Working Paper Series

Asset Issuance in Over-The-Counter Markets

WP 17-13R

Zachary Bethune
University of Virginia

Bruno Sultanum
Federal Reserve Bank of Richmond

Nicholas Trachter
Federal Reserve Bank of Richmond

FEDERAL RESERVE BANK OF RICHMOND
Richmond ▪ Baltimore ▪ Charlotte

Asset Issuance in Over-the-Counter Markets*

Zachary Bethune
University of Virginia

Bruno Sultanum
Federal Reserve Bank of Richmond

Nicholas Trachter
Federal Reserve Bank of Richmond

November 12, 2018

Working Paper No. 17-13R

Abstract

We model asset issuance in over-the-counter markets. Investors buy newly issued assets in a primary market and trade existing assets in a secondary market, where trade in both markets is over-the-counter (OTC). We show that the level of asset issuance and its efficiency depend on how investors split the surplus in secondary market trade. If buyers get most of the surplus, then sellers do not have incentives to participate in the primary market in order to intermediate assets and the economy has a low level of assets. On the other hand, if sellers get most of the surplus, buyers have strong incentives to participate in the primary market and the economy has a high level of assets. The decentralized equilibrium is inefficient for any splitting rule. The result follows from a double-sided hold-up problem in which it is impossible for all investors to take into account the full social value of an asset when trading. We propose a tax/subsidy scheme and show how it restores efficiency. We also extend the model in several dimensions and study the robustness of the inefficiency result. Finally, we explore the effects of the inefficiency using numerical examples. We study how bargaining power and trading speed in the secondary market affect the efficiency result, and we notice some interesting implications for policy interventions aimed to restore efficiency to OTC markets.

JEL Classification: D53, D82, G14

Keywords: Decentralized markets, bilateral trade, asset issuance, liquidity, hold-up.

*Contact: Zachary Bethune: zab2t@virginia.edu, Bruno Sultanum: bruno@sultanum.com, Nicholas Trachter: trachter@gmail.com. We benefited from comments on earlier drafts at SED 2017, Search and Matching Workshop (Philadelphia), Federal Reserve Bank of Richmond, Vienna Macro Workshop, and RED conference. We thank Maryam Farboodi and Benjamin Lester for particularly useful comments. Any opinions or views expressed in this article are not necessarily those of the Federal Reserve Bank of Richmond or the Federal Reserve System.

1 Introduction

Many assets, both real and financial, are traded in secondary over-the-counter (OTC) markets after their initial issuance (e.g. real estate, municipal bonds, treasuries, asset-backed securities, etc.). Further, many of these markets experienced severe volatility during the 2008 financial crisis, and several policies were enacted that aimed to directly support the issuance of new assets. For example, the Federal Reserve created the Term Asset-Backed Securities Loan Facility (TALF) to support the issuance of asset-backed securities collateralized by different types of private loans, and the Commercial Paper Funding Facility (CPFF) to support the issuance of commercial papers.[1] While there is a large literature studying OTC markets (see Duffie et al. (2005), Duffie et al. (2007), Lagos and Rocheteau (2009), and Hugonnier et al. (2014), among others), most studies in this literature assume a fixed supply of assets, a nonstarter in understanding the effects of policies aimed to spur issuance. In this paper, we study how the trading of seasoned assets in secondary OTC markets affects their primary issuance and, in turn, aggregate asset supply and welfare.

We emphasize two frictions in OTC trade: (i) searching for counterparties to trade, which takes time, and (ii) conditional on a trade opportunity, the terms of the trade are determined by bargaining. We explore the canonical economy of Duffie et al. (2005) but for two differences: (i) we abstract from competitive market-makers in order to make the model more tractable (as in Duffie et al. (2007)), and (ii) we introduce the notion of issuers—agents who have a technology to issue new assets. In the model, trade occurs in pairwise meetings and these meetings are subject to frictions. We interpret meetings between an investor and an issuer as occurring in the primary market since they involve the potential issuance of a new asset. Likewise, we interpret meetings between two investors as occurring in the secondary market as they involve a transfer of a previously issued asset. When two agents meet, either in the primary or secondary market, the terms of trade are determined by Nash bargaining.

We solve for the decentralized equilibrium in the economy and compare it with the constrained efficient allocation, which is the welfare-maximizing allocation constrained by the search frictions. Surprisingly, we find that the decentralized equilibrium allocation is never constrained efficient. This conclusion holds even though trade in the secondary market is constrained efficient when the asset level is fixed, consistent with the literature cited above. Under search and bargaining, the prices at which investors trade in the secondary market do not reflect the social return of assets. When asset supply is fixed, this mispricing is irrelevant in determining the allocation—assets flow from low-valuation agents to high-

[1] See www.newyorkfed.org/markets/funding_archive/index.html for details.

valuation agents—and equilibrium is constrained efficient. When we introduce issuance, this mispricing affects investors' incentives to buy assets in the primary market, which distorts the asset allocation across investors. This distortion of asset allocation further affects the mispricing of assets in the secondary market and in turn affects issuance. In the end, we show that the allocation of assets across investors is inefficient in any decentralized equilibrium—regardless of investors bargaining weights when trading.

The inefficiency we find can be interpreted as the result of a double-sided hold-up problem. A hold-up problem, as first described by Willianson (1975) and Klein et al. (1978), arises when one party must bear the entire cost of an investment while others share in the payoff. In markets with trading frictions, hold-up problems arise often because investments must be made ex-ante, before agents meet. For instance, in monetary search models (e.g., Lagos and Wright (2005); Rocheteau and Wright (2005); Aruoba et al. (2007)) agents acquire money balances before trading with sellers, and in labor search models (e.g. Acemoglu (1996); Masters (1998); Acemoglu and Shimer (1999)) firms or workers invest in capital before negotiating wages.

In our environment, the hold-up problem is two-sided, faced by both buyers and sellers in the secondary market. This occurs as both buyers and sellers must make specific "investments" before trade. Sellers in the secondary market create surplus when they buy assets from issuers and resell them to high valuation investors, or buyers, in the secondary market. That is, these agents create surplus by intermediation. However, because bargaining in the secondary market happens ex-post –after the agent acquires the asset, sellers only receive a share of the gains from trade, resulting in a hold-up problem. In fact, this is the standard hold-up problem. To fix it, the bargaining outcome needs to assign all the trade surplus to the seller in the secondary market.

While sellers in the secondary market create surplus when they buy assets from issuers in the primary market, buyers in the secondary market can destroy surplus when they buy assets from issuers instead of waiting to buy them from sellers in the secondary market. This occurs as, when buying the asset from issuers, this agent is destroying the surplus that could be created by intermediation. Because bargaining in the secondary market happens ex-post, these buyers only receive a share of the gains from trade, resulting again in a hold-up problem. This is a less standard hold-up problem. Here, unlike with more standard hold-up problems, the sunk cost underlying the inefficiency does not come from making an early investment –such as acquiring an asset from issuers, but from *not* making it. Although it looks different, the inefficiency is essentially the same. Fixing this inefficiency requires to assign all the trade surplus to the buyer in the secondary market.

The difference between a one-sided hold-up problem and our two-sided hold-up problem is that the inefficiency that follows from the former can be solved by an appropriate choice of the trade surplus sharing rule, while the inefficiency that follows from the latter cannot be solved by any sharing rule of the trade surplus: While the one-sided hold-up problem requires to assign all trade surplus to sellers, the double-sided hold-up problem that we discuss requires to assign full trade surplus to both buyers and sellers, which is clearly not feasible. In other words, when faced with the double-sided hold-up problem, there is no trade surplus sharing rule that makes the decentralized equilibrium contrained-efficient.

Trade is inefficient for any bargaining rule, however the direction of the inefficiency crucially depends on the way buyers and sellers split the surplus in the secondary market. We show that when the secondary-market sellers have all the bargaining power, investors overvalue assets, and issuance and intermediation are inefficiently high; and, when the secondary-market buyers have all the bargaining power, there are little gains to buy newly created assets in order to resell, and issuance and intermediation are inefficiently low.

Our result provides a rationale for the types of intervention in OTC markets that were observed during the 2008 financial crisis. This rationale is independent of additional frictions, such as private information (see Chang (2017), Chiu and Koeppl (2015), and Bethune et al. (2016), among others). To highlight the role of intervention, we propose a simple government policy that individually corrects the double-sided hold-up problem and decentralizes the constrained efficient solution. Since low-value investors do not fully internalize the gain in intermediating assets, the government subsides their asset holdings. Likewise, since high-value investors do not fully internalize their outside option value of waiting to buy assets in the future, the government taxes their asset holdings. The budget is balanced through lump-sum taxation.

We find it natural to interpret low-valuation investors in our economy as intermediaries, and high-valuation investors as customers. Viewed trough this lens, it is pertinent to ask the efficiency properties of markets where customers do not access primary issuance and have to buy assets trough brokers or underwriters. This alternative market configuration also helps us to better understand the nature of the double-sided hold-up problem. We show that the decentralized equilibrium can be made efficient by an appropriate choice of bargaining power in the secondary market in economies where only low-valuation investors –that is, intermediaries– access the primary market. This follows because, once only the low-type investors access the primary market, the double-sided hold-up problem reduces to a single-sided hold-up problem, which can be handled by an appropriate design of the institutions governing trade.

We explore the effects of the inefficiency using numerical examples. First we study how the magnitude of the inefficiency, or the gains from intervention, depend on the way the surplus is split between buyers and sellers in secondary market trade. We find a U-shaped pattern; the double-sided hold-up problem is the most severe when the gains from trade are shared unevenly in the secondary market. For intermediate values of the bargaining power, the inefficiency is reduced, but does not vanish. We then study how the gains from intervention depend on trading speed in secondary markets. We find that the inefficiency is hump-shaped in trading speed; there is no role for intervention when secondary market trade is shut down or goes to infinity and the inefficiency is most severe in slow markets. However, importantly, we show that the gains from intervention converge slowly to zero as trading speed goes to infinity. Hence policies solely focused on increasing the speed of trade may have limited effects as a result of bilateral trade and bargaining.

