufl.edu. For helpful comments, we thank Doug Diamond, Willi Fuchs, Victoria Ivashina, David Hirshelifer, and seminar participants at Chicago Booth finance seminar, HBS-Sloan joint seminar, Stanford FRILL Lunch Seminar, USC Marshall School of Business, HKUST Guangzhou campus, ESSFM 2024 at Gerzensee, and University of Florida. Zhiguo He acknowledges financial support from the John E. Jeuck Endowment at the University of Chicago Booth School of Business as the majority of this paper was completed while He was working at University of Chicago. All errors are our own.

1. Introduction

The past two decades have witnessed a significant shift in financial intermediation, marked by the growing importance of intermediation in traditional *originate-to-distribute* businesses. For instance, as illustrated in Figure 1, increasing volume (panel a) and share (panel b) of asset backed securities (ABS, broadly defined by Finsight to include CLOs, credit card loans, equipment loans and CMBS, etc) are being sold via financial intermediaries (such as Apollo Global Management Inc and Blackstone), as opposed to by originators (such as Bank of America, Fleet Mortgages Ltd, and LoanDepot.com LLC) of the underlying financial claims.[1] In particular, collateralized loan obligations (CLO), a representative type of ABS with corporate loans repackaged and sold by intermediaries like Blackstone, rose from less than 25% to more than 50% of total ABS outstanding between 2000 and 2020 (Bord and Santos, 2012; Cordell et al., 2023). Consistent with this trend, nonbank financial institutions like private equity and asset management firms are increasingly involved in the originate-to-distribute model as commercial banks endeavor to align with stricter regulatory capital requirements (Jiang et al., 2020).[2] This extra layer of *intermediation* amplifies the significance of the originate-to-distribute model.

In the meantime, the growing prominence of such intermediated originate-to-distribute model is accompanied by increased intensity in financial intermediaries' technology investments. Based on data from Pitchbook, total venture capital investment involving financial intermediaries has surged from around \$20 billion in the early 2000s to more than \$200 billion in 2021.[3]

Despite the clear comovement between information technology and intermediation, the economics behind intermediaries' increasingly important role in the originate-to-distribute model is unclear. After all, if investors were to trade with someone who has superior information, there is no obvious explanation for why an informed intermediary should be a more "credible" seller than the originator.[4] More specifically, if investors are concerned

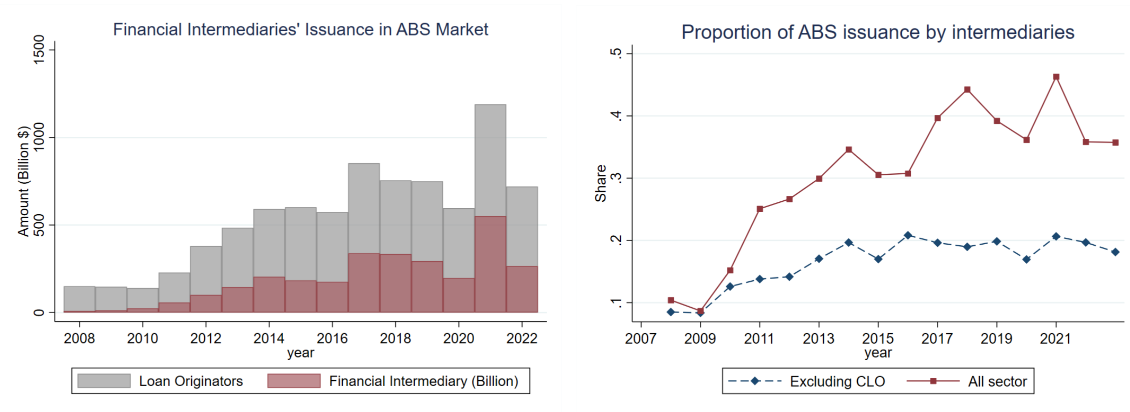
[1]Interestingly, there are some overlaps between originators and intermediaries in the ABS market. For instance, JP Morgan Chase & Co is an originator of CLOs' underlying corporate loans, and in the MBS market, it shows up as a financial intermediary selling mortgage loans originated by other financial institutions.

[2]The article from Wall Street Journal also describes these trending "risk-unloading" transactions between commercial banks and private fund managers. Relatedly, Jiang (2023) documents that banks finance shadow banks who conduct loan origination in the mortgage market.

[3]We aggregate the dollar amount of all VC investment deals with participation by financial intermediaries in Pitchbook. Similar to Figure 1, we classify financial intermediaries as financial institutions with 2-digit SIC code between 62 and 67; and start-ups are defined as fintech start-ups if their Pitchbook industry "verticals" belong to "Artificial Intelligence & Machine Learning," "Big Data," "CloudTech & DevOps," "Fintech," "SaaS," and "TMT." Deal types include those in venture capital, M&A, private equity, and IPO.

[4]While the continuation value driven by intermediaries' reputation concerns might be a candidate explanation, the same logic should also apply to the original asset producers (or claim issuers) had they also

Fig. 1. Growing Share of Financial Intermediaries in the Asset Backed Securities Market



(a) (b)

Source: Finsight, Zoominfo. The figure shows the growing share of financial intermediaries (as opposed to originators) as issuers in the ABS market. Categories of ABS in the figure include student loans, credit card loans, CLOs, auto loans, equipment loans, CMBS, RMBS and esoteric receivables. We take the identity information of issuer from Finsight and classify them into two categories. *Financial intermediaries* are financial institutions with 2-digit SIC code between 62 and 67 (such as Blackstone). *Loan originators* include financial institutions with SIC code 60 (depository institutions, say Citigroup) and 61 (non-depository institutions, say LendingClub) and non-financial companies with industry codes such as 70 and 51.

that asset originators may exploit information gaps by selling them "lemons," they should have similar concerns when purchasing assets from intermediaries who also hold an informational advantage.

By developing a novel general equilibrium framework under information asymmetry, our paper seeks to provide economic insights behind intermediaries' operation in the originate-to-distribute process and its connection with technology development. Our economy consists of an asset origination sector, an intermediary sector, and an investor sector. Assets are produced by agents in the origination sector with a costly retention; high- (low-) type originators produce high- (low-) quality assets. Intermediaries, who are also subject to retention costs, receive (partially) informative signals on asset types; this provides a venue for originators to offload their produced assets, as intermediaries could buy assets with certain "labels" in corresponding signal markets. Finally, in the asset market a group of competitive risk-neutral investors without retention costs buy assets from either originators or intermediaries. As such, a socially efficient arrangement in this economy should only produce high-quality assets and redistribute all produced assets from the origination sector to the investor sector before assets pay off.

Yet an information gap prevents such socially efficient outcomes from being achieved, as outside investors can observe neither underlying asset types, nor signals generated

sought profit maximization over a longer horizon.

by intermediaries. Endogenous production decisions by originators—with privately observed types—worsen information asymmetry, resulting in inefficiencies.

Similar to direct disciplining from trading counterparties in the canonical signaling setting, a market illiquidity–based mechanism provides disciplining in our analysis where an informed seller also makes endogenous production decisions.[5] Following Guerrieri and Shimer (2014), in which adverse selection gives rise to illiquid asset trading with rationing, we model asset market illiquidity as each unit of assets brought to the asset market for sale ends up with $\lambda \in (0,1)$ units of retention. Importantly, this forced (partial) on-balance-sheet retention causes retention costs.[6]

We start with analyzing the market equilibrium in a benchmark "direct trading" economy without intermediaries; this corresponds to the traditional originate-to-distribute model. Two economic inefficiencies plague the market equilibrium in this economy. First, a *production inefficiency* arises as, in equilibrium, both types of originators are producing. While costly retention induced by market illiquidity disciplines low-type originators and hence alleviates the production inefficiency, it entails *allocative inefficiency*—a positive wedge of equilibrium marginal retention costs that exists between two types of originators. Section 3.2 shows that both inefficiencies are fully eliminated when we hypothetically introduce a frictionless intertype market; there, low-type originators who can identify asset quality naturally cease their own production and essentially serve as "intermediaries" between high-type originators and outside investors.

Motivated by the perfect efficiency restoration of intertype trading, Section 3.3 explores a more empirically relevant market solution where intermediaries leverage their signal-generating technology (which generates high or low signals that inform asset types) to serve as middlemen between originators and investors—i.e., an intermediated originate-to-distribute model. We first establish a key property in the equilibrium: assets traded in the signal market(s) must be homogeneous, with either identical asset quality in two signal markets or endogenous closure of the l signal market. Combined, we refer to this result as "the endogenous closure of the lemon signal market," and relate it to the classic "monitoring the monitor" issue (à la banks as delegated monitor in Diamond (1984)) behind intermediaries' operation.

Here is the key intuition. Just as low-type originators would like to produce and sell

[5] In the previous signaling-based literature (Leland and Pyle, 1977; DeMarzo and Duffie, 1999; DeMarzo, 2005; Vanasco, 2017), production is often held as exogenously fixed.

[6] In the situation of CLO arrangement, CLO managers typically retain equity tranches to signal quality to investors or to comply with regulations (Benmelech et al., 2012; Cordell et al., 2023; Kundu, 2023). For instance, the "Risk Retention Rule" applying to CLOs proposed in 2016 once required CLO managers to retain 5% of equity tranches or 5% of all vertical tranches.

lemons in a direct trading economy, intermediaries also have the option to purchase low signal assets (presumably at a lower price) to substitute for high signal ones, as both are sold at the same price in the asset market. However, because (the retention of) both high and low signal assets will sit on a common balance sheet, intermediaries are naturally induced to "cherry pick" in their asset trading. This, in turn, effectively enables them to "commit to" not purchasing any low signal assets from originators unless assets traded in both signal markets are homogeneous.

With the aid of this powerful result, Section 3.3.2 fully characterizes the market equilibria in an intermediated economy and Section 4 reveals a critical role played by the technology development level in this economy. We show that, at the low end of the technology level, intermediaries' technology has no impact on equilibrium outcomes—the intermediated equilibrium falls into a range we refer to as "technology irrelevant." In this range, uninformed intermediaries could potentially harm social welfare, despite the fact that their involvement increases the economy's balance sheet capacity.

Intuitively, the operation of uninformed intermediaries lowers the production efficiency in the economy. In particular, in such tech-irrelevant equilibria, originators' *production wedge* (i.e., the production gap between two types of originators) stays the same as that in a direct trading economy; consequently, the elevated production levels of both originators due to intermediated sales worsens the average quality of the total production. When the lemons problem is severe, this negative effect on production efficiency dominates its positive effect on allocation efficiency as intermediaries bring in more balance sheet capacity to the economy. Furthermore, we are able to provide a sharp parameter condition—both necessary and sufficient—on the severity of the lemons problem, under which "uninformed intermediaries" hurt the surplus of the economy.

Progress in intermediaries' technology starts to transmit into equilibrium outcomes once the technology level surpasses a certain threshold. In such a "technology relevant" range of the intermediated equilibrium, an "*eligible selling*" constraint—that originators cannot sell more in h signal market than the amount of their produced assets that receive the favorable (h) signal—becomes binding for low-type originators (but slack for high-type ones). This asymmetric impact on originators' selling in the signal market translates into an enlarged production wedge in the economy (compared to that in a direct trading economy) and gets further widened as intermediaries' technology improves. As such, the operation of informed intermediaries generates a socially efficient "cleansing" effect on asset production in the economy: the widened production wedge improves average asset quality traded in the asset market, supporting an elevated trading price and total production quantity (hence, input capital price) in the economy, which further disciplines "lemon" production by low-type originators.

We ask a question of important regulatory implication in Section 4.3: if a regulator can add new balance sheet capacity into the economy, should these new balance sheet capacity go to intermediaries or originators? Our analysis delivers a clear policy rule: when intermediaries are relatively uninformed, new capacity should be added to originators to ensure efficient (extra) production; however, when intermediaries are sufficiently informed, it is superior to allocate the additional capacity towards intermediaries as doing so entails a more efficient allocation of the newly produced assets in the system.

We consider extensions of our model along several dimensions and study the implications under these alternative specifications in Section 5. In particular, we provide a microfoundation following Guerrieri and Shimer (2014) for a key economic variable, the market illiquidity λ, which is held as exogenous for most of our analysis. We also consider extensions with imperfectly observed seller occupation identity (so that asset market sales by originators and intermediaries could be executed at separate prices) or indivisible trading in asset market (so market illiquidity amounts to a probabilistically asset sale).

The rest of the article is organized as follows. After a brief literature review, Section 2 introduces our model of (intermediated) asset origination and distribution under information asymmetry where asset sales are subject to market illiquidity. We analyze its market equilibrium in Section 3, and investigate the welfare implications in Section 4. We discuss model extensions in Section 5 and conclude in Section 6.

Related Literature By studying the traditional originate-to-distribute lending model with information asymmetry in a modern context with ever-improving information technology, our analysis in this paper connects several strands of literature.

Signalling models via costly retention. Since the seminal work by Leland and Pyle (1977), an extensive literature has developed on asset sales in the presence of adverse selection, with the common feature that informed asset sellers signal their types via retention.[7] Our analysis departs from this literature in the following important respects.

First, despite a close connection, our notion of market illiquidity differs from that in the classic signaling literature on asset sales under information asymmetry, which typically features an upward-sloping pricing curve against retention under a fully separating equilibrium. Our modeling of illiquidity and trading protocol is closer to the directed search setting developed by Guerrieri and Shimer (2014), in which assets brought to the

[7] To name a few, Gorton and Pennacchi, 1990; DeMarzo and Duffie, 1999; DeMarzo, 2005; Vanasco, 2017; Asriyan et al., 2017; Fuchs et al., 2024. For instance, DeMarzo and Duffie (1999) study the ex ante optimal security design problem for the informed seller who retains certain assets for signaling purposes ex post, and Vanasco (2017) shows that costly retention of cash flows is essential to implement ex ante asset screening.

same market are subject to the same selling illiquidity (modeled as the probability of sale being rationed). This market illiquidity–based mechanism allows us to conduct analysis of asset sales with information asymmetry under environments where other endogenous decisions by asset sellers (say, production) potentially limit sellers' commitment capability—which is key to the retention-based signaling mechanism.

Second, by studying asset originators' endogenous production decisions, our paper explicitly addresses concerns over the "originate-to-distribute" model featured in a large body of both theoretical and empirical studies (Gorton and Pennacchi, 1995; Parlour and Plantin, 2008; Drucker and Puri, 2008; Keys et al., 2010). Notably, in contrast to considering particular contract features (which involve assuming implicit guarantees by the loan originating bank) that mitigate banks' moral hazard when a fraction of their originated loans are sold as in Gorton and Pennacchi (1995), our analysis does not rely on imposing any implicit guarantees (or other such commitments) by asset originators in resolving the moral hazard problem.[8] Relatedly, compared to the typical over-the-counter setting assumed in this literature (say, DeMarzo, 2005; Vanasco, 2017), the Walrasian-style trading protocol adopted by this paper could further exacerbate such incentive problems associated with the production of "lemons." Also, unlike DeMarzo (2005), who studies a monopolistic intermediary, we focus on a competitive intermediary sector and highlight the welfare impact of intermediation when information technology improves.

Third, focusing on pooling equilibria, we investigate the macroeconomic consequences of asset production and trading under adverse selection. In this regard, our paper is closely related to Eisfeldt (2004); Kurlat (2013, 2019); Guerrieri and Shimer (2014), all of whom study pooling equilibria arising in lemon markets to understand the macroeconomic implications of adverse selection.

Financial intermediation. Our paper is also related to the literature on financial intermediation with information asymmetry.[9] When private information ownership is transferred to an "intermediary," the issue of *"monitoring the monitor"* generally arises (Diamond, 1984), which also applies to our setting with (informed) intermediaries.[10] Our analysis differs in that we focus on market-based solutions, rather than contracting, in resolving incentive

[8]Gorton and Pennacchi (1995) consider two contract features in loan sale arrangements that incentivize the bank to provide credit service: (i) an agreement by the bank to sell only part of the loan, and (ii) a guarantee by the bank to repurchase the loan in certain situations.

[9]For instance, Diamond, 1984; DeMarzo, 2005; Glode and Opp, 2016; Dang et al., 2017.

[10]Just as the banker in Diamond (1984) can misreport the cash flow she privately observed to outside investors, informed intermediaries in our analysis also have the option to purchase low signal assets (presumably at low prices) and sell these low signal assets to uninformed buyers, who cannot observe the type/signal of assets purchased by intermediaries.

issues.[11] Mechanism-wise, "cherry-picking" by intermediaries in our model is reminiscent of "winner-picking" by corporate headquarters in Stein (1997).

Unlike Dang et al. (2017) where intermediaries serve the role as "secret keepers" who prevent the dissemination of new information, our analysis includes only one layer of information asymmetry—once the production type shock is realized and assets have been produced, there is no further "state uncertainty" regarding asset payoff, which is crucial for the "Hirshleifer effect" in Dang et al. (2017).[12]

Intermediation chains. By emphasizing an "intermediated" originate-to-distribute model, our paper also relates to the literature on intermediation chains, which includes over-the-counter search models (Atkeson et al., 2015; Hugonnier et al., 2019; Sambalaibat, 2021; Shen et al., 2021) and credit chains (He and Li, 2023; Glode and Opp, 2023).

Like our paper, Glode and Opp (2016) point out that placing a moderately informed intermediary between trading counterparties who are subject to information asymmetry can improve trading efficiency. There are at least two salient differences. First, while Glode and Opp (2016) study a setting involving a monopolistic seller, our analysis is placed in a competitive setting within a macro context where costly retention is a critical ingredient affecting trading and production efficiency in the economy. Second, perhaps more importantly, in Glode and Opp (2016) intermediaries play no role if they are either fully informed or uninformed, while we deliver a monotonically increasing effect of intermediaries' informedness on trading efficiency.

2. An Intermediated Originate-to-Distribute Model

This section introduces our model formally (Section 2.1) and defines the market equilibrium in this economy (Section 2.2). Section 2.3 then solves for the constrained efficient allocation as a benchmark.

2.1. Model Setup

Consider an economy with three dates ($t = 0, \frac{1}{2}, 1$). As shown in Figure 2, the model consists of three distinctive economic sectors: an asset origination sector, an intermediary sector, and an investor sector. Both the asset origination and intermediation sectors are

[11] In Diamond (1984) debt contracts are the solution to incentive provision for delegated monitors to conduct monitoring and report truthfully while incentive provision in Park (2000) hinges on claim seniority.

[12] In Dang et al. (2017), keeping this second layer of information regarding the state realization from being unfolded is the key to ex post insurance between early and late consumers—an effect articulated by Hirshleifer (1971).

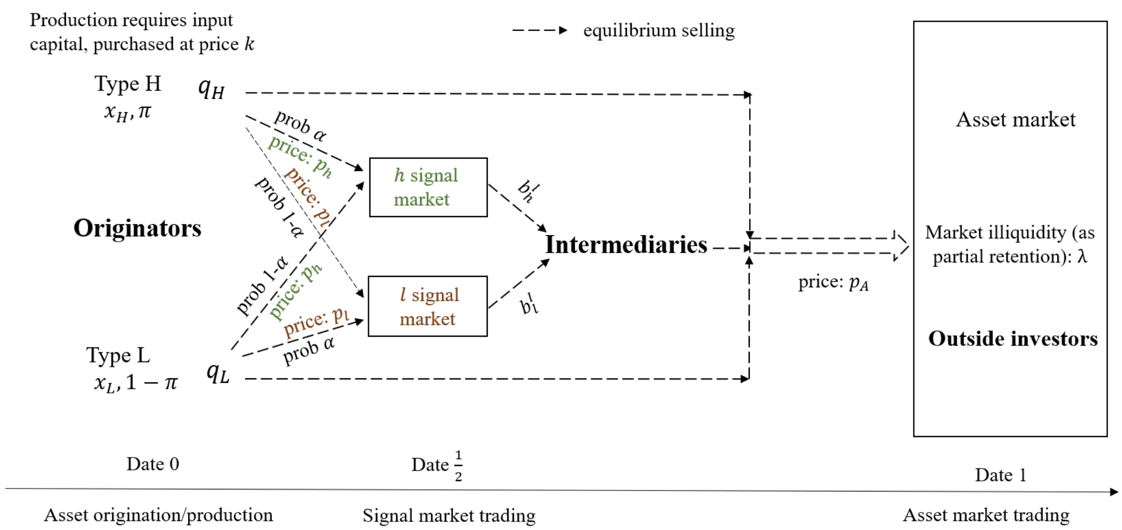


Fig. 2. Model Scheme

run by infinitesimal agents with unit measure.

2.1.1. *Agents, preferences, and technologies*

Asset originators. A unit measure of asset originators are equipped with production technology to originate certain assets that generate payoff on date 1. In our main application these assets are financial assets, although our model could also apply to real goods (e.g., cars, luxury goods).

The economy is populated with two types of originators; a measure of $\pi \in (0,1)$ $(1-\pi \in (0,1))$ are H-type (L-type) originators. Type-θ originators, with $\theta \in \{H, L\}$ can produce assets that generate a per-unit payoff of x_θ at date 1, with $x_H > x_L$. Each unit of asset originators' production requires a unit of input capital.[13] The aggregate supply of input capital in this economy is captured by a convex curve $K(Q)$, where $Q > 0$ is the total input capital with $K'(Q) > 0$, $K''(Q) > 0$, implying a competitive input capital price $k = K'(Q)$.[14] For simplicity we set $K(Q) = Q + \frac{\kappa}{2}Q^2$, such that capital production involves a linear marginal cost (normalized to unity without loss of generality) and a quadratic adjustment cost with $\kappa > 0$ à la Hayashi (1982).

Throughout the analysis, without loss of generality we specify $x_H = X > 1$ and $x_L = 0$ to simplify the notation. Given the capital production function, we know it is socially efficient to let only H-type originators produce in this economy.

[13] In the context of the origination of financial claims/assets, the input capital could be viewed as deposits or other funding obtained from money markets.

[14] The upward sloping supply curve of input capital could be realistically mapped into the setting where the local deposit market serves as the major source of (cheap) funding for loan origination performed by banks operating in a local economy.

We assume that asset originators are subject to an important constraint: it is costly for asset originators to hold the assets they have produced until the moment when asset return is generated. We capture this costly holding of assets for originators by a retention cost function $R(r) = \frac{r^2}{2\rho}$, where r is the amount of assets remaining on the asset originator's own balance sheet before the assets pay off.[15] Practically, this convex asset retention cost can be understood as stemming from inventory cost (Ho and Stoll, 1980), risk aversion (Leland and Pyle, 1977; He and Krishnamurthy, 2013), and/or opportunity cost of capital (DeMarzo and Duffie, 1999; Opler et al., 1999).

The convex holding cost incentivizes originators to sell their assets, with the following two options. They can trade either directly with outside investors in the *asset market* on date 1, or indirectly by going through intermediaries in the *signal markets* on date $\frac{1}{2}$.

Intermediaries and information technology. Facilitating the redistribution of originated assets, the competitive intermediary sector with a unit mass purchases produced assets from originators and then brings them to the asset market for sale; see Figure 2. Like originators, intermediaries incur a quadratic retention cost $R_I(r) = \frac{r^2}{2\rho_I}$ if they retain r on their balance sheets. Here, $\rho_I \geq 0$ denotes the intermediaries' risk-bearing capacity.

Intermediaries in our model specialize in identifying assets types, with a common information technology that generates informative signals on assets produced by originators. Following Hauswald and Marquez (2003) and He et al. (2023), we model the intermediary's information technology by a partially informative signal $j \in \{h, l\}$ for any asset with type $\theta \in \{H, L\}$, with i.i.d. realizations across all assets:

$$\begin{aligned} \alpha_{Hh} &\equiv \Pr(j = h|\theta = H) = \alpha_{Ll} \equiv \Pr(j = l|\theta = L) = \alpha, \\ \alpha_{Hl} &\equiv \Pr(j = l|\theta = H) = \alpha_{Lh} \equiv \Pr(j = h|\theta = L) = 1 - \alpha. \end{aligned} \tag{1}$$

Under this symmetric signal specification, the parameter $\alpha \in [\frac{1}{2}, 1]$ captures the level of information technology in this economy, the scope of which includes but is not limited to big data, machine learning, and artificial intelligence. In our model, a higher α allows intermediaries to generate more accurate signals on asset quality and thus make asset purchases in more informed manners.

Aided by signals, the intermediation sector fosters two potential signal markets in-

[15] This is the certainty equivalent of risk aversion–driven retention costs under a CARA-normal setting as in Leland and Pyle (1977); there, an informed seller with an constant absolute risk aversion $\frac{1}{\rho}$ (or risk-bearing capacity ρ) retains an endogenous fraction of her assets to signal the asset quality of what she is selling. Most of our analysis goes through with a general convex retention cost, except that the quadratic cost does carry one extra property: marginal cost is linear in holdings. This nice analytical feature plays a role later in Section 3.3 when we invoke the sequential equilibrium refinement à la Kreps and Wilson (1982).

dexed by $j \in \{h, l\}$ where assets labelled by the same signal are pooled and sold from originators to intermediaries who then resell to investors in the asset market. Denote the price in the signal market j by $p_j \geq 0$.

Asset buyers. Finally, in the asset market, a group of competitive risk-neutral investors purchase assets sold either by originators or intermediaries. Without retention costs, these investors are the most efficient holders of the originated assets in this economy; this is where the gain from trade lies—as has also been recognized by previous studies of originate-to-distribute lending models. However, unlike Gorton and Pennacchi (1995), we assume that no implicit guarantees (e.g., repurchasing problem loans at prespecified prices) can be granted in asset market transactions. As a result, L-type originators (intermediaries) tend to overproduce (overpurchase) given the absence of contract-based commitments.

2.1.2. *Information structure, trading, and the illiquid date-1 asset market*

As shown in Figure 2, on date 0 an asset originator decides his production size q_θ *after* privately observing his own type θ. On date $\frac{1}{2}$, each asset receives a noisy signal $j \in \{h, l\}$ from the intermediary sector as given in (1) in an i.i.d. manner, so that a fraction $\alpha_{\theta h}$ ($\alpha_{\theta l}$) of the originators' produced assets receive an h (l) signal.[16] The originator then decides to sell $s_{\theta j} \in [0, \alpha_{\theta j} q_\theta]$ of his produced assets labeled by j signal to intermediaries at each signal $j \in \{h, l\}$ market, which leaves $q_\theta - \sum_j s_{\theta j} \geq 0$ in his hand.

We highlight that the asset originator faces an "eligibility constraint in signal market selling" ("eligible selling constraint" for short) that says $s_{\theta j} \leq \alpha_{\theta j} q_\theta$. A higher information technology parameter α, which amounts to decreasing (or increasing) $\alpha_{\theta j}$ as in (1), thus tightens (or loosens) the eligible selling constraint for a θ-type originator when selling in j signal market. As we will show later, such tightening/loosening of eligible selling constraints then affects originators' production decisions, through which the progress in intermediary technology α gets transmitted to equilibrium outcomes.

On date 1, both originators and intermediaries sell their holdings at an endogenous price p_A to outside investors in the asset market, which is illiquid (to be discussed shortly). For a θ-type originator whose retention is $\sum_j r_{\theta j}$ after trading, where $r_{\theta j}$ is his post-trading asset holding with signal j, he has sold $q_\theta - \sum_j s_{\theta j} - \sum_j r_{\theta j} \geq 0$ at price p_A in the asset market. Intermediaries who have purchased $b_{Ij} \geq 0$ in signal markets bring $s_{Ij} \in [0, b_{Ij}]$

[16] The signal is at the asset level—i.e., for a given asset, every intermediary in the entire intermediary sector receives the same signal. This can be justified that the intermediary sector shares similar information technology and allows us to avoid the unnecessary complication that intermediaries face a winner's curse issue in purchasing the assets from originators.

to the asset market for sale, at the same price p_A as originators. We call p_A the asset market price, and we will discuss the assumption of a single selling price for both originators and intermediaries shortly.

Following the spirit of Guerrieri and Shimer (2014), we assume that the asset market on date 1 has uninformed investors as asset buyers and is "illiquid," in the sense that for any unit of asset put on sale, there is some rationing and with an exogenous probability $\lambda > 0$ this asset remains on the seller's balance sheet, causing retention costs. Since traders can always diversify by splitting their orders, this assumption amounts to market illiquidity where for each unit of asset put on sale, sellers have to retain λ fraction of it, capturing the essence of Leland and Pyle (1977). Otherwise, markets operate in a "Walrasian" manner; that is, price-taking sellers (originators or intermediaries) submit their orders for any arbitrary amount of assets for sale.

Our analysis centers around this market illiquidity-based disciplining—modeled in reduced form as a minimum retention fraction of λ per unit of asset brought for sale. In the context of CLOs, such retention includes the equity tranches that are often retained by CLO managers due to signaling or regulatory purposes (Benmelech et al., 2012).[17] Furthermore, the same market illiquidity λ applies to both originators and intermediaries only in the date-1 asset market, while trading between originators and intermediaries on date $\frac{1}{2}$ in the signal markets is free from this friction; they rule out any mechanical (dis)advantage of intermediation.

It is also worthy noting that the frictionless trading between originators and intermediaries might give rise to certain "pathological equilibrium" due to the lack of signaling devices (e.g., the trading illiquidity in asset market), particularly when intermediaries' retention capacity is high enough.[18] For this reason, in our later analysis we impose parametric restriction on ρ_I (specified in Section 2.2) to ensure the immunity to such issues.

Finally, we impose the following two technical assumptions for our equilibrium to be well-behaved, although they are not central to our mechanism. First, for equilibrium uniqueness, we follow the equilibrium refinement approach as in Selten (1975) and Kreps and Wilson (1982) whenever needed. Specifically, we introduce a positive sequence $\{\epsilon\}$ with $\epsilon \to 0$, so that asset sales in each signal market fail with probability $\epsilon > 0$. The

[17] In practice CLO managers often keep equity portion; this could map to an effective retention share that is greater than their dollar-amount retention as equity tranches provides more incentives.