The paper is structured as follows. Section 2 introduces the environment. Section 3 defines a decentralized equilibrium and discusses how bargaining determines equilibrium asset allocations. Section 4 describes the constrained efficient benchmark and compares it with the decentralized equilibrium. Section 5 discuss a government intervention to decentralize the constrained efficient outcome. Section 7 provides a numerical exploration of the model. Finally, Section 8 concludes. We provide proofs of all results in the paper in Appendix A.

Literature Following Duffie et al. (2005), the OTC literature has mostly focused on studying trading dynamics in decentralized markets with no meaningful issuance margin and an exogenous supply of assets. For example, Lagos and Rocheteau (2007) and Gârleanu (2009) feature unrestricted asset holdings but leave the aggregate asset supply constant. He and Milbradt (2014) include debt maturity but assume that firms reissue assets to replace maturing debt. Recent work has started to include a meaningful role for asset issuance in an OTC setting. Arseneau et al. (2016) examine similar questions as we do in a three-period model in which assets are created in a primary market subject to a costly state verification problem. Alternatively, we characterize equilibrium dynamics in infinite time in which assets are allocated in a frictional primary market. Geromichalos and Herrenbrueck (2016) also consider an environment with asset issuance and decentralized secondary markets, but their focus is on the determination of liquidity and not on efficiency or policy.

Our efficiency results have a similar flavor to those in the OTC literature that introduce some endogenous extensive margin in trade. For instance, Lagos and Rocheteau (2007) consider an environment in which traders' asset holdings are unrestricted and assets are reallocated between traders through a competitive inter-dealer market.[2] They find that un-

[2] Also see Lagos and Rocheteau (2009) for a similar environment.

der free entry by dealers, the decentralized equilibrium cannot implement the constrained-efficient solution. Efficient entry requires that dealers' bargaining power equal their impact on matching, a la Hosios (1990). However, positive dealer bargaining power leads to a hold-up problem by traders as a result of ex-post bargaining.[3]

Gofman (2011) also considers the efficiency of OTC markets but in an environment in which agents can trade bilaterally according to an exogenous network structure. If the network is complete, in that all traders can trade directly with each other, then the equilibrium is efficient for any set of bargaining powers. However, if the network is incomplete and bargaining powers are strictly between zero and one, a hold-up problem arises. Traders do not internalize the full gain of transferring the asset on valuations further along the network and, as a result, assets may not end up with the highest valuation traders, lowering welfare. Efficiency can be restored for any connected network when sellers possess all the bargaining power. That happens because only sellers face a hold-up problem. We show, however, that introducing asset issuance necessarily introduces an inefficiency that cannot be restored for any set of bargaining powers because both buyers and sellers face a hold-up problem.

Double-sided hold-up problems have been studied in the context of the labor market in which firms and workers make investment decisions before matching and determining wages. Acemoglu (1996) shows that random search and ex-post bargaining lead to social increasing returns in the production technology that create an externality since the gains from trade must be split. This leads to a similar result that efficiency cannot be restored by choosing the right bargaining power. Masters (1998) shows this type of inefficiency always leads to underinvestment in physical and human capital. Alternatively, in the context of our asset market there could be under or overinvestment since there are not generally increasing returns to investment on both sides of the market. One-side of the market tends to underinvest in assets (sellers in the secondary market) and the other side tends to overinvest (buyers in the secondary market). We show that bargaining power has an important role in determining the shape of inefficiency.

Recent work has highlighted how the presence of intermediaries in OTC markets with random search can sometimes lead to inefficiency. Farboodi et al. (2017) endogenize contact rates in Duffie et al. (2005) and also find that the equilibrium is, in general, inefficient. Traders inefficiently invest in contact rates as a result of a search externality: they do not internalize that increasing their contact rate affects the distribution of other traders' contacts. A Pigouvian tax that charges traders when they make contact and uses the revenue

[3]The hold-up problem also arises in Lagos et al. (2011), who study dynamic equilibria in which dealers can hold inventories.

to supplement the gains from trade decentralizes the Pareto optimum. Our efficiency result is similar in that the planner would like to increase the size of the surplus in any trade, however the tax/transfer scheme in Farboodi et al. (2017) only achieves the optimum in the case with symmetric bargaining weights. Further, Farboodi et al. (2017) do not consider issuance or endogenous asset supply.

In Menzio et al. (2016), traders meet randomly but differ with respect to their ability to commit to take-it-or-leave-it offers, and as a result, their bargaining power. If types, or bargaining powers, are exogenous, then equilibrium is efficient since (i) bilateral trade is efficient and (ii) all traders will meet each other, almost surely.[4] However, if bargaining types are endogenous and commitment requires a sunk cost, equilibrium is inefficient.

Also related to our work, Nosal et al. (2014) and Nosal et al. (2016) study an environment with endogenous intermediaries where efficiency only arises if bargaining weights satisfy a version of the Hosios (1990) condition.

2 Environment

Time is continuous and goes from zero to infinity. There are two types of agents: issuers and investors. All agents are infinitely lived, risk-neutral, and discount the future at rate $r > 0$. Investors have measure two and issuers have measure one. These choice of measures simplify the notation but are not necessary for our results. There are objects called assets that issuers issue (or produce) and investors value for their flow of a good called dividends. An asset matures with Poisson arrival rate $\mu > 0$ and pays a unit flow of dividends until maturity. The asset disappears upon maturity and has no terminal payment. Dividends are not tradable and investors must hold an asset to consume its dividends.

Each issuer has to pay an issuance cost to issue an asset, and they do not value asset dividends. The issuance cost is heterogeneous among issuers. It follows a uniform distribution, with density $g(c) = \frac{1}{\bar{c}-\underline{c}}$ and cumulative distribution $G(c) = \frac{c-\underline{c}}{\bar{c}-\underline{c}}$ in the interval $[\underline{c}, \bar{c}]$. Since issuers do not value dividends, they do not hold assets in equilibrium.

Investors are one of two types, low or high, and types are fixed over time.[5] Half of the investors are low type and the remaining half are high type. An investor's type is associated with their utility from consuming dividends. Low-type investors have utility

[4]Related to Gofman (2011), in a random search environment with a fixed supply of assets, the trading network is almost surely complete and equilibrium is efficient even if bargaining powers are inside the unit interval.

[5] Much of the OTC literature following Duffie et al. (2005) and Lagos and Rocheteau (2007) uses preference shocks to a steady-state with trade in the secondary market. In our setting, although preference are time-invariant, a steady-state with trade occurs as a result of asset maturity and asset creation. The mismatch between these two forces ensures that there is trade in steady-state.

$\nu_l \geq 0$, while high-type investors have utility $\nu_h > \nu_l$. In Appendix B, we allow for a continuum of investor types and show that our main results are robust to this extension. Investors' asset holdings are discrete, either zero or one. We call investors holding an asset *owners* and those not holding an asset *non owners*. Let $\phi_l^0 \in [0,1]$ and $\phi_h^0 \in [0,1]$ denote the measures of low and high-type owner investors that are given in period $t = 0$.

Assumption 1. *(i) $\nu_l/(r+\mu) = \underline{c}$, and (ii) $\nu_h/(r+\mu) \leq \bar{c}$.*

We impose Assumption 1 throughout the paper. Part (i) implies that low-type investors only hold assets if they profit from reselling the asset in the secondary market.[6] To see this, notice that $\nu_l/(r+\mu)$ is the discounted present value that a low-type investor would obtain from holding an asset until maturity. Because this discounted flow equals the lowest production cost, $\underline{c}$, low-type investors never find it profitable to buy the asset only to consume its dividends until maturity. This assumption permits us to isolate the intermediation channel in the model and thus allows us to study how the gains from intermediation shape asset issuance and allocations. Part (ii) guarantees that the allocation is interior, which allows us to take derivatives when needed.

Issuers contact investors at random with Poisson arrival rate $2\lambda_p > 0$, and investors contact other investors at random with Poisson arrival rate $\lambda_s/2 > 0$. Note that an investor expects to meet with an issuer at Poisson arrival rate λ_p; an issuer contacts an investor at rate $2\lambda_p$, and those contacts are distributed among the measure two of investors. Note also that an investor expects to meet with another investor at Poisson arrival rate λ_s; at rate $\lambda_s/2$ he contacts another investor, and at rate $\lambda_s/2$ another investor contacts him. We call the market in which issuers sell newly issued assets to investors *the primary market*, and the market in which investors sell existing assets to each other *the secondary market*. Meetings are bilateral, and utility is transferable across investors and issuers.

To facilitate the analysis here, it is useful to anticipate some equilibrium trading patterns. First, low-type investors sell assets to high-type investors and never the other way around. Thus, low-type investors act as intermediaries. Second, if an issuer with cost c issues to a particular investor—either of high or low type—then all issuers with cost below c also issue to this particular investor. As a result, when defining an allocation, we define the issuance policy only in terms of issuance thresholds c_l and c_h. We later verify that these restrictions are without loss of generality in terms of characterizing decentralized equilibria and efficient allocations in our economy.

An asset allocation and issuance policy define an allocation of the economy. An asset allocation is a map $\boldsymbol{\phi} = \{\phi_l(t), \phi_h(t)\}_t$, where $\phi_l(t)$ and $\phi_h(t)$ denote the measure of low-

[6] Low-type investors resemble " flippers" as described in Green et al. (2007a).

and high-type owner investors at time t, respectively. An issuance policy is a map $\boldsymbol{c} = \{c_l(t), c_h(t)\}_t$, where $c_l(t)$ and $c_h(t)$ denote thresholds for the issuance cost such that issuers with cost below $c_l(t)$ and $c_h(t)$ issue to low- and high-type non owner investors in time t meetings. We drop the argument t from an allocation in order to keep the notation simple whenever it does not cause confusion.