[18] For instance, it is possible that H type originators in equilibrium coordinate within themselves to only sell in the l signal market, while L type sell in both h and l signal markets. There, all assets in the h signal market are lemons and the equilibrium price in l signal assets is *higher* than that in the h market. However, we can rule out this type of pathological equilibrium for sufficiently small ρ_I; in this case, the price in the signal markets will be so low that originators (in particular, H type) would prefer selling directly in the asset market. For more details, see the proof of Lemma 3 in Appendix A3.1.

model with perfectly liquid signal market trading is thus the limiting case of $\epsilon \to 0$. Second, to determine "shadow" prices in signal markets with potentially zero volume in equilibrium, we simply assume that a $\nu \to 0$ measure of originators randomly bring their assets to the corresponding signal market.[19]

2.1.3. *Discussion on model assumptions*

Illiquidity and trading protocol of asset market. The asset market illiquidity in our model, captured by a positive probability of failed asset sale and hence forced retention, is closely related to but differs from those in the previous signaling literature (e.g., Leland and Pyle, 1977; DeMarzo and Duffie, 1999; DeMarzo, 2005; Vanasco, 2017) that focuses on fully separating equilibria. From this perspective, the illiquidity-implied retention in our model can be viewed as a pooling equilibrium in Leland and Pyle (1977) such that both types sell a fraction $1 - \lambda$ of their assets, with the off-equilibrium belief that only low-type deviate. When studying the role of informed intermediaries in a general equilibrium framework, our approach of treating the illiquidity-implied retention as an exogenous parameter offers a clear advantage over the more classic microfounded signaling setting.[20] In particular, exogenously fixing the (minimum) retention fraction per unit of asset allows us to study the aggregate quantity of assets produced—and their quality composition—in the market equilibrium.

As explained, our modeling of illiquidity and trading protocol is closer to the directed search setting developed by Guerrieri and Shimer (2014). As an extension, in Section 5.1 we follow their framework to endogenize market illiquidity λ as the probability of an asset sale being rationed when total demand of assets falls short of total supply. In Section 5.3 we also consider an alternative setting of indivisible assets; this alternative setting is equivalent to a random search framework with matching probability λ and is perhaps more suitable for the application of real goods (as opposed to financial assets).

Asset market trading price. For most of our analysis, we assume that the identities of asset sellers (i.e., originators or intermediaries) are concealed from buyers. Consequently, both sellers are pooled, and investors buy assets at the same price denoted by p_A.[21] In

[19] They behave like "noise traders" and can be justified by some unmodeled trading demand. Note, because the produced quantities by two types of originators are endogenous, the "shadow" prices are endogenous as well.

[20] In this literature, the retention-based signaling mechanism crucially relies on the seller's commitment technology, which is restrictive given the seller's other endogenous decisions in our macro-based economy. For instance, with endogenous production, a retention-based signaling mechanism becomes ineffective when uninformed buyers cannot directly observe the total assets owned by the seller. More generally, the point of unobserved endowment is highlighted by He (2009) who studies sales of two correlated assets.

[21] In practice, outside investors in the asset market might not have a good idea about whether an asset seller is behaving as an originator or an intermediary; our baseline can be considered this friction's extreme

Section 5.2 we relax this assumption by considering a generalized setting where seller identity is (imperfectly) observed in asset market trading. More specifically, we assume that with certain exogenous probability an intermediary's selling could be executed at the price p_A, which pools with orders from originators as in our benchmark analysis, or at a separate selling price p_A^I only for intermediaries.

"Signal" generated by intermediaries. Because we are interested in how intermediaries (with the aid of information technology) overcome the classic "monitoring the monitor" problem à la Diamond (1984), we assume that "signals" generated by intermediaries in our model cannot be observed by outside investors. In this regard, such signals are best interpreted as "internal ratings" (yet highly correlated among intermediaries given the common technology) that guide intermediaries' purchasing decisions. For example, in the ABS market, intermediaries develop so-called proprietary credit scoring models that leverage alternative data and machine learning tools to evaluate the default probability of underlying loans.

The role played by intermediaries in our model differs from that of rating agencies whose ratings are often "public" news in the market. Unlike rating agencies, whose profits come solely from generating and selling information to potential users of the information, intermediaries in our model utilize their own balance sheets to leverage private information they obtain through generating informative signals. Our model can be interpreted as the game conditional on the public rating released by rating agencies.

2.2. *Market Equilibrium*

Given a strategy $\{q_\theta, s_{\theta j}, r_{\theta j}\}_{j\in\{h,l\}}$, a type-$\theta$ originator has an expected payoff of

$$v_\theta\left(\{q_\theta, s_{\theta j}, r_{\theta j}\}_{j\in\{h,l\}}\right) = -\underbrace{q_\theta k}_{\text{date 0 production cost}} + \underbrace{\sum_j p_j s_{\theta j}}_{\text{date } \frac{1}{2} \text{ payoff}} + \underbrace{\left[x_\theta \sum_j r_{\theta j} + \left(q_\theta - \sum_j s_{\theta j} - \sum_j r_{\theta j}\right) p_A - R\left(\sum_j r_{\theta j}\right)\right]}_{\text{date 1 payoff net of retention cost}}, \tag{2}$$

case. Relatedly, as mentioned in footnote 1, the fact that some large commercial banks (e.g., JP Morgan) actively serve as both originator and intermediary in certain loan markets (see this news report) also implies practical difficulties for investors pricing the asset contingent on seller identity. Consistent with this, JP Morgan Chase & Co serves as an originator of the underlying corporate loans of CLOs but simultaneously shows up as financial intermediaries selling mortgage loans originated by other financial institutions in the MBS market, based on our Finsight data for Figure 1.

where he produces q_θ on date 0, sells $s_{\theta j}$ in the signal j market on date $\frac{1}{2}$, and chooses the post-trading retention $r_{\theta j}$ of his signal j assets—by bringing $\frac{1}{1-\lambda}\left(\alpha_{\theta j}q_\theta - s_{\theta j} - r_{\theta j}\right)$ of his remaining j-signal assets for sale in the asset market on date 1.

Throughout the paper, we require that neither originators nor intermediaries can bring more than what they own to the asset market.[22] Given signal generating technology as in Eq. (1) and market illiquidity such that each unit of assets brought to sale in the asset market will lead to λ units of asset retention, this requirement implies a "minimum retention" constraint, $r_{\theta j} \geq \lambda(\alpha_{\theta j}q_\theta - s_{\theta j})$, for originators. Therefore taking the equilibrium prices $\{p_A, p_j, c\}_{j\in\{h,l\}}$ as given, the originator solves

$$v_\theta \equiv \max_{\{q_\theta, s_{\theta j}, r_{\theta j}\}_{j\in\{h,l\}}, r_{\theta j}\geq\lambda(\alpha_{\theta j}q_\theta - s_{\theta j})} v_\theta\left(\{q_\theta, s_{\theta j}, r_{\theta j}\}_{j\in\{h,l\}}\right), \tag{3}$$

Also, we implicitly require all variables (e.g., q_θ, $s_{\theta j}$) to be weakly positive; nonnegativity constraints are omitted to avoid cumbersome notation.

An intermediary, who observes informative signals $\{h, l\}$ on originated assets, chooses her asset purchasing and trading strategy $\{b_{Ij}, r_{Ij}\}_{j\in\{h,l\}}$ in which she purchases b_{Ij} ($j \in \{h, l\}$) from signal market j, and chooses post-trading holding $\{r_{Ij}\}_{j\in\{h,l\}}$ of purchased assets given the illiquid asset market, with a similar minimum retention constraint $r_{Ij} \in [\lambda b_{Ij}, b_{Ij}]$ for both $j \in \{h, l\}$. The payoff to an intermediary with strategy $\{b_{Ij}, r_{Ij}\}_{j\in\{h,l\}}$ is

$$v_I \equiv \max_{\{b_{Ij}, r_{Ij}\}_{j\in\{h,l\}}} \underbrace{\left[\sum_j r_{Ij}x_j + p_A\sum_j\left(b_{Ij} - r_{Ij}\right) - R_I\left(\sum_j r_{Ij}\right)\right]}_{\text{date 1 payoff net of retention cost}} - \underbrace{\sum_j p_j b_{Ij}}_{\text{date }\frac{1}{2}\text{ purchase cost}}, \tag{4}$$

where x_j ($j \in \{h, l\}$) is the average asset quality in signal market j:

$$x_j = \frac{\sum_\theta x_\theta s_{\theta j}}{\sum_\theta s_{\theta j}}, \tag{5}$$

given the equilibrium signal market trading strategies $\{s_{\theta j}\}_{\theta\in\{H,L\}, j\in\{h,l\}}$ adopted by asset originators and the signal generating technology as specified in Eq. (1).

The average quality (and hence intermediaries' belief of it) in a j signal market as characterized by Eq. (5) is not well defined given zero trading volume (which could occur in equilibrium for the l signal market in Section 3.3). To properly define the off-equilibrium belief in such situations, recall toward the end of Section 2.1.2 we assumed

[22] This requirement is essentially the same as imposing a no-short-selling constraint on agents in this economy.

that an infinitesimal $\nu > 0$ measure of originators behave as "noise traders," such that they bring all their eligible (j-signal) assets to the j signal market for sale.[23] Therefore, intermediaries' belief for this j signal market with zero volume is

$$x_j = \frac{\sum_\theta x_\theta \pi_\theta \alpha_{\theta j} q_\theta}{\sum_\theta \pi_\theta \alpha_{\theta j} q_\theta}. \tag{6}$$

Note, this off-equilibrium belief about the average quality is still endogenously determined as it depends on the equilibrium quantities $\{q_\theta\}$.

Given the equilibrium strategies, the equilibrium asset market price p_A is determined according to Bayesian updating by risk-neutral outside investors,

$$p_A = \frac{\sum_\theta x_\theta \left[\sum_j \pi_\theta \left(\alpha_{\theta j} q_\theta - s_{\theta j} - r_{\theta j}\right)\right] + \sum_j x_j \left(b_{Ij} - r_{Ij}\right)}{\sum_\theta \left[\sum_j \pi_\theta \left(\alpha_{\theta j} q_\theta - s_{\theta j} - r_{\theta j}\right)\right] + \sum_j \left(b_{Ij} - r_{Ij}\right)}, \tag{7}$$

where a type-θ originator (intermediary) brings $\frac{\alpha_{\theta j} q_\theta - s_{\theta j} - r_{\theta j}}{1-\lambda}$ $(\frac{b_{Ij} - r_{Ij}}{1-\lambda})$ of his (her) produced (purchased) j-signal assets for sale in the asset market.

On date $\frac{1}{2}$, each signal market $j \in \{h, l\}$ must clear, which requires

$$\sum_\theta \pi_\theta s_{\theta j} = b_{Ij}. \tag{8}$$

At date 0, the total asset quantity, $Q \equiv \sum_\theta \pi_\theta q_\theta$, pins down the equilibrium input price k:

$$k = K'(Q) = K'\left(\sum_\theta \pi_\theta q_\theta\right) = 1 + \kappa \left(\sum_\theta \pi_\theta q_\theta\right). \tag{9}$$

We now formally define the market equilibrium in the economy.

Definition 1. ***(Market equilibrium).*** *A market equilibrium consists of the asset production and trading strategy $\{q_\theta, s_{\theta j}, r_{\theta j}\}_{\theta \in \{H,L\}, j \in \{h,l\}}$ of asset originators, the purchasing and retention strategy $\{b_{Ij}, r_{Ij}\}_{j \in \{h,l\}}$ of intermediaries, and equilibrium prices $\{p_A, p_j, k\}_{j \in \{h,l\}}$ such that*

1. ***Optimization of agents:*** *Given equilibrium prices $\{p_A, p_j, k\}_{j \in \{h,l\}}$, originators' equilibrium strategy $\{q_\theta, s_{\theta j}, r_{\theta j}\}_{\theta \in \{H,L\}, j \in \{h,l\}}$ solves (3) and intermediaries' equilibrium strategy $\{b_{Ij}, r_{Ij}\}_{j \in \{h,l\}}$ solves (4);*
2. ***Bayesian consistency:*** *Given equilibrium strategies $\{q_\theta, s_{\theta j}, r_{\theta j}, b_{Ij}, r_{Ij}\}_{\theta \in \{H,L\}, j \in \{h,l\}}$, asset market price p_A satisfies the Bayesian updating rule as in (7);*

[23] For the signal market with strictly positive volume in equilibrium, this is an innocuous assumption as the resulting equilibrium is the limiting equilibrium when $\nu \to 0$.

3. ***Market clearing:*** *Prices* $\{p_j\}_{j\in\{h,l\}}$ *clear the signal markets as in* (8) *and* k *is the input capital price as in* (9).

Parameter assumptions. We impose two parameter assumptions throughout the paper. First, $X > 1$; that is, H-type production has a positive NPV without capital adjustment cost, so that H-type originators produce in the planner's solution. Furthermore, for certain convenient equilibrium properties (e.g., immunity to the issue of pathological equilibrium as described in Section 2.1.2), we restrict intermediaries' risk bearing capacity ρ_I to be relatively small, so that

$$\rho_I \leq \min\left\{\frac{\lambda Q^d}{2(1-\pi)X}, \lambda\rho\right\}, \tag{10}$$

where Q^d is a function of primitive parameters (independent of ρ_I) defined in Eq. (22) in Section 3.1.[24]

2.3. *Constrained Efficient Allocations*

As a benchmark, we characterize the (constrained) efficient allocation in this economy. The planner maximizes the total payoff from produced assets, net of the cost associated with input capital and the retention of assets, subject to the same illiquidity in the asset market and minimum retention constraint. Denote post-trading retentions across both originators and intermediaries as $r_\theta \equiv \sum_j r_{\theta j}$ and $r_I \equiv \sum_j r_{Ij}$. The constrained efficient allocation is defined as follows.

Definition 2. ***(Constrained efficient allocation).*** *The constrained efficient allocation* $\{q_\theta^\star, r_\theta^\star, r_I^\star\}$ *for* $\theta \in \{H, L\}$ *solves the following problem:*

$$S^\star \equiv \max_{\{q_\theta, r_\theta, r_I\}_{\theta\in\{H,L\}}} \underbrace{\sum_\theta \pi_\theta q_\theta x_\theta}_{\text{Production}} - \underbrace{K\left(\sum_\theta \pi_\theta q_\theta\right)}_{\text{Production cost}} - \underbrace{\left[\sum_\theta \pi_\theta R\left(r_\theta\right) + R_I\left(r_I\right)\right]}_{\text{Retention cost}}, \tag{11}$$

with the minimum retention constraint $\sum_\theta \pi_\theta r_\theta + r_I \geq \lambda \sum_\theta \pi_\theta q_\theta$.

[24] Our analysis does not necessarily require that $\rho_I < \rho$, although in practice given the commercial banking sector's advantages (e.g., cheap deposits funding) over the shadow banking sector, it seems indeed true that it is way costlier for intermediaries to retain assets on their balance sheet than originators. Indeed, as shown by Jiang et al. (2020), shadow banks retain 5.6% while banks retain around 40% of their originated home mortgage loans (one year after origination).

Under constrained efficient allocation, only H-type originators should produce; the optimal retention policy equalizes marginal retention costs across all agents:

$$r_H^\star = r_L^\star = \frac{\rho}{\rho + \rho_I} \cdot \lambda \pi q_H^\star,\ r_I = \frac{\rho_I}{\rho + \rho_I} \cdot \lambda \pi q_H^\star. \tag{12}$$

Type-H asset originators produce $q_H^\star$ so that $x_H - K'\left(\pi q_H^\star\right) - \lambda R'\left(r_H^\star\right) = 0$, and the constrained-efficient production $Q^\star = \frac{q_H^\star}{\pi}$ solves

$$X = x_H = K'(Q^\star) + \lambda R'\left(\frac{\rho \lambda Q^\star}{\rho_I + \rho}\right) \iff \underbrace{\left(\kappa + \frac{\lambda^2}{\rho + \rho_I}\right) Q^\star}_{\text{marginal cost}} = \underbrace{X - 1}_{\text{marginal gain (of } H \text{ type)}}. \tag{13}$$

Here, the left hand side is total marginal costs including input adjustment cost and retention, and the right hand side is marginal gain from production (of the H type). Note here π does not play a role since the planner will guarantee that all input capital is used for only type-H production. Solving for the constrained-efficient level Q^*, we have the following proposition.

Proposition 1. *The constrained efficient allocation that solves Eq. (11) is characterized by*

1. *efficient asset production: only type-H asset originators produce, i.e., $q_L^\star = 0$, and the quantity of production by each type-H originator is $q_H^\star = \frac{Q^\star}{\pi}$; and*
2. *efficient asset allocation: marginal asset retention cost is equalized across both types of originators and intermediaries, i.e., $r_H^\star = r_L^\star = \frac{\rho}{\rho_I + \rho} \lambda Q^\star$ and $r_I^\star = \frac{\rho_I}{\rho_I + \rho} \lambda Q^\star$.*

3. Equilibrium Characterization

This section first characterizes the equilibrium with direct originate-to-distribute as a benchmark (Section 3.1), based on which two distinct economic inefficiencies are highlighted (Section 3.2). Section 3.3 then delivers a full characterization of our model where an intermediation sector is allowed to participate in the originate-to-distribute process.

3.1. *Equilibrium with Direct Originate-to-Distribute*

This section studies a benchmark economy where intermediaries are "banned" from trading (i.e, setting $\rho_I = 0$), so that asset originators can only sell directly to investors; see Figure 3 for the model scheme indicated by superscript "d." As mentioned, this economy corresponds to the traditional originate-to-distribute model. We will refer to it as a *direct originate-to-distribute* economy, or a *direct trading* economy interchangeably.

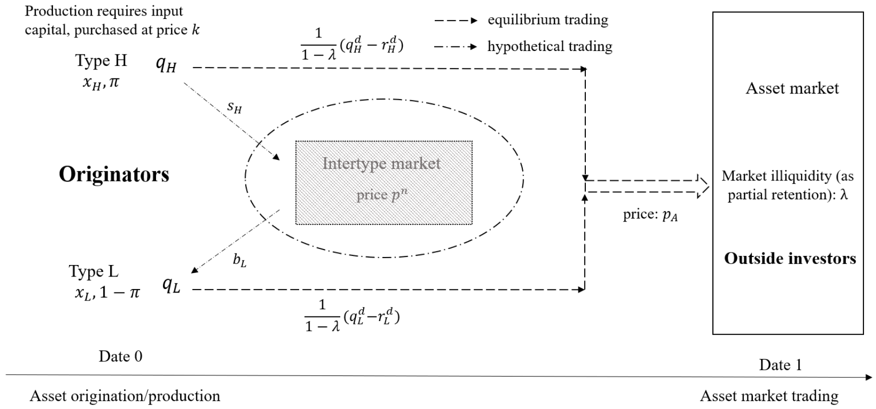


Fig. 3. Economy with direct originate-to-distribute (without intermediation)
This diagram illustrates the production and trading scheme in a direct trading economy. A fictional intertype market, represented by a shaded box inside the dashed circle, will be added in Section 3.2.2.

No voluntary retention (of originators). Consider first a type-θ originator's selling strategy. Given date 0 production q_θ^d, post-trading retention satisfies $r_\theta^d \in [\lambda q_\theta^d, q_\theta^d]$ due to market illiquidity. Therefore optimal post-trading retention r_θ^d is determined as

$$r_\theta^d(q_\theta^d) \begin{cases} = \lambda q_\theta^d, & \text{if } x_\theta - p_A^d - R'\left(r_\theta^d\right) < 0; \\ \in (\lambda q_\theta^d, q_\theta^d], & \text{if } x_\theta - p_A^d - R'\left(r_\theta^d\right) \geq 0 \text{ with equality only if binds at } q_\theta. \end{cases} \tag{14}$$

Intuitively, originators try either to sell all their produced assets (so retention r_θ binds at λq_θ), or to retain some assets voluntarily so that post-trading retention $r_\theta > \lambda q_\theta$. The next lemma shows that in equilibrium only the first case in (14) arises.

Lemma 1. *In equilibrium, $p_A^d - k^d > 0$. Asset originators bring all produced assets to the asset market for sale, with equilibrium post-trading asset retention $r_\theta^d = \lambda q_\theta^d$ for $\theta \in \{H, L\}$.*

To see the intuition, if the originators' post-trade asset holdings are ever in the second case $r_\theta^d \in (\lambda q_\theta^d, q_\theta^d]$, then they can always produce and sell δ more by putting $\frac{\delta}{1-\lambda}$ more on sale (which keeps the total holding r_θ^d fixed). This deviation strategy allows any originator to earn an additional payoff of $p_A^d - k^d$, which is strictly positive in equilibrium.[25] This contradicts individual originators' optimization.

Thanks to Lemma 1, an originator with type θ solves

$$v_\theta^d = \max_{q_\theta^d \geq 0} v_\theta^d(q_\theta^d) \equiv \max_{q_\theta^d \geq 0} x_\theta \lambda q_\theta^d + p_A^d(1-\lambda)q_\theta^d - R\left(\lambda q_\theta^d\right) - k^d q_\theta^d, \tag{15}$$

[25]The intuition for $p_A^d - k^d > 0$ is as follows. If $p_A^d \leq k^d$, the L-type originators will not produce, since the marginal payoff to their production is bounded by $\lambda x_L + (1-\lambda)p_A^d \leq p_A^d$. But this means that p_A^d equals x_H, which must be strictly higher than k^d to ensure that the H-type are indeed producing in equilibrium.

with the following optimality condition:

$$q_\theta^d \begin{cases} > 0, & \text{if } (1-\lambda)p_A^d + \lambda x_\theta - \lambda R'\left(\lambda q_\theta^d\right) = k^d, \\ = 0, & \text{if } (1-\lambda)p_A^d + \lambda x_\theta - \lambda R'\left(\lambda q_\theta^d\right) < k^d. \end{cases} \tag{16}$$

Moving on to prices, since π $(1-\pi)$ measure of H- (L-)type originators bring in q_H (q_L) units of H (L) assets, the asset market price in Eq. (7) becomes:

$$p_A^d = \frac{\pi q_H^d x_H + (1-\pi) q_L^d x_L}{\pi q_H^d + (1-\pi) q_L^d} = \frac{\pi q_H^d X}{\pi q_H^d + (1-\pi) q_L^d}, \tag{17}$$

and the equilibrium input capital price $k^d = K'(Q^d) \equiv K'\left(\pi q_H^d + (1-\pi) q_L^d\right)$ as in Eq. (9).

Equilibrium characterization. First, from originators' optimization (16), when both are producing we obtain the *production wedge* between two types of originators:

$$\Delta q \equiv q_H^d - q_L^d = \frac{\rho X}{\lambda}, \tag{18}$$

which is independent of production costs (i.e., adjustment cost κ). Intuitively, since both types of originators purchase input capital and sell produced assets at the same prices, the indifference condition requires that the wedge in marginal retention cost should be exactly offset by the asset quality difference, that is,

$$\underbrace{\Delta R'(\lambda q_\theta) = \frac{\lambda}{\rho}\Delta q}_{\text{wedge in marginal retention cost}} = \underbrace{\Delta x_\theta = X}_{\text{asset quality difference}}. \tag{19}$$

Combining originators' optimality conditions (16), one can show that, system-wise, the marginal cost and benefit of asset production equalize in equilibrium. Further, with investors' risk-neutral pricing (17) that relates p_A to Δq in (19), we have

$$\underbrace{\left(\kappa + \frac{\lambda^2}{\rho}\right) Q^d}_{\text{marginal cost}} = \underbrace{(1-\lambda)p_A^d + \lambda \pi X - 1}_{\text{marginal gain}} = \underbrace{\pi X - 1}_{\text{baseline gain}} + \underbrace{(1-\lambda)\pi(1-\pi)\frac{\rho X^2}{Q^d \lambda}}_{\text{incentive due to retention}}. \tag{20}$$

In Eq. (20), we have the same marginal cost of production due to input capital adjustment and retention (except $\rho_I = 0$) as Eq. (13), but the marginal gain from production differs from the constrained-efficient level $X - 1$: it consists of a baseline gain $\pi X - 1$ determined by the projects' average quality with noncontingent production and an incentive-based gain $(1-\lambda)\pi(1-\pi)\frac{\rho X^2}{Q^d\lambda}$ driven by originators' production (and hence retention) wedge.

In equilibrium H-type originators always produce (i.e., $q_H > 0$) while L-type originators' production could be binding at $q_L = 0$. When $q_L > 0$, the equilibrium total production Q^d solves the quadratic equation (20), which always admits a unique positive solution. Proposition 2 characterizes the market equilibrium with direct trading.

Proposition 2. ***(Characterization of equilibrium with direct trading).*** *A unique market equilibrium exists in a direct originate-to-distribute economy and is always stable.*

1. *If $\frac{X-1}{X} \leq \frac{\rho\kappa\pi}{\lambda} + \lambda$, L-type originators do not produce, i.e., $q_L = 0$, while H-type originators produce $q_H = \frac{\rho(X-1)}{\kappa\pi\rho+\lambda^2} > 0$; equilibrium prices are given by*

$$p_A^d = X, \quad k^d = \frac{\kappa\pi\rho X + \lambda^2}{\kappa\pi X + \lambda^2} > 1. \tag{21}$$

2. *Otherwise, the equilibrium features an interior solution with input capital price given by:*

$$Q^d = \frac{B_d + \sqrt{B_d^2 - 4A_d C_d}}{2A_d} \in (0, \infty), \tag{22}$$

where $A_d \equiv \lambda\left(\kappa + \frac{\lambda^2}{\rho}\right)$, $B_d \equiv \lambda(\pi X - 1)$, and $C_d \equiv -(1-\lambda)\rho(1-\pi)\pi X^2$. The equilibrium input capital price k and asset market price p_A^d are given by

$$k^d = 1 + \kappa Q^d \text{ and } p_A^d = \frac{1}{1-\lambda}\left[\left(\kappa + \frac{\lambda^2}{\rho}\right) Q^d - \lambda\pi X + 1\right] \in (\pi X, X). \tag{23}$$

The originators' equilibrium strategies $\{q_\theta^d, r_\theta = \lambda q_\theta^d\}_{\theta\in\{H,L\}}$ are given by

$$q_L^d = Q^d - \frac{\rho\pi X}{\lambda} > 0, \;\; q_H^d = Q^d + \frac{\rho(1-\pi)X}{\lambda}. \tag{24}$$

In a direct trading economy, L-type originators produce a positive amount given a sufficiently small retention cost ρ; in our subsequent analysis, we focus on the scenario of positive production from L-type originators. In addition, the asset market price p_A^d is always above the average asset return without contingent production, πX, which reflects the disciplining effect of market illiquidity λ.

Finally, we note that the equilibrium in Proposition 2 must be stable: any price deviation from the equilibrium level p_A^d always leads to a new average quality in the asset market (via affecting originators' production), which cannot keep up with the initial price change, pushing the system back to the original equilibrium. Trading illiquidity in the asset market plays a key role in guaranteeing equilibrium stability in direct trading economies;[26] as analyzed later in Section 3.3, unstable equilibria potentially arise in an intermediated economy where trading between originators and intermediaries is immune from illiquidity problems.

3.2. *Efficiency Analysis of Market Equilibrium with Direct Trading*

We first demonstrate two distinctive sources of inefficiencies suffered by market equilibrium in a direct originate-to-distribute economy, which can be addressed by *ex ante commitment* and *ex post trading*, respectively. We then show that frictionless intertype trading can fully restore efficiency.

3.2.1. *Two economic inefficiencies in market equilibrium with direct trading*

First, both types of originators produce in the market equilibrium in a pooling equilibrium, reflecting the standard lemon problem; recall L-type should produce nothing in the planner's solution (Proposition 1). This captures *production inefficiency*.

Second, while retention caused by market illiquidity disciplines asset production, it necessarily leads to another distinctive source of *allocative inefficiency* given the convex retention costs. The marginal retention cost for the type-θ originator is $R'(\lambda q_\theta)$; as H-type originators produce more given their superior production technology (see Eq. (24)), they have a higher equilibrium marginal retention cost.[27] The wedge between marginal retention cost, $R'(\lambda q_H) - R'(\lambda q_L)$, is thus positive and creates room for mutually beneficial trading between asset originators of different types ex post, to which we turn next.

3.2.2. *Intertype trading and efficiency restoration*

Before analyzing the role of informed intermediation, we study one interesting variant of our direct trading economy by allowing for intertype trading. More specifically, as in Figure 3, suppose that originators can trade their assets among themselves in a frictionless intertype market before they sell in the illquid asset market. Importantly, "frictionless" here means that the intertype trading market is free from information asymmetry (say,

[26]Through its disciplining effect, the market illiquidity in the asset market ensures that the equilibrium asset quality traded in the market is high enough, such that a positive price deviation cannot induce further asset quality improvement in the asset market.

[27]The wedge between equilibrium marginal retention costs is related to studies in which incompleteness of markets or contracts hinders the equalization of economic agents' wealth across states. This large body of literature includes Caballero and Krishnamurthy (2003), Lorenzoni (2008), He and Kondor (2016), and Dávila and Korinek (2018).

because asset originators are experts in this business). In Appendix A2.2 we show that L-type originators purchase assets produced by H-types in the intertype market and resell them in the asset market, so much so that it restores the efficiency fully. We have the following proposition for this economy indicated by superscript "d" and "hat."

Proposition 3. ***(Efficiency restoration by ex post intertype trading).*** *Suppose $\rho_I = 0$. The market equilibrium with intertype trading implements constrained efficient allocation with efficient production, i.e., $\hat{q}_H^d = \frac{Q^\star}{\pi}$ with $Q^\star$ given in Eq. (13), and $\hat{q}_L^d = 0$; and efficient balance-sheet allocation, i.e., $\hat{q}_H^d - \hat{s}_H^d = \hat{q}_L^d + \hat{b}_L^d$.*

Since frictionless intertype trading allows both types of originators to reshuffle their holdings, it is not surprising that all asset originators end up with the same holding, resulting in efficient allocation. Somewhat surprisingly, asset production also becomes efficient—both the composition of asset production (no lemons being produced) and the total quantity of assets produced ($Q^\star$) are identical to those given in Proposition 1.

To understand the production efficiency, note that the intertype trading opportunity (on H-type assets) lowers an L-type originator's incentive to produce lemons. On the other hand, an H-type originator—now with an additional venue to offload assets—is encouraged to produce more, and the combination of both forces "cleanse" the asset market. This cleansing process is self-fulfilling under endogenous asset market prices. In equilibrium, L-type originators cease their own production completely and provide intermediation services only, which facilitate H-type originators' originate-to-distribute.[28]

As we will see, similar economic forces emerge in Section 3.3 where we formally study how the informed intermediary sector operates in this economy; note, here L-type originators essentially serve as informed intermediaries.