An issuance policy $\boldsymbol{c}$ implements an asset allocation $\boldsymbol{\phi}$ if the measures of owner investors, ϕ_l and ϕ_h, are consistent with the issuance thresholds, c_l and c_h, the trade pattern in the secondary market, and the initial asset holdings ϕ_l^0 and ϕ_h^0. That is, the law of motion for ϕ_l and ϕ_h solve the system of differential equations

$$\dot{\phi}_l = \lambda_p G(c_l)(1-\phi_l) - \mu\phi_l - \lambda_s\phi_l(1-\phi_h) \quad \text{and} \tag{1}$$

$$\dot{\phi}_h = \lambda_p G(c_h)(1-\phi_h) - \mu\phi_h + \lambda_s\phi_l(1-\phi_h), \tag{2}$$

given the initial conditions $\phi_l(0) = \phi_l^0$ and $\phi_h(0) = \phi_h^0$. We say that an asset allocation $\boldsymbol{\phi}$ is feasible if there is an issuance policy $\boldsymbol{c}$ that implements $\boldsymbol{\phi}$.

Equation (1) describes the law of motion for low-type owner investors at time t. On the right hand side, the first term captures the inflow of low-type non owner investors that meet issuers with cost below c_l and thus buy assets from these issuers. The second term captures the outflow of low-type owners due to asset maturity. The third term captures the outflow of low-type owners due to the fact that they meet high-type non owners and sell their assets to them. The law of motion for high-type owner investors described in (2) follows similarly, except that the last term represents the inflow from trade between investors. This difference between the asset flow equations for low- and high-type investors follows from the different roles played by both types of investors in the secondary market: low-type investors are sellers, while high-type investors are buyers.

3 Decentralized equilibrium

We characterize a decentralized equilibrium in terms of investors' reservation value. The reservation value is the absolute change in value function—that is, in expected present value of utility—induced by acquiring or giving up an asset. Let $V_l(q)$ denote the value function of low-type investors, where $q \in \{0, 1\}$ denotes the asset holdings of the investor. Then $\Delta_l = V_l(1) - V_l(0)$ is the reservation value of low-type investors. Analogously, let $V_h(q)$ denote the value function of high-type investors. Then $\Delta_h = V_h(1) - V_h(0)$ is the reservation value of high-type investors.

Nash bargaining determines trade and price in the primary market, where we assume that buyers—the investors—have all the bargaining power. We make this assumption for two reasons. First, abstracting from general equilibrium effects, the fact that buyers have full bargaining power makes trade in the primary market efficient. Second, and more importantly, the assumption simplifies the environment, and has no qualitative implications for our results regarding the inefficiency of decentralized equilibrium.[7] Low- and high-type non owner investors buy an asset if their reservation value is greater than the issuer's issuance cost. Because buyers hold all the bargaining power, the price equals the issuance cost. These trade outcomes imply the trade pattern in the primary market we anticipated earlier. Namely, if an issuer with cost c issues to a particular investor, either high or low type, then all issuers with cost below c must also issue to this particular investor. The reservation values, Δ_l and Δ_h, specify the thresholds for asset issuance.

Nash bargaining also determines trade and price in the secondary market, where buyers have bargaining power $\theta \in [0, 1]$ and sellers have bargaining power $1 - \theta$. To maximize the surplus in a trade, as required by Nash bargaining, the investor with the higher reservation value buys the asset from the investor with the lower reservation value. We anticipate a trade pattern where low-type investors sell assets to high-type investors, as discussed earlier in the paper. To obtain this pattern, we conjecture that the reservation value of high-type investors is strictly higher than the one of low-type investors—that is $\Delta_h > \Delta_l$. We verify this inequality later using equilibrium equations. Denote by x the price a non owner high-type investor pays to buy an asset from a low-type owner investor. By Nash bargaining, x maximizes $(x - \Delta_l)^{1-\theta}(\Delta_h - x)^{\theta}$, which implies $x = \Delta_h - \theta[\Delta_h - \Delta_l]$.

Given these outcomes in bilateral trade, the value functions for low-type owner and non owner investors satisfy

$$rV_l(0) = \dot{V}_l(0) + \lambda_p \int_{\underline{c}}^{\Delta_l} (\Delta_l - c) g(c) dc \quad \text{and} \tag{3}$$

$$rV_l(1) = \dot{V}_l(1) + \nu_l - \mu\Delta_l + \lambda_s(1 - \phi_h)(1 - \theta)(\Delta_h - \Delta_l). \tag{4}$$

Equation (3) describes the law of motion for the value function of low-type non owner investors. The first term is the change in utility at a point in time and the second term is the expected gain in utility from meeting and purchasing an asset from an issuer in the primary market. Equation (4) describes the law of motion for the value function of low-type owner investors. The first term is the change in utility at a point in time, the second term is the utility flow from holding the asset, the third term is the expected loss in utility due to asset maturity, and the last term is the gain in utility from meeting and selling an

[7] We can provide a formal proof upon request.

asset to a high-type non owner investor in the secondary market.

Similarly, the value function of high-type owner and non owner investors satisfy

$$rV_h(0) = \dot{V}_h(0) + \lambda_p \int_{\underline{c}}^{\Delta_h} (\Delta_h - c) g(c) dc + \lambda_s \phi_l \theta (\Delta_h - \Delta_l) \quad \text{and} \tag{5}$$

$$rV_h(1) = \dot{V}_h(1) + \nu_h - \mu \Delta_h \, . \tag{6}$$

Equation (5) describes the law of motion for the value function of high-type non owner investors. The first term is the change in utility at a point in time, the second term is the expected gain in utility from meeting and purchasing an asset from an issuer in the primary market, and the last term is the expected gain in utility from meeting and purchasing an asset from a low-type owner investor in the secondary market. Equation (6) describes the law of motion for the value function of high-type owner investors. The first term is the change in utility at a point in time, the second term is the utility flow from holding the asset, and the last term is the expected loss in utility due to asset maturity.

We obtain a differential equation for the reservation value of low-type investors by taking the difference between equations (3) and (4), and for the reservation value of high-type investors by taking the difference between equations (5) and (6). The reservation values of low- and high-type investors satisfy

$$(r+\mu)\Delta_l = \dot{\Delta}_l + \nu_l - \lambda_p \int_{\underline{c}}^{\Delta_l} (\Delta_l - c) g(c) dc + \lambda_s (1-\phi_h)(1-\theta)(\Delta_h - \Delta_l) \quad \text{and} \tag{7}$$

$$(r+\mu)\Delta_h = \dot{\Delta}_h + \nu_h - \lambda_p \int_{\underline{c}}^{\Delta_h} (\Delta_h - c) g(c) dc - \lambda_s \phi_l \theta (\Delta_h - \Delta_l). \tag{8}$$

Investors' reservation values can be decomposed into three components: a fundamental value and two option values. To understand this decomposition, consider the reservation value equations (7) and (8) in steady state:

$$\Delta_l = \frac{\nu_l}{r+\mu} - \lambda_p \frac{\int_{\underline{c}}^{\Delta_l} (\Delta_l - c) g(c) dc}{r+\mu} + \lambda_s \frac{(1-\phi_h)(1-\theta)(\Delta_h - \Delta_l)}{r+\mu} \quad \text{and}$$

$$\Delta_h = \frac{\nu_h}{r+\mu} - \lambda_p \frac{\int_{\underline{c}}^{\Delta_h} (\Delta_h - c) g(c) dc}{r+\mu} - \lambda_s \frac{\phi_l \theta (\Delta_h - \Delta_l)}{r+\mu}.$$

The first term in the first equation represents the fundamental value of the asset to low-type investors—the expected discounted utility flow from consuming the dividend, ν_l, until maturity. If there was no trade (for instance, if λ_p and λ_s were equal to zero), the fundamental value alone would determine the reservation value Δ_l. However, since there is trade, the options of buying and selling assets in the future play an important role in determining the reservation value. Consider the reservation value for low-type investors. When purchasing an asset, they lose the option of waiting to purchase the asset later in the primary market but gain the option of selling the asset in the secondary market to a high-type investor. These are represented as the second and third terms in the first equation above. Similarly, when high-type investors purchase an asset, they lose the option of waiting to purchase the asset later in the primary market and also lose the option of waiting to purchase the asset later in the secondary market. These are represented as the second and third terms in the second equation above.

Without loss of generality, we define a decentralized equilibrium in terms of reservation values instead of value functions. For any pair of reservation values, the differential equations (3)-(6) yield the associated value functions. Additionally, we do not include the value function of issuers in our equilibrium definition since it is always zero because they have no bargaining power and are paid their cost when issuing an asset.

Definition 1. *A decentralized equilibrium is an asset allocation and reservation values for investors,* $\{\boldsymbol{\phi}, \boldsymbol{\Delta}\} = \{\phi_l, \phi_h, \Delta_l, \Delta_h\}$*, that solve the differential equations*

$$(r+\mu)\Delta_l = \dot{\Delta}_l + \nu_l - \lambda_p \int_{\underline{c}}^{\Delta_l} (\Delta_l - c) g(c) dc + \lambda_s (1-\phi_h)(1-\theta)(\Delta_h - \Delta_l) \tag{9}$$

$$\dot{\phi}_l = \lambda_p (1-\phi_l) G(\Delta_l) - \mu \phi_l - \lambda_s \phi_l (1-\phi_h) \tag{10}$$

$$(r+\mu)\Delta_h = \dot{\Delta}_h + \nu_h - \lambda_p \int_{\underline{c}}^{\Delta_h} (\Delta_h - c) g(c) dc - \lambda_s \phi_l \theta (\Delta_h - \Delta_l) \tag{11}$$

$$\dot{\phi}_h = \lambda_p (1-\phi_h) G(\Delta_h) - \mu \phi_h + \lambda_s \phi_l (1-\phi_h) \tag{12}$$

with initial conditions $\phi_l(0) = \phi_l^0$ *and* $\phi_h(0) = \phi_h^0$*.*

Notice that the reservation values Δ_l, Δ_h are bounded as the value functions $V_l(0)$, $V_l(1)$ and $V_h(0)$, $V_h(1)$ are bounded. The value functions for the high-type investor are bounded because they (i) must be below the value function for a high-type investor that it is holding the asset with no maturity, and (ii) must be above zero, as they can always discard the asset. That is, $0 \leq V_h(0) < V_h(1) < \nu_h/r$. As a result, Δ_h is bounded. Likewise, for the low-type investor we have that $0 \leq V_l(0) < V_l(1) < \nu_h/r$ and thus Δ_l is bounded.

A decentralized equilibrium is at a steady state when the asset allocation and reservation value of investors are constant.

Definition 2. *A decentralized steady-state equilibrium is an asset allocation and bounded reservation values for dealers and investors,* $\{\boldsymbol{\phi}, \boldsymbol{\Delta}\} = \{\phi_l, \phi_h, \Delta_l, \Delta_h\}$*, that solve the system of equations* (9)-(12) *with time derivative* $\dot{\phi}_l = \dot{\phi}_h = \dot{\Delta}_l = \dot{\Delta}_h = 0$ *for all* t*.*

The existence of a decentralized equilibrium (in and out of steady state) follows from standard methods used to solve non linear differential equations.[8]

Finally, using (7) and (8), Lemma 1 verifies that the reservation value of high-type investors is strictly higher than the reservation value of low-type investors—that is $\Delta_h > \Delta_l$.

Lemma 1. *The reservation value of high-type investors is strictly higher than the reservation value of low-type investors at any point in time. That is, $\Delta_h(t) > \Delta_l(t)$ for all t.*

High-type investors enjoy a higher flow utility from consuming dividends, $\nu_h > \nu_l$. As a result, the fundamental value component of the reservation value for the high-type investors is higher than that for the low-type investors. If reservation values are not consistent with the fundamental component, that is if $\Delta_h \leq \Delta_l$, then they must be consistent with beliefs about future reservation values. Low-type investors must believe that their reservation value will keep increasing. This generates an explosive path for the reservation value that is inconsistent with bounded payoffs, a contradiction that implies that reservation values are consistent with their fundamental-value component.[9]

3.1 Bargaining power and equilibrium asset allocations

Bargaining power in the secondary market plays an important role in shaping the incentives of investors to buy or sell assets, which makes it key in determining equilibrium asset allocations. This is not clear from the existing literature. For example, in the standard DGP model, the equilibrium asset allocation is independent of the bargaining power—which determines only transfers of utils. In our model, the equilibrium asset allocation is a function of the bargaining power. In this section, we study this function.

We compare three state-steady economies: the sellers' economy, the buyers' economy, and the interior economy. The three economies are the same in all primitives except for the bargaining power of buyers and sellers in the secondary market. In the sellers' economy, sellers have all the bargaining power, that is $\theta = 0$. In the buyers' economy, buyers have all the bargaining power, that is $\theta = 1$. In the interior economy, neither sellers or buyers have all the bargaining power, that is $0 < \theta < 1$. We order the economies from A to C according to the bargaining power, that is $0 = \theta^A < \theta^B < \theta^C = 1$.

[8] We omit an existence proof here but we provide it upon request.

[9] Lemma 1 confirms our assumption that $\Delta_h > \Delta_l$; however, it does not show that an alternative equilibrium with $\Delta_h \leq \Delta_l$ does not exist. That is, if we assume $\Delta_h \leq \Delta_l$, the reservation value equations (7) and (8) would be different because Nash bargaining would imply trade in the opposite direction—-low-type investors buying assets from high-type investors. These new equations could be consistent with $\Delta_h \leq \Delta_l$, however we show that this is not the case. The trade pattern associated with $\Delta_h < \Delta_l$ implies that $\Delta_h > \Delta_l$, which is a contradiction, and $\Delta_h = \Delta_l$ can easily be ruled out because $\nu_l \neq \nu_h$. We discuss these results in the Appendix with the proof of Lemma 1.

Proposition 1. *If the asset allocations and reservation values* $\{\phi_l^A, \phi_h^A, \Delta_l^A, \Delta_h^A\}$, $\{\phi_l^B, \phi_h^B, \Delta_l^B, \Delta_h^B\}$, *and* $\{\phi_l^C, \phi_h^C, \Delta_l^C, \Delta_h^C\}$ *are associated with steady-state decentralized equilibria for the sellers' economy, the interior economy, and the buyers' economy, then*

(i) $\phi_l^A > \phi_l^B > \phi_l^C = 0$,

(ii) $\phi_h^A > \phi_h^B > \phi_h^C$,

(iii) $\Delta_l^A > \Delta_l^B > \Delta_l^C = \frac{v_l}{\mu+r}$, *and*

(iv) $\Delta_h^A = \Delta_h^C > \Delta_h^B$.

Part (i) and (ii) of Proposition 1 provide that low- and high-type investors hold the least amount of assets when buyers possess all the bargaining power and hold the most when sellers possess all the bargaining power. An immediate implication is that aggregate asset supply, $\phi_l + \phi_h$, follows a similar pattern. These results follow by noticing that asset holdings move in the same direction as the reservation value of low-type investors (part (iii) of the proposition). Low-type investors serve as natural intermediates—buying assets from issuers and selling them to high-type investors. In the buyers' economy, low-type investors have a low reservation value because they do not get any of the gains from trade when selling assets. As a result, low-type investors have no incentive to intermediate asset issuance and hold no assets, which reduces the buying options of high-type investors and leads them to hold less assets in equilibrium. The sellers' economy features the opposite pattern. Low-type investors have a high reservation value because they obtain all the gains from trade when selling assets in the secondary market. As a result, low-type investors have higher incentives to intermediate assets, increasing the option value of buying assets for high-type investors and leading them to hold more assets in equilibrium. Finally, part (iv) of the proposition provides that the reservation value of high-type investors is maximized for bargaining powers in the extremes, $\theta \in \{0, 1\}$. This follows because when $\theta = 0$ or $\theta = 1$ the option value of buying seasoned assets in the secondary market is zero for high-type investors, and positive option values reduce the value of holding, and thus acquiring, an asset. When $\theta = 0$, the option value is zero because all of the gains from trade are captured by the seller—the low-type investor. When $\theta = 1$, the option value is zero because there is no trade of seasoned assets as low-type investors have no incentives to intermediate.

4 Efficient asset allocation

We now turn to solving for the constrained-efficient allocation, constrained in the sense that the allocation has to satisfy the search frictions of the economy. We label the problem of finding a constrained-efficient asset allocation the planner's problem. The planner is allowed to choose the issuance policy in the primary market and to choose the trade pattern in the secondary market. As we did with the decentralized equilibrium, we anticipate a few results. First, when an owner low-type investor meets a non owner high-type investor, the planner will transfer the asset from the low-type investor to the high-type investor. Second, when an owner high-type investor meets a non owner low-type investor, the planner will not transfer the asset from the high-type investor to the low-type investor.

An asset allocation, $\boldsymbol{\phi}$, is constrained-efficient if there is an issuance policy, $\boldsymbol{c}$, such that $\boldsymbol{\phi}$ and $\boldsymbol{c}$ maximize aggregate utility,

$$\int_0^\infty e^{-rt}\left\{\phi_l\nu_l+\phi_h\nu_h-\lambda_p\left[(1-\phi_l)\int_{\underline{c}}^{c_l}cg(c)dc+(1-\phi_h)\int_{\underline{c}}^{c_h}cg(c)dc\right]\right\}dt, \tag{13}$$

subject to the feasibility conditions (1) and (2).[10] Lemma 2 provides the first order conditions that are necessary for an asset allocation to be constrained-efficient.

Lemma 2. *If an asset allocation $\boldsymbol{\phi}$ is constrained-efficient, then there exist co-state variables γ_l, $\gamma_h \geq 0$ such that γ_l, γ_h, ϕ_l, and ϕ_h solve the system of differential equations given by*

$$r\gamma_l=\dot{\gamma}_l+\nu_l-\mu\gamma_l-\lambda_p\int_{\underline{c}}^{\gamma_l}(\gamma_l-c)g(c)dc+\lambda_s(1-\phi_h)(\gamma_h-\gamma_l) \tag{14}$$

$$\dot{\phi}_l=\lambda_p(1-\phi_l)G(\gamma_l)-\mu\phi_l-\lambda_s\phi_l(1-\phi_h) \tag{15}$$

$$r\gamma_h=\dot{\gamma}_h+\nu_h-\mu\gamma_h-\lambda_p\int_{\underline{c}}^{\gamma_h}(\gamma_h-c)g(c)dc-\lambda_s\phi_l(\gamma_h-\gamma_l) \tag{16}$$

$$\dot{\phi}_h=\lambda_p(1-\phi_h)G(\gamma_h)-\mu\phi_h+\lambda_s\phi_l(1-\phi_h) \tag{17}$$

with boundary conditions $\phi_l(0)=\phi_l^0$, $\phi_h(0)=\phi_h^0$, and $\lim e^{-rt}\gamma_l=\lim e^{-rt}\gamma_h=0$.

The co-state variables, γ_l and γ_h, represent the social value of having low- and high-type investors holding an asset—we call them the social value of a low- and high-type investor. The social value of an investor is analogous to their reservation value, only from the standpoint of aggregate welfare. First, the planner understands that there is a fundamental value in giving an asset to an investor coming from the dividend valuations, ν_l and

[10] Note that in this definition we give Pareto weight one to each agent. This is without loss of generality because the model has transferable utility—differences in Pareto weights determine transfers but do not distort the asset allocation.

ν_h. The planner also understands that there are two option values for each type of investor. For low-type investors, there is the option value of buying the asset later in the primary market, $\lambda_p \int_0^{\gamma_l} (\gamma_l - c) g(c) dc$, which he loses by acquiring an asset, and the option value of transferring the asset later to a high-type investor, $\lambda_s(1-\phi_h)(\gamma_h - \gamma_l)$, which he gains by acquiring an asset. For high-type investors, there is the option value of buying the asset later in the primary market, $\lambda_p \int_0^{\gamma_h} (\gamma_h - c) g(c) dc$, which he loses by acquiring an asset, and the option value of buying the asset later from a low-type investor, $\lambda_s \phi_l (\gamma_h - \gamma_l)$, which he also loses by acquiring an asset.