3.3. *Equilibrium with Intermediated Originate-to-Distribute*

The hypothetically postulated intertype trading in Section 3.2.2 provides useful insights on efficiency restoration with market solutions. In this section, we explore a more empirically relevant market solution where an informed intermediary sector can buy assets in the signal markets and resell them to the asset market—that is, an *intermediated originate-to-distribute* model.

[28]Whenever a type-L originator is still producing ($q_L > 0$), this cleansing process, during which an L-type originator cuts his own production by $\delta > 0$ while purchasing δ more from H-type originators, could continue. Given intertype market price $\hat{p}_{in}^d$, this deviation strategy yields a marginal gain of $\lambda(x_H - x_L)\delta - \hat{p}_{in}^d \delta + \hat{k}^d \delta$, which has to be nonpositive (i.e., $\hat{p}_{in}^d - \hat{k}^d \geq \lambda(x_H - x_L) > 0$) in equilibrium. But seeing $\hat{p}_{in}^d > \hat{k}^d$, those type-$H$ originators would want to produce an infinite amount, which cannot hold in equilibrium.

3.3.1. *Properties of intermediated equilibria*

We first establish two important lemmas in the intermediated economy, which not only help simplify the exposition, but also provide useful insights on the structure of potential intermediated equilibria. First, Lemma 2 generalizes Lemma 1 in the setting of an intermediated originate-to-distribute economy.

Lemma 2. ***(No voluntary retention).*** *Both types of originators bring all assets (after trading in signal markets) to the asset market, so post-trading asset holdings are* $r_\theta = \lambda(q_\theta - s_\theta)$ *for* $\theta \in \{H, L\}$. *When* ρ_I *is small enough (a sufficient condition being* $\rho_I \leq \frac{\lambda Q^d}{2(1-\pi)X}$, *guaranteed by assumption* (10)*), intermediaries also bring all purchased assets to the asset market for sale, i.e.,* $r_I = \lambda b_I$.

The next lemma, which establishes the endogenous closure of the "lemon" signal market in our model, is new and is key to the understanding of our paper. In a nutshell, it shows that the intermediaries' trading problem can be simplified: in equilibrium intermediaries purchase assets with homogeneous quality x_s at a single price p_s (from either one or two signal markets), and they then resell them at price p_A in the asset market.

Lemma 3. ***(Endogenous closure of the "lemon" signal market).*** *In the market equilibrium, the prices in the signal markets satisfy* $p_h \geq p_l$, *and we have the following properties:*

1. *Suppose that* $p_h = p_l$. *Both signal markets have positive trading volume with the same trading price and average quality, i.e.,* $p_h = p_l = k$ *and* $x_h = x_l$;
2. *Suppose that* $p_h > p_l$. *Then positive trading could occur in the h signal market, while there exists no trading in the l signal market, i.e.,* $b_{Il} = 0$ *and* $s_{\theta l} = 0$ *for* $\theta \in \{H, L\}$.

The first result that homogeneous asset prices imply equal asset quality in signal markets is straightforward. With positive trading volume in both signal markets, the optimality condition for intermediaries implies that

$$p_h - p_l = \lambda(x_h - x_l). \tag{25}$$

To see this, note that assets sold to the asset market via intermediaries—regardless of their quality type θ or signal j—share a common balance sheet cost and a common selling price. Eq. (25) then follows from the fact that the price differential in two signal markets must compensate for the quality difference due to illiquidity-induced retention. And, for originators who are active on both signal markets, assets sold in both signal markets share a common linear production function. Thus when signal market prices are equal, we must have $p_h = p_l = k$ in equilibrium.[29]

The second result of no l-market trading is more surprising. Suppose that, counterfactually, both signal markets have positive trading volume; and consider the situation where $x_h > x_l$ and hence $p_h > p_l$.[30] But the L-types' optimality condition says $k = (1-\alpha)\, p_h + \alpha p_l$, implying that the H-type would like to produce strictly more as

$$\underbrace{k}_{\text{production cost}} = \underbrace{(1-\alpha)\, p_h + \alpha p_l}_{\text{average selling revenue of } L\text{-type}} < \underbrace{\alpha p_h + (1-\alpha)\, p_l}_{\text{average selling revenue of } H\text{-type}} . \tag{26}$$

Intuitively, relative to L-type originators, the H-type have greater advantage in producing h-signal assets which are sold at a high price; then if L-type originators are indifferent at producing/selling, it must be that H-type peers find it strictly optimal to produce/sell. But this pushes up the input price k, eventually crowding out L-type production, which contradicts the premise that L-type originators are selling in the l signal market.

Economics of intermediaries' operation. It is interesting to connect this endogenous closure of the "lemon" market to the "winner/cherry picking" behavior of headquarters as highlighted in Stein (1997). From Eq. (25) we have seen that intermediaries' exposure to asset market illiquidity gives them an incentive to "cherry pick" the signal market with a higher quality. Then, "cherry picking" by intermediaries, combined with a competitive input capital market (hence a common marginal cost for all assets), squeezes the l signal market out completely in equilibrium.[31]

In a broader context, the economics behind the operation of intermediaries in our model echo insights from the delegated monitoring literature. In providing intermediation services in this originate-to-distribute economy—acquiring originated assets from signal markets in an informed manner and subsequently reselling them to uninformed investors—intermediaries face a classic "monitoring the monitor" problem à la Diamond (1984). Just as L-type originators have an incentive to overproduce and sell "lemons,"

[29] This is because the option to either increase or decrease production for an originator who is active in both signal markets implies that in equilibrium $k = \alpha p_h + (1-\alpha) p_l$ or $k = (1-\alpha) p_h + \alpha p_l$.

[30] This is a more relevant case; fixing the same l market, H-type originators who can produce good assets should have a lower selling incentive, relative to L-type originators; so likely in the end only L-type was selling in the l market. For an exhaustive analysis of all possible cases in which both markets have positive trading volume, see Appendix A3.1.

[31] It is worthwhile to compare the "pooled" asset sales through intermediaries to those in the literature. In contrast to depriving the option of exploiting asset-specific information advantages as in DeMarzo (2005) or He (2009), pooled asset sales via informed intermediaries in our analysis preserve flexibility in making signal-specific purchases, which has the benefit of "winner picking" in Stein (1997) in the context of an internal capital market when corporate headquarters allocate scarce resources to competing projects.

informed intermediaries are effectively subject to the same agency problem—as signals are privately observed only by intermediaries, they have the option on date $\frac{1}{2}$ to purchase more l signal assets instead of h signal ones (with the same sale price on date 1).

Nevertheless, in the second case in Lemma 3, informed intermediaries behave in a self-disciplining manner without purchasing any "lemons" from the signal markets. Unlike solutions that rely on explicit *contracting* in the previous literature (e.g., debt contract in Diamond (1984), claim seniority in Park (2000)), the self-disciplining of the delegated monitor in our model (i.e., informed intermediaries) arises endogenously as a *market* outcome. As discussed above, the common balance sheet on which assets acquired by a intermediary from different signal markets are placed, coupled with a competitive input capital market, are key to this market-based solution.

3.3.2. *Equilibrium characterization of the intermediated economy*

This section provides a full characterization of market equilibria in an intermediated originate-to-distribute economy. The analysis proceeds as follows: we first present some preliminary analyses grounded in the equilibrium properties established in Section 3.3.1 and discuss the possible structure of intermediated equilibria (part A), which reveals two distinct realms into which the market equilibrium in an intermediated economy could potentially fall; we then analyze these two scenarios separately in part B and part C.

A. Preliminary analysis and structure of intermediated equilibria

Decisions by agents. Consider a θ-type asset originator who produces q_θ with $\alpha_{\theta h} q_\theta$ ($\alpha_{\theta l} q_\theta$) receiving a h (l) signal. Thanks to Lemma 2 and Lemma 3, his optimization problem given in Eq. (2) can be simplified to:

$$v_\theta(q_\theta) \equiv \max_{s_{\theta j} \in \left[0, \alpha_{\theta j} q_\theta\right]} p_s s_\theta + (q_\theta - s_\theta)\left[(1-\lambda)p_A + \lambda x_\theta\right] - R(\lambda(q_\theta - s_\theta)) - k q_\theta, \tag{27}$$

where $s_\theta = \sum_j s_{\theta j}$, with its solution given by

$$s_\theta(q_\theta) \begin{cases} > 0, & \text{if } p_s = \lambda x_\theta + (1-\lambda)p_A - \lambda R'\left(\lambda(q_\theta - s_\theta)\right); \\ = 0, & \text{if } p_s < \lambda x_\theta + (1-\lambda)p_A - \lambda R'\left(\lambda(q_\theta - s_\theta)\right). \end{cases} \tag{28}$$

In words, he sells his eligible assets in the signal market(s) until the marginal value of selling in the asset market equals the signal market price p_s. This originator then produces until the marginal value of production—a weighted average of p_s and the marginal value of selling (ineligible) assets via the asset market—equals the input capital price.

Moving on to intermediaries, given Lemma 2 and Lemma 3, they solve the following problem by choosing the quantity of assets b_I to purchase in signal markets, all of which they then optimally bring for sale in the asset market at price p_A:

$$v_I \equiv \max_{b_I \geq 0} (1-\lambda) p_A b_I + \lambda x_s b_I - R_I(\lambda b_I) - p_s b_I, \tag{29}$$

where x_s is the (common) equilibrium average asset quality in signal markets. The associated optimality condition is:[32]

$$(1-\lambda) p_A + \lambda \left[x_s - R'(\lambda b_I) \right] = p_s. \tag{30}$$

Production wedge Δq. The optimality conditions of originators in (28) imply that the equilibrium production wedge in an intermediated originate-to-distribute economy is

$$\Delta q = q_H - q_L = \underbrace{\frac{\rho X}{\lambda}}_{\text{asset market induced wedge}} + \underbrace{s_H - s_L}_{\text{signal market induced wedge}}, \tag{31}$$

which consists of the base wedge $\frac{\rho X}{\lambda}$ induced by asset market trading as in Eq. (19) plus the wedge induced by signal market trading $s_H - s_L$. As will be clear shortly, the signal market trading induced production wedge plays a key role in characterizing the intermediated equilibrium and the welfare implication of intermediaiton.

Equilibrium prices. First, Lemma 2 implies that all assets produced on date 0—either by originators directly, or via intermediaries indirectly—will be brought to the asset market on date 1. As such, the equilibrium asset market price set by risk neutral investors is the average quality of the total production in the economy:

$$p_A = \frac{x_H \pi q_H + x_L (1-\pi) q_L}{\pi q_H + (1-\pi) q_L} = \frac{\pi X q_H}{\pi q_H + (1-\pi) q_L}. \tag{32}$$

In the signal market at date $\frac{1}{2}$, the equilibrium trading price p_s equalizes the aggregate purchase from the intermediary sector to the total sale by the originators (i.e., $b_I = \pi s_H + (1-\pi) s_L$) and the equilibrium input capital price is similarly determined as in Eq. (9).

Structure of market equilibria. Lemma 3 has important implications for our equilibrium construction and characterization in later sections. We show that two distinctive

[32] Intermediaries always buy some positive amount ($b_I > 0$) since $p_A > k$ always holds in equilibrium.

equilibrium classes, defined based on the structure of signal markets, can emerge given the same parameterization. Subsequently we refer to equilibria where only the h signal market has trading while the l signal market has zero trading volume with (shadow) prices satisfying $p_h > p_l$ as *class-I* equilibria, and those where both signal markets have positive trading with $p_h = p_l$ as *class-II* equilibria.

Tech-relevance and stability of intermediated equilibria. To organize our subsequent analysis, it is helpful to highlight several important features of the market equilibrium, in terms of its dependence on intermediary technology α ("tech-relevance") and the equilibrium stability, before we characterize the equilibrium algebraically

First, we will show that for any given $\rho_I > 0$ both classes of equilibria feature a tech-relevance cutoff in the technology parameter α, so that when α is below this threshold equilibrium productions and prices do not vary with α. We call that region *tech-irrelevant*, with cut-off technology levels— $\hat{\alpha}_1(\rho_I)$ and $\hat{\alpha}_2(\rho_I)$ for class-I and class-II respectively—to be characterized for both equilibrium classes. In part B of this section, we show that the intermediated equilibria in this tech-irrelevant range are essentially unique.[33]

Second, unlike in the direct trading economy where illiquid asset market trading guarantees the market equilibrium is always stable (see discussions after Proposition 2), perfectly liquid trading between originators and intermediaries could potentially render the market equilibrium in the intermediated economy unstable. In the tech-irrelevant range the equilibrium is always stable, but this is no longer the case when α lies above a certain threshold and the economy enters the *tech-relevant* range. In part C of this section we show that, unlike class-I equilibria, class-II equilibria with positive trading in both signal markets are always unstable in the tech-relevant range.

In what follows, we characterize the intermediated equilibrium in the tech-irrelvant range (part B) and tech-relevant range (part C) respectively. For reasons discussed above, we focus on class-I equilibria, which are both stable and featuring "endogenous closure" of the l signal market, in studying an intermediated originate-to-distribute economy.[34]

B. Intermediated equilibrium in the tech-irrelevant range: $\alpha \in \left(\frac{1}{2}, \hat{\alpha}_1(\rho_I)\right)$

In the range with relatively low intermediary technology α, the h signal market selling is slack for L-type originators (who have relatively fewer h-signal assets), i.e., $s_{Lh} \in$

[33] That is, despite the distinct classes that the equilibrium might fall into, the equilibrium outcomes (production, trading prices of assets, and input capital) are uniquely determined in the tech-irrelevant range in the intermediated economy for a given pair (ρ_I, α).

[34] In sum, the class-II equilibria with positive trading volume in both signal markets are identical to class-I equilibria in the tech-irrelvant range while always unstable in the tech-relevant range.

$[0,(1-\alpha)q_L)$. As explained shortly, selling from H-type originators (who have a higher fraction of h-signal assets) must also be slack in the h signal market. In this range of the intermediated equilibrium, both types of originators are indifferent about taking the last unit of produced assets for sale in either signal markets at date $\frac{1}{2}$, or in the asset market at date 1, which implies that for both $\theta \in \{H, L\}$:

$$\underbrace{(1-\lambda)p_A + \lambda\left[x_\theta - R'\left(\lambda(q_\theta - s_\theta)\right)\right] = p_h}_{\text{indifference between two selling options}} \underbrace{= k}_{\text{production}} . \tag{33}$$

Here, the first equality is from choosing between two selling methods while the second is from originators' linear production technology.

The perfectly liquid signal market(s) leads to certain equilibrium indeterminancy due to linearity,[35] and we focus on the limiting equilibrium when the signal market illiquidity $\epsilon \to 0$ following the sequential equilibrium refinement (Kreps and Wilson, 1982; see Section 2.1.2). For any arbitrary $\epsilon > 0$ that breaks the linearity, we show that both types of originators aim to sell the exact same amount of assets in the signal market, independent of their asset quality x_θ, i.e., $s_H = s_L$. The quadratic functional form of agents' retention cost plays a key role in this result. With linear marginal retention cost, the difference in the marginal value of assets between the two states (successful sales vs. failed sales) depends solely on the quantity of assets that originators attempt to sell in the signal markets—it is not affected by the quality of originators' assets or the quantity of assets they produce.[36]

This observation has several implications. First, when h signal market selling is not binding for either type, equilibrium sales for both assets follow the measures of originator types; this implies the average quality in h signal markets is $x_h = \pi X$. Second, equal h signal market selling from both types also implies that if in equilibrium the selling of L-type originators is slack in the h signal market, then that of the H-type must be slack too.[37] Accordingly, the market clearing condition of L-types' selling their h-signal asset

[35] In such equilibrium, the h signal market price (in a class-I equilibrium; $p_h = p_l$ if in a class-II equilibrium) equals the input capital price. Both originators thus face a linear problem when their selling in signal market(s) has yet to bind (which is the case in the tech-irrelevant range), and the inherent indifference allows for any $x_h \in [0, X]$ to be supported as equilibrium asset quality in the signal market. Any infinitesimal chance that trading in the signal market could fail breaks the linearity and eliminates the indeterminancy.

[36] With retention cost being quadratic, the marginal value of produced assets is linear in the holding of assets. As a result, the particular asset quality x_θ cancels out in the difference between the marginal value of successful or failed sales; see the proof of Proposition 4. As another feature arising in our linear-quadratic setting, the difference in marginal values between successful versus failed selling states is also independent of the equilibrium production quantity, which would otherwise be affected by the asset quality x_θ.

[37] To see this, note that a larger fraction of produced assets are eligible to be sold in the h signal market for H-type originators as $\alpha > \frac{1}{2}$. Therefore, with a looser eligibility constraint and higher quality of produced assets, the equilibrium production of H-type originators cannot be lower than that of a L-type one. This

determines the cutoff level $\hat{\alpha}_1(\rho_I)$.

More central to the economic mechanism of tech-driven intermediation, given the equal selling by both types of originators in signal market, the production wedge stays constant at $\Delta q = q_H - q_L = \frac{\rho X}{\lambda}$ (as defined in (19)), just like with no intermediaries (i.e., $\rho_I = 0$). In this regard, when intermediaries' technology α is sufficiently low such that the intermediated equilibrium resides in the tech-irrelevant range, the operation by intermediaries provides *no disciplining* on lemon production in the economy.

With the production wedge identical to that in a direct originate-to-distribute economy, in the intermediated equilibrium where both types of originators are slack in selling their h-signal assets, the total production Q is similarly determined by

$$\underbrace{\left(\kappa + \frac{\lambda^2}{\rho + \rho_I}\right) Q}_{\text{marginal cost}} = \underbrace{\pi X - 1}_{\text{baseline gain}} + \underbrace{(1-\lambda)\pi(1-\pi)\frac{\rho X^2}{Q\lambda}}_{\text{incentive due to retention}}, \tag{34}$$

which differs from the direct economy equilibrium condition (20) only in that the system-wise marginal cost of production now becomes $\left(\kappa + \frac{\lambda^2}{\rho+\rho_I}\right) Q$ as opposed to $\left(\kappa + \frac{\lambda^2}{\rho}\right) Q^d$, thanks to the extra retention capacity ρ_I brought in by intermediaries. The following proposition formally characterizes the market equilibrium in the tech-irrelevant range.

Proposition 4. ***(Technology-irrelevant range).*** *Suppose that $\alpha \leq \hat{\alpha}_1(\rho_I)$. In the unique (sequential) equilibrium with signal market illiquidity $\epsilon \to 0$, both types of originators sell $s_H = s_L < (1-\alpha)\min\{q_H, q_L\}$ in the signal market. The equilibrium is stable and the equilibrium outcomes are independent of the technology parameter α:*

$$Q = \frac{B_u + \sqrt{B_u^2 - 4A_uC_u}}{2A_u},\ s_H = s_L = \frac{\rho_I Q}{\rho + \rho_I}, \tag{35}$$

$$x_h = \pi X,\ p_h = k = 1 + \kappa Q,\ \text{and } p_A = \frac{1}{1-\lambda}\left[\left(\kappa + \frac{\lambda^2}{\rho+\rho_I}\right) Q - \lambda \pi X + 1\right] \tag{36}$$

where $A_u \equiv \kappa + \frac{\lambda^2}{\rho_I + \rho}$, $B_u \equiv \frac{(1-\lambda)\rho}{\rho+\rho_I} + \lambda\pi X - 1$, and $C_u \equiv -\frac{\rho\kappa(1-\lambda)(1-\pi)X^2}{\lambda}$. The threshold $\hat{\alpha}_1(\rho_I)$ determining technology relevance is given by

$$\hat{\alpha}_1(\rho_I) = \max\left\{\frac{1}{2}, \left[\frac{\lambda^2 \frac{\rho_I Q}{\rho+\rho_I}}{\rho\left[(1-\lambda)p_A - k\right]} + 1\right]^{-1}\right\}. \tag{37}$$

implies $s_H = s_L \leq (1-\alpha)q_L < \alpha q_H$.

Note that the tech-irrelevant equilibrium as characterized in Proposition 4 could also be supported under a class-II equilibrium, where both signal markets have positive trading volumes. In fact, the intermediated equilibrium is *essentially unique* in the tech-irrelevant range; that is, the tech-irrelevant equilibrium constructed under both equilibrium classes have identical equilibrium outcomes (e.g., total production, asset market and signal market prices, allocation of asset retention).

C. Intermediated equilibrium in tech-relevant range: $\alpha \in [\hat{\alpha}_1(\rho_I), 1]$

When α surpasses the cutoff $\hat{\alpha}_1(\rho_I)$, the market equilibrium enters into the technology relevant range where further changes in intermediaries' technology α will start to transmit into equilibrium outcomes. Under the assumption throughout our analysis that intermediaries' risk-bearing capacity is small enough, we have the following proposition characterizing the intermediated equilibrium in this range of the (ρ_I, α) space.

Proposition 5. ***(Technology-relevant range).*** *When $\alpha \geq \hat{\alpha}_1(\rho_I)$, the unique stable (sequential) class-I equilibrium with signal market illiquidity $\epsilon \to 0$ varies with intermediation technology α. In this equilibrium, type-L originators sell all their h-signal assets, i.e., $s_L = (1-\alpha)q_L$, while type-H originators' selling is slack, i.e., $s_H < \alpha q_H$, in their h signal market trading. Furthermore, in equilibrium $s_H > s_L$, and the asset quality in the h signal market satisfies $x_h > \pi X$. Details for equilibrium characterization are provided in the Appendix A3.3.*

In this range of the intermediated equilibrium, L-type originators face the eligible selling constraint $s_L \leq (1-\alpha)q_L$ in trading with intermediaries, as only a $1-\alpha$ fraction of their produced assets are eligible for sale in signal market(s)—which can only be h signal market under class-I equilibria. In such equilibria, changes in technology level (say a higher α) will be transmitted to equilibrium outcomes via affecting the tightness of the eligibility constraint $s_L \leq (1-\alpha)q_L$ of L-type originators' selling in h signal market—information technology now becomes relevant. Proposition 5 establishes a strictly positive selling wedge $s_H - s_L > 0$ in the tech-relevant range; later Proposition 6 further shows that this wedge increases with α, leading to an widened production wedge Δq in this economy. We defer further discussions on this important point to Section 4.1, where we formally investigate the impact of intermediary technology on equilibrium outcomes.

For our later welfare analysis, the following corollary provides the closed form characterization of the equilibrium with perfectly informed intermediation ($\alpha = 1$).

Corollary 1. ***(Perfectly informed intermediaries).*** *Suppose $\alpha = 1$. The intermediated equilibrium is always in the technology-relevant range (i.e., $\hat{\alpha}_1(\rho_I) < 1$). In the unique (sequential)* *equilibrium with signal market illiquidity $\epsilon \to 0$, only h signal market has positive trading volume and equilibrium outcomes are as follows:*

$$Q = \frac{B_i + \sqrt{B_i^2 - 4A_iC_i}}{2A_i}, \; \pi s_H = b_I = \frac{\rho_I}{\rho + \rho_I}\left(Q + \frac{(1-\pi)\,X\rho}{\lambda}\right) \tag{38}$$

$$x_h = X, \; p_h = k = 1 + \kappa Q, \; \textit{and } p_A = \frac{1}{1-\lambda}\left[\left(\kappa + \frac{\lambda^2}{\rho + \rho_I}\right)Q - \frac{\lambda X(\rho\pi + \rho_I)}{\rho_I + \rho} + 1\right], \tag{39}$$

where $A_i \equiv \kappa + \frac{\lambda^2}{\rho_I + \rho}$, $B_i \equiv \frac{\rho\pi X + \rho_I X}{\rho_I + \rho} - 1$, *and* $C_i \equiv -\frac{(1-\lambda)\kappa\rho(1-\pi)(\rho\pi+\rho_I)X^2}{(\rho_I+\rho)\lambda}$.

In this extreme case with perfectly informed intermediaries, only H-type assets will be traded in the signal markets—so intermediaries "commit to" not purchasing any lemons produced in the economy (i.e., type L selling $s_L = 0$). The disciplining by perfectly informed intermediaries effectively increases the equilibrium production wedge by $s_H = b_I/\pi > 0$. Section 4.2 further analyzes this originate-to-distribute equilibrium with perfectly informed intermediation, where an economic efficiency gain is demonstrated.

Remark: Instability of class-II equilibria. While our characterization of the intermediated economy in the tech-relevant range only focuses on class-I equilibria, it can be shown that class-II equilibria are always *unstable* in this range. To see this, note in the tech-relevant range type-L (-H) originators are binding in selling their h (l) signal assets. Suppose h signal market price $\tilde{p}_h$ deviates from its equilibrium level p_h positively by $\delta > 0$, i.e., $\tilde{p}_h = p_h + \delta$, while all other prices are fixed at the equilibrium levels. All type-H originators will then respond by moving part (or all) of their selling in l signal market (which was binding at $(1-\alpha)q_H$) to sell in h signal market instead. In contrast, type-L originators cannot respond (although they wish to) because their selling in h signal market has already been binding. Critically, any infinitesimal deviation $\delta > 0$ induces an upward jump in H-type's selling of their h signal assets, given the linearity in their optimization regarding selling allocation across the two signal markets.[38]

The originators' selling strategy in signal markets affects how intermediaries perceive the quality of assets sold in the two signal markets. Importantly, given the discrete jump in H-type's response discussed above, intermediaries' perception of asset quality $\tilde{x}_j$ in signal markets with the same prices $p_h = p_l$ is guaranteed to satisfy $\tilde{x}_h - \tilde{x}_l > \delta = (p_h + \delta) - p_l$ for sufficiently small $\delta > 0$. Consequently, intermediaries will respond by

[38] In contrast, such an upward discrete jump in H-type's selling in the h signal market following an infinitesimal positive price deviation will not occur in class-I equilibria, where the trading volume in l signal market is zero. In such cases, there will not be reallocation among signal markets and any production by originators will create a positive amount of ineligible assets that can only be sold in the asset market.

only purchasing assets from the h signal market. This leads to a substantial increase in intermediaries' demand for h signal assets and hence a positive excess demand for assets sold in h signal market, rendering the equilibrium unstable.

4. Economic Implications of Tech-Driven Intermediation

Focusing on the role played by intermediation technology α, we discuss the key economics behind intermediation's operation intermediation in our originate-to-distribute economy. In what follows, our analysis focuses on class-I equilibria with endogenously closed l signal market for reasons discussed in Section 3.3.2.

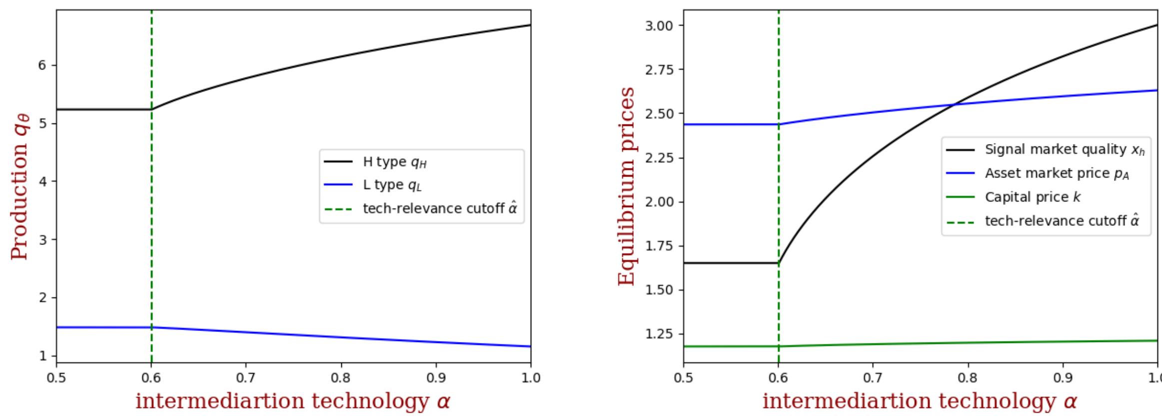
4.1. Positive Analysis of Intermediation Technology α

We start with a set of comparative statics analyses—with regard to the intermediation technology α—on the equilibrium outcomes characterized in Section 3.3.2. As shown in Figure 5(a), intermediation technology α has no impact on equilibrium outcomes until it exceeds the threshold $\hat{\alpha}_1(\rho_I)$ at which point the economy enters the tech-relevant range. There, further improvement in α starts to generate a "cleansing effect"—as α ascends, L-type originators' production shrinks while H-type originators' expands. As discussed in Part C in Section 3.3.2, this cleansing effect arises due to a tightened eligible selling constraint on L-type originators' signaling market trading (i.e., $s_L \leq (1-\alpha)q_L$), without any (or, if any, in fact a loosening) effect on H-type's selling s_H in h signal market.

This asymmetric impact of technology progress on two types' selling s_θ in h signal market immediately implies an enlarged production wedge Δq in the economy; recall that by Proposition 5, $s_H > s_L$ (and hence $\Delta q > \frac{\rho X}{\lambda}$, which is the wedge in the tech-irrelevant range) in the tech-relevant range. The enlarged production wedge also brings up the average asset market quality and total production in the economy, translating into higher equilibrium prices as shown in Figure 5(b).

Regarding the type-dependent production levels, note that improved asset quality in the signal market encourages more purchases by intermediaries, which boosts production from both types of originators. Taken together, as α increases, the loosened eligible selling constraint effect and the elevated intermediary demand effect work in the same direction for H-type originators, who therefore expand their production q_H unambiguously. For L-type originators, however, these two forces are opposed, but the tightened eligible selling constraint dominates so q_L goes down with α, as shown in Figure 5(a).

The comparative statics results on α shown in Figure 4 are robust to all the parameterizations we have tried. In fact, we can formally prove the following proposition that



(a) Asset production (b) Equilibrium prices

Fig. 4. Intermediated equilibrium outcomes as functions of α

Market equilibrium outcomes in an intermediated economy as functions of intermediation technology α. Panel 5(a) plots the equilibrium asset production q_θ; Panel 5(b) plots the asset quality in signal market (x_h) and equilibrium prices (p_A and k). In both panels, the cutoff $\hat{\alpha}_1(\rho_I)$ determining the relevance of intermediary's technology are highlighted. Parameterization: $\pi = 0.55$, $X = 3$, $\lambda = 0.4$, $\rho = 0.5$, $\rho_I = 0.1$ and $\kappa = 0.05$.

states these results in the vicinity around the direct economy (where $\rho_I = 0$) and under a mild condition regarding the corresponding equilibrium production Q^d.