There is a key difference between the social values in (14) and (16) and private reservation values in (9) and (11). When the planner evaluates the social value of allocating an asset to a low-type investor, he takes into account that the low-type investor will, with some probability, transfer the asset to a high-type investor. In doing so, the planner takes into account the entire surplus generated by transfering an asset from a low-type investor to a high-type investor, or the entire surplus from trade in the secondary market. When the low-type investor evaluates his reservation value, however, he takes into account only a fraction $1-\theta$, his bargaining power, of this surplus. The reason is because he faces a hold-up problem—his decision to invest in buying an asset occurs before meeting with a buyer for the asset in the secondary market. As usual in hold-up problems, the only way that the equations for the social value and reservation value of a low-type investor coincide is if the low-type investor has all the bargaining power when selling an asset. That is, $1-\theta = 1$.

Similarly, when the planner evaluates the social value of a high-type investor, he takes into account that the high-type investor will, with some probability, meet with a low-type owner investor in the future in which he could have bought the asset from. In doing so, the planner takes into account the entire loss of surplus generated by passing an asset from a low-type investor to a high-type investor. When the high-type investor evaluates his reservation value, he takes into account only a fraction θ, his bargaining power, of this surplus. The high-type investors also faces a hold-up problem—his decision not to invest in buying an asset occurs before meeting with a seller in the secondary market. The only way that the equations for the social value and reservation value of a high-type investor coincide is if the high-type investor has all the bargaining power when buying an asset. That is, $\theta = 1$.

Fixing together, by an appropriate choice of the bargaining power θ, the hold-up problem of both low- and high-type investors is not possible, as this would require that both buyers and sellers have all the surplus generated by a trade. Investors will never value the gains from trade in the same way the planner does, and as a result, the outcome of a decentralized equilibrium cannot replicate the planner's solution—no matter how investors bargain over gains from trade. We prove Proposition 2 with a formal version of this argument.

Proposition 2. *A decentralized-equilibrium asset allocation is never efficient.*

The way in which investors split the surplus when trading is irrelevant in concluding that decentralized trade is inefficient. For any surplus splitting rule, investors cannot both fully internalize the social value of trade that leads to inefficiency. However, the surplus splitting rule does matter in determining the direction the equilibrium allocation is distorted. To illustrate this, consider again the sellers' and buyers' economy associated with bargaining powers $\theta^A = 0$ and $\theta^C = 1$ from Section 3.1. In the following proposition, we show how the bargaining power determines equilibrium allocations.

Proposition 3. *If the asset allocation and reservation values $\{\phi_l^A, \phi_h^A, \Delta_l^A, \Delta_h^A\}$ are associated with a steady-state decentralized equilibria for the sellers' economy, $\{\phi_l^C, \phi_h^C, \Delta_l^C, \Delta_h^C\}$ are associated with a steady-state decentralized equilibria for the buyers' economy, and the asset allocation (ϕ_l^*, ϕ_h^*) is efficient, in steady state, and has social values γ_l and γ_h, then*

(i) $\phi_l^A > \phi_l^* > \phi_l^C = 0$,

(ii) $\phi_h^A > \phi_h^* > \phi_h^C$,

(iii) $\Delta_l^A > \gamma_l > \Delta_l^C = \frac{\nu_l}{\mu+r}$, *and*

(iv) $\Delta_h^A = \Delta_h^C > \gamma_h$.

In the sellers' economy, high-type investors do not internalize the option value of buying assets in the secondary market because they gain no surplus. As a result, they overvalue purchasing assets through the primary market and $\Delta_h^A > \gamma_h$. Low-type investors also overvalue assets since the secondary market surplus reflects the over valuation of high-type investors, $\Delta_l^A > \gamma_l$. In equilibrium, there is over issuance and the asset supply is too high, $\phi_l^A + \phi_h^A > \phi_l^* + \phi_h^*$. The opposite is true in the buyers' economy. In this case, low-type investors do not gain any surplus from reselling and, since $\underline{c} = \nu_l/(r+\mu)$, they do not hold assets, $\phi_l^C = 0$ and $\Delta_l^C < \gamma_l$. Since there is no secondary market trade, high-type investors have no secondary market option value, which again implies they overvalue primary issuance, $\Delta_h^C > \gamma_h$. In equilibrium, there is under issuance and the asset supply is too low, $\phi_l^C + \phi_h^C < \phi_l^* + \phi_h^*$.

Two comments are relevant before concluding this section. The first comment is that the inefficiency result is independent of the type of bilateral bargaining protocol that we use in the model. What is key for the result is that buyers and sellers in a meeting need to make costly actions prior to the meeting, thus allowing a double hold-up problem to arise.[11]

The second comment pertains to the way we split the surplus in the primary market. We assigned full bargaining power to buyers in the primary market for two reasons. Assigning the full bargaining power to the buyer is optimal if we abstract from the secondary market, and the inefficiency result survives even if we allow the splitting rule in the primary market to differ. Because of these reasons, we decided to abstract from having a generic splitting rule in the primary market to keep the model as simple as possible. This allows us to focus on how the way trade surpluses are split in the secondary market between buyers and sellers affects the economy and generates the inefficiency we discuss in Proposition 2. Still, a proof of the inefficiency result with a generic splitting rule in the primary market is available for the interested reader.

5 A government intervention

Decentralized trade in the secondary market with ex-post bargaining necessarily leads to wedges between the investors' reservation values and the social values coming from investors not internalizing the full surplus generated by their trade. To correct for these wedges, we propose a simple tax-subsidy government intervention. Since low-type investors, when they buy an asset, do not internalize the full gain in the option value of selling later in the secondary market, we propose subsidizing their asset holdings. Since high-type investors, when they buy an asset, do not internalize the full loss in the option value of buying later in the secondary market, we propose a tax to their asset holdings. We achieve a balanced budget through lump-sum taxation. We show that this simple policy fully corrects for the double hold-up problem and achieves efficiency.

Formally, a government intervention, or just an intervention to keep it simple, is a triple $\boldsymbol{\tau} = \{\tau_l(t), \tau_h(t), \bar{\tau}(t)\}_t$, where τ_l is a subsidy on asset holdings of low-type investors, τ_h is a tax on asset holdings of high-type investors, and $\bar{\tau}$ is a lump-sum tax on all investors.

[11] Note, however, that a more general mechanism or trading protocol could, in principle, approximate the efficient outcome. For instance, if trades and transfers are history dependent, efficiency could arise due to a form of folk theorem which holds in this environment. Alternatively, price posting and directed search have been shown to solve double-sided hold-up problems in environments with constant returns to scale matching (which is not true in the benchmark OTC environment we use). We do not allow for more general mechanisms or trading protocols because we interpret these markets as spot trading and take seriously the notion that investors are limited in their information about potential trading partners, as is broadly considered the case in the OTC literature.

Given an asset allocation $\boldsymbol{\phi}$, an intervention $\boldsymbol{\tau} = \{\tau_l(t), \tau_h(t), \bar{\tau}(t)\}_t$ is feasible if

$$\int_0^\infty e^{-rt}[2\bar{\tau}(t) + \phi_h \tau_h(t) - \phi_l \tau_l(t)]dt \geq 0. \tag{18}$$

We adjust the decentralized equilibrium equations to account for the intervention and get the following equilibrium definition.

Definition 3. *A decentralized equilibrium with intervention $\boldsymbol{\tau}$ is an asset allocation and bounded reservation values for investors, $\{\boldsymbol{\phi}, \boldsymbol{\Delta}\} = \{\phi_l, \phi_h, \Delta_l, \Delta_h\}$, that solve the equations*

$$(r+\mu)\Delta_l = \dot{\Delta}_l + \nu_l + \tau_l - \lambda_p \int_{\underline{c}}^{\Delta_l} (\Delta_l - c) g(c) dc + \lambda_s (1-\phi_h)(1-\theta)(\Delta_h - \Delta_l) \tag{19}$$

$$\dot{\phi}_l = \lambda_p (1-\phi_l) G(\Delta_l) - \mu \phi_l - \lambda_s \phi_l (1-\phi_h) \tag{20}$$

$$(r+\mu)\Delta_h = \dot{\Delta}_h + \nu_h - \tau_h - \lambda_p \int_{\underline{c}}^{\Delta_h} (\Delta_h - c) g(c) dc - \lambda_s \phi_l \theta (\Delta_h - \Delta_l) \tag{21}$$

$$\dot{\phi}_h = \lambda_p (1-\phi_h) G(\Delta_h) - \mu \phi_h + \lambda_s \phi_l (1-\phi_h) \tag{22}$$

$$\int_0^\infty e^{-rt}[2\bar{\tau}(t) + \phi_h \tau_h(t) - \phi_l \tau_l(t)]dt = 0 \tag{23}$$

with initial conditions $\phi_l(0) = \phi_l^0$ and $\phi_h(0) = \phi_h^0$.

Note that the lump-sum tax $\bar{\tau}$ does not appear in the reservation-value equations (19) and (21). This is because investors pay $\bar{\tau}$ both when they are holding or not holding an asset, so $\bar{\tau}$ does not have a direct impact on the gain of holding an asset.

Proposition 4. *Consider an efficient asset allocation, $\boldsymbol{\phi}^*$, associated with bounded social values γ_l and γ_h. Define $\boldsymbol{\tau} = \{\tau_l, \tau_h, \bar{\tau}\}$ as $\tau_l = \theta \lambda_s (1-\phi_h^*)(\gamma_h - \gamma_l)$, $\tau_h = (1-\theta)\lambda_s \phi_l^* (\gamma_h - \gamma_l)$, and $\bar{\tau} = (\phi_l \tau_l - \phi_h \tau_h)/2$. Then $\{\phi_l^*, \phi_h^*, \gamma_l, \gamma_h\}$ is a decentralized equilibrium with intervention $\boldsymbol{\tau}$.*

An immediate implication of Proposition 4 is that the intervention policy simplifies when the bargaining power is either zero or one. If sellers have all the bargaining power ($\theta = 0$), the intervention restores efficiency simply with a tax to asset holdings of high-type investors. In this case, there is a unique source of inefficiency to be solved: buyers fail to internalize the option value lost when they acquire an asset. Likewise, if buyers have all the bargaining power ($\theta = 1$), the intervention restores efficiency by subsidizing asset holdings of low-type investors. The unique source of inefficiency to be solved is that sellers fail to internalize the option value gained when they acquire an asset. The next corollary formalizes these claims.