Proposition 6. ***(Comparative statics of α).*** *Focus on the tech-relevant range $\alpha \geq \hat{\alpha}_1(\rho_I)$. Consider infinitesimal intermediation capacity $\rho_I > 0$, and suppose that the direct economy production in (22) satisfies $Q^d \geq \max\left\{1, \frac{1-\pi}{\pi-1+\lambda/\rho}\right\}$. As intermediaries' technology level α improves,*

1. *in the h signal market, originators' selling wedge widens, i.e., $\frac{\partial(s_H - s_L)}{\partial\alpha} > 0$, asset quality increases, i.e., $\frac{\partial x_h}{\partial\alpha} > 0$, and intermediaries increase their asset purchase, i.e., $\frac{\partial b_I}{\partial\alpha} > 0$;*
2. *H-type originators' production expands while L type originators' shrinks, i.e., $\frac{\partial q_H}{\partial\alpha} > 0$ and $\frac{\partial q_L}{\partial\alpha} < 0$, which implies asset market price improves, i.e., $\frac{\partial p_A}{\partial\alpha} > 0$;*
3. *the total production of assets and input capital price increase, i.e., $\frac{\partial Q}{\partial\alpha} > 0$ and $\frac{\partial k}{\partial\alpha} > 0$.*

4.2. *Normative Analysis of Intermediation Technology α*

We now formally establish the welfare implications of intermediation in this economy of asset origination and distribution. Our main interest is on the intermediation technology α. We focus on two "corners" of the intermediated equilibria—one with relatively uninformed intermediation in the tech-irrelevant range (Proposition 4), and the other with perfectly informed intermediation with $\alpha = 1$, a special case of the tech-relevant economy (Corollary 1).

Welfare function and key results. To compare Proposition 4 or Corollary 1 for an intermediated originate-to-distribute economy (with $\rho_I > 0$) to Proposition 2 for a direct trading economy (with $\rho_I = 0$), we define the (ex ante) payoff to an originator as $v_O \equiv \pi v_H(q_H) + (1-\pi)v_L(q_L)$, with $v_\theta(q_\theta)$ given by Eq. (27) and evaluated at the equilibrium production levels q_θ. Further, define the social welfare in an intermediated economy, which is indexed by the intermediary sector's risk-bearing capacity ρ_I, as

$$w(\rho_I) \equiv v_O + v_I + v_K, \tag{40}$$

where v_I is the equilibrium payoff to an intermediary in Eq. (29) and $v_K \equiv kQ - K(Q)$ is the equilibrium profit by input capital producers.[39] Accordingly, the social welfare in a benchmark direct trading economy without intermediaries is

$$w(\rho_I = 0) \equiv w^d \equiv v_O^d + v_K^d. \tag{41}$$

We have the following proposition.

Proposition 7. *We have the following two welfare results on tech-driven intermediation.*

1. ***(Potential) Welfare impairment by less informed intermediation.*** *In the tech-irrelevant range $\alpha \leq \hat{\alpha}_1(\rho_I)$, which includes the uninformed case $\alpha = \frac{1}{2}$, the social surplus in an intermediated economy is below that in a direct trading economy, i.e., $w(\rho_I) < w^d$, if and only if $\pi X < 1$.*
2. ***Welfare improvement by perfectly informed intermediation.*** *Suppose $\alpha = 1$. The social surplus is always higher in an intermediated economy than that in a direct trading economy, i.e., $w(\rho_I) > w^d$, for any $\rho_I > 0$.*

As explained in Sections 2.3 and 3.2.1, the welfare in this economy hinges on two key factors: i) whether production is mostly conducted by H-type originators (production efficiency), and ii) how assets retained due to market illiquidity are allocated (allocative efficiency). Intermediation performed by uninformed or informed agents impacts welfare depending on how it affects efficiency in these two respects.

(Potential) welfare impairment by uninformed intermediation. In the case of uninformed intermediation (part 1 of Proposition 7), in Appendix A4.2 we show that social

[39] One could think of all agents (i.e., originators, intermediaries and capital producers) in the model as firms in an economy, so that maximization of social welfare requires maximizing the total firm profits, which are $w(\rho_I)$ as defined in Eq. (40).

surplus attained by the intermediated economy in the tech-irrelevant range is

$$w = \underbrace{\left(\kappa + \frac{\lambda^2}{\rho + \rho_I}\right)\frac{Q^2}{2}}_{\text{quantity-based social gain}} + \underbrace{\frac{\rho}{2}\pi(1-\pi)X^2}_{\text{quality-based social gain}} = \underbrace{(\pi X - 1)\frac{Q}{2}}_{\text{baseline gain}} + \underbrace{\frac{\rho(1-\pi)\pi}{2\lambda}X^2}_{\text{incentive gain}}, \tag{42}$$

where the second equality is based on the equilibrium condition Eq. (34). The welfare expressions given by Eq. (42), which only involve the endogenous total production Q and other exogenous parameters, are strikingly simple; in the second expression, we see that ρ_I affects welfare w only through the endogenous total production Q.

To understand (42), the welfare in this economy consists of a *production quantity*–based component that stems from quadratic costs in capital adjustment and holding retention (which thus implies an equilibrium profit that is quadratic in optimal production or retention volume), and a *production quality*–based component from disciplining lemon production. Our analysis reveals that when intermediation is relatively less informed, the equilibrium relation between production scale Q and production quality p_A leads to a constant social gain $\frac{\rho(1-\pi)\pi}{2\lambda}X^2$ from retention based incentives, together with a linear contribution from production $(\pi X - 1)\frac{Q}{2}$ as the baseline gain (see detailed derivation in Appendix A4.2). Therefore, the welfare impact of allowing uninformed intermediaries to operate—which effectively increases intermediaries' retention capacity ρ_I from zero (as in a direct trading economy) to positive and hence leads to more production Q—simply hinges on whether the average return without contingent production, πX, exceeds 1.

Economically, introducing uninformed intermediation has a mixed effect on economic efficiencies: while ex post homogeneous asset retention by intermediaries improves the (allocative) distribution efficiency, it entails lowering the production efficiency in the economy. Specifically, the operation of uninformed intermediaries fosters an h signal market where both types of asset originators sell the exact same quantity of their produced assets, which in turn hurts production efficiency as asset production in the economy is effectively made to be less state/type contingent. In our model, uninformed intermediaries later bring these relatively low-quality assets purchased in the signal markets to sell in the asset market, which effectively weakens the disciplining and brings down the equilibrium trading price in the asset market.

Welfare improvement by informed intermediation. In the technology-relevant range with $\alpha \geq \hat{\alpha}_1(\rho_I)$, informed intermediation in general brings about welfare improvement, with the sharpest result for perfectly informed intermediaries (part 2 of Proposition 7). When $\alpha = 1$, no low-quality assets receive the h signal, implying no lemons being brought to the signal market since the l signal market endogenously closes. More importantly, intermediaries' "voluntarily committing" to only purchasing assets from the h signal market helps achieve a socially efficient cleansing in asset production and hence effectively facilitates the state/type contingency of asset production in the economy. As such, perfectly informed intermediation always improves social welfare by not only smoothing the distribution of asset retention, but also facilitating more efficient utilization of input capital in asset production.

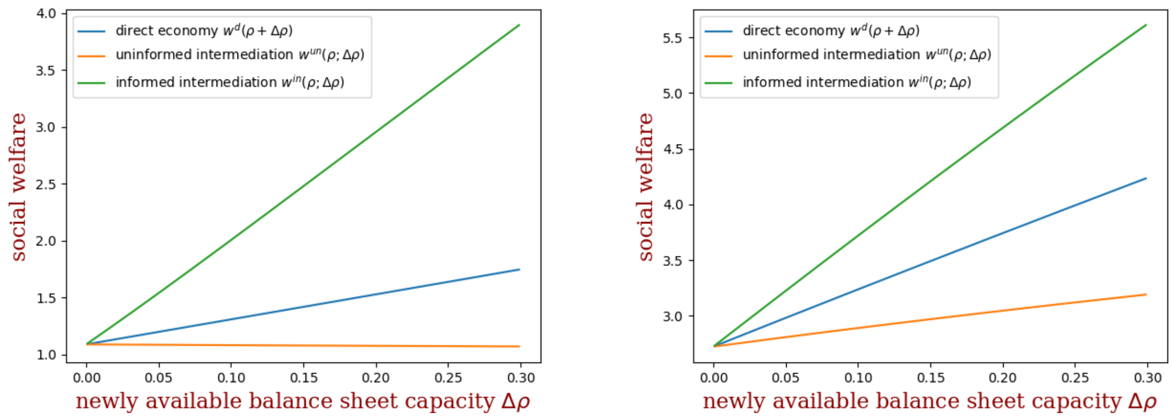
One can also understand the welfare implications of perfectly informed intermediation ($\alpha = 1$) by comparing it to the direct originate-to-distribute economy with a hypothetical intertype market, where L-type originators (like perfectly informed intermediaries) purchase assets from the H-type and then resell them in the asset market (Section 3.2.2). Resembling the full efficiency restoration in this fictitious economy (Proposition 3), perfectly informed intermediaries ensure that only H-type assets are traded in the signal market (and resold in the asset market), pushing the economy towards constrained efficiency. Nevertheless, such efficiency improvement limited as L-type originators still make positive production (as opposed to ceasing production completely under a frictionless intertype market).

4.3. *Regulation Implication: How New Balance Sheets Should Be Allocated?*

If new balance sheets capacity were to be made available in the economy (e.g., through loosened regulations on equity injection or acquisition activities), should these new balance sheet capacity go to intermediaries or originators? The answer to this question has direct and important regulatory implications—for instance, it provides useful insights to regulators regarding whether merging between originators (e.g., JP Morgan) and intermediaries (e.g., Blackstone) should be allowed.

Consider an direct trading economy where originators have risk retention capacity ρ (and intermediaries have capacity $\rho_I = 0$). Now suppose an extra $\Delta\rho > 0$ capacity is available to be added to the economy. Define $w^d(\rho + \Delta\rho)$ as the welfare attained in a direct trading economy where originators have a capacity of $\rho + \Delta\rho$; and $w^I(\rho; \Delta\rho)$ stands for the welfare in an intermediated economy where the capacity of originators and intermediaries is ρ and $\Delta\rho$ respectively.

Let $w^I(\rho; \Delta\rho) \in \{w^{un}(\rho; \Delta\rho), w^{in}(\rho; \Delta\rho)\}$, where intermediaries are either uniformed



(a) $\pi = 0.3$ (b) $\pi = 0.6$

Fig. 5. How new balance sheets should be allocated?
This figure plots functions $w^d(\rho + \Delta\rho)$, $w^{un}(\rho; \Delta\rho)$ and $w^{in}(\rho; \Delta\rho)$ against the newly added balance sheet capacity $\Delta\rho$, in blue, orange and green colors respectively. The exercise is preferred for two different values of H type share: $\pi = 0.3$ and $\pi = 0.6$. Parameterization: $X = 3$, $\lambda = 0.4$, $\rho = 0.5$, and $\kappa = 0.05$.

"*un*" or perfectly informed "*in*."[40] We are interested in the following comparison:

$$w^d(\rho + \Delta\rho) \lesseqgtr w^I(\rho; \Delta\rho). \tag{43}$$

While analytical proof is challenging, our numerical analysis delivers a robust ordering of the social welfare associated with different allocations of $\Delta\rho$:

$$w^{un}(\rho; \Delta\rho) < w^d(\rho + \Delta\rho) < w^{in}(\rho; \Delta\rho). \tag{44}$$

That is, when there is extra balance sheet capacity $\Delta\rho > 0$ available to be added to the financial system, it is more efficient to allocate it to intermediaries than to originators.

Figure 5 plots functions $w^d(\rho + \Delta\rho)$, $w^{un}(\rho; \Delta\rho)$ and $w^{in}(\rho; \Delta\rho)$ against $\Delta\rho$ for different levels of H type fraction π. Consistent with Proposition 7, the welfare impact of allocating new balance sheets to uninformed intermediaries $w^{un}(\rho; \Delta\rho)$ is negative when $\pi X < 1$ (e.g., in Panel *a* with $\pi = 0.3$) and is positive when $\pi X > 1$ (e.g., in Panel *b* with $\pi = 0.6$). In both panels, the ordering described in Eq. (44) is consistently observed.

Intuition of $w^{un}(\rho; \Delta\rho) < w^d(\rho + \Delta\rho)$. One critical factor determining the welfare impact of introducing new balance sheet capacity into the originate-to-distribute system is the extra production brought about by the newly added balance sheet. When uninformed

[40]In particular, uninformed intermediaries have an α satisfying $\alpha \leq \hat{\alpha}_1(\rho_I)$ as in Proposition 4, while perfectly informed intermediaries have an $\alpha = 1$.

intermediaries are holding these newly added balance sheets, the extra production in the system will be determined by the incentives of the H/L type originators to dump assets to intermediaries. Since selling assets to uninformed intermediaries is at the same price without retention, both types of originators have an equalized incentive of producing and then dumping assets to uninformed intermediaries (as suggested by Proposition 4). In contrast, when new balance sheet is added to originators (who have to sell in the illiquid asset market directly), the quality difference between the extra assets sit on H/L type originators' balance sheets induce them to respond differently in terms of making extra production. As a consequence, allocating the newly available balance sheet capacity to originators themselves would induce more efficient (extra) production in the originate-to-distribute system.

Intuition of $w^d(\rho + \Delta\rho) < w^{in}(\rho; \Delta\rho)$. This result is driven by the other factor regarding efficient asset allocation that affects the welfare impact of a larger balance sheet (Proposition 1, item 2). We have seen that new balance sheets directly allocated to originators induces more efficient (extra) production. On the other hand, adding balance sheet to perfectly informed intermediaries also induces relatively efficient (extra) production. But these two ways differ in how the newly produced assets are being distributed in the system. In $w^d(\rho + \Delta\rho)$, $1 - \pi$ fraction of these newly added capacity (i.e., those held by L type originators) is not fully exploited, while in $w^{in}(\rho; \Delta\rho)$ all newly added capacity (allocated to informed intermediaries) is fully exploited.

5. Model Extensions and Discussions

By studying several extensions of our model, this section discusses the robustness of the results and messages delivered by our analysis under alternative model specifications.

5.1. *Endogenizing Illiquidity λ in a Directed Search Framework*

We provide a microfoundation for the key exogenous parameter in our analysis—the minimum retention ratio or the market illiquidity λ. Specifically, we follow the directed search framework developed in Guerrieri and Shimer (2014) to derive a positive probability of failed trading endogenously. For illustration purposes, our analysis focuses on the direct trading economy as in Section 3.1.

Consider the direct trading economy, but the unit measure of risk neutral investors are with endowment e and reservation return of 1. Two asset markets $m \in \{\mathcal{H}, \mathcal{L}\}$ could potentially exist in equilibrium, each captured by the trading price p_m and illiquidity (one minus buyer-seller ratio) λ_m. As in Guerrieri and Shimer (2014) where high-quality assets are taken to the illiquid market for signaling purposes, we focus on equilibria where exactly one of the two markets is perfectly liquid, i.e., $\lambda_{\mathcal{L}} = 0$ while $\lambda_{\mathcal{H}} > 0$. It can be shown that the illiquid $\mathcal{H}$ market, which corresponds to the asset market in our main model specification (with exogenous λ), always exists in equilibrium.[41]

A θ-type originator optimally chooses the amount of produced assets $s_{\theta m}$ taken to asset market m for sale, taking as given $\{p_m, \lambda_m\}$ in both markets. In equilibrium, originators are indifferent between taking the last unit of their produced assets to either market (if both markets have positive trading volume), based on a trade-off of selling pricing against trading illiquidity à la Guerrieri and Shimer (2014). The average quality of assets trading in market m is thus $x_m = \frac{\sum_\theta x_\theta \pi_\theta s_{\theta m}}{\sum_\theta \pi_\theta s_{\theta m}}$, and risk-neutral investors with unity reservation return set the asset price $p_m = x_m$.

In any market, whenever the total value of assets brought in by sellers exceeds the endowment buyers bring to the market, sellers are rationed to clear the market so that the probability of selling orders being executed equals

$$\Theta_m = \min\left\{\frac{b_m}{p_m \sum_\theta \pi_\theta s_{\theta m}}, 1\right\}, \tag{45}$$

where $b_m \in [0, e]$ is the total endowment brought by buyers to market m. Intuitively, from the perspective of sellers, market liquidity is reflected by the equilibrium buyer-seller ratio (if less than one). Connecting back to our setting, λ_m, which is the illiquidity of market m, is defined as $1 - \Theta_m$; the higher the equilibrium buyer-seller ratio, the greater the chance that sell orders get executed, the lower the illiquidity. The above rationing rule in (45) also implies the market clearing condition:

$$\sum_m \left\{ (1-\lambda_m) p_m \cdot \underbrace{\sum_\theta \pi_\theta s_{\theta m}}_{\text{total sale in market } m} \right\} = \underbrace{\sum_m b_m = e}_{\text{total buying orders in both markets}} . \tag{46}$$

Similar to our analysis of the intermediated equilibrium in Section 3.3.2, potential equilibrium indeterminancy arises when both originators have access to trading in a perfectly liquid asset market. As with our analysis there, we adopt the exact same sequential equilibrium refinement by introducing a sequence of equilibria such that asset trading in the liquid $\mathcal{L}$ market fails with probability $\nu > 0$ (where $\nu \to 0$). Analogously, under this

[41] The liquid asset market $\mathcal{L}$ with $\lambda_{\mathcal{L}} = 0$ might have zero trading in equilibrium under certain parameterizations; see details in Appendix A5.1.

sequential equilibrium refinement, the equilibrium average quality of assets trading in the perfectly liquid market is $x_{\mathcal{L}} = \pi X$.

One can follow Guerrieri and Shimer (2014) to uniquely determine the equilibrium with directed search, which consists of endogneous asset market price and illiquidity $\{\lambda_m, p_m\}_{m \in \{\mathcal{H}, \mathcal{L}\}}$ and other equilibrium outcomes (e.g., production q_θ, asset selling $s_{\theta m}$). For details, see Appendix A5.1.

5.2. *Observable Seller Identity*

Our analysis thus far has assumed away the observability of sellers' occupation identity—i.e., asset originators or intermediaries—to outside buyers in the asset market; therefore assets sold by originators or intermediaries are pooled together as described by Eq. (7).[42]

We now extend the analysis to a generalized setting where investors in the asset market can (imperfectly) observe seller identity and thus potentially separately price assets sold by originators or intermediaries respectively. Specifically, after intermediaries make their asset purchasing decisions (at date $\frac{1}{2}$), we assume that each individual intermediary is subject to an idiosyncratic "identity observability" shock in her date 1 trading in the asset market: with probability z her selling is executed at an "intermediary" asset market price $p_A^I \geq 0$, while with probability $1 - z \in [0, 1]$ her selling is executed at same pooling price $p_A \geq 0$ as that of originators. Our baseline model maps into the case with $z = 0$; detailed analysis for general z is in Appendix A5.2.

The pricing specification in the asset market affects how the economic surplus gets allocated between originators and intermediaries. With a single prevailing asset trading price, intermediaries' technology α (which determines the quality of their purchased assets) can affect the originators' equilibrium payoff via the selling price that commonly applies to both agents. More specifically, intermediation activities performed by relatively uninformed intermediaries hurt the originators' surplus via a lowered asset market price, while sufficiently informed intermediation generates a positive (pecuniary) externality benefiting originators. In contrast, these spillover effects on originators' selling price are absent if investors can perfectly separately price the assets sold by intermediaries ($z = 1$).

Interestingly, as shown in the next proposition (focusing on the other corner case with $z = 1$), unlike in our benchmark analysis ($z = 0$), less informed intermediation can never hurt social welfare when seller identity is perfectly observable in asset market trading. Intuitively, the observability of sellers' occupations provides a natural disciplining on inter-

[42] One could also interpret this pooled asset selling price for originators and intermediaries by linking to the direct trading economy intertype market as characterized in Section3.2.2. There L-type originators are effectively performing the role of intermediaries but always sell assets at the same price as H-type ones in the asset market.

mediaries' trading when their purchased assets are relatively low quality (which occurs when α is low). In fact, it can be shown that with sellers' occupation identity observable in the asset market, in the tech-irrelevant range, intermediaries will not make any purchase from signal market(s) if the lemons problem in the economy is sufficiently severe such that $\pi X < 1$. In this case, no gain is generated from trading between originators and intermediaries as assets purchased by intermediaries have a value of πX (which is the equilibrium originator's selling price in the asset market), while a cost of $k > 1$ for originators to produce. In contrast, with a pooling asset market price as in our main model, less informed intermediaries will always actively acquire assets and then sell, which could hurt social welfare when $\pi X < 1$ (Proposition 7).[43]

Proposition 8. ***(Welfare impact of intermediation with observable occupation identity).*** *When $z = 1$ so that asset sales by originators and intermediaries in the asset market are executed at perfectly separated prices, intermediaries' operation always (weakly) improves social welfare, i.e., $w(\rho_I) \geq w^d$ for $\rho_I > 0$.*

As a policy implication, this result thus suggests more transparent seller identity disclosure when intermediaries' information technology is not sufficiently developed.

5.3. Indivisible Trading in Asset Market

Market illiquidity in our analysis thus far amounts to a "partial retention" in which any arbitrary amount q of assets brought to the market for sale ends up with λq post-trading retention. While this specification of market illiquidity is suitable for modeling the trading of financial assets, it is less appealing for the application of real goods/assets.

In an alternative setting of indivisible assets, market illiquidity is captured by a positive probability λ of an asset sale failing to be executed, in which case the seller bears the additional retention cost applied to the entire amount of assets he brings to the market. All of our analyses stay qualitatively robust (see Appendix A5.3). Intuitively, under indivisible asset trading, each sellers' exposure to market illiquidity effectively increases due to the convex retention cost; this leads to a stronger disciplining than can be supported under divisible trading. One should thus expect the importance of intermediation (and its technology level α) to be dampened when assets are less divisible, due to a diminished disciplining value of informed intermediation when originators are already effectively disciplined (by market illiquidity itself).

[43]Being able to pool with originators, the selling price of the intermediaries' purchased assets is $p_A > k$, so a gain from "intermediation" always exists.

6. Conclusion

Development of information technology in recent decades has greatly reshaped the way economic activities and financial transactions are carried out in the economy. Among others, the "recommerce" of a wide variety of "goods" ranging from financial assets (e.g., asset-backed securities) to real goods (e.g., clothing, vehicles, luxury items) has been drastically expanding, behind which the operation of intermediaries as "middlemen" plays a critical role. This paper focuses on an important representation of these growing "recommerce" practices in the economy—the fast expansion of intermediation services in the traditional originate-to-distribute models operated by the banking sector. Our analysis provides a novel understanding of the economics beneath such intermediation in the origination and distribution processes of assets subject to the lemons problem.

Adopting a market equilibrium approach, our model highlights the operation of intermediaries as a market solution to economic inefficiencies identified in the originate-to-distribute process. Allowing intermediaries to operate has a mixed social welfare effect: while uninformed intermediation can potentially impair social surplus by weakening market illiquidity's disciplining effect, with a common balance sheet, sufficiently informed intermediaries are induced to "cherry pick" in their asset trading, effectively rendering them incentive-compatible delegated monitors who improve social surplus.

Our analysis of the intermediated economy contributes to a better understanding of the social value and welfare implications of the ever-growing financial expertise represented by intermediation activities (among others), an issue of significant policy and regulatory importance also studied in recent work such as Kurlat (2019). Further, our paper also sheds light on how asset trading may endogenously evolve as technology progresses in the economy—transactions previously reliant on direct disciplining from trading counterparties (e.g., via OTC-based trading) could now be implemented in a Walrasian manner, thanks to the operation of informed intermediaries.

Finally, although our model is motivated by the banking sector's originate-to-distribute model with underlying assets being typically loans or financial claims, the framework can speak to the more general "recommerce" practice involving real goods transactions.[44] Like financial intermediaries in our model, the middlemen in these real goods "recommerce" practice often rely on cutting-edge information technology to authenticate and screen the quality of goods purchased from original owners, and our model helps understand the economics behind these businesses.

[44] In these "recommerce" practices, important technology-based real goods middlemen include second-hand car retailers (e.g., Carmax and TrueCar), second-hand luxury goods (e.g., TheRealReal and Fashionphile), and second-hand clothing (e.g., ThredUP).

References

Asriyan, V., Fuchs, W., and Green, B. (2017). Information spillovers in asset markets with correlated values. *American Economic Review*, 107(7):2007–2040.

Atkeson, A. G., Eisfeldt, A. L., and Weill, P.-O. (2015). Entry and exit in OTC derivatives markets. *Econometrica*, 83(6):2231–2292.

Benmelech, E., Dlugosz, J., and Ivashina, V. (2012). Securitization without adverse selection: The case of clos. *Journal of Financial Economics*, 106(1):91–113.

Bord, V. and Santos, J. (2012). The rise of the originate-to-distribute model and the role of banks in financial intermediation. *Economic Policy Review*, 18:21–34.

Caballero, R. J. and Krishnamurthy, A. (2003). Excessive dollar debt: Financial development and underinsurance. *Journal of Finance*, 58(2):867–893.

Cordell, L., Roberts, M. R., and Schwert, M. (2023). Clo performance. *The Journal of Finance*, 78(3):1235–1278.

Dang, T. V., Gorton, G., Holmström, B., and Ordoñez, G. (2017). Banks as secret keepers. *American Economic Review*, 107(4):1005–29.

Dávila, E. and Korinek, A. (2018). Pecuniary externalities in economies with financial frictions. *Review of Economic Studies*, 85(1):352–395.

DeMarzo, P. and Duffie, D. (1999). A liquidity-based model of security design. *Econometrica*, 67(1):65–99.

DeMarzo, P. M. (2005). The pooling and tranching of securities: A model of informed intermediation. *Review of Financial Studies*, 18(1):1–35.

Diamond, D. W. (1984). Financial intermediation and delegated monitoring. *Review of Economic Studies*, 51(3):393–414.

Drucker, S. and Puri, M. (2008). On loan sales, loan contracting, and lending relationships. *Review of Financial Studies*, 22(7):2835–2872.

Eisfeldt, A. L. (2004). Endogenous liquidity in asset markets. *Journal of Finance*, 59(1):1–30.

Fuchs, W., Fukuda, S., and Neuhann, D. (2024). Rules versus disclosure: Prudential regulation and market discipline. *Available at SSRN 4799706*.

Glode, V. and Opp, C. (2016). Asymmetric information and intermediation chains. *American Economic Review*, 106(9):2699–2721.

Glode, V. and Opp, C. C. (2023). Private renegotiations and government interventions in credit chains. *Review of Financial Studies*, 36(11):4502–4545.

Gorton, G. and Pennacchi, G. (1990). Financial intermediaries and liquidity creation. *Journal of Finance*, 45(1):49–71.

Gorton, G. and Pennacchi, G. (1995). Banks and loan sales marketing nonmarketable assets. *Journal of Monetary Economics*, 35(3):389–411.

Guerrieri, V. and Shimer, R. (2014). Dynamic adverse selection: A theory of illiquidity, fire sales, and flight to quality. *American Economic Review*, 104(7):1875–1908.

Hauswald, R. and Marquez, R. (2003). Information technology and financial services competition. *The Review of Financial Studies*, 16(3):921–948.

Hayashi, F. (1982). Tobin's marginal q and average q: A neoclassical interpretation. *Econometrica*, 50(1):213–224.

He, Z. (2009). The sale of multiple assets with private information. *Review of Financial Studies*, 22(11):4787–4820.

He, Z., Huang, J., and Parlatore, C. (2023). Bank competition with specialized lenders. Working Paper 30403, National Bureau of Economic Research.

He, Z. and Kondor, P. (2016). Inefficient investment waves. *Econometrica*, 84(2):735–780.

He, Z. and Krishnamurthy, A. (2013). Intermediary asset pricing. *American Economic Review*, 103(2):732–70.

He, Z. and Li, J. (2023). Intermediation via credit chains. Technical report, National Bureau of Economic Research.

Hirshleifer, J. (1971). The private and social value of information and the reward to inventive activity. *American Economic Review*, 61(4):561–574.

Ho, T. and Stoll, H. R. (1980). On dealer markets under competition. *Journal of Finance*, 35(2):259–267.

Hugonnier, J., Lester, B., and Weill, P.-O. (2019). Frictional intermediation in over-the-counter markets. *Review of Economic Studies*, 87(3):1432–1469.

Jiang, E. X. (2023). Financing Competitors: Shadow Banks' Funding and Mortgage Market Competition. *The Review of Financial Studies*, 36(10):3861–3905.

Jiang, E. X., Matvos, G., Piskorski, T., and Seru, A. (2020). Banking without deposits: Evidence from shadow bank call reports. *NBER Working Paper No. w26903*.

Keys, B. J., Mukherjee, T., Seru, A., and Vig, V. (2010). Did securitization lead to lax screening? evidence from subprime loans. *Quarterly Journal of Economics*, 125(1):307–362.

Kreps, D. M. and Wilson, R. (1982). Sequential equilibria. *Econometrica*, 50(4):863–894.

Kundu, S. (2023). The anatomy of corporate securitizations and contract design. *Journal of Corporate Finance*, 81:102195. Private Equity.

Kurlat, P. (2013). Lemons markets and the transmission of aggregate shocks. *American Economic Review*, 103(4):1463–1489.

Kurlat, P. (2019). The social value of financial expertise. *American Economic Review*, 109(2):556–90.

Leland, H. E. and Pyle, D. H. (1977). Informational asymmetries, financial structure, and financial intermediation. *Journal of Finance*, 32(2):371–387.

Lorenzoni, G. (2008). Inefficient credit booms. *Review of Economic Studies*, 75(3):809–833.

Opler, T., Pinkowitz, L., Stulz, R., and Williamson, R. (1999). The determinants and implications of corporate cash holdings. *Journal of Financial Economics*, 52(1):3–46.

Park, C. (2000). Monitoring and structure of debt contracts. *Journal of Finance*, 55(5):2157–2195.

Parlour, C. A. and Plantin, G. (2008). Loan sales and relationship banking. *Journal of Finance*, 63(3):1291–1314.

Sambalaibat, B. (2021). A theory of liquidity spillover between bond and cds markets. *Review of Financial Studies*, 35(5):2525–2569.

Selten, R. (1975). Reexamination of the perfectness concept for equilibrium points in extensive games. *International Journal of Game Theory*, 4:25–55. 10.1007/BF01766400.

Shen, J., Wei, B., and Yan, H. (2021). Financial intermediation chains in an over-the-counter market. *Management Science*, 67(7):4623–4642.

Stein, J. C. (1997). Internal capital markets and the competition for corporate resources. *Journal of Finance*, 52(1):111–133.

Vanasco, V. (2017). The downside of asset screening for market liquidity. *Journal of Finance*, 72(5):1937–1982.

Internet Appendix

A1. ***Proofs and Calculations for Constrained Efficient Allocation***

This section provides algebraic proof and detailed calculations for the analysis of a planner's constrained efficient allocation.

Proof of Proposition 1

In this proof, we prove the Proposition 1 which characterizes the constrained efficient allocation as defined in Definition 2.

To show that in the planner's constrained efficient allocation only H-type assets are produced, suppose instead $q_L^\star > 0$ in the planner's solution to Eq. (11). Now consider the following deviation: $\tilde{q}_L = q_L^\star - \delta$ and $\tilde{q}_H = q_H^\star + \frac{1-\pi}{\pi}\delta$, where $\delta > 0$ is infinitesimal such that $\tilde{q}_L > 0$. Following this deviation, it can be seen that the quantity of total production $Q \equiv \sum_\theta \pi_\theta q_\theta$ is held constant. As such, the production cost of input capital $K(Q)$ is unaffected while the original allocation of asset retention $\{r_\theta^O, r^I\}$ is still feasible with the new production scheme.

Therefore following the deviation as described above, the minimum retention constraint still holds under original allocation of asset retention and the social surplus as defined in Eq. (11) is increased by

$$\Delta S = \pi x_H \left(\frac{1-\pi}{\pi}\right)\delta - (1-\pi)x_L\delta = (1-\pi)X\delta > 0$$

It thus follows that in planner's constrained efficient allocation, L-type originators should not produce any thing. Further, the convexity in retention cost function $R(\cdot)$ implies that the post-trading allocation of asset retention must be equalized across the balance sheets of all agents. This implies that in the planner's constrained efficient allocation,

$$r_\theta^O = r^I = \lambda Q^\star \tag{A.1}$$

for $\theta \in \{H, L\}$, where $Q^\star \equiv \pi q_H^\star$ is the total quantity of assets produced (all of which are H-type ones) and is determined by equation

$$x_H = K'(Q) + \lambda R'(\lambda Q). \tag{A.2}$$

From Eq. (A.2) it can be seen that aggregate production quantity $Q^\star$ implied by the planner's constrained efficient allocation does not depend on π or x_L. ■

A2. ***Proofs and Calculations for Direct Trading Economy***

A2.1. ***Characterization of the benchmark direct trading economy***

This section provides algebraic proofs and detailed calculations for our analysis of the benchmark direct trading economy (Section 3.1).

Proof of Lemma 1

In this proof, we show that equilibrium trading prices satisfy $p_A^d - k^d > 0$ and both types of asset originators bring all produced assets to the asset market for sale.

First, if $p_A^d \leq k^d$, it then immediately follows that L-type originators must be not producing anything in equilibrium (as $k^d \geq p_A^d \geq x_L$, so marginal benefit of L-types production is no higher than k). Therefore $p_A^d = x_H$ given no lemons are being produced or traded in the economy. But this implies that $k^d \geq x_H$, which hence leads to H-type originator to be also not producing anything in equilibrium. Contradiction.

Next, suppose in equilibrium a H-type originator is not taking all his produced assets to the asset market for sale. Note we focus on H-type originators as they have stronger incentives to keep assets from taking to the asset market, if any, than their L-type counterparts. Consider the following deviation strategy for this H-type originator: increasing the amount of assets brought to the asset market by an arbitrarily small positive $\delta > 0$, while increasing the production by $(1-\lambda)\delta$ and then hold these newly produced assets. Following this deviation, the change in the post-trading retention is

$$\underbrace{(1-\lambda)\delta}_{\text{new production}} - \underbrace{(1-\lambda)\delta}_{\text{increased sale}} = 0.$$

Therefore, following the above deviation the payoff change is

$$\Delta(\delta) = \underbrace{(1-\lambda)\delta(p_A^d - x_H)}_{\text{increased sale}} + \underbrace{(1-\lambda)\delta(x_H - k^d)}_{\text{new production}} = (1-\lambda)\delta(p_A^d - k^d).$$

By definition, $\Delta(\delta) \leq 0$, which thus implies $p_A^d \leq k^d$. But this implies L-type originators are not producing as discussed above. This in turn leads to the no production of H-type, which hence gives rise to a contradiction. ■

Proof of Proposition 2

In this proof, we provide detailed calculations for the characterization of market equilibrium in a direct trading economy.

To begin with, first consider the case with pure lemon trading. Suppose in equilibrium only L-type assets are traded in the asset market, it then follows that $p_A^d = x_L$. But since market illiquidity $\lambda > 0$, it must be that $k^d < x_L$ so that L-type originators are producing (and hence trading) in equilibrium. But at this equilibrium input capital price, it can be shown that H-type originators are able to make strictly positive profit by deviating to simultaneously produce and bring to asset market for sale infinitesimal more amount of assets at certain properly chosen ratio.

To see this, note that with equilibrium $p_A^d = x_L = 0$, if in equilibrium a H-type originator produces q_H^d, then it must be that $R'(q_H^d) \leq x_H - p_A^d$ since he is not bringing anything to the asset market. With $k^d < p_A^d$, a type H originator must be producing positive q_H and if in equilibrium he does not sell anything in the asset market, he can always be better off by making the following deviation: bring δ units of already produced assets for sale in the asset market and produce $(1-\lambda)\delta$ units of new assets which he holds by himself. Following this strategy, the H-type originator's post-trading retention is held fixed and the payoff change is: $(1-\lambda)\delta(p_A^d - k^d) > 0$. Hence in equilibrium H-type originator must also sell in the asset market. This contradiction rules out the possibility of a pure "lemon" trading equilibrium.

Next, consider the other corner case with $p_A^d = x_H = X$, we show that a market equilibrium with no lemon trading can be sustained if and only if $\frac{X-1}{X} \leq \frac{\rho\kappa\pi}{\lambda} + \lambda$. To see this, note that in such a no "lemon" trading equilibrium a L-type asset originator would like to produce a bit (and tries to sell) if and only if $(1-\lambda)p_A^d > k^d$, thanks to the Inada condition ($\lim_{x\to 0^+} R'(x) = 0$). When $p_A^d = X$, this condition corresponds to

$$(1-\lambda)X > K'\left(\hat{Q}\right), \tag{A.3}$$

where $\hat{Q}$ solves $X = K'\left(\hat{Q}\right) + \lambda R'\left(\frac{\lambda\hat{Q}}{\pi}\right)$; and with $K(Q) = Q + \frac{\kappa}{2}Q^2$ this condition can be rewritten as $\frac{X-1}{X} > \frac{\rho\kappa\pi}{\lambda} + \lambda$. When this condition holds, the market equilibrium in which $p_A^d = x_H$ cannot be sustained as L-type would always want to bring in positive amount of lemons, which then drives the equilibrium price p_A^d below x_H. In such a corner equilibrium with no lemon production and trading, asset market price is $p_A = X$ and equilibrium capital price is $k = K'(\hat{Q})$ where $\hat{Q}$ is as defined above. Combined with H-type's optimality condition $X - \lambda R'(\lambda q_H) = k^d$, we obtain $k^d = \frac{\kappa\pi\rho X + \lambda^2}{\kappa\pi X + \lambda^2} > 1$ and $q_H = \frac{\rho(X-1)}{\kappa\pi\rho + \lambda^2} > 0$.

Finally, when $\frac{X-1}{X} > \frac{\rho\kappa\pi}{\lambda} + \lambda$, as discussed above, the market equilibrium in a direct trading economy is an interior one and is determined by Eq. (20). In this case, the unique positive solution of equilibrium capital price Q^d is given by Eq. (22). Furthermore, combining originators' optimality condition Eq. (16) with the equilibrium capital price Eq. (9), we get equilibrium asset market price p^A as an affine function of k^d, i.e., Eq. (23). By Eq. (20), it follows that

$$\left(\kappa + \frac{\lambda^2}{\rho}\right) Q^d > \pi X - 1.$$

Thus from Eq. (23) it follows that

$$p_A^d = \frac{1}{1-\lambda}\left[\left(\kappa + \frac{\lambda^2}{\rho}\right) Q^d - \lambda\pi X + 1\right] > \pi X$$

Lastly, to show that $q_L^d = \frac{k^d-1}{\kappa} - \frac{\rho\pi X}{\lambda} > 0$, it is equivalent to show that $\lambda(k^d-1) > \kappa\rho\pi X$. Rewrite Eq. (20) as

$$\left(\frac{1}{\lambda} + \frac{\lambda}{\rho\kappa}\right)\lambda(k^d-1) + 1 - \pi X = \frac{\kappa\rho(1-\lambda)(1-\pi)\pi X^2}{\lambda(k^d-1)}. \tag{A.4}$$

Note that the LHS of the above equation is increasing in $\lambda(k^d-1)$ while the RHS is decreasing in $\lambda(k^d-1)$. Thus we just need to show

$$\left(\frac{1}{\lambda} + \frac{\lambda}{\rho\kappa}\right)\kappa\rho\pi X + 1 - \pi X < \frac{\kappa\rho(1-\lambda)(1-\pi)\pi X^2}{\kappa\rho\pi X}, \tag{A.5}$$

which can be reduced to $\frac{\kappa\rho\pi X}{\lambda} + \lambda\pi X + 1 - \pi X < (1-\lambda)(1-\pi)X$ or $\frac{X-1}{X} > \frac{\rho\kappa\pi}{\lambda} + \lambda$. ■

A2.2. ***Characterization of Direct Trading economy with Intertype Trading***

In this part, we provide detailed calculations for the characterization of the market equilibrium arising in a direct trading economy with (hypothetically introduced) frictionless inter-type trading between originators. Importantly, we show that the equilibrium outcomes in such a direct trading economy with inter-type trading coincide with the constrained efficient allocation, as stated in Proposition 3.

Market equilibrium in a direct trading economy with inter-type market

Recall that in the market equilibrium with direct trading as characterization in Proposition 2, type H originators are retaining more assets and hence are incuring a higher marginal retention cost in equilibrium. As such, in what follows, we conjecture and verify that H-type originators offload some of their produced assets to L-type ones at certain market price $\hat{p}_{in}$ in the inter-type trading market. As in Lemma 1, it can be similarly shown that after trading in the inter-type market both originators will bring all remaining assets for sale in the asset market, in which the trading price is $\hat{p}_A^d$.

As such, a type-H originator in this economy solves

$$\max_{\{\hat{q}_H^d \geq 0, \hat{s}_H^d \geq 0\}} \hat{p}_{in}\hat{s}_H^d + x_H\lambda\left(\hat{q}_H^d - \hat{s}_H^d\right) + (1-\lambda)\hat{p}_A^d\left(\hat{q}_H^d - \hat{s}_H^d\right) - R\left(\lambda\left(\hat{q}_H^d - \hat{s}_H^d\right)\right) - \hat{k}^d\hat{q}_H^d, \tag{A.6}$$

in which he decides the production quantity $\hat{q}_H^d$ on date 0 and the selling $\hat{s}_H^d$ in the inter-type market on date $\frac{1}{2}$, and then brings remaining assets to the asset market for sale on date 1. Similarly, a L-type originator in this economy, who purchase good assets from the H-type in the inter-type market and resell them in the asset market, solves

$$\max_{\{\hat{q}_L^d \geq 0, \hat{b}_L^d \geq 0\}} x_H\lambda\hat{b}_L^d + x_L\lambda\hat{q}_L^d + (1-\lambda)(\hat{q}_L^d + \hat{b}_L^d)\hat{p}_A^d - R\left(\lambda(\hat{b}_L^d + \hat{q}_L^d)\right) - \hat{p}_{in}\hat{b}_L^d - \hat{k}^d\hat{q}_L^d \tag{A.7}$$

in choosing his optimal production and trading strategy. Likewise, a L-type originator chooses his production $\hat{q}^d_L$ and purchasing $\hat{b}^d_L$ from the inter-type trading market, both of which he then brings to the asset market for sale.

As in Section 3.1, the equilibrium asset market price, p^d_A, are competitively set by risk-neutral investors as the average quality of assets sold in the market:

$$\hat{p}^d_A = \frac{\left[\pi(\hat{q}^d_H - \hat{s}^d_H) + (1-\pi)\hat{b}^d_L\right] x_H + (1-\pi)\hat{q}^d_L x_L}{\left[\pi(\hat{q}^d_H - \hat{s}^d_H) + (1-\pi)\hat{b}^d_L\right] + (1-\pi)\hat{q}^d_L}, \tag{A.8}$$

where $\pi(\hat{q}^d_H - \hat{s}^d_H) + (1-\pi)\hat{b}^d_L$ is the total quantity of type-H assets brought to the asset market and $(1-\pi)\hat{q}^d_L$ is the total quantity of the type-L assets.

In the inter-type trading market, the equilibrium trading prices p^n clears the market by equating the aggregate demand with the aggregate supply:

$$\pi\hat{s}^d_H = (1-\pi)\hat{b}^d_L. \tag{A.9}$$

Finally, the equilibrium input capital price is determined as in Equation (9).

Formally, the market equilibrium in a direct trading economy with frictionless inter-type market can be defined as follows.

Definition 3. ***Direct trading equilibrium with inter-type market***
In a direct trading economy with frictionless ex-post inter-type trading, a market equilibrium is consisted of equilibrium prices $\{\hat{p}^d_A, \hat{p}_{in}, \hat{k}^d\}$, *the asset origination and trading strategy* $\{\hat{q}^d_\theta, \hat{s}^d_H, \hat{b}^d_L\}_{\theta\in\{H,L\}}$ *by originators satisfying*

1. ***Agents optimization:*** *Given equilibrium prices* $\{\hat{p}^d_A, \hat{p}_{in}, \hat{k}^d\}$, $\{\hat{q}^d_H, \hat{s}^d_H\}$ *solves Eq.* (A.6) *and* $\{\hat{q}^d_L, \hat{b}^d_L\}$ *solves Eq.* (A.7)*;*
2. ***Bayesian consistency:*** *Given equilibrium strategies* $\{\hat{q}^d_\theta, \hat{s}^d_H, \hat{b}^d_L\}_{\theta\in\{H,L\}}$, *asset market trading price* p^d_A *satisfies Bayesian updating rule as in Eq.* (A.8)*;*
3. ***Market clearing:*** $\hat{p}_{in}$ *clears the inter-type market as in Eq.* (A.9) *and* $\hat{k}^d$ *is the input capital price as in Eq.* (9)*.*

Proof of Proposition 3

We now show that in an economy with $\rho_I = 0$, introducing a frictionless inter-type market into a direct trading economy on date $\frac{1}{2}$ can fully restore efficiency by achieving the constrained efficient allocation as characterized in Proposition 1.

First we show that no "lemon" will be produced in the market equilibrium in a direct trading economy with frictionless inter-type trading. Suppose in equilibrium an L-type originator is producing positive amount of assets, i.e., $\hat{q}^d_L > 0$. Consider the following deviation strategy for this L-type originator: decreasing his own production by an arbitrarily small amount $\delta > 0$ and increasing his purchasing from the inter-type market by δ, which he then brings to the asset market for sale. Following this deviation strategy, the payoff change to this L-type originator is

$$\Delta(\delta) = \underbrace{\hat{k}^d \delta - \hat{p}_{in}\delta}_{\text{saved cost}} + \underbrace{\lambda(x_H - x_L)\delta}_{\text{quality change of retained assets}}, \tag{A.10}$$

since the post-trading retention is held fixed following this deviation strategy. In equilibrium the payoff change $\Delta(\delta)$ cannot be strictly positive, which thus implies

$$\hat{p}_{in} - \hat{k}^d \geq \lambda(x_H - x_L) > 0. \tag{A.11}$$

However, with equilibrium prices satisfying $\hat{p}_{in} > \hat{k}^d$ each H-type originator would want to keep producing and selling in the inter-type market. Therefore, in equilibrium no "lemons" are being produced in the economy, i.e., $\hat{q}^d_L = 0$.

In equilibrium, an L-type originator must be indifferent to purchase a marginal unit of asset from the inter-type market and then brings to the asset market for sale. This implies

$$\lambda x_H + (1-\lambda)\hat{p}^d_A - \lambda R'\left(\lambda \hat{b}^d_L\right) = 0. \tag{A.12}$$

Similarly, an H-type originator must be indifferent to produce a marginal unit of assets, which he then brings to either the inter-type market or the asset market. This implies

$$\hat{p}_{in} - \hat{k}^d = 0, \text{ and } \lambda x_H + (1-\lambda)\hat{p}^d_A - \lambda R'\left(\lambda(\hat{q}^d_H - \hat{s}^d_H)\right) = 0 \tag{A.13}$$

Combine Eq. (A.12) and Eq. (A.13), we get

$$\hat{q}^d_H - \hat{s}^d_H = \hat{b}^d_L. \tag{A.14}$$

That is, in equilibrium the post-trading retention are equalized across both types of originators. As such, with no "lemon" production and equalized post-trading asset retention—the two conditions that guarantee the constrained efficient allocation as identified in Proposition 1, it thus follows that (constrained) efficiency is fully restored in a direct trading economy with frictionless inter-type trading. ■

A3. ***Proofs and Calculations for Intermediated Equilibrium***

This section provides algebraic proofs and detailed calculations for our analysis of the intermediated equilibrium (Section 3.3).

A3.1. ***Two properties of the intermediated equilibrium (Lemma 2 and Lemma 3)***

Proof of no voluntary retention (Lemma 2)

In this proof, we show that in the market equilibrium that arises in an intermediated economy, it must be true that all originators are bringing all their produced assets (or remaining ones after trading in the signal markets on date $\frac{1}{2}$) to the asset market for sale on date 1, and when intermediaries' retention capacity ρ_I is sufficiently small, all intermediaries also bring all purchased assets to the asset market for sale.

Suppose in equilibrium after trading in the signal market, H-type originators are keeping a positive quantity of produced assets from bringing to the asset market for sale on date 1. This implies that the equilibrium post-trading retention for a H-type originator $r_H > \lambda(q_H - s_H)$. Without loss of generality, assume type H trade positive volume in h signal market and $p_h \geq p_l$.

Case 1: $p_h \leq p_A$

In the first case, the equilibrium trading price in the h signal market is no greater than that in the asset market. Now consider the following deviation strategy by a H-type originator:

1. increase selling to the asset market by an arbitrarily small amount $\delta > 0$ (feasible whenever the originator is keeping a positive quantity from bringing to the asset market for sale);
2. increase the asset production by $(1-\lambda)\delta$ and keep these newly produced assets.

Following the deviation strategy as described above, the change in the post-trading retention of the H-type originator is:

$$\Delta r_H = -(1-\lambda)\delta + (1-\lambda)\delta = 0, \tag{A.15}$$

and hence the payoff change following the strategy is:

$$\Delta w_H = p_A(1-\lambda)\delta - x_H(1-\lambda)\delta + (1-\lambda)\delta(x_H - k) = (1-\lambda)\delta(p_A - k).$$

By definition, in equilibrium the above payoff change cannot be positive, this implies $p_A \leq k$. But since $p_h \leq p_A$, it thus follows that $p_h \leq k$.

However, when both p_h and p_A are no greater than k, L-type originators will not be producing in the economy. This means that there will be no lemon in the asset market and hence $p_A = x_H$. But when $k \geq p_A = x_H$, H-type originators will also not be producing.

Case 2: $p_h > p_A$

In the second case, trading price in the h signal market is strictly higher than that in the asset market. Now consider the following deviation strategy for a H-type originator:

1. increase selling to the asset market by an arbitrarily small amount $(1-\alpha)\delta > 0$ (feasible whenever the originator is keeping a positive quantity from bringing to the asset market for sale);
2. increase production by $(1-\lambda)\delta > 0$, among which $\alpha(1-\lambda)\delta$ will receive an h signal and $(1-\alpha)(1-\lambda)\delta$ will receive an l signal;
3. Sell these $\alpha(1-\lambda)\delta$ h-signal assets in the h signal market, while keep $(1-\alpha)(1-\lambda)\delta$ by himself.

Following the deviation strategy as described above, the change in the post-trading retention of the H-type originator is:

$$\Delta r_H = -(1-\alpha)\delta + \lambda(1-\alpha)\delta + (1-\alpha)(1-\lambda)\delta = 0, \tag{A.16}$$

and hence the payoff change following the deviation strategy is:

$$\begin{aligned}\Delta w_H &= (1-\lambda)(1-\alpha)\delta(p_A - x_H) + \alpha(1-\lambda)\delta p_h + (1-\alpha)(1-\lambda)\delta x_H - (1-\lambda)\delta k \\ &= (1-\lambda)\delta\left[\alpha p_h + (1-\alpha)p_A - k\right]. \end{aligned} \tag{A.17}$$

By definition, in equilibrium the above payoff change cannot be positive, this implies:

$$\alpha p_h + (1-\alpha)p_A \leq k. \tag{A.18}$$

Since in this case $p_h > p_A$, given $\alpha > \frac{1}{2}$, we have

$$(1-\alpha)p_h + \alpha p_A < k. \tag{A.19}$$

This again implies that L-type originators will not produce, therefore in equilibrium there is no lemons in either signal market or the asset market. It then follows that $p_A = x_H$. However, this means that $k > x_H$ and hence H-type originator also will not be producing.

Therefore, in equilibrium originators must be holding no voluntary retention such that they bring all remaining assets (after signal market trading) to the asset market for sale.

Now we consider intermediaries. First, note that when $p_s < p_A$, it follows that intermediaries in equilibrium must be taking all purchased asset for sale in the asset market. Otherwise, an intermediaries can follow a similar deviation strategy as constructed above for type H originators:

1. increase selling to the asset market by an arbitrarily small amount $\delta > 0$ (feasible whenever the intermediary is keep a positive quantity from bringing to the asset market for sale);
2. increase the asset purchase by $(1-\lambda)\delta$ and keep these newly produced assets.

Following this deviation strategy, the change in the post-trading retention for the intermediary is zero and the payoff change is

$$\begin{aligned}\Delta w_I &= p_A(1-\lambda)\delta - x_h(1-\lambda)\delta + (1-\lambda)\delta(x_h - p_h) \\ &= (1-\lambda)\delta(p_A - p_h). \end{aligned} \tag{A.20}$$

Since $p_h < p_A$, this deviation gives a strictly positive payoff change.

Now consider the case where $p_h \geq p_A$. In this case, both types of originators must be selling all their h signal assets in the signal market, i.e., $s_H = \alpha q_H$ and $s_L = (1-\alpha)q_L$. Thus the equilibrium asset quality in the signal market is

$$x_h = \frac{\alpha\pi q_H X}{\alpha\pi q_H + (1-\alpha)(1-\pi)q_L}.$$

If in equilibrium an intermediary who has purchased b_I units of assets in the signal markets is not taking all her purchases assets for sale in the asset market, then we must have

$$\underbrace{x_h - R'(\lambda b_I)}_{\text{voluntary retention}} > \underbrace{(1-\lambda)p_A + \lambda\left[x_h - R'(\lambda b_I)\right]}_{\text{bringing to asset market}}, \tag{A.21}$$

in which the LHS is the marginal benefit of retaining a marginal unit of purchased asset when the intermediary has no voluntary retention (i.e., $r_I = \lambda b_I$), and the RHS is the marginal benefit of bringing a marginal unit to the asset market for sale. Condition (A.21) is equivalent to

$$x_h - p_A > \frac{\lambda b_I}{\rho_I} \Leftrightarrow \rho_I > \frac{\lambda b_I}{x_h - p_A} \tag{A.22}$$

It can be shown that in equilibrium we always have $p_A > \pi X$, thus from (A.22), we have

$$\rho_I > \frac{\lambda b_I}{x_h - \pi X} \geq \frac{\lambda\left(\alpha q_H + (1-\alpha)q_L\right)}{\frac{\alpha\pi q_H X}{\alpha\pi q_H + (1-\alpha)(1-\pi)q_L} - \pi X} \geq \frac{\lambda\alpha q_H}{(1-\pi)X},$$

where the second inequality uses the fact that $b_I \geq \alpha q_H + (1-\alpha)q_L$ and the last inequality uses the fact that $q_L \geq 0$.

Finally, since $q_H > Q$ (because $q_H > q_L$) and $Q > Q^d$ for any $\rho_I > 0$ where Q^d is the equilibrium total production in a direct economy as described by (22), it thus follows that

$$\rho_I > \frac{\lambda\alpha Q^d}{(1-\pi)X} \geq \frac{\lambda Q^d}{2(1-\pi)X}, \tag{A.23}$$

where the second inequality uses $\alpha \geq \frac{1}{2}$. Therefore, a sufficient condition that guarantees that intermediaries bring all purchased assets for sale is

$$\rho_I \leq \frac{\lambda Q^d}{2(1-\pi)X}. \tag{A.24}$$

This completes the proof for Lemma 2. ■

Proof of no "lemon" market (Lemma 3)

In this proof, we show that in the market equilibrium that arises in an intermediated system with technology parameter $\alpha \in (\frac{1}{2}, 1]$, the "lemon" market endogenously closes (i.e., the intermediated equilibrium is described by either of the two items in Lemma 3). We start our proof by first establishing the following useful lemma.

Lemma A1. *If in equilibrium there exists a type $\theta \in \{H, L\}$ originator who sells positive amount in signal j market but does not sell anything in the signal j' market, then the equilibrium trading prices must satisfy $p_j \geq p_{j'}$.*

Proof of Lemma A1

Suppose in equilibrium a θ-type originator is selling $s_{\theta j} > 0$ in signal market j but selling $s_{\theta j'} = 0$ in signal market j'. Then consider the following deviation strategy that is feasible for this θ-type originator: decreasing his selling in the signal j market by an arbitrarily small $\delta > 0$ while increasing his selling in the signal j' market by δ. Since by Lemma 2 this originator must be bringing all remaining assets for sale in the asset market, it thus follows that after such a deviation, the change in this originators' post-trading retention is zero.

Therefore, the payoff change following this deviation strategy for the originator is

$$\Delta(\delta) = -(p_j - x_\theta)\delta + (p_{j'} - x_\theta)\delta = (p_{j'} - p_j)\delta,$$

where we used the fact that both signal j and signal j' assets produced by this type θ originator have the same quality x_θ. In equilibrium, this deviation cannot give the originator a strictly positive payoff change, which thus allows us to conclude that $p_j \geq p_{j'}$. ■

In what follows, we structure the remainder of the proof into three steps.

Step 1

In the first step, we establish the following important result: if in equilibrium both signal markets have strictly positive trading volume, then it follows that both H-type and L-type originators must be simultaneously trading positive amount in both h and l signal markets.

We prove this result following the way of contradiction method, in which we rule out the following three possible cases that either one of H-type or L-type is not simultaneously trading in both signal markets. Before we proceed, it is worthy noting that when both signal markets have positive trading volume (with trading prices being p_h and p_l), in equilibrium we must have:

$$\begin{aligned} k &\geq \alpha p_h + (1-\alpha) p_l, \quad \text{for type } H \\ k &\geq (1-\alpha) p_h + \alpha p_l, \quad \text{for type } L \end{aligned} \tag{A.25}$$

since both types of originators can always produce and then sell. Furthermore, if both types of originators are actively selling in both h and l signal markets, it then follows that

$$c = \alpha p_h + (1-\alpha) p_l = (1-\alpha) p_h + \alpha p_l \tag{A.26}$$

since it is always a feasible for both originators to cut their production by $\delta > 0$ and reduce their selling in the signal markets by $(\alpha\delta, (1-\alpha)\delta)$ for type H and by $((1-\alpha)\delta, \alpha\delta)$ for type L.

Now we consider the following three cases that we seek to rule out.

Case 1: *Both types of originators are active in h signal market, but only one type is active in l signal market.*

In this case, exactly one of the inequalities in Eq. (A.25) holds with equality. First suppose in equilibrium we have $p_h > p_l$. Then we must have:

$$\begin{aligned} k &= \alpha p_h + (1-\alpha)p_l, \\ k &> (1-\alpha)p_h + \alpha p_l, \end{aligned} \tag{A.27}$$

which implies that L-type originator is not selling in l signal market. This further implies that the average quality in the l signal market is $x_l = x_H$ since there are no lemons sold in it. But since both types are actively selling in the h signal market, it follows that $x_h < x_H = x_l$. From intermediaries optimization problem it is easy to show when both signal markets have positive trading volume, in equilibrium the signal market prices must satisfy:

$$p_h - p_l = \lambda(x_h - x_l), \tag{A.28}$$

otherwise there exist deviation strategies with which intermediaries can make strictly positive profits. But $x_l > x_h$ then implies $p_l > p_h$. Contradiction.

Next suppose that $p_l > p_h$. In this case we must have

$$\begin{aligned} k &> \alpha p_h + (1-\alpha)p_l, \\ k &= (1-\alpha)p_h + \alpha p_l, \end{aligned} \tag{A.29}$$

which implies that H-type are not active in the l market. As a consequence, all assets being sold in the l signal market are lemons and hence $x_l = x_L$. With both types are actively selling in the h signal market, it then follows $x_h > x_L = x_l$, which then imlies $p_h > p_l$ by intermediaries' optimization condition (as in A.28). Again, contradiction arises.

This allows us to rule out possibility of case 1.

Case 2: *Each type of originators sell in exactly one signal market.*

In the first scenario of case 2, suppose H-types are only selling in the h signal market and L-types are only selling in the l signal market. Then from intermediaries' optimization it follows $p_h > p_l$—because h signal market has no lemons and l signal market only has lemons. But since in equilibrium L-type is only selling in l signal market while not selling in h signal market, from Lemma A1, it implies that $p_l \geq p_h$. Contradiction.

In the second scenario of case 2, suppose H-types are only selling in the l signal market and L-types are only selling in the h signal market. Again, intermediaries' optimization implies $p_h < p_l$. However, since in equilibrium L-type originator is only selling in h signal market, it thus follows that $p_h \geq p_l$ based on Lemma A1. Again contradiction is derived.

This allows to conclude that case 2 is impossible.