Corollary 1. *Consider the buyer's and seller's economy described before. Then, the following holds:*

(i) in the seller's economy, that is $\theta = 0$, if $\boldsymbol{\phi}^*$ is an efficient asset allocation associated with bounded social values γ_l and γ_h, then $\{\phi_l^*, \phi_h^*, \gamma_l, \gamma_h\}$ is a decentralized equilibrium with intervention $\tau_l = 0$, $\tau_h = \lambda_s \phi_l^*(\gamma_h - \gamma_l)$ and $\bar{\tau} = -\phi_h^* \tau_h/2$; and

(ii) in the buyer's economy, that is $\theta = 1$, if $\boldsymbol{\phi}^*$ is an efficient asset allocation associated with bounded social values γ_l and γ_h, then $\{\phi_l^*, \phi_h^*, \gamma_l, \gamma_h\}$ is a decentralized equilibrium with intervention $\tau_l = \lambda_s(1 - \phi_h^*)(\gamma_h - \gamma_l)$, $\tau_h = 0$ and $\bar{\tau} = \phi_l^* \tau_l/2$.

The implementation of the tax-subsidy scheme $\boldsymbol{\tau} = \{\tau_l, \tau_h, \bar{\tau}\}$ associated with Proposition 4 can be challenging. The main reason for this is that taxes conditioned on investors types may not be available to the government, either due to legal restrictions or lack of information on the utility types of the investors.

One can think about alternative ways to implement the efficient asset allocation that would not have these issues. For example, the government could subsidize the trade itself, not asset holdings. In fact, we can show that, by using a variation of the Vickrey-Clarke-Groves mechanism to subsidize trade, the government can implement the efficient outcome in our environment without investor specific taxes or knowledge of investors valuations.[12] On the other hand, this intervention will create other challenges in terms of implementation. In particular, the government would have to intervene in every trade to have investors trading using the Vickrey-Clarke-Groves mechanism, which may not be feasible either.

Given the complexity of the problem, and how different issues may occur in different markets, we prefer to think of the policy discussed in this section as a general guideline to help us understand what wedges must be corrected by the policy in order to solve the inefficiency that follows from the double-sided hold-up problem.

6 Efficiency with rationed primary market

Our economy resembles a market with three readily recognizable agents: issuers, intermediaries (low-type investors), and customers (high-type investors). Moreover, there are some market where only intermediaries can access the primary market and buy assets from issuers, and where customers can only acquire assets by obtaining them from intermediaries. With this market configuration in mind, in this section we study the efficiency properties of markets in which all primary trade occurs through intermediaries. We show that, with the appropriate choice of bargaining power in the secondary market, the equilibrium is constrained efficient. This implies that observed market configurations where only inter-

[12] A formal proof of this statement is available upon request.

mediaries buy from issuers, and customers only buy from intermediaries, can be viewed as efficient market configurations.

The equilibrium in an economy where only low-type investors have access to issuers is characterized by

$$\begin{aligned}
(r+\mu)\Delta_l &= \dot{\Delta}_l + \nu_l - \lambda_p \int_{\underline{c}}^{\Delta_l} (\Delta_l - c) g(c) dc + \lambda_s (1-\phi_h)(1-\theta)(\Delta_h - \Delta_l) \\
\dot{\phi}_l &= \lambda_p (1-\phi_l) G(\Delta_l) - \mu \phi_l - \lambda_s \phi_l (1-\phi_h) \\
(r+\mu)\Delta_h &= \dot{\Delta}_h + \nu_h - \lambda_s \phi_l \theta (\Delta_h - \Delta_l) \\
\dot{\phi}_h &= -\mu \phi_h + \lambda_s \phi_l (1-\phi_h) \ .
\end{aligned}$$

Similarly, the constrained-efficient allocation is characterized by

$$\begin{aligned}
(r+\mu)\gamma_l &= \dot{\gamma}_l + \nu_l - \lambda_p \int_0^{\gamma_l} (\gamma_l - c) g(c) dc + \lambda_s (1-\phi_h)(\gamma_h - \gamma_l) \\
\dot{\phi}_l &= \lambda_p (1-\phi_l) G(\gamma_l) - \mu \phi_l - \lambda_s \phi_l (1-\phi_h) \\
(r+\mu)\gamma_h &= \dot{\gamma}_h + \nu_h - \lambda_s \phi_l (\gamma_h - \gamma_l) \\
\dot{\phi}_h &= -\mu \phi_h + \lambda_s \phi_l (1-\phi_h)
\end{aligned}$$

Notice that because the high-type investor does not access the primary market, he is a passive agent: he will buy assets from low-type investors for any $\theta \in [0,1]$. Moreover, because all assets in the secondary market are channeled through the low-type investor, efficiency of the equilibrium reduces to finding a θ for which $\Delta_l = \gamma_l$ and $\dot{\Delta}_l = \dot{\gamma}_l$, while Δ_h does not need to be equal to γ_h. The strategy of our proof is therefore the following. We show that $\Delta_l > \gamma_l$ for $\theta = 0$ and $\Delta_l < \gamma_l$ for $\theta = 1$. Then, because Δ_l is continuous in θ, we conclude that it exists a value for $\theta \in (0,1)$ for which $\Delta_l = \gamma_l$, and thus the equilibrium is constrained-efficient.

Suppose that $\theta = 1$. Inspection of the first equation of each of the two systems presented above provides that $\Delta_l < \gamma_l$. This occurs because, as we concluded earlier, it is always the case that $\gamma_h > \gamma_l$. Suppose now that $\theta = 0$. In this case the first two equations of each system are identical, and thus $\Delta_l = \gamma_l$ only if $\Delta_h = \gamma_h$. However, if we compare the third equation of each system we conclude that $\Delta_h > \gamma_h$, given that $\gamma_h > \gamma_l$. Now, because $\Delta_h > \gamma_h$ when $\theta = 0$, it follows that Δ_l is not equal to γ_l. Rather, $\Delta_l > \gamma_l$. Because $\Delta_l < \gamma_l$ when $\theta = 1$ and $\Delta_l > \gamma_l$ when $\theta = 0$ and Δ_l is continuous in θ, we conclude that there exists a $\theta \in (0,1)$ such that $\Delta_l = \gamma_l$. Thus, when only low-type investors are allowed to access the primary market, there exists a value for the bargaining power in the secondary market that makes the equilibrium constrained-efficient.

Although here we only discuss efficiency of the rationed economy when there are only two-types of investors, the result is more general. For example, consider an economy populated by a large set of heterogeneous investors with asset valuation in the set $[\underline{\nu}, \bar{\nu}]$. If access to the primary market where to be rationed so that only the lowest valuation investors could access it, then the rationed economy can be made constrained-efficient by an appropriate choice of bargaining powers. However, if all investors can access the primary market, the decentralized equilibrium is always constrained-inefficient, as we discuss in Appendix B.

7 Numerical exploration of the model

In this section, we use our baseline model to numerically explore how the inefficiency caused by the double-sided holdup problem depends on (i) how the gains from trade are split in secondary markets, θ, and (ii) on the speed of trade, λ_s. Besides these two parameters, the model has 7 others to set: r, ν_l, ν_h, $\underline{c}$, $\bar{c}$, μ, and λ_p. Our baseline model is simple, only including enough to illustrate the how the inefficiency we study manifests. We do not claim the model captures all the features of trade in OTC markets and so, as a result, we do not attempt calibrate the model to any particular asset market. Instead we choose parameters that yield reasonable moments when compared to data. We set the discount rate $r = 0.05$, which is associated with a time length of one year. We assume that low-type investors derive no utility from the dividends of assets, $\nu_l = 0$, and set $\nu_h = 1$. We follow Assumption 1 and impose that the issuance cost parameters follow $\underline{c} = \nu_l/(r+\mu)$ and $\bar{c} = \nu_h/(r+\mu)$. We set $\mu = 0.1$ which gives an average maturity of assets of 10 years, roughly that of corporate bonds (9.9 years) or municipal bonds (13 years). Finally, we choose $\lambda_p = 1.0$, which leads to an annual issuance rate relative to the value of assets outstanding of 12%, in line that with that for municipal and corporate bonds, 11% and 14%, respectively.[13]

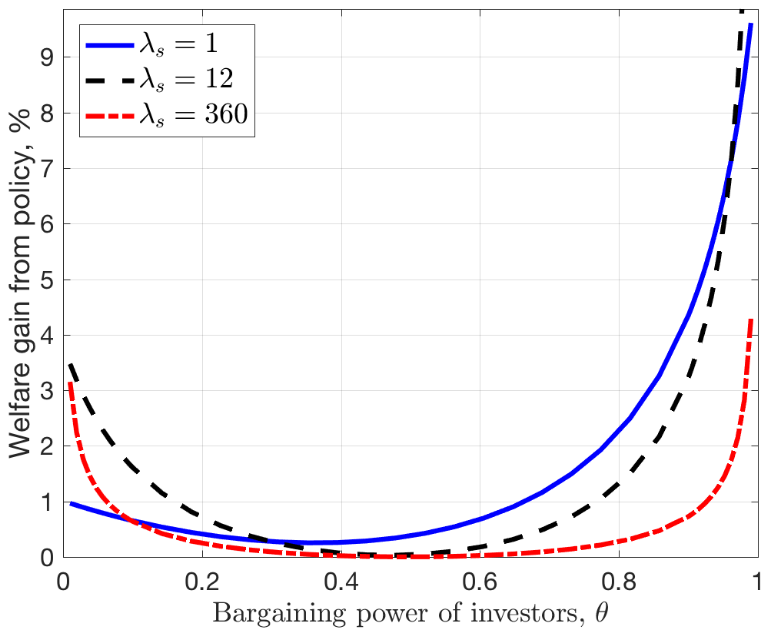
7.1 Effects of bargaining power

As described in Proposition 3, the way the surplus is split between agents in the secondary market is key in influencing the direction of misallocation and inefficient asset supply. Giving all the bargaining power to low-type investors – the natural intermediaries – implies

[13]The Securities Industry and Financial Markets Association (SIFMA) provides data about the volume of issuance and the value of outstanding debt for municipal and corporate bonds. Maturity comes from the Municipal Securities Rulemaking Board (MSRB) for municipal bonds and Di Maggio et al. (2017) for corporate bonds.

asset supply is too high. Alternatively, giving all the bargaining to high-type investors – the natural end-buyers – eliminates the incentive to intermediate and implies asset supply is too low. To get a sense of the magnitude of these effects, Figure 1 illustrates the impact θ has on the efficiency gains from introducing the tax/subsidy scheme, for varying trading speeds.