Case 3: *Both types are active in l signal market and only one type active in h signal market.*

As in our analysis of case 1, first suppose that in equilibrium we have: $p_h > p_l$. Then we must have:

$$\begin{aligned} k &= \alpha p_h + (1-\alpha)p_l, \\ k &> (1-\alpha)p_h + \alpha p_l, \end{aligned} \tag{A.30}$$

which implies that L-type is only active in l signal market while H-type is active in both signal markets. From Lemma A1, the fact that L-type only sell in the l signal market implies that $p_l \geq p_h$. This gives us a contradiction.

Next suppose that in equilibrium we have: $p_l > p_h$. Then we must have:

$$\begin{aligned} k &> \alpha p_h + (1-\alpha) p_l, \\ k &= (1-\alpha) p_h + \alpha p_l, \end{aligned} \tag{A.31}$$

which implies that H-type is only selling in the l signal market while L-type is selling in both market. Unlike in previous cases, here H-type only selling in l signal market implies that $p_l \geq p_h$ from Lemma A1, which does not contradict with the superior asset quality sold in l signal market given only L-type is selling in the h signal market. We therefore pursue contradiction following a different route.

First it can be shown that H-type's selling in the l signal market must be binding, i.e., $s_{Hl} = (1-\alpha) q_H$. Otherwise, a type H originator can make a profitable deviation by following the strategy below:

1. increasing selling in the l signal market by an infinitesimal amount $\delta > 0$ (feasible when H-type is not binding in l signal market selling);
2. increasing production by δ and take all these newly produced assets for sale directly in the asset market.

Step 1 reduces the post-trading retention by $\lambda\delta$ while step 2 adds $\lambda\delta$. Therefore, the deviation strategy described above leaves the post-trading retention unaffected for the type H originator while gives a payoff change

$$\Delta w_H(\delta) = \underbrace{\delta p_l - [(1-\lambda)\delta p_A + \lambda \delta x_H]}_{\text{step 1}} + \underbrace{(1-\lambda)\delta p_A + \lambda \delta x_H - \delta k}_{\text{step 2}} = \delta(p_l - k).$$

But since $k = (1-\alpha) p_h + \alpha p_l < p_l$ (given $p_l > p_h$), it thus follows $\Delta w_H(\delta) > 0$, which cannot be true in equilibrium. Thus H-type must be binding in l signal market. Similarly, L-types' selling must also be binding in l signal market, i.e., $s_{Ll} = \alpha q_L$.

As such, with both types of originators being binding in selling their l signal assets, it implies that in equilibrium the total asset purchase by intermediaries from the signal market must satisfy:

$$b_I > (1-\alpha) q_H + \alpha q_L. \tag{A.32}$$

Since intermediaries are purchasing a positive amount in the h signal market in which originators do not bind in selling, it follows that

$$\lambda x_h + (1-\lambda) p_A - \lambda R_I'(\lambda b_I) = p_h. \tag{A.33}$$

For type-L originators, since they are selling a positive amount in the h signal market, it follows from their optimization that

$$\lambda x_L + (1-\lambda) p_A - \lambda R'(r_L) = p_h \tag{A.34}$$

But since only L type sell in the h signal market, it thus follows that $x_h = x_L$. Therefore, Eq. (A.33) and (A.34) thus imply

$$\frac{\lambda b_I}{\rho_I} = \frac{r_L}{\rho}. \tag{A.35}$$

However, given L-type originators sell all their l signal assets and a positive amount of their h assets, it thus follows $r_L < (1-\alpha)q_L$. As such, using (A.32) we have

$$\frac{\rho_I}{\rho} = \frac{\lambda b_I}{r_L} > \frac{\lambda(1-\alpha)q_H + \alpha q_L}{(1-\alpha)q_L} > \frac{\lambda\alpha}{1-\alpha},$$

which cannot hold when intermediaries' retention capacity ρ_I is sufficiently small (e.g., whenever $\rho_I \leq \frac{\lambda\alpha\rho}{1-\alpha}$, which can be guaranteed by $\rho_I \leq \lambda\rho$ given $\alpha \geq \frac{1}{2}$). The above analysis thus implies that under our specification that intermediaries' retention capacity ρ_I being sufficiently small, case 3 can also be ruled out in an intermediated equilibrium.

To sum up, our analysis above establishes the following fact: whenever both signal markets have positive trading volume, it follows that both types of originators must be simultaneously selling positive amount in both signal markets.

Step 2

Now we prove that an equilibrium in which both signal markets have positive trading volume must satisfy: $p_h = p_l = k$ and $x_h = x_l$.

To show this, suppose such an equilibrium exists and the prices in the two actively trading signal markets are p_h and p_l respectively. Then from our analysis above (Step 1 of the proof), it must follow that these equilibrium prices satisfy:

$$k = \alpha p_h + (1-\alpha)p_l = (1-\alpha)p_h + \alpha p_l \tag{A.36}$$

which in turn implies that $p_h = p_l = k$ whenever $\alpha > \frac{1}{2}$. However, with $p_h = p_l$, in equilibrium both types of originators are indifferent between selling in two signal markets and their aggregate selling mixture in the two signal markets must satisfy $x_h = x_l$, so that intermediaries can also be indifferent between purchasing from either signal market.

Step 2 thus implies that if in equilibrium both signal markets have positive trading volume, then the two signal markets must have the same trading price and average asset quality. This maps to the first type of equilibrium described in Lemma 3.

Step 3

Our proof thus far has shown that if in equilibrium the (shadow) prices in the two signal markets are different, i.e., $p_h \neq p_l$, then there can be only one signal market that has positive trading volume. In the last step of our proof, we show that if $p_h \neq p_l$ in equilibrium, then $p_h > p_l$ and the l signal market must have zero trading volume.

Again, following a way of contradiction procedure, assume that in equilibrium only l signal market has a positive trading volume while the h signal market has zero trading.

In proving the above result, we first show that the following must be true. That is, the equilibrium satisfying the following two conditions is impossible to exist:

1. In the equilibrium, only l signal market has positive trading volume (at price p_l), while the h signal market has zero trading (at some "shadow price" p_h that leads to zero trading);
2. Each of the L-type originators sell more than $(1-\alpha)$ fraction of their produced assets in the l signal market—-which is possible, since the total quantity of their l signal assets is $\alpha q_L > (1-\alpha)q_L$.

To prove that the equilibrium satisfying the above two conditions does not exist, we need to show there does not exist a p_h such that at this shadow price p_h in the h signal market, originators want to sell zero quantity and intermediaries want to buy zero quantity of assets from the h signal market.

The condition that guarantees that the selling from originators in the h signal market to be zero is $p_h \leq p_l$. The condition that guarantees the purchasing from intermediaries in the h signal market to be zero is $p_h \geq p_l + \lambda(x_h - x_l)$, where x_l is the equilibrium average quality of assets sold in the l signal market and x_h is the intermediaries' belief of average quality in the h signal market, which is

$$x_h = \frac{x_H \pi\alpha q_H + x_L(1-\pi)(1-\alpha)q_L}{\pi\alpha q_H + (1-\pi)(1-\alpha)q_L} = \frac{\pi\alpha q_H X}{\pi\alpha q_H + (1-\pi)(1-\alpha)q_L}, \tag{A.37}$$

under our specification of intermediaries' off-equilibrium belief regarding asset quality in signal market with zero trading volume. Recall this off-equilibrium belief could be understood as an arbitrarily small mass ϵ of originators randomly coming into the h signal markets sell all their eligible assets: each H-type has αq_H of h signal assets and each L-type has $(1-\alpha)q_L$ of h signal assets.

But since in the l signal market each H-type originators sells more than $(1-\alpha)q_L(i.e., s_{Ll} > (1-\alpha)q_L)$, and each H-type originators' selling in the l signal market is bounded by $x_{Hl} \in [0, (1-\alpha)q_H]$ (which is the total amount of l signal assets in their hands), we have

$$x_l = \frac{\pi s_{Hl} X}{\pi s_{Hl} + (1-\pi)s_{Ll}} \leq \frac{\pi(1-\alpha)q_H X}{\pi(1-\alpha)q_H + (1-\pi)(1-\alpha)q_L} < \frac{\pi\alpha q_H X}{\pi\alpha q_H + (1-\pi)(1-\alpha)q_L} = x_h,$$

where the second inequality is due to $\alpha > \frac{1}{2}$. But for the equilibrium with equilibrium prices $\{p_h, p_l\}$ and associated trading strategies to be indeed an equilibrium, we need the following set of the "shadow price" p_h to be an non-empty set:

$$p_l + \lambda(x_h - x_l) \leq p_h \leq p_l. \tag{A.38}$$

However, our above calculation suggests $p_l + \lambda(x_h - x_l) > p_l$, which means that the above set is bound to be an empty set. This proves the non-existence of such equilibrium.

The above result allows us to conclude that when in equilibrium only the l signal market has positive trading volume, it must be true that in equilibrium the L-type originators' selling of their l signal assets satisfy $s_{Ll} \leq (1-\alpha)q_L$. However, in this case, this equilibrium can be equivalently viewed as an equilibrium in which only the h signal market has positive trading volume. This is because when L-type originators' selling s_{Ll} of their "l

signal" assets does not exceed $(1-\alpha)q_L$, which is the total quantity of assets produced by an L-type originator that receive a h signal, one can equivalently view these assets as the h signal assets produced by L-type originators. With this "label swapping," the l signal assets sold by H-type originators in equilibrium, which is bounded by $(1-\alpha)q_H$ and hence smaller than αq_H (since $\alpha > \frac{1}{2}$), can be equivalently viewed as their selling of produced assets that receive a h signal.

In this way, we end up with an equilibrium where only h signal market has positive trading volume with the equilibrium (shadow) trading prices in the two signal markets satisfying $p_h > p_l$. This maps to the second type of equilibrium described in Lemma 3. ■

A3.2. ***Intermediated equilibrium in technology-irrelevant range***

Proof of Proposition 4

In this proof, we provide detailed calculations for the characterization of the intermediated equilibrium in the tech-irrelevant range.

We begin by establishing an important property of the sequential equilibrium arising in an intermediated economy with neither types of originators bind in their h signal market selling. To characterize the sequential equilibrium with perfectly liquid trading in the h signal market, we characterize a sequence of assessments with positive market illiquidity $\epsilon > 0$ (modeled as the probability of asset sale being failed) in the h signal market, where $\epsilon \to 0$.

Now with a $\epsilon > 0$ probability of asset sale failure in the h signal market, the post-trading retention born by θ-type originator is

$$r_\theta = \begin{cases} r_{\theta Y} \equiv \lambda(q_\theta - s_\theta) & \text{if sale succeeds in signal market} \\ r_{\theta N} \equiv \lambda(q_\theta - s_\theta) + s_\theta & \text{if sale fails in signal market} \end{cases}, \tag{A.39}$$

if this θ-type originator chooses to bring s_θ for sale in the h signal market and $q_\theta - s_\theta$ for sale in the asset market.

If in equilibrium this θ-type originator is selling interior in the h signal market, then it follows that this originator must be indifferent between bringing the last unit of produced assets to either the h signal market or the asset market. In equilibrium the marginal cost of production for θ-type originator should equal to the marginal payoff of bringing the last unit of produced assets to the h signal market. This requires:

$$\underbrace{(1-\epsilon)p_h}_{\text{signal market sale succeeds}} + \underbrace{\epsilon\left[x_\theta - R'(r_{\theta N})\right]}_{\text{signal market sale fails}} = k, \tag{A.40}$$

where $r_{\theta N} \equiv \lambda(q_\theta - s_\theta) + s_\theta$ is the equilibrium retention when signal market sale fails.

With interior selling, the marginal cost of production for θ-type originator should also equal to the marginal payoff of bringing the last unit of produced assets to the asset market. This requires:

$$\underbrace{(1-\epsilon)\left[(1-\lambda)p_A+\lambda x_\theta-\lambda R'(r_{\theta Y})\right]}_{\text{signal market sale succeeds}}+\underbrace{\epsilon\left[(1-\lambda)p_A+\lambda x_\theta-\lambda R'(r_{\theta N})\right]}_{\text{signal market sale fails}}=k, \quad \text{(A.41)}$$

where $r_{\theta Y} \equiv \lambda(q_\theta - s_\theta)$ is the equilibrium retention when signal market sale succeeds and $r_{\theta N}$ is defined as above.

From Eq. (A.40) and (A.41) we can get

$$x_\theta - R'(r_{\theta N}) = \frac{1}{\epsilon}\left[k-(1-\epsilon)p_h\right], \text{ and } \; x_\theta - R'(r_{\theta Y}) = p_h - \frac{1-\lambda}{\lambda(1-\epsilon)}\left(p_A-k\right),$$

from which we get the equilibrium selling $s_\theta(\epsilon)$ in the signal h market:

$$\begin{aligned}\rho s_\theta(\epsilon) &= p_h - \frac{1-\lambda}{\lambda(1-\epsilon)}\left(p_A-k\right) - \frac{1}{\epsilon}\left[k-(1-\epsilon)p_h\right] \\ &= \frac{1}{\epsilon}\left(p_h-k\right) - \frac{1-\lambda}{\lambda(1-\epsilon)}\left(p_A-k\right), \quad \text{(A.42)}\end{aligned}$$

for both $\theta \in \{H, L\}$.

Critically, from Eq. (A.42) it follows that for any $\epsilon > 0$ both types of originator bring the exact same amount of assets to sell in the h signal market, if neither type bind in selling their h-signal assets. As such, if technology α is sufficiently low such that in equilibrium L-type originator is not bring all his h-signal assets to the signal market, it must follow that H-type originators are also not binding in selling their h-signal assets since $\alpha \geq \frac{1}{2}$ and in equilibrium we must have $q_H \geq q_L$.

Furthermore, for any market illiquidity $\epsilon > 0$ in the h signal market, we have equilibrium average quality $x_h(\epsilon)$ in the h signal market is:

$$x_h(\epsilon) = \frac{x_H \pi s_H(\epsilon) + x_L(1-\pi)s_L(\epsilon)}{\pi s_H(\epsilon) + (1-\pi)s_L(\epsilon)} = \pi X$$

It thus follows that when the market illiquidity ϵ in the h market is taken to zero, the average asset quality in the signal market must converge to $\lim_{\epsilon\to 0} x_h(\epsilon) = \pi X$.

Therefore, in the sequential equilibrium with perfectly liquid signal market trading, if trading volume is positive in the h signal market and neither type bind in selling their h-signal assets, then in equilibrium the average quality of assets traded in the h signal market must be $x_h = \pi X$. This then allows us to uniquely pin down the sequential equilibrium with both types of originators being slack in selling their h-signal assets, which is characterized by a single quadratic equation system of $\{q_H, q_L, s_h, b_I, x_h, p_A, k\}$ as follows:

$$
\begin{aligned}
\lambda X + (1-\lambda) p_A - \frac{\lambda^2}{\rho}(q_H - s_h) &= k, && \text{(A.43)} \\
(1-\lambda) p_A - \frac{\lambda^2}{\rho}(q_L - s_h) &= k, && \text{(A.44)} \\
\lambda x_h + (1-\lambda) p_A - \frac{\lambda^2}{\rho_I} b_I &= k, && \text{(A.45)} \\
s_h &= b_I, && \text{(A.46)} \\
x_h &= \pi X, && \text{(A.47)} \\
1 + \kappa\left[\pi q_H + (1-\pi) q_L\right] &= k, && \text{(A.48)} \\
\frac{\pi q_H X}{\pi q_H + (1-\pi) q_L} &= p_A, && \text{(A.49)}
\end{aligned}
$$

in which Eq. (A.43, A.44 and A.45) are from optimization by originators and intermediaries, Eq. (A.46) is from h signal market clearing, Eq. (A.47, A.48 and A.49) are equilibrium prices.

Combining equations, we can get the equilibrium sell $s_h = \frac{\rho_I(k-1)}{(\rho+\rho_I)\kappa}$ by both types of originators in the h signal market and the asset market price $p_A = \frac{1}{1-\lambda}\left[k + \frac{\lambda^2(k-1)}{(\rho_I+\rho)\kappa} - \lambda\pi X\right]$. Plug in Eq. (A.49), we get

$$
\frac{(1-\pi)\rho\kappa X}{\lambda(k-1)} = \frac{1}{\pi X(1-\lambda)}\left[k + \frac{\lambda^2(k-1)}{(\rho_I+\rho)\kappa} - \pi X\right]. \tag{A.50}
$$

It can be easily shown that Eq. (A.50) always exists a unique solution on $k \in (1, \infty)$, given that the LHS of Eq. (A.50) monotonically decrease in k while the RHS monotonically increases in k for $k \in (1, \infty)$. Solving this quadratic equation in k, we get the solution as in Proposition 4.

It is worthy noting that sequential equilibrium characterized above is independent of technology level parameter α. In this regard, the intermediated equilibrium in which both types of originators are not binding in selling their h-signal assets is tech-irrelevant. To pin down the cutoff level $\hat{\alpha}_1(\rho_I)$ in intermediation technology α, note that at $\alpha = \hat{\alpha}_1(\rho_I)$ the type L originators just start to be binding in selling their l signal assets. As such, the tech-relevance cutoff $\hat{\alpha}_1(\rho_I)$ is determined by condition:

$$
s_h \geq (1-\hat{\alpha}_1) q_L \;\Leftrightarrow\; (1-\lambda) p_A - k \leq \frac{\alpha\lambda^2 s_h}{\rho(1-\alpha)}, \tag{A.51}
$$

which implies

$$
\hat{\alpha}_1(\rho_I) = \max\left\{\frac{1}{2}, \left[\frac{\lambda^2 s_\theta}{\rho\left[(1-\lambda)p_A - k\right]} + 1\right]^{-1}\right\}, \tag{A.52}
$$

or Eq. (37). This completes the proof. ■

A3.3. ***Intermediated equilibrium in technology-relevant range***

Proof of Proposition 5

In this proof, we provide detailed calculations for the characterization of the intermediated equilibrium in the tech-irrelevant range.

As we discussed in Section 3.3.2, with intermediaries' retention capacity ρ_I being sufficiently small, it can be guaranteed that H-type originators never bind in selling their h signal assets in an intermediated equilibrium. To see this, suppose both types of originators are binding in selling their h signal assets. Since type H is taking his last unit of h signal asset to the h signal market rather than the asset market, it follows that

$$\lambda X + (1-\lambda)p_A - \frac{\lambda^2}{\rho}(1-\alpha)q_H \le p_h. \tag{A.53}$$

In equilibrium, intermediaries' purchase b_I in h signal market satisfies:

$$\lambda x_h + (1-\lambda)p_A - \frac{\lambda^2}{\rho_I}b_I = p_h. \tag{A.54}$$

Since $x_h < X$ in equilibrium, it thus follows that

$$\frac{\lambda^2}{\rho}(1-\alpha)q_H \ge \lambda X + (1-\lambda)p_A - p_h \ge \lambda x_h + (1-\lambda)p_A - p_h = \frac{\lambda^2}{\rho_I}b_I.$$

Therefore we have

$$\frac{\rho_I}{\rho} \ge \frac{b_I}{(1-\alpha)q_H} = \frac{\pi\alpha q_H + (1-\pi)s_{Lh}}{(1-\alpha)q_H} > \frac{\pi\alpha}{1-\alpha},$$

which can be guaranteed to fail to hold under our parameterization assumption (10).

Now we know that in the tech-relevant range of the intermediated equilibrium, L-type originators bind in their h signal market selling while H-type do not. The equilibrium production and trading by an H-type originator thus imply

$$\lambda X + (1-\lambda)p_A - \frac{\lambda^2}{\rho}(q_H - s_{Hh}) = p_h = k. \tag{A.55}$$

In equilibrium, L-type originators sell all their h signal assets, i.e., $s_{Lh} = (1-\alpha)q_L$, and equilibrium production by L-type implies

$$(1-\alpha)p_h + \alpha\left[(1-\lambda)p_A - \frac{\lambda^2}{\rho}\alpha q_L\right] = k. \tag{A.56}$$

Further, with L-type binding in their h signal market selling, the equilibrium asset quality in the signal market is determined as

$$x_h = \frac{X\pi s_{Hh}}{\pi s_{Hh} + (1-\pi)(1-\alpha)q_L} \tag{A.57}$$

Again, trading prices in the asset market and equilibrium capital price are given by

$$p_A = \frac{X\pi q_H}{\pi q_H + (1-\pi)q_L} \text{ and } k = 1 + \kappa\left[\pi q_H + (1-\pi)q_L\right]. \tag{A.58}$$

Finally, h signal market clearing requires

$$\pi s_{Hh} + (1-\pi)(1-\alpha)q_L = b_I \tag{A.59}$$

and intermediary optimization implies

$$\lambda x_h + (1-\lambda)p_A - \frac{\lambda^2}{\rho_I}b_I = p_h \tag{A.60}$$

After some combination of equations, we get an equation system of $\{x_h, b_I, k\}$:

$$b_I = \frac{\rho_I(k-1)}{(\rho+\rho_I)\kappa} + \frac{(x_h - \pi X)\rho\rho_I}{\lambda(\rho+\rho_I)} \tag{A.61}$$

$$\frac{x_h b_I}{X} - \frac{k-1}{\kappa} = -\frac{\rho}{\lambda^2}\left[\pi\lambda X + \left(\frac{\lambda^2}{\rho_I}b_I - \lambda x_h\right)\left(\pi + \frac{1-\pi}{\alpha}\right)\right] \tag{A.62}$$

$$\left(\frac{\lambda^2}{\rho_I} + k - \lambda x_h\right)\frac{k-1}{(1-\lambda)\kappa} = \frac{X(k-1)}{\kappa} - \frac{(1-\pi)\rho X}{\alpha\lambda}\left(\frac{\lambda}{\rho_I}b_I - x_h\right) \tag{A.63}$$

Characterization of intermediated equilibrium with $\alpha = 1$ (Corollary 1)

In this section, we provide detailed calculations for the special case with perfectly informed intermediation, i.e., technology parameter $\alpha = 1$.

With $\alpha = 1$ all assets produced by H-type originators will receive a h signal and hence H-type cannot be binding in selling their h signal assets. In addition, since none of the assets produced by a L-type originator will receive a h signal, it follows that $x_h = X$.

The equilibrium production and trading by a H-type originator is described as in Eq. (A.55), while the equilibrium production by a L-type satisfies

$$(1-\lambda)p_A - \frac{\lambda^2}{\rho}q_L = k. \tag{A.64}$$

In this case, intermediaries' asset purchasing and relatedly the h signal market clearing become

$$\lambda X + (1-\lambda) p_A - \frac{\lambda^2}{\rho_I} b_I = p_h \tag{A.65}$$

$$\pi s_{Hh} = b_I, \tag{A.66}$$

and equilibrium asset market price and input capital price are still as given by Eq. (A.58). Combine equations we get intermediary's purchase

$$b_I = \frac{\rho_I (k-1)}{(\rho + \rho_I)\,\kappa} + \frac{(1-\pi)\, X \rho \rho_I}{\lambda\,(\rho + \rho_I)}, \tag{A.67}$$

and the equilibrium input capital price k determined by the following quadratic equation:

$$\frac{\lambda^2 (k-1)}{(\rho_I + \rho)\kappa} + k = \frac{(1-\lambda)(1-\pi)\rho X^2}{\lambda} \cdot \left(\frac{k-1}{\kappa}\right)^{-1} + \frac{(\rho_I + \pi\rho) X}{\rho_I + \rho} \tag{A.68}$$

It is easy to see that the LHS of Eq. (A.68) is monotonically increasing in k while the RHS of Eq. (A.68) is monotonically decreasing k for $k \in (1, \infty)$, which thus implies that the equilibrium with perfectly informed intermediation uniquely exists with $k > 1$. Solving this quadratic equation gives us equilibrium capital price k (or total production $Q = \frac{k-1}{\kappa}$) as in Eq. (38). ■

A4. Proofs and Calculations for Economic Implications of Tech-driven Intermediation

A4.1. Comparative statics analysis

Proof of Proposition 6

In this section, we show that when the technology level α improves in the economy, the average quality of assets produced in the economy improves and intermediaries play a (weakly) more significant role in the asset trading market, captured by (weakly) increasing trading volume ratio $\frac{vol_I}{vol_O}$.

First, note that in the tech-irrelevant range, increasing in technology level α has no impact on equilibrium outcomes. In the tech-relevant range where L-type's selling in the h signal market is binding while H-type are not binding in their selling, then an increase in α has no direct effect on H-type's equilibrium production and trading, while it has a direct effect on tightening L-type's eligibility constraint. As such, in equilibrium L-type respond by cutting their production q_L. The reduced production by L-type transmit to the production decision by H-type through its impact on equilibrium prices: holding q_H fixed, a reduced q_L leads to a lowered input capital price k and a heightened asset trading price p_A, both of which then encourage H-type originators to expand their production.

Mathematically, note that the intermediated equilibrium in the tech-relevant is characterized by equation system Eq. (A.61), (A.62) and (A.63) of $\{x_h, b_I, k\}$. We can write this equation system as an equation system of two unknowns $\{x_h, k\}$:

$$\frac{x_h}{X}\left[\frac{(x_h-\pi X)\rho\rho_I}{\lambda(\rho+\rho_I)}+\frac{\rho_I(k-1)}{\kappa(\rho+\rho_I)}\right]-\frac{k-1}{\kappa}=\rho\left(\pi+\frac{1-\pi}{\alpha}\right)\left[\frac{\rho_I x_h+\pi\rho X}{(\rho+\rho_I)\lambda}-\frac{k-1}{(\rho+\rho_I)\kappa}\right]-\frac{\rho\pi X}{\lambda} \tag{A.69}$$

and

$$\left(\frac{\lambda^2}{\rho_I}+k-\lambda x_h\right)\frac{k-1}{(1-\lambda)\kappa}=\frac{X(k-1)}{\kappa}-\frac{(1-\pi)\rho X}{\alpha\lambda}\left[\frac{\lambda(k-1)}{(\rho+\rho_I)\,\kappa}+\frac{(x_h-\pi X)\,\rho}{\rho+\rho_I}-x_h\right] \tag{A.70}$$

To apply the implicit function theorem, write the equation system (A.69) and (A.70) as

$$F(x_h,k,\alpha)\equiv\begin{pmatrix}F_1(x_h,k,\alpha)\\ F_2(x_h,k,\alpha)\end{pmatrix}=\begin{pmatrix}0\\ 0\end{pmatrix} \tag{A.71}$$

Then by implicit function theorem,

$$\begin{pmatrix}\frac{\partial x_h}{\alpha}\\ \frac{\partial k}{\partial\alpha}\end{pmatrix}=-\begin{bmatrix}\frac{\partial F_1}{\partial x_h} & \frac{\partial F_1}{\partial k}\\ \frac{\partial F_2}{\partial x_h} & \frac{\partial F_2}{\partial k}\end{bmatrix}^{-1}\begin{bmatrix}\frac{\partial F_1}{\partial\alpha}\\ \frac{\partial F_2}{\partial\alpha}\end{bmatrix}. \tag{A.72}$$

Plug in the functional forms from (A.69) and (A.70), we have

$$\begin{aligned}
\frac{\partial F_1}{\partial x_h} &= \frac{\rho\rho_I}{\lambda(\rho+\rho_I)}\left[\frac{2x_h-\pi X}{X}+\frac{\lambda(k-1)}{\kappa\rho}-\left(\pi+\frac{1-\pi}{\alpha}\right)\right]\\
\frac{\partial F_1}{\partial k} &= \frac{\rho_I x_h}{\kappa(\rho+\rho_I)X}-\frac{1}{\kappa}+\left(\pi+\frac{1-\pi}{\alpha}\right)\frac{\rho}{(\rho+\rho_I)\kappa}\\
\frac{\partial F_2}{\partial x_h} &= -\frac{\lambda(k-1)}{(1-\lambda)\kappa}-\frac{(1-\pi)\rho\rho_I X}{\alpha\lambda(\rho+\rho_I)}\\
\frac{\partial F_2}{\partial k} &= \frac{k-1}{(1-\lambda)\kappa}+\left(\frac{\lambda^2}{\rho_I}+k-\lambda x_h\right)\frac{1}{(1-\lambda)\kappa}-\frac{X}{\kappa}+\frac{(1-\pi)\rho X}{\alpha\,(\rho+\rho_I)\,\kappa}
\end{aligned}$$

and

$$\begin{aligned}
\frac{\partial F_1}{\partial\alpha} &= \frac{(1-\pi)\,\rho}{\alpha^2}\left[\frac{\rho_I x_h+\pi\rho X}{(\rho+\rho_I)\lambda}-\frac{k-1}{(\rho+\rho_I)\kappa}\right]\\
\frac{\partial F_2}{\partial\alpha} &= -\frac{(1-\pi)\rho X}{\alpha^2\lambda}\left[\frac{\lambda(k-1)}{(\rho+\rho_I)\,\kappa}+\frac{(x_h-\pi X)\,\rho}{\rho+\rho_I}-x_h\right]
\end{aligned}$$

By Eq. (A.72), we have

$$\begin{aligned}
\frac{\partial x_h}{\partial \alpha} &= \left(\frac{\partial F_1}{\partial k}\cdot\frac{\partial F_2}{\partial x_h}-\frac{\partial F_1}{\partial x_h}\cdot\frac{\partial F_2}{\partial k}\right)^{-1}\left(\frac{\partial F_2}{\partial k}\cdot\frac{\partial F_1}{\partial \alpha}-\frac{\partial F_1}{\partial k}\cdot\frac{\partial F_2}{\partial \alpha}\right)\\
\frac{\partial k}{\partial \alpha} &= \left(\frac{\partial F_1}{\partial k}\cdot\frac{\partial F_2}{\partial x_h}-\frac{\partial F_1}{\partial x_h}\cdot\frac{\partial F_2}{\partial k}\right)^{-1}\left(-\frac{\partial F_2}{\partial x_h}\cdot\frac{\partial F_1}{\partial \alpha}+\frac{\partial F_1}{\partial x_h}\cdot\frac{\partial F_2}{\partial \alpha}\right)
\end{aligned}$$