Figure 1: Welfare gains from intervention: the effect of θ.



Notes: The figure presents the welfare gains of the optimal tax/subsidy scheme, measured as percentage gains relative to the steady state with no tax/subisdies.

The gains from intervention is U-shaped in θ. If the share of the surplus is weighted heavily in one direction of the trade, then the size of the double-sided hold-up problem is magnified. However, if the gains from trade are split more evenly between parties, then the inefficiency is greatly reduced, especially so in fast markets. Since many OTC markets are characterized by high trading costs for investors as a result of the size of markups dealers earn on intermediating assets, one may think dealers posses high market power.[14] In fact, Hugonnier et al. (2018) finds that the bargaining power of dealers is substantiable: under our definition of bargaining power, they find that $\theta < 0.1$, and even around 0.03, depending on modeling assumptions. Under their calibration, the welfare gains of fixing the double-

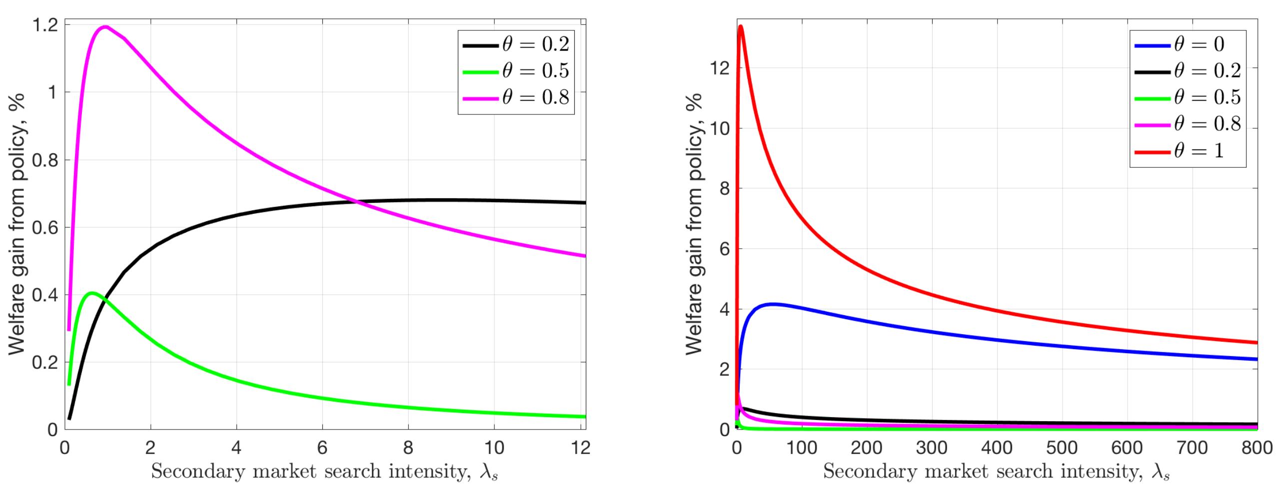
[14]Markups are typically defined as the percentage gain in sale price over purchase price of an asset and range from 0.5% for corporate bonds to 1.7% for municipal bonds and 2.9% for collateralized mortgage obligations. For evidence, see Di Maggio et al. (2017) for corporate bonds, Green et al. (2007b) for municipal bonds, and Hollifield et al. (2017) for securitizations.

sided hold-up problem are sizable, even when search frictions are small. However, policies designed to drive markups to zero, equivalent to setting $\theta = 1$, would make markets less efficient and hurt welfare.

7.2 Effects of trading speed

Despite the decentralized nature of OTC markets, average trading delays are often minimal. For instance, in the municipal bond market the median time it takes dealers to intermediate assets between two investors is approximately 5 days (Green et al., 2007b). In the market for corporate bonds or securities it takes slightly longer to intermediate assets at 12 and 37 days, respectively (Di Maggio et al., 2017; Hollifield et al., 2017), but one could argue that these are all indicative of markets in which search frictions are low. Despite this, we argue that the double-sided holdup problem resulting from bilateral trade and bargaining may still imply there is a significant role for policy, even in fast markets. Policies aimed at solely improving trading speed may be limited by the inefficiency of bilateral trade. We highlight these limitations in Figure 2, which illustrates the effect of trading speed on the welfare gains of the policy introduced in Section 5, for varying bargaining powers.

Figure 2: Welfare gains from intervention: the effect of λ



Notes: The figure presents the welfare gains of the optimal tax/subsidy scheme, measured as percentage gains relative to the steady state with no tax/subisdies.

The left panel illustrates the welfare gain from intervention for low values of λ_s and intermediate bargaining powers while the right panel illustrates the gain in fast markets and when bargaining power is zero or one. Generally, the gains from intervention are hump-shaped in trading speed. When trading speed is zero, there is no scope for policy.

Markets with a slow trading speed have the greatest need for intervention, or the most severe hold-up problems. If the surplus is split equally among investors the inefficiency is less severe, peaking around 0.4% when $\theta = 0.5$. However if the gains from trade fall heavily towards one side of the market, the inefficiency worsens. For instance in the buyer's economy, $\theta = 1$, the welfare gains from intervention can be as high as 13.3%.

As trading speed increases, the inefficiency generally becomes less severe. However, even for high values of λ_s, there is still a considerable gain correcting the double-sided holdup problem, especially so when the gains from trade are weighted heavily towards one side of the market. Fast markets do not imply that the inefficiency of decentralized trade is small and policies singularly designed to increase the speed in trade may be limited.

8 Concluding remarks

We showed that the issuance of assets that retrade in OTC secondary markets is always distorted as a result of a double-sided hold-up problem that cannot be solved by correctly splitting the gains from trade. Low-value investors, acting as natural intermediaries, require the full surplus from trade in order to fully internalize the social value of transferring the asset to higher-valuation investors. On the other hand, high-value investors require the full surplus from trade in order to fully internalize their outside option of waiting to purchase the asset in the future. Both conditions can never be simultaneously satisfied, leading to an inefficient aggregate asset supply and misallocation among investors. Further, the direction of inefficiency and misallocation depends on the way the surplus is split. If secondary market buyers posses all the bargaining power, intermediation and asset supply are depressed. If secondary market sellers posses all the bargaining power, then intermediation and asset supply are too high.

We also show that the inefficiency result persists even when we introduce heterogeneity in bargaining powers in the primary market. Key to the inefficiency result is the fact that investors cannot commit to an investment rule in the primary market, thus strategic complementaries arise. A natural solution to this problem is to restrict access to the primary market to either of the investor types. In this case, there is no double-sided hold-up problem. Rather, just the more standard single-sided hold-up problem which can be handled with an appropriate choice of bargaining power. Not surprisingly, there exists a way to split the surplus in the secondary market for which the equilibrium in this case becomes efficient, however the solution is not to give all the bargaining power to one agent.

Efficiency can be restored by levying a tax on the asset holdings of high-valuation investors and subsidizing asset holdings of low-valuation investors. Through numerical examples, we illustrate that the gains from intermediation are high when the surplus is shared unevenly between buyers and sellers in the secondary market and when trading speed is low. The inefficiency goes to zero as trading speed goes to infinity, however the speed of convergence is slow. Even in fast markets, e.g. those in which traders meet multiple times a day, the inefficiency caused by bilateral trade and bargaining persists.

While we chose to illustrate the inefficiency result in a simple environment with two types, discrete asset holdings, and no competitive dealers, we conjecture that these results are quite general. To illustrate some of these, we also studied a model with a continuum of types, in the spirit of Hugonnier et al. (2014). We show that the inefficiency result still holds there, even if the bargaining parameter θ is allowed to vary with (i) the identity of a particular investor, or (ii) the identities of the two parties involved in each trade meeting. Again, the key ingredient needed is having a two-sided investment decision (which introducing asset issuance implies) with ex-post bargaining.

Of course, there are also some limitations in our analysis. The capacity constraint (the constraint that investors hold at most one asset) plays an important role. The inefficiency would disappear if investors could hold an unlimited amount of assets so buying/not buying assets in the primary market would not destroy trade surplus in the secondary market. However, we conjecture that the inefficiency would persist if the utility function of investors is concave in the number of assets (or if the costs of holding assets is convex in the number of assets), so that the gains from holding a second unit of the asset are lower than the gains from holding one unit of the asset, the gains from holding a third unit of the asset are lower than the gains from holding two units of the asset, and so on. A more crucial limitation is the way we model the primary market. In many financial asset markets, the secondary market is over the counter, but the primary market is organized in a different way. Considering different ways to organize the primary market is definitely important to better understand the consequences of the double hold-up problem we discussed in this paper, and we leave such extensions for future research.

References

Acemoglu, D. (1996). A microfoundation for social increasing returns in human capital accumulation. *The Quarterly Journal of Economics*, 111(3):779–804.