Plug into partial derivatives, we have

$$\begin{aligned}
&\frac{\partial F_2}{\partial k}\cdot\frac{\partial F_1}{\partial \alpha}-\frac{\partial F_1}{\partial k}\cdot\frac{\partial F_2}{\partial \alpha}\\
\propto\;& \left[\frac{k-1}{(1-\lambda)\kappa}+\left(\frac{\lambda^2}{\rho_I}+k-\lambda x_h\right)\frac{1}{(1-\lambda)\kappa}-\frac{X}{\kappa}+\frac{(1-\pi)\rho X}{\alpha\,(\rho+\rho_I)\,\kappa}\right]\cdot\left[\frac{\rho_I x_h+\pi\rho X}{(\rho+\rho_I)\lambda}-\frac{k-1}{(\rho+\rho_I)\kappa}\right]\\
&+\left[\frac{\rho_I x_h}{\kappa(\rho+\rho_I)X}-\frac{1}{\kappa}+\left(\pi+\frac{1-\pi}{\alpha}\right)\frac{\rho}{(\rho+\rho_I)\kappa}\right]\cdot\frac{X}{\lambda}\left[\frac{\lambda(k-1)}{(\rho+\rho_I)\,\kappa}+\frac{(x_h-\pi X)\,\rho}{\rho+\rho_I}-x_h\right]\\
\propto\;& \frac{1}{(1-\lambda)\kappa}\left(\frac{\lambda^2}{\rho_I}+2k-1-\lambda x_h\right)\cdot\left[\frac{\rho_I x_h+\pi\rho X}{\lambda}-\frac{k-1}{\kappa}\right]\\
&+\frac{(\rho_I x_h+\pi\rho X)\,(k-1)}{(\rho+\rho_I)\kappa^2}-\frac{(\rho_I x_h+\pi\rho X)^2}{\lambda(\rho+\rho_I)\kappa}\\
=\;& \left[\frac{1}{(1-\lambda)\kappa}\left(\frac{\lambda^2}{\rho_I}+2k-1-\lambda x_h\right)-\frac{\rho_I x_h+\pi\rho X}{(\rho+\rho_I)\kappa}\right]\cdot\left[\frac{\rho_I x_h+\pi\rho X}{\lambda}-\frac{k-1}{\kappa}\right]
\end{aligned}$$

Note that $b_I=\frac{\rho_I(k-1)}{(\rho+\rho_I)\kappa}+\frac{(x_h-\pi X)\rho\rho_I}{\lambda(\rho+\rho_I)}$ by Eq. (A.61), thus

$$\begin{aligned}
\frac{k-1}{\kappa}-\frac{\rho_I x_h+\pi\rho X}{\lambda} &= \left(\frac{\rho_I+\rho}{\rho_I}\right)b_I-\frac{\rho_I x_h}{\lambda}>b_I-\frac{\rho_I x_h}{\lambda}\\
&= \frac{\rho_I}{\lambda^2}\left(\frac{\lambda^2}{\rho_I}b_I-\lambda x_h\right)=\frac{\rho_I}{\lambda^2}\left[(1-\lambda)p_A-k\right]=\frac{\rho_I}{\lambda^2}\cdot\frac{\lambda^2}{\rho}\alpha q_L>0
\end{aligned}$$

and it is easy to see that

$$\frac{1}{(1-\lambda)\kappa}\left(\frac{\lambda^2}{\rho_I}+2k-1-\lambda x_h\right)-\frac{\rho_I x_h+\pi\rho X}{(\rho+\rho_I)\kappa}>0$$

under our specification with ρ_I being small. Therefore, it follows that $\frac{\partial F_2}{\partial k}\cdot\frac{\partial F_1}{\partial \alpha}-\frac{\partial F_1}{\partial k}\cdot\frac{\partial F_2}{\partial \alpha}<$

0. Furthermore, we have

$$
\begin{aligned}
& \frac{\partial F_1}{\partial k}\cdot\frac{\partial F_2}{\partial x_h}-\frac{\partial F_1}{\partial x_h}\cdot\frac{\partial F_2}{\partial k} \\
\propto \quad & -\left[\frac{\rho_I x_h}{X}-(\rho+\rho_I)+\left(\pi+\frac{1-\pi}{\alpha}\right)\rho\right]\left[\frac{\lambda(k-1)}{(1-\lambda)\kappa}+\frac{(1-\pi)\rho\rho_I X}{\alpha\lambda(\rho+\rho_I)}\right] \\
& -\frac{\rho\rho_I}{\lambda}\left[\frac{2x_h-\pi X}{X}+\frac{\lambda(k-1)}{\kappa\rho}-\left(\pi+\frac{1-\pi}{\alpha}\right)\right]\cdot\left[\frac{1}{1-\lambda}\left(\frac{\lambda^2}{\rho_I}+2k-1-\lambda x_h\right)-X+\frac{(1-\pi)\rho X}{\alpha\left(\rho+\rho_I\right)}\right] \\
= \quad & \frac{(1-\pi)\rho\rho_I}{\alpha(\rho+\rho_I)\lambda}\left[-\rho_I x_h+(\rho+\rho_I)X-\rho(2x_h-\pi X)-\lambda X\left(\frac{k-1}{\kappa}\right)\right] \\
& -\frac{\lambda(k-1)}{(1-\lambda)\kappa}\left[\frac{\rho_I x_h}{X}-(\rho+\rho_I)+\left(\pi+\frac{1-\pi}{\alpha}\right)\rho\right] \\
& -\frac{\rho\rho_I}{\lambda}\left[\frac{2x_h-\pi X}{X}+\frac{\lambda(k-1)}{\kappa\rho}-\left(\pi+\frac{1-\pi}{\alpha}\right)\right]\cdot\left[\frac{1}{1-\lambda}\left(\frac{\lambda^2}{\rho_I}+2k-1-\lambda x_h\right)-X\right]
\end{aligned}
$$

To simplify notations, denote

$$
\begin{aligned}
A \quad \equiv \quad & \frac{(1-\pi)\rho\rho_I}{\alpha(\rho+\rho_I)\lambda}\left[-\rho_I x_h+(\rho+\rho_I)X-\rho(2x_h-\pi X)-\lambda X\left(\frac{k-1}{\kappa}\right)\right] \\
B \quad \equiv \quad & -\frac{\lambda(k-1)}{(1-\lambda)\kappa}\left[\frac{\rho_I x_h}{X}-(\rho+\rho_I)+\left(\pi+\frac{1-\pi}{\alpha}\right)\rho\right] \\
& -\frac{\rho\rho_I}{\lambda}\left[\frac{2x_h-\pi X}{X}+\frac{\lambda(k-1)}{\kappa\rho}-\left(\pi+\frac{1-\pi}{\alpha}\right)\right]\cdot\left[\frac{1}{1-\lambda}\left(\frac{\lambda^2}{\rho_I}+2k-1-\lambda x_h\right)-X\right]
\end{aligned}
$$

For infinitesimal $\rho_I > 0$, we have

$$
\begin{aligned}
A \quad \propto \quad & -\rho_I x_h+(\rho+\rho_I)X-\rho(2x_h-\pi X)-\lambda X\left(\frac{k-1}{\kappa}\right)\approx \rho X-\rho(2x_h-\pi X)-\lambda X Q^d \\
< \quad & \rho X-\rho(2\pi X-\pi X)-\lambda X Q^d=\left[(1-\pi)\rho-\lambda Q^d\right]X
\end{aligned}
$$

where the inequality uses the fact that the stable equilibrium in the tech-relevant range must have signal asset quality $x_h > \pi X$. Similarly, with $\rho_I > 0$ being taken to zero, we have

$$
\begin{aligned}
B \quad \approx \quad & -\frac{\lambda Q^d}{1-\lambda}\left[-\rho+\left(\pi+\frac{1-\pi}{\alpha}\right)\rho\right]-\frac{\rho\rho_I}{\lambda}\left[\frac{2x_h-\pi X}{X}+\frac{\lambda(k-1)}{\kappa\rho}-\left(\pi+\frac{1-\pi}{\alpha}\right)\right]\cdot\frac{\lambda^2}{(1-\lambda)\rho_I} \\
= \quad & -\frac{\lambda\rho}{1-\lambda}\left[-Q^d+\left(\pi+\frac{1-\pi}{\alpha}\right)Q^d\right]-\frac{\lambda\rho}{1-\lambda}\left[\frac{2x_h-\pi X}{X}+\frac{\lambda Q^d}{\rho}-\left(\pi+\frac{1-\pi}{\alpha}\right)\right] \\
< \quad & -\frac{\lambda\rho}{1-\lambda}\left[-Q^d+\left(\pi+\frac{1-\pi}{\alpha}\right)Q^d\right]-\frac{\lambda\rho}{1-\lambda}\left[\frac{2\pi X-\pi X}{X}+\frac{\lambda Q^d}{\rho}-\left(\pi+\frac{1-\pi}{\alpha}\right)\right] \\
= \quad & -\frac{\lambda\rho}{1-\lambda}\left[-Q^d+\left(\pi+\frac{1-\pi}{\alpha}\right)Q^d+\frac{\lambda Q^d}{\rho}-\frac{1-\pi}{\alpha}\right] \\
\leq \quad & -\frac{\lambda}{1-\lambda}\left[\lambda Q^d-(1-\pi)\rho\right]=\frac{\lambda}{1-\lambda}\left[(1-\pi)\rho-\lambda Q^d\right]
\end{aligned}
$$

where the first inequality uses the fact that in stable equilibrium $x_h > \pi X$ and the second inequality uses the fact that $\alpha \leq 1$.

Therefore, under the condition that

$$(1-\pi)\rho - \lambda Q^d < 0, \tag{A.73}$$

where the total production in a direct trading economy Q^d is as given by Eq. (22), we have $A < 0$ and $B < 0$. It is easy to see that (A.73) is guaranteed under $Q^d \geq \max\left\{1, \frac{1-\pi}{\pi-1+\lambda/\rho}\right\}$, thus we have

$$\frac{\partial F_1}{\partial k} \cdot \frac{\partial F_2}{\partial x_h} - \frac{\partial F_1}{\partial x_h} \cdot \frac{\partial F_2}{\partial k} \propto A + B < 0,$$

and thus

$$\frac{\partial x_h}{\partial \alpha} = \left(\frac{\partial F_1}{\partial k} \cdot \frac{\partial F_2}{\partial x_h} - \frac{\partial F_1}{\partial x_h} \cdot \frac{\partial F_2}{\partial k}\right)^{-1} \left(\frac{\partial F_2}{\partial k} \cdot \frac{\partial F_1}{\partial \alpha} - \frac{\partial F_1}{\partial k} \cdot \frac{\partial F_2}{\partial \alpha}\right) > 0$$

Similarly, for the comparative statics regarding k, note that when ρ_I is close to 0, we have

$$-\frac{\partial F_2}{\partial x_h} \cdot \frac{\partial F_1}{\partial \alpha} + \frac{\partial F_1}{\partial x_h} \cdot \frac{\partial F_2}{\partial \alpha} \quad \approx \quad \frac{\lambda Q^d}{1-\lambda} \cdot \frac{(1-\pi)\rho}{\alpha^2} \left(\frac{\pi X}{\lambda} - \frac{Q^d}{\rho}\right).$$

But since $Q^d > \frac{\pi \rho X}{\lambda}$, it thus implies that when ρ_I is close to 0, it is guaranteed that

$$-\frac{\partial F_2}{\partial x_h} \cdot \frac{\partial F_1}{\partial \alpha} + \frac{\partial F_1}{\partial x_h} \cdot \frac{\partial F_2}{\partial \alpha} < 0$$

and thus

$$\frac{\partial k}{\partial \alpha} = \left(\frac{\partial F_1}{\partial k} \cdot \frac{\partial F_2}{\partial x_h} - \frac{\partial F_1}{\partial x_h} \cdot \frac{\partial F_2}{\partial k}\right)^{-1} \left(-\frac{\partial F_2}{\partial x_h} \cdot \frac{\partial F_1}{\partial \alpha} + \frac{\partial F_1}{\partial x_h} \cdot \frac{\partial F_2}{\partial \alpha}\right) > 0$$

To show that increasing intermediation technology α also translates into an increased trading price in asset market p_A (which thus reflects an improved production composition $\frac{q_H}{q_L}$), note that the assets traded in the asset markets consists of two parts—one part that is brought in by intermediaries and the other part that is brought in directly by originators. The part brought in by intermediaries has an average quality x_h, which is increasing in α. Suppose by way of contradiction that after an increase in α, the trading price in the asset market p_A decreases. Then it must be that the part of assets brought in directly by originators has a deteriorating average quality. But by Eq. (A.55) and (A.56), we can get the asset mixture brought into by originators directly to the asset market

$$m_O(\alpha) \equiv \frac{q_H - s_{Hh}}{\alpha q_L} = \frac{\lambda X + (1-\lambda)p_A - k}{(1-\lambda)p_A - k} = 1 + \frac{\lambda X}{(1-\lambda)p_A - k}. \tag{A.74}$$

But since $\frac{\partial k}{\partial \alpha} > 0$, hence $\frac{\partial p_A}{\partial \alpha} < 0$ immediately implies that $\frac{\partial m_O}{\partial \alpha} > 0$, which contradicts that the average quality of assets brought in by originators must be deteriorating. Therefore, after an increase in the intermediation technology, the equilibrium asset market trading price must increase, i.e., $\frac{\partial p_A}{\partial \alpha} > 0$.

In the last part of the proof, we show that under the conditions provided in Proposition 6, we have $\frac{\partial q_L}{\partial \alpha} < 0$. First note that by L type originators' optimality condition, we have

$$(1-\lambda)p_A - \frac{\lambda^2}{\rho}\alpha q_L = k.$$

Differentiating the above equation w.r.t. α, we get

$$(1-\lambda)\frac{\partial p_A}{\partial \alpha} - \frac{\partial k}{\partial \alpha} = \frac{\lambda^2}{\rho}\left(q_L + \alpha\frac{\partial q_L}{\partial \alpha}\right).$$

Therefore, $\frac{\partial q_L}{\partial \alpha} < 0$ is equivalent to

$$(1-\lambda)\frac{\partial p_A}{\partial \alpha} - \frac{\partial k}{\partial \alpha} < \frac{\lambda^2}{\rho}q_L. \tag{A.75}$$

Next, note that by Eq. (A.60) and Eq. (A.61), we have

$$\lambda x_h + (1-\lambda)p_A - \frac{\lambda^2 Q}{\rho+\rho_I} - \frac{\lambda\left(x_h - \pi X\right)\rho}{\rho+\rho_I} = k,$$

which implies

$$(1-\lambda)\frac{\partial p_A}{\partial \alpha} - \frac{\partial k}{\partial \alpha} = \frac{\lambda^2}{(\rho+\rho_I)\kappa}\frac{\partial k}{\partial \alpha} + \frac{\lambda\rho}{\rho+\rho_I}\frac{\partial x_h}{\partial \alpha} - \lambda\frac{\partial x_h}{\partial \alpha}. \tag{A.76}$$

For ρ_I close to zero, we have $(1-\lambda)\frac{\partial p_A}{\partial \alpha} - \frac{\partial k}{\partial \alpha} \approx \frac{\lambda^2}{\rho\kappa}\frac{\partial k}{\partial \alpha}$, and hence it suffices to prove $\frac{\partial k}{\partial \alpha} < \kappa q_L^d$.

From our calculations above, when ρ_I is close to zero we have

$$
\begin{aligned}
\frac{\partial k}{\partial \alpha} &= \left(\frac{\partial F_1}{\partial k}\cdot\frac{\partial F_2}{\partial x_h} - \frac{\partial F_1}{\partial x_h}\cdot\frac{\partial F_2}{\partial k}\right)^{-1}\left(-\frac{\partial F_2}{\partial x_h}\cdot\frac{\partial F_1}{\partial \alpha} + \frac{\partial F_1}{\partial x_h}\cdot\frac{\partial F_2}{\partial \alpha}\right) \\
&\approx \left[\frac{\lambda}{(1-\lambda)\kappa}\left(\frac{2x_h - \pi X}{X} + \frac{\lambda}{\rho}Q^d + (\pi + \frac{1-\pi}{\alpha})(Q^d - 1) - Q^d\right)\right]^{-1} \\
&\quad \cdot\left[\frac{\lambda Q^d}{1-\lambda}\cdot\frac{(1-\pi)\rho}{\alpha^2}\left(-\frac{\pi X}{\lambda} + \frac{Q^d}{\rho}\right)\right] \\
&\leq \left[\frac{\lambda}{(1-\lambda)\kappa}\left(\frac{\lambda}{\rho}Q^d - (1-\pi)\right)\right]^{-1}\cdot\left[\frac{\lambda(1-\pi)Q^d}{1-\lambda}\left(Q^d - \frac{\rho\pi X}{\lambda}\right)\right] \\
&= \frac{(1-\pi)\kappa Q^d q_L^d}{\frac{\lambda}{\rho}Q^d - (1-\pi)},
\end{aligned}
$$

where the inequality is using the fact that $\alpha \leq 1$ and $x_h \geq \pi X$ as well as the condition $Q^d \geq 1$, and the last equality is using the fact that $Q^d = \frac{\pi\rho X}{\lambda} + q_L^d$.

Thus to show $\frac{\partial k}{\partial \alpha} < \kappa q_L^d$, it suffices to show that $\frac{(1-\pi)\kappa Q^d q_L^d}{\frac{\lambda}{\rho}Q^d - (1-\pi)} < \kappa q_L^d$, which is equivalent to $(1-\pi)Q^d < \frac{\lambda}{\rho}Q^d - (1-\pi)$. Thus under the condition that $Q^d \geq \max\left\{1, \frac{1-\pi}{\frac{\lambda}{\rho} - 1 + \pi}\right\}$, it is guaranteed that $\frac{\partial k}{\partial \alpha} < \kappa q_L^d$, which further implies that $\frac{\partial q_L}{\partial \alpha} < 0$.

This completes the proof for Proposition (6). ■

*A4.2. **Welfare analysis***

Proof of Proposition 7

In this proof, we show that uninformed intermediation could potentially impair social surplus while informed intermediation always improves social surplus. Particularly, for the low technology α range, we show that when $\alpha \leq \hat{\alpha}_1(\rho_I)$ so that the intermediated equilibrium is in the tech-irrelevant range, for any $\rho_I > 0$ we have $w(\rho_I) < w^d$ if and only if $\pi X < 1$. For the high technology α range, we show that when $\alpha = 1$ we have $w(\rho_I) > w^d$ always holds for any $\rho_I > 0$.

Part 1. Welfare implication of uninformed intermediation ($\alpha \leq \hat{\alpha}_1(\rho_I)$)

We start with calculating the social welfare attained in a direct trading economy, as characterized in Section 3.1. Given equilibrium outcomes characterized in Proposition 2 the equilibrium payoff to an originator in a direct trading economy is

$$\begin{aligned} v_O^d &= \pi \left[\lambda X q_H + (1-\lambda) p_A q_H - \frac{1}{2\rho} (\lambda q_H)^2 - k q_H \right] + (1-\pi) \left[(1-\lambda) p_A q_L - \frac{1}{2\rho} (\lambda q_L)^2 - k q_L \right] \\ &= \frac{\lambda^2}{2\rho} \left[\pi q_H^2 + (1-\pi) q_L^2 \right] = \frac{\lambda^2}{2\rho} \left[\frac{(k^d-1)^2}{\kappa^2} + \frac{\rho^2 X^2 \pi (1-\pi)}{\lambda^2} \right] \\ &= \frac{\lambda^2 (k^d-1)^2}{2\rho\kappa^2} + \frac{\rho}{2} \pi (1-\pi) X^2, \end{aligned} \tag{A.77}$$

and the equilibrium payoff to input capital producer in a direct trading economy is

$$\begin{aligned} v_K^d &= k^d \left[\pi q_H + (1-\pi) q_L \right] - K \left(\pi q_H + (1-\pi) q_L \right) = \frac{k^d (k^d-1)}{\kappa} - \frac{k^d-1}{\kappa} - \frac{\kappa}{2} \left(\frac{k^d-1}{\kappa} \right)^2 \\ &= \frac{(k^d-1)^2}{2\kappa} \end{aligned} \tag{A.78}$$

Therefore, the social welfare attained in a direct trading economy is

$$w^d = v_O^d + v_K^d = \left(1 + \frac{\lambda^2}{\kappa\rho} \right) \frac{(k^d-1)^2}{2\kappa} + \frac{\rho}{2} \pi (1-\pi) X^2 \tag{A.79}$$

Furthermore, by Eq. (20), we have

$$\left(1 + \frac{\lambda^2}{\rho\kappa} \right) \left(k^d - 1 \right) + 1 - \pi X = \frac{\kappa\rho (1-\lambda)(1-\pi)\pi X^2}{\lambda (k^d-1)}. \tag{A.80}$$

Therefore, we can further get

$$\begin{aligned} w^d &= \frac{\rho (1-\lambda)(1-\pi)\pi X^2}{2\lambda} + \left(\frac{\pi X - 1}{2\kappa} \right) (k^d-1) + \frac{\rho}{2} \pi (1-\pi) X^2 \\ &= \left(\frac{\pi X - 1}{2\kappa} \right) (k^d-1) + \frac{\rho \pi (1-\pi) X^2}{2\lambda}, \end{aligned} \tag{A.81}$$

in which equilibrium input capital price k^d is given by Eq. (23).

Move on to the intermediated equilibrium in the tech-irrelevant range characterized by Proposition 4, the equilibrium payoff to an asset originator in this economy is

$$\begin{aligned} v_O^u &= \pi \left(q_H - s_h \right) \left[\lambda X + (1-\lambda) p_A - \frac{\lambda^2}{2\rho} \left(q_H - s_h \right) - k \right] \\ &\quad + (1-\pi) \left(q_L - s_h \right) \left[(1-\lambda) p_A - \frac{\lambda^2}{2\rho} \left(q_L - s_h \right) - k \right] \\ &= \frac{\pi\rho}{2\lambda^2} \left[\lambda X + (1-\lambda) p_A - k \right]^2 + \frac{(1-\pi)\rho}{2\lambda^2} \left[(1-\lambda) p_A - k \right]^2 \end{aligned} \tag{A.82}$$

Using Eq. (A.50), we get

$$\begin{aligned}
v_O^u &= \frac{\pi\rho}{2\lambda^2}\left[\lambda X - k + k + \frac{\lambda^2(k-1)}{(\rho_I+\rho)\,\kappa} - \lambda\pi X\right]^2 + \frac{(1-\pi)\rho}{2\lambda^2}\left[k + \frac{\lambda^2(k-1)}{(\rho_I+\rho)\,\kappa} - \lambda\pi X - k\right]^2 \\
&= \frac{\pi\rho}{2\lambda^2}\left[\frac{\lambda^2(k-1)}{(\rho_I+\rho)\,\kappa} + (1-\pi)\lambda X\right]^2 + \frac{(1-\pi)\rho}{2\lambda^2}\left[\frac{\lambda^2(k-1)}{(\rho_I+\rho)\,\kappa} - \pi\lambda X\right]^2 \\
&= \frac{\pi\rho}{2\lambda^2}\left[\frac{\lambda^4(k-1)^2}{(\rho_I+\rho)^2\,\kappa^2} + 2(1-\pi)\lambda X\cdot\frac{\lambda^2(k-1)}{(\rho_I+\rho)\,\kappa} + (1-\pi)^2\lambda^2X^2\right] \\
&\quad + \frac{(1-\pi)\rho}{2\lambda^2}\left[\frac{\lambda^4(k-1)^2}{(\rho_I+\rho)^2\,\kappa^2} - 2\pi\lambda X\cdot\frac{\lambda^2(k-1)}{(\rho_I+\rho)\,\kappa} + \pi^2\lambda^2X^2\right] \\
&= \frac{\rho\lambda^2(k-1)^2}{2\,(\rho_I+\rho)^2\,\kappa^2} + \frac{\rho}{2}\pi(1-\pi)X^2 \qquad \text{(A.83)}
\end{aligned}$$

The equilibrium payoff to an intermediary in this economy is

$$\begin{aligned}
v_I^u &= b_I\left[\lambda x_h + (1-\lambda)p_A - \frac{\lambda^2}{2\rho_I}b_I - k\right] = \frac{\rho_I}{2\lambda^2}\left[\lambda\pi X + (1-\lambda)p_A - k\right]^2 \\
&= \frac{\rho_I}{2\lambda^2}\left[\lambda\pi X + k + \frac{\lambda^2(k-1)}{(\rho_I+\rho)\,\kappa} - \lambda\pi X - k\right]^2 \\
&= \frac{\rho_I\lambda^2(k-1)^2}{2\,(\rho_I+\rho)^2\,\kappa^2} \qquad \text{(A.84)}
\end{aligned}$$

Finally, the equilibrium payoff to input capital producer in this economy is

$$v_K^u = k\left[\pi q_H + (1-\pi)q_L\right] - K\left(\pi q_H + (1-\pi)q_L\right) = \frac{(k-1)^2}{2\kappa} \qquad \text{(A.85)}$$

Therefore, the social welfare attained by an intermediated economy in the tech-irrelevant range is

$$\begin{aligned}
w^u &= v_O^u + v_I^u + v_K^u = \frac{\rho\lambda^2(k^u-1)^2}{2\,(\rho_I+\rho)^2\,\kappa^2} + \frac{\rho}{2}\pi(1-\pi)X^2 + \frac{\rho_I\lambda^2(k^u-1)^2}{2\,(\rho_I+\rho)^2\,\kappa^2} + \frac{(k^u-1)^2}{2\kappa} \\
&= \frac{(k^u-1)^2}{2\kappa}\left[1 + \frac{\lambda^2}{(\rho_I+\rho)\,\kappa}\right] + \frac{\rho}{2}\pi(1-\pi)X^2, \qquad \text{(A.86)}
\end{aligned}$$

in which we use k^u to denote the equilibrium input capital price in such an economy with uninformed intermediation. Furthermore, by Eq. (A.50), we have

$$(k^u-1)^2\left[1 + \frac{\lambda^2}{(\rho_I+\rho)\,\kappa}\right] + (1-\pi X)(k^u-1) = \left(\frac{1-\lambda}{\lambda}\right)\pi(1-\pi)\rho\kappa X^2 \qquad \text{(A.87)}$$

Therefore, we can get

$$
\begin{aligned}
w^u &= \frac{(\pi X - 1)(k^u - 1)}{2\kappa} + \left(\frac{1-\lambda}{2\lambda}\right)\pi(1-\pi)\rho X^2 + \frac{\rho}{2}\pi(1-\pi)X^2 \\
&= \left(\frac{\pi X - 1}{2\kappa}\right)(k^u - 1) + \frac{\rho\pi(1-\pi)X^2}{2\lambda} \qquad \text{(A.88)}
\end{aligned}
$$

Comparing the social welfare of the direct economy (w^d) to that of the intermediated economy in the tech-irrelevant range (w^u), we have

$$
w^u(\rho_I) - w^d = \left(\frac{\pi X - 1}{2\kappa}\right)\left[k^u(\rho_I) - k^d\right], \qquad \text{(A.89)}
$$

in which we index the intermediated equilibrium with intermediaries' retention capacity $\rho_I > 0$.

Note that $k^u(0) = k^d$. For $\rho_I > 0$, it can be shown that $\frac{\partial k^u(\rho_I)}{\partial \rho_I} > 0$. To see this, recall that by Eq. (A.50) we have

$$
\frac{(1-\pi)\rho\kappa X}{\lambda(k^u - 1)} = \frac{1}{\pi X(1-\lambda)}\left[k^u + \frac{\lambda^2(k^u-1)}{(\rho_I + \rho)\kappa} - \pi X\right]. \qquad \text{(A.90)}
$$

The LHS of the above equation is monotonically decreasing in k^u while the RHS is monotonically increasing in k^u on $k^u \in (1, \infty)$. An increase in ρ_I leaves the LHS unaffected while lowers the RHS, pushing the solution k^u towards right.

Therefore, with $\frac{\partial k^u(\rho_I)}{\partial \rho_I} > 0$ for any $\rho_I > 0$, it thus follows that

$$
k^u(\rho_I) > k^u(0) = k^d. \qquad \text{(A.91)}
$$

As such, when the lemon's problem is sufficiently severe in the economy such that $\pi X < 1$, we have

$$
w^u(\rho_I) - w^d = \left(\frac{\pi X - 1}{2\kappa}\right)\left[k^u(\rho_I) - k^d\right] < 0 \qquad \text{(A.92)}
$$

for any $\rho_I > 0$. Conversely, if $\pi X > 1$, when intermediation always improves social welfare, even if the intermediated equilibrium is in the tech-irrelevant range. ■

Part 2. Welfare implication of perfectly informed intermediation ($\alpha = 1$)

We now prove the second part of Proposition 7. Namely, the social welfare attained in an economy with perfectly informed intermediation, as characterized by Corollary 1, is always higher than that attained by a direct trading economy. The equilibrium payoff to an asset originator in this economy is

$$
\begin{aligned}
v_O^{in} &= \pi\left(q_H - s_h\right)\left[\lambda X + (1-\lambda)p_A - \frac{\lambda^2}{2\rho}\left(q_H - s_h\right) - k\right] + (1-\pi)q_L\left[(1-\lambda)p_A - \frac{\lambda^2}{2\rho}q_L - k\right] \\
&= \frac{\pi\rho}{2\lambda^2}\left[\lambda X + (1-\lambda)p_A - k\right]^2 + \frac{(1-\pi)\rho}{2\lambda^2}\left[(1-\lambda)p_A - k\right]^2 \\
&= \frac{\pi\rho}{2\lambda^2}\left[\lambda X + \frac{\lambda^2(k-1)}{(\rho_I+\rho)\kappa} + k - \frac{\lambda X(\rho\pi+\rho_I)}{\rho_I+\rho} - k\right]^2 + \frac{(1-\pi)\rho}{2\lambda^2}\left[\frac{\lambda^2(k-1)}{(\rho_I+\rho)\kappa} + k - \frac{\lambda X(\rho\pi+\rho_I)}{\rho_I+\rho} - k\right]^2 \\
&= \frac{\pi\rho}{2\lambda^2}\left[\frac{\lambda^2(k-1)}{(\rho_I+\rho)\kappa} + \frac{\lambda X(1-\pi)\rho}{\rho_I+\rho}\right]^2 + \frac{(1-\pi)\rho}{2\lambda^2}\left[\frac{\lambda^2(k-1)}{(\rho_I+\rho)\kappa} - \frac{\lambda X(\rho\pi+\rho_I)}{\rho_I+\rho}\right]^2 \\
&= \frac{\rho\lambda^2(k-1)^2}{2(\rho_I+\rho)^2\kappa^2} - \frac{(1-\pi)\lambda\rho\rho_I X(k-1)}{(\rho_I+\rho)^2\kappa} + \frac{\pi(1-\pi)\rho^2X^2}{2(\rho_I+\rho)} + \frac{(1-\pi)\rho\rho_I(\pi\rho+\rho_I)X^2}{2(\rho_I+\rho)^2} \qquad \text{(A.93)}
\end{aligned}
$$

The equilibrium payoff to intermediaries in this economy is

$$
\begin{aligned}
v_I^{in} &= b_I\left[\lambda X + (1-\lambda)p_A - \frac{\lambda^2}{2\rho_I}b_I - k\right] = \frac{\rho_I}{2\lambda^2}\left[\lambda X + (1-\lambda)p_A - k\right]^2 \\
&= \frac{\rho_I}{2\lambda^2}\left[\lambda X + \frac{\lambda^2(k-1)}{(\rho_I+\rho)\kappa} + k - \frac{\lambda X(\rho\pi+\rho_I)}{\rho_I+\rho} - k\right]^2 = \frac{\rho_I}{2\lambda^2}\left[\frac{\lambda^2(k-1)}{(\rho_I+\rho)\kappa} + \frac{\lambda X(1-\pi)\rho}{\rho_I+\rho}\right]^2 \\
&= \frac{\rho_I\lambda^2(k-1)^2}{2(\rho_I+\rho)^2\kappa^2} + \frac{\lambda\rho_I\rho X(k-1)}{(\rho_I+\rho)^2\kappa} + \frac{\rho_I\rho^2X^2(1-\pi)^2}{2(\rho_I+\rho)^2} \qquad \text{(A.94)}
\end{aligned}
$$

Adding the equilibrium payoff to input capital producer $v_K^{in} = \frac{(k-1)^2}{2\kappa}$, the welfare in this economy with perfectly informed intermediation is $w^{in} = v_O^{in} + v_I^{in} + v_K^{in}$:

$$
\begin{aligned}
w^{in} &= \frac{\rho\lambda^2(k-1)^2}{2(\rho_I+\rho)^2\kappa^2} - \frac{(1-\pi)\lambda\rho\rho_I X(k-1)}{(\rho_I+\rho)^2\kappa} + \frac{\pi(1-\pi)\rho^2X^2}{2(\rho_I+\rho)} + \frac{(1-\pi)\rho\rho_I(\pi\rho+\rho_I)X^2}{2(\rho_I+\rho)^2} \\
&\quad + \frac{\rho_I\lambda^2(k-1)^2}{2(\rho_I+\rho)^2\kappa^2} + \frac{\lambda\rho_I\rho X(k-1)}{(\rho_I+\rho)^2\kappa} + \frac{\rho_I\rho^2X^2(1-\pi)^2}{2(\rho_I+\rho)^2} \\
&\quad + \frac{(k-1)^2}{2\kappa} \\
&= \left[\frac{\lambda^2}{2(\rho_I+\rho)\kappa^2} + \frac{1}{2\kappa}\right](k-1)^2 + \frac{\pi\rho\rho_I X(k-1)}{(\rho_I+\rho)^2\kappa} + \frac{(1-\pi)\rho(\rho_I+\pi\rho)X^2}{2(\rho_I+\rho)} \qquad \text{(A.95)}
\end{aligned}
$$

Using Eq. (A.68), we have

$$
\left[\frac{\lambda^2}{(\rho_I+\rho)\kappa^2} + \frac{1}{\kappa}\right](k-1)^2 = \frac{(1-\lambda)(1-\pi)\rho(\rho_I+\rho\pi)X^2}{\lambda(\rho_I+\rho)} + \left(\frac{(\rho_I+\pi\rho)X}{\rho_I+\rho} - 1\right)\frac{k-1}{\kappa}
$$

Thus we have

$$\begin{aligned} w^{in}(\rho_I) &= \frac{(1-\lambda)(1-\pi)\rho(\rho_I+\rho\pi)X^2}{2\lambda(\rho_I+\rho)} + \left(\frac{(\rho_I+\pi\rho)X}{\rho_I+\rho} - 1\right)\frac{k^{in}-1}{2\kappa} \\ &\quad + \frac{\pi\rho\rho_I X(k^{in}-1)}{(\rho_I+\rho)^2\kappa} + \frac{(1-\pi)\rho(\rho_I+\pi\rho)X^2}{2(\rho_I+\rho)} \\ &= \left[\frac{2\pi\rho\rho_I X}{(\rho_I+\rho)^2} + \frac{(\rho_I+\pi\rho)X}{\rho_I+\rho} - 1\right]\left(\frac{k^{in}-1}{2\kappa}\right) + \frac{(1-\pi)\rho(\rho_I+\rho\pi)X^2}{2\lambda(\rho_I+\rho)}, \\ &= \left[\frac{2\pi\rho\rho_I X}{(\rho_I+\rho)^2} + \frac{(\rho_I+\pi\rho)X}{\rho_I+\rho} - 1\right]\frac{Q}{2} + \frac{(1-\pi)\rho(\rho_I+\rho\pi)X^2}{2\lambda(\rho_I+\rho)}. \end{aligned}$$

To show that $w^{in}(\rho_I) > w^d$ for any $\rho_I > 0$, first note that Eq. (A.68) can be rewritten as:

$$0 = F(\rho_I, Q) \equiv \left(\kappa + \frac{\lambda^2}{\rho_I+\rho}\right)Q^2 + \left[1 - \frac{(\rho_I+\pi\rho)\,X}{\rho_I+\rho}\right]Q - \frac{(1-\lambda)(1-\pi)\rho X^2}{\lambda}, \quad \text{(A.96)}$$

where we used the fact $\frac{k-1}{\kappa} = Q$. Applying implicit function theorem to $F(\rho_I, Q)$, we have

$$\begin{aligned} \frac{\partial Q}{\partial \rho_I} &= -\frac{\partial F}{\partial \rho_I}\Big/\frac{\partial F}{\partial Q} \\ &= \left[\frac{\lambda^2 Q^2}{(\rho_I+\rho)^2} + \frac{(1-\pi)\rho X Q}{(\rho_I+\rho)^2}\right] \Big/ \left[2\left(\kappa + \frac{\lambda^2}{\rho_I+\rho}\right)Q + 1 - \frac{(\rho_I+\pi\rho)\,X}{\rho_I+\rho}\right] > 0. \end{aligned}$$

Next denote

$$\hat{w}^{in}(\rho_I) \equiv \left[\frac{(\rho_I+\pi\rho)X}{\rho_I+\rho} - 1\right]\frac{Q}{2} + \frac{(1-\pi)\rho(\rho_I+\rho\pi)X^2}{2\lambda(\rho_I+\rho)}. \quad \text{(A.97)}$$

Then it is easy to see that $w^{in}(\rho_I) > \hat{w}^{in}(\rho_I)$ for any $\rho_I > 0$, and $\hat{w}^{in}(0) = w^d$ (see Eq. (A.81)). In what follows, we show $\frac{\partial \hat{w}^{in}(\rho_I)}{\partial \rho_I} > 0$ for any $\rho_I > 0$. We have:

$$\begin{aligned} \frac{\partial \hat{w}^{in}(\rho_I)}{\partial \rho_I} &= \frac{(1-\pi)\rho X}{(\rho_I+\rho)^2}\cdot\frac{Q}{2} + \left[\frac{(\rho_I+\pi\rho)X}{\rho_I+\rho} - 1\right]\cdot\frac{1}{2}\cdot\frac{\partial Q}{\partial \rho_I} + \frac{(1-\pi)^2\rho^2X^2}{2\lambda(\rho_I+\rho)^2} \\ &\propto \underbrace{\left[(1-\pi)\rho X Q + \frac{(1-\pi)^2\rho^2 X^2}{\lambda}\right]\left[2\left(\kappa + \frac{\lambda^2}{\rho_I+\rho}\right)Q + 1 - \frac{(\rho_I+\pi\rho)\,X}{\rho_I+\rho}\right]}_{A} \\ &\quad + \underbrace{\left[\frac{(\rho_I+\pi\rho)X}{\rho_I+\rho} - 1\right]\left[\lambda^2 Q^2 + (1-\pi)\rho X Q\right]}_{B}, \end{aligned} \quad \text{(A.98)}$$

We would like to show $A + B > 0$. First, by Eq. (A.96), we have

$$A = \left[(1-\pi)\rho X Q + \frac{(1-\pi)^2\rho^2 X^2}{\lambda}\right]\left[\left(\kappa + \frac{\lambda^2}{\rho_I + \rho}\right)Q + \frac{(1-\lambda)(1-\pi)\rho X^2}{\lambda Q}\right] > 0.$$

There are two cases to consider.

1. Suppose $\frac{(\rho_I+\pi\rho)X}{\rho_I+\rho} - 1 > 0$. Then we know $B > 0$ immediately and the claim of $A + B > 0$ follows.
2. Suppose $\frac{(\rho_I+\pi\rho)X}{\rho_I+\rho} - 1 < 0$. Expand Eq. (A.98), we have $A + B$ equals

$$\underbrace{\left[2(1-\pi)\rho X\left(\kappa + \frac{\lambda^2}{\rho_I+\rho}\right) + \lambda^2\left(\frac{(\rho_I+\pi\rho)X}{\rho_I+\rho} - 1\right)\right]}_{C} Q^2 +$$
$$\tfrac{(1-\pi)^2\rho^2X^2}{\lambda}\cdot 2\left(\kappa + \tfrac{\lambda^2}{\rho_I+\rho}\right)Q + \tfrac{(1-\pi)^2\rho^2X^2}{\lambda}\left(1 - \tfrac{(\rho_I+\pi\rho)X}{\rho_I+\rho}\right) \qquad (A.99)$$

Note that with $\frac{(\rho_I+\pi\rho)X}{\rho_I+\rho} - 1 < 0$ the second line is positive. And, "C" in the first line of (A.99) satisfies

$$\begin{aligned} C &> (1-\pi)\rho X\left(\kappa + \frac{\lambda^2}{\rho_I+\rho}\right) + \lambda^2\left(\frac{(\rho_I+\pi\rho)X}{\rho_I+\rho} - 1\right) \\ &> (1-\pi)\rho X\frac{\lambda^2}{\rho_I+\rho} + \lambda^2\left(\frac{(\rho_I+\pi\rho)X}{\rho_I+\rho} - 1\right) \\ &= \lambda^2(X-1) > 0 \end{aligned}$$

Hence it follows that $A + B > 0$.

Recall that $A + B > 0$ implies $\frac{\partial \hat{w}^{in}(\rho_I)}{\partial \rho_I} > 0$. Therefore, for any $\rho_I > 0$, we have $w^{in}(\rho_I) > \hat{w}^{in}(\rho_I) > w^d$ which completes the proof. ■

A5. *Detailed Analysis for Model Robustness and Implications*

This section provides detailed calculations for our analysis of the micro-foundation of market illiquidity λ and model robustness under alternative model specifications.

A5.1. *Calculations for endogenizing λ in a directed search setting*

In this section, we provide calculations for the characterization of generalized equilibrium with endogenous market illiquidity λ. For illustration, consider the case where trading volume in both $\mathcal{H}$ market and $\mathcal{L}$ market is positive. From our analysis in Section 5.1, both types of originators have equalized selling $s_{\mathcal{L}}$ in the liquid market and equilibrium asset quality in the two assets markets are given by

$$x_{\mathcal{H}} = \frac{x_H \pi s_{H\mathcal{H}} + x_L(1-\pi)s_{L\mathcal{H}}}{\pi s_{H\mathcal{H}} + (1-\pi)s_{L\mathcal{H}}}, \;\; x_{\mathcal{L}} = \pi X, \tag{A.100}$$

and equilibrium prices in asset markets are thus determined as

$$p_{\mathcal{H}} = x_{\mathcal{H}}, \;\; p_{\mathcal{L}} = x_{\mathcal{L}}, \tag{A.101}$$

so that assets in both markets offer a reservation return of 1.

On equilibrium production and trading, H-type originators brings $s_{H\mathcal{H}}$ of produced assets to sell in the illiquid market and the production q_H is determined as

$$s_{H\mathcal{H}} + s_{\mathcal{L}} = q_H, \text{ and } \; \lambda_{\mathcal{H}} x_H + (1-\lambda_{\mathcal{H}})p_{\mathcal{H}} - \frac{1}{\rho}\lambda_{\mathcal{H}}^2 s_{H\mathcal{H}} = k, \tag{A.102}$$

under our quadratic functional form specification. Similarly, an L-type originator produces q_L and bring $s_{Lm} \in (0, q_L)$ of produced assets to market m for sale, such that

$$s_{L\mathcal{H}} + s_{\mathcal{L}} = q_L, \;\text{ and }\; \lambda_{\mathcal{H}} x_L + (1-\lambda_{\mathcal{H}})p_{\mathcal{H}} - \rho\lambda_{\mathcal{H}}^2 s_{L\mathcal{H}} = k. \tag{A.103}$$

Furthermore, if the trading volume is positive in the perfect liquid $\mathcal{L}$ market, we must have in equilibrium

$$k = p_{\mathcal{L}} = \pi X. \tag{A.104}$$

Finally, the clearing condition (under the rationing rule as in Guerrieri and Shimer (2014)) of investors' endowment requries

$$(1-\lambda_{\mathcal{H}})p_{\mathcal{H}}\left[\pi s_{H\mathcal{H}} + (1-\pi)s_{L\mathcal{H}}\right] + p_{\mathcal{L}}(1-\pi)s_{\mathcal{L}} = e, \tag{A.105}$$

and equilibrium input capital price is given by

$$1 + \kappa\left[\pi q_H + (1-\pi)q_L\right] = k. \tag{A.106}$$

As such, we obtain a just-identified equation system consisting of Eq. (A.100-A.106) and unknowns $\{p_{\mathcal{H}}, p_{\mathcal{L}}, \lambda_{\mathcal{H}}, x_{\mathcal{H}}, x_{\mathcal{L}}, q_H, q_L, s_{H\mathcal{H}}, s_{L\mathcal{H}}, s_{\mathcal{L}}, k\}$.

As a relevant special case, it is useful to describe how the equilibrium is characterized when the liquid market (the $\mathcal{L}$ market) has zero trading volume in equilibrium. In this case, with only $\mathcal{H}$ has trading, the equilibrium production and trading for both type $\theta \in \{H, L\}$ of originators becomes

$$s_{\theta\mathcal{H}} = q_\theta, \;\text{ and }\; \lambda_{\mathcal{H}} x_\theta + (1-\lambda_{\mathcal{H}})p_{\mathcal{H}} - \rho\lambda_{\mathcal{H}}^2 s_{\theta\mathcal{H}} = k \tag{A.107}$$

The equilibrium asset quality and trading price in $\mathcal{H}$ market are determined by

$$p_{\mathcal{H}} = x_{\mathcal{H}} = \frac{x_H \pi s_{H\mathcal{H}} + x_L(1-\pi)s_{L\mathcal{H}}}{\pi s_{H\mathcal{H}} + (1-\pi)s_{L\mathcal{H}}},, \tag{A.108}$$

and paralleling Eq. (A.105), with trading volume being zero in the $\mathcal{L}$ market we have

$$(1-\lambda_{\mathcal{H}})p_{\mathcal{H}}\left[\pi s_{H\mathcal{H}}+(1-\pi)s_{L\mathcal{H}}\right]=e. \tag{A.109}$$

As such, we have a just-identified equation system consisting of Eq. (A.106-A.109) and unknowns $\{p_{\mathcal{H}}, \lambda_{\mathcal{H}}, x_{\mathcal{H}}, q_H, q_L, s_{H\mathcal{H}}, s_{L\mathcal{H}}, k\}$.

A5.2. ***Analysis for observable seller occupation identity***

In this section, we provide calculations for model analysis under the alternative specification that sellers' occupation identity is observable to the outside investors in the asset market.

Generalized framework with (partially) observable seller occupation identity

To formalize our discussion in Section 5.2, we assume that each individual intermediary is subject to an idiosyncratic "identity observability" shock $id \in \{P, S\}$ in her date 1 trading in the asset market: with probability $1-z$, her selling is executed at same pooling price p_A as those by originators (i.e., $id = P$); and with probability $z \in [0,1]$ her selling is executed at a separate "intermediary" price p_A^I (i.e., $id = S$). Importantly, we assume that intermediaries make their asset purchasing decisions before knowing the realization of shock id.

As such, while the equilibrium characterization for originators' problem stays largely unaffected, there are now two separate contingencies that intermediaries' optimization problem might involve. For an intermediary who has purchased $\{b_{Ij}\}_{j\in\{h,l\}}$ from signal markets on date $\frac{1}{2}$, she solves the optimal trading problem on date 1 contingent on the shock $id \in \{P, S\}$ received. Specifically, if the originator receives a shock $id = P$, she gets to sell her purchased assets at the same price pooled with those sold by originators, in which case she solves the following problem (paralleling Eq. (4) in the benchmark analysis):

$$v_I^P\left(\{b_{Ij}\}\right) \equiv \max_{\left\{r_{Ij}^P\right\}_{j\in\{h,l\}}} \underbrace{\left[\sum_j r_{Ij}^P x_j + p_A \sum_j \left(b_{Ij}-r_{Ij}^P\right) - R_I\left(\sum_j r_{Ij}^P\right)\right]}_{\text{date 1 payoff net of retention cost}} - \underbrace{\sum_j p_j b_{Ij}}_{\text{date } \frac{1}{2} \text{ purchase cost}}.$$

Similarly, in the case where the intermediary receives a shock $id = S$, she solves problem:

$$v_I^S\left(\{b_{Ij}\}\right) \equiv \max_{\left\{r_{Ij}^S\right\}_{j\in\{h,l\}}} \underbrace{\left[\sum_j r_{Ij}^S x_j + p_A^I \sum_j \left(b_{Ij}-r_{Ij}^S\right) - R_I\left(\sum_j r_{Ij}^S\right)\right]}_{\text{date 1 payoff net of retention cost}} - \underbrace{\sum_j p_j b_{Ij}}_{\text{date } \frac{1}{2} \text{ purchase cost}}.$$

In either cases, the intermediary takes as given her asset selling price p_A or p_A^I. On date $\frac{1}{2}$, the intermediary determines her optimal asset purchasing $\{b_{Ij}\}_{j\in\{h,l\}}$ in signal markets by solving

$$v_I \equiv \max_{\{b_{Ij}\}_{j\in\{h,l\}}} z v_I^P \left(\{b_{Ij}\}\right) + (1-z) v_I^S \left(\{b_{Ij}\}\right),$$

before knowing the realization of shock *id*. Accordingly, risk neutral investors set asset market prices as

$$p_A^P = \frac{\sum_\theta x_\theta \left[\sum_j \pi_\theta \left(q_\theta - s_{\theta j} - r_{\theta j}\right)\right] + z \sum_j x_j \left(b_{Ij} - r_{Ij}^P\right)}{\sum_\theta \left[\sum_j \pi_\theta \left(q_\theta - s_{\theta j} - r_{\theta j}\right)\right] + z \sum_j \left(b_{Ij} - r_{Ij}^P\right)} \text{ and } p_A^S = \frac{\sum_j x_j \left(b_{Ij} - r_{Ij}^S\right)}{\sum_j \left(b_{Ij} - r_{Ij}^S\right)}$$

for asset selling that involves pooling among both originators and intermediaries sellers (p_A^P) and that only involves intermediary sellers (p_A^S) respectively.

Proof of Proposition 8

It can be easily shown that perfectly informed intermediation still always strictly improves social welfare. In what follows, we show that unlike in the setting with unobserved seller occupation and pooled asset sale with a common price p_A where uninformed intermediation could potentially hurt social welfare (Proposition 7), now uninformed intermediation always (weakly) improves social welfare.

Similar as in Section 3.3.2, it can be shown that in the tech-irrelevant range both type of originators still sell the same amount in the signal market and thus the signal market asset quality is $x_h = \pi X$. In this case, with separated equilibrium trading prices p_O and p_I in the asset market, a θ-type originators' equilibrium production and trading satisfy

$$\lambda x_\theta + (1-\lambda) p_A - \frac{\lambda^2}{\rho}(q_\theta - s_h) = k, \tag{A.110}$$

and an intermediary's equilibrium optimization now becomes

$$\lambda x_h + (1-\lambda) p_I - \frac{\lambda^2}{\rho_I} b_I = k. \tag{A.111}$$

Bayesian consistency of the equilibrium trading prices in the asset market now becomes

$$p_I = x_h = \pi X, \text{ and } p_O = \frac{\pi X (q_H - s_h)}{\pi (q_H - s_h) + (1-\pi)(q_L - s_h)} \tag{A.112}$$

for intermediaries' selling price p_I and originators' selling price p_O respectively.

First, it can be seen that for intermediaries to be actively operating in this economy (i.e., $b_I > 0$), it must be true that the selling price by intermediaries is strictly higher than the equilibrium input capital price, i.e., $\pi X = p_I > k$. But since $k > 1$, thus it follows that if intermediaries are operating in the economy, we must have $\pi X > 1$.

Now suppose $\pi X > 1$ and intermediaries are actively trading in this economy with uninformed intermediation (such that the equilibrium is located in the tech-irrelevant range), the equilibrium prices satisfy:

$$\pi X\left(\lambda+\frac{\rho_I}{\rho}\right)+(1-\lambda)p_O=\left(1+\frac{\rho_I}{\rho}\right)k+\frac{\lambda^2(k-1)}{\rho\kappa} \tag{A.113}$$

and

$$p_O=\pi X+\frac{\rho\lambda\pi(1-\pi)X^2}{\left(\frac{\lambda^2}{\kappa}+\rho_I\right)(k-1)-\rho_I(\pi X-1)} \tag{A.114}$$

In this equilibrium, the payoff to an asset originator is

$$\begin{aligned}
v_O^u &= \pi\left(q_H-s_h\right)\left[\lambda X+(1-\lambda)p_O-\frac{\lambda^2}{2\rho}\left(q_H-s_h\right)-k\right]+(1-\pi)\left(q_L-s_h\right)\left[(1-\lambda)p_O-\frac{\lambda^2}{2\rho}\left(q_L-s_h\right)-k\right]\\
&= \frac{\rho\pi}{2\lambda^2}\left[\lambda X+(1-\lambda)p_O-k\right]^2+\frac{\rho(1-\pi)}{2\lambda^2}\left[(1-\lambda)p_O-k\right]^2\\
&= \frac{\rho\pi}{2\lambda^2}\left[\frac{\lambda^2(k-1)}{\rho\kappa}+\frac{\rho_I}{\rho}(k-\pi X)-(1-\pi)\lambda X\right]^2+\frac{\rho(1-\pi)}{2\lambda^2}\left[\frac{\lambda^2(k-1)}{\rho\kappa}+\frac{\rho_I}{\rho}(k-\pi X)-\pi\lambda X\right]^2\\
&= \frac{\rho}{2\lambda^2}\left[\left(\frac{\lambda^2}{\rho\kappa}+\frac{\rho_I}{\rho}\right)(k-1)+\frac{\rho_I}{\rho}(1-\pi X)\right]^2+\frac{\rho\pi(1-\pi)X^2}{2}
\end{aligned} \tag{A.115}$$

Using Eq. (A.113) and (A.114), we can get

$$\begin{aligned}
\left[\left(\frac{\lambda^2}{\rho\kappa}+\frac{\rho_I}{\rho}\right)(k-1)+\frac{\rho_I}{\rho}(1-\pi X)\right]^2 &= (1-\lambda)\lambda\pi(1-\pi)X^2\\
&\quad -(k-\pi X)\left[\left(\frac{\lambda^2}{\rho\kappa}+\frac{\rho_I}{\rho}\right)(k-1)+\frac{\rho_I}{\rho}(1-\pi X)\right]
\end{aligned}$$

Plug into Eq. (A.115), we have

$$\begin{aligned}
v_O^u &= \frac{\rho(\pi X-k)}{2\lambda^2}\left[\left(\frac{\lambda^2}{\rho\kappa}+\frac{\rho_I}{\rho}\right)(k-1)+\frac{\rho_I}{\rho}(1-\pi X)\right]+\frac{\rho(1-\lambda)\pi(1-\pi)X^2}{2\lambda}+\frac{\rho\pi(1-\pi)X^2}{2}\\
&= \frac{\pi X-k}{2\lambda^2}\left[\left(\frac{\lambda^2}{\kappa}+\rho_I\right)(k-1)+\rho_I(1-\pi X)\right]+\frac{\rho\pi(1-\pi)X^2}{2\lambda}
\end{aligned} \tag{A.116}$$

The equilibrium payoff to an intermediary in this economy is

$$v_I^u = b_I\left(\pi X-\frac{\lambda^2}{2\rho_I}b_I-k\right)=\frac{\rho_I}{2\lambda^2}\left(\pi X-k\right)^2 \tag{A.117}$$

and the equilibrium payoff to the input capital producer is $v_K^u=\frac{(k-1)^2}{2\kappa}$. Therefore, the social welfare attained in this economy with uninformed intermediation is

$$\begin{aligned}
w^u(\rho_I) &= v_O^u + v_I^u + v_K^u \\
&= \frac{\pi X - k^u}{2\lambda^2}\left[\left(\frac{\lambda^2}{\kappa} + \rho_I\right)(k^u - 1) + \rho_I(1 - \pi X)\right] + \frac{\rho\pi(1-\pi)X^2}{2\lambda} \\
&\quad + \frac{\rho_I}{2\lambda^2}(\pi X - k^u)^2 + \frac{(k^u - 1)^2}{2\kappa} \\
&= \frac{(\pi X - k^u)(k^u - 1)}{2\kappa} + \frac{(k^u - 1)^2}{2\kappa} + \frac{\rho\pi(1-\pi)X^2}{2\lambda} \\
&= \frac{(\pi X - 1)(k^u - 1)}{2\kappa} + \frac{\rho\pi(1-\pi)X^2}{2\lambda}, \qquad \text{(A.118)}
\end{aligned}$$

where the equilibrium input capital price k^u is determined by Eq. (A.113) and (A.114). Similar to our proof for Proposition 7, it can be easily shown that higher ρ_I pushes up equilibrium input capital price, i.e., $\frac{\partial k^u}{\partial \rho_I} > 0, \forall \rho_I > 0$. Therefore, we have

$$\begin{aligned}
w^u(\rho_I) - w^d &= \frac{(\pi X - 1)(k^u - 1)}{2\kappa} + \frac{\rho\pi(1-\pi)X^2}{2\lambda} - \left[\frac{(\pi X - 1)(k^d - 1)}{2\kappa} + \frac{\rho\pi(1-\pi)X^2}{2\lambda}\right] \\
&= \frac{(\pi X - 1)(k^u - k^d)}{2\kappa} \\
&> 0, \qquad \text{(A.119)}
\end{aligned}$$

where the last line uses the fact that $\pi X > 1$ and $k^u(\rho_I) > k^u(0) = k^d$ for any $\rho_I > 0$. Therefore, we can conclude that under observable sellers' occupation identity in the assets market, uninformed intermediation weakly improves social welfare—when intermediaries' technology α is low, they do not operate if $\pi X < 1$; when uninformed intermediaries do operate in the economy, the social welfare is strictly improve. ■

A5.3. ***Calculations for indivisible asset trading***

In this section, we provide calculations for model analysis under the alternative specification that assets produced and traded in the economy is indivisible. In contrast to the partial retention of the assets brought to the asset market for sale in our model specification in the main text, market illiquidity in trading indivisible assets is captured by the occurrence of probabilistic retention of all assets brought in for sale.

For illustration, we analyze the market equilibrium arising in a direct trading economy under indivisible asset trading. Now with indivisible asset trading, for any amount $s_\theta \in [0, q_\theta]$ that a θ-type originator brings to the asset market for sale, with probability $1 - \lambda$ selling is successful and all assets brought in are sold, while with probability λ sale fails and the originator retains all these s_θ amount of assets.

Let us start with a L-type originator, who in equilibrium brings all produced assets to the asset market for sale, i.e., $s_L = q_L$, as under the divisible trading setting. Upon a

L-type realization, an asset originator solves

$$v_L = \max_{q_L} (1-\lambda) \underbrace{p_A^d q_L}_{\text{sale succeeds}} + \lambda \underbrace{[x_L q_L - R(q_L)]}_{\text{sale fails}} - k q_L, \tag{A.120}$$

taking as given the equilibrium prices $\{p_A^d, c\}$. The optimal production q_L chosen by an L-type originator is then determined as

$$(1-\lambda) p_A^d + \lambda \left[x_L - R'(q_L)\right] - k = 0. \tag{A.121}$$

Unlike in the setting with divisible asset trading, now an H-type will voluntarily hold certain amount of produced and refrain from taking them to the asset market for sale. More specifically, an asset originator who gets a H-type realization solves

$$v_H = \max_{q_H, s_H} (1-\lambda) \underbrace{\left[x_H(q_H - s_H) + p_A^d s_H - R(q_H - s_H)\right]}_{\text{sale succeeds}} + \lambda \underbrace{[x_H q_H - R(q_H)]}_{\text{sale fails}} - k q_H, \tag{A.122}$$

in choosing his production and trading strategy $\{q_H, s_H\}$. The optimal sale amount s_H, or equivalently the voluntary retention $q_H - s_H$ chosen by a type H originator who produced q_H on date 0 is then determined as

$$R'(q_H - s_H) = x_H - p_A^d, \tag{A.123}$$

and the optimal production q_H is determined by

$$(1-\lambda) p_A^d + \lambda \left[x_H - R'(q_H)\right] - k = 0. \tag{A.124}$$

Accordingly, Bayesian consistency in the asset market trading implies that the equilibrium selling price p_A^d is

$$p_A^d = \frac{\pi s_H x_H + (1-\pi) q_L x_L}{\pi s_H + (1-\pi) q_L}. \tag{A.125}$$

Under our quadratic functional forms, the optimal production by θ-type as characterized by Eq. (A.121) and (A.124) becomes

$$(1-\lambda) p_A^d + \lambda x_\theta - \lambda \frac{q_\theta}{\rho} = k, \tag{A.126}$$

and the optimal selling by H-type is

$$\frac{1}{\rho}(q_H - s_H) = x_H - p_A^d. \tag{A.127}$$

It can be shown that the equilibrium trading price with indivisible asset trading is higher than that associated with divisible assets. Intuitively, when produced assets are indivisible and retention cost is born on the entire amount of assets brought in once the sale fails, the market-illiquidity based disciplining mechanism is even stronger due to the convexity of retention cost function.