Acemoglu, D. and Shimer, R. (1999). Holdups and efficiency with search frictions. *International Economic Review*, 40(4):827–849.

Arseneau, D., Rappoport, D., and Vardoulakis, A. (2016). Secondary market liquidity and the optimal capital structure.

Aruoba, S. B., Rocheteau, G., and Waller, C. (2007). Bargaining and the value of money. *Journal of Monetary Economics*, 54(8):2636–2655.

Bethune, Z., Sultanum, B., and Trachter, N. (2016). Private information in over-the-counter markets. *FRB Richmond Working Paper*.

Chang, B. (2017). Adverse selection and liqudity distortion. *Review of Economic Studies*, 85(1):275–306.

Chiu, J. and Koeppl, T. (2015). Trading dynamics with adverse selection and search: Market freeze, intervention and recovery. *Review of Economic Studies*, 00:1–35.

Di Maggio, M., Kermani, A., and Song, Z. (2017). The value of trading relations in turbulent times. *Journal of Financial Economics*, 124(2):266–284.

Duffie, D., Garleanu, N., and Pedersen, L. H. (2005). Over-the-counter markets. *Econometrica*, 73:1815–1847.

Duffie, D., Garleanu, N., and Pedersen, L. H. (2007). Valuation in over-the-counter markets. *The Review of Financial Studies*, 20(6):1865–1900.

Farboodi, M., Jarosch, G., and Shimer, R. (2017). The Emergence of Market Structure. NBER Working Papers 23234, National Bureau of Economic Research, Inc.

Gârleanu, N. (2009). Portfolio choice and pricing in illiquid markets. *Journal of Economic Theory*, 144(2):532–564.

Geromichalos, A. and Herrenbrueck, L. (2016). The strategic determination of the supply of liquid assets. Technical report, Working Papers, University of California, Department of Economics.

Gofman, M. (2011). A network-based analysis of over-the-counter markets.

Green, R., Hollifield, B., and Schurhoff, N. (2007a). Dealer intermediation and price behavior in the aftermarket for new bond issues. *Journal of Financial Economics*, 86:643–682.

Green, R., Hollifield, B., and Schurhoff, N. (2007b). Financial intermediation and the costs of trading in an opaque market. *The Review of Financial Studies*, 20(2):275–314.

He, Z. and Milbradt, K. (2014). Endogenous liquidity and defaultable bonds. *Econometrica*, 82(4):1443–1508.

Hollifield, B., Neklyudov, A., and Spatt, C. (2017). Bid-ask spreads, trading networks, and the pricing of securitizations. *The Review of Financial Studies*, 30(9):3048–3085.

Hosios, A. (1990). On the efficiency of matching and related models of search unemployment. *Review of Economic Studies*, 57:279–298.

Hugonnier, J., Lester, B., and Weill, P.-O. (2014). Heterogeneity in decentralized asset markets. Technical report, National Bureau of Economic Research.

Hugonnier, J., Lester, B., and Weill, P.-O. (2018). Frictional Intermediation in Over-the-counter Markets. NBER Working Papers 24956, National Bureau of Economic Research, Inc.

Klein, B., Crawford, R. G., and Alchian, A. A. (1978). Vertical integration, appropriable rents, and the competitive contracting process. *The journal of Law and Economics*, 21(2):297–326.

Lagos, R. and Rocheteau, G. (2007). Search in asset markets: Market structure, liquidity, and welfare. *American Economic Review*, 97(2):198–202.

Lagos, R. and Rocheteau, G. (2009). Liquidity in asset markets with search frictions. *Econometrica*, 57:403–426.

Lagos, R., Rocheteau, G., and Weill, P.-O. (2011). Crises and liquidity in over-the-counter markets. *Journal of Economic Theory*, 146(6):2169–2205.

Lagos, R. and Wright, R. (2005). A unified framework for monetary theory and policy analysis. *Journal of Political Economy*, 113(3):463–484.

Masters, A. M. (1998). Efficiency of investment in human and physical capital in a model of bilateral search and bargaining. *International Economic Review*, pages 477–494.

Menzio, G., Jarosch, G., and Farboodi, M. (2016). Tough Middlemen. 2016 Meeting Papers 1371, Society for Economic Dynamics.

Nosal, E., Wong, Y.-Y., and Wright, R. (2014). More on Middlemen: Equilibrium Entry and Efficiency in Intermediated Markets. Working paper series, Federal Reserve Bank of Chicago.

Nosal, E., Wong, Y.-Y., and Wright, R. (2016). Who wants to be a middleman? Technical report, Mimeo, University of Wisconsin Madison.

Rocheteau, G. and Wright, R. (2005). Money in search equilibrium, in competitive equilibrium, and in competitive search equilibrium. *Econometrica*, 73(1):175–202.

Willianson, O. E. (1975). Markets and hierarchies: analysis and antitrust implications. *New York*.

A Proofs

A.1 Proof of Lemma 1

The difference $\Delta_h - \Delta_l$ implies by the reservation-value equations (7) and (8) is

$$\begin{aligned}(r+\mu+2\rho)(\Delta_h-\Delta_l) = {} & (\dot{\Delta}_h-\dot{\Delta}_l)+(\nu_h-\nu_l) \\ & -\lambda_p\left[\int_{\underline{c}}^{\Delta_h}(\Delta_h-c)g(c)dc-\int_{\underline{c}}^{\Delta_l}(\Delta_l-c)g(c)dc\right] \\ & -\lambda_s\Big[\phi_l\theta+(1-\phi_h)(1-\theta)\Big](\Delta_h-\Delta_l).\end{aligned}$$

We prove the lemma by a contradiction argument. Suppose, by the way of contradiction, that $\Delta_h(t)-\Delta_l(t)\le 0$ for some time period t. Then the left hand side of the above equation is smaller or equal to zero. However, in the right hand side, all the terms but $\dot{\Delta}_h-\dot{\Delta}_l$ are non-negative, and at least $\nu_h-\nu_l$ is strictly positive. Therefore, $\dot{\Delta}_h-\dot{\Delta}_l\le-(\nu_h-\nu_l)<0$ is strictly negative in time t, $\Delta_h-\Delta_l$ stays negative, and $\dot{\Delta}_h-\dot{\Delta}_l\le-(\nu_h-\nu_l)<0$ for all $t'\ge t$. But this implies an explosive path for either Δ_h or Δ_l, which contradicts that both functions are bounded. Hence, $\Delta_h(t)-\Delta_l(t)>0$ for all time periods t.

Further, if we suppose that $\Delta_l>\Delta_h$, and derive the reservation value of low and high type investors associated with the trade pattern implied by $\Delta_l>\Delta_h$, the reservation value equations would be

$$(r+\mu)\Delta_l=\dot{\Delta}_l+\nu_l+\rho(\Delta_h-\Delta_l)-\lambda_p\int_{\underline{c}}^{\Delta_l}(\Delta_l-c)g(c)dc-\lambda_s\phi_h\theta(\Delta_l-\Delta_h) \tag{24}$$

$$(r+\mu)\Delta_h=\dot{\Delta}_h+\nu_h-\rho(\Delta_h-\Delta_l)-\lambda_p\int_{\underline{c}}^{\Delta_h}(\Delta_h-c)g(c)dc+\lambda_s(1-\phi_l)(1-\theta)(\Delta_l-\Delta_h) \tag{25}$$

Which implies that

$$
\begin{aligned}
(r+\mu+2\rho)(\Delta_h-\Delta_l) = {} & (\dot{\Delta}_h-\dot{\Delta}_l)+(\nu_h-\nu_l) \\
& -\lambda_p\left[\int_{\underline{c}}^{\Delta_h}(\Delta_h-c)g(c)dc-\int_{\underline{c}}^{\Delta_l}(\Delta_l-c)g(c)dc\right] \\
& -\lambda_s\Big[(1-\phi_l)(1-\theta)+\phi_h\theta\Big](\Delta_h-\Delta_l)
\end{aligned}
$$

This equation is analogous to the one we had before, and, in the same way, implies $\Delta_h > \Delta_l$, a contradiction of our assumption that $\Delta_l < \Delta_h$. □

A.2 Proof of Proposition 1

The general equilibrium effects of changing θ complicate the proof. We deal with it by first arguing that, in the sellers economy, the reservation value of high-type investors is independent of the other equilibrium objects, and in the buyers economy, the reservation value of low-type investors is independent of the other equilibrium objects. That is, in these two-limit cases we eliminate one of the four equilibrium equations, which reduces the general equilibrium effects we need to account for. Then we use the remaining three equations to sign the effects of moving θ between θ^A, θ^B, and θ^C in the equilibrium outcomes.

The equilibrium equations in steady state are

$$0=(r+\mu)\Delta_l+\lambda_p\int_{\underline{c}}^{\Delta_l}(\Delta_l-c)g(c)dc-\lambda_s(1-\phi_h)(1-\theta)(\Delta_h-\Delta_l)-\nu_l \tag{26}$$

$$0=\lambda_p(1-\phi_l)G(\Delta_l)-\mu\phi_l-\lambda_s\phi_l(1-\phi_h) \tag{27}$$

$$0=(r+\mu)\Delta_h+\lambda_p\int_{\underline{c}}^{\Delta_h}(\Delta_h-c)g(c)dc+\lambda_s\phi_l\theta(\Delta_h-\Delta_l)-\nu_h \tag{28}$$

$$0=\lambda_p(1-\phi_h)G(\Delta_h)-\mu\phi_h+\lambda_s\phi_l(1-\phi_h) \tag{29}$$

Proof of the buyer's economy *vs* the interior economy

$\Delta_l^C=\frac{\nu_l}{\mu+r}$: The equilibrium equation (26) implies

$$(\mu+r)\Delta_l^C+\lambda_p\int_{\underline{c}}^{\Delta_l^C}(\Delta_l^C-c)g(c)dc=\nu_l.$$

This implies that low-types' reservation value in the buyer's economy is $\Delta_l^C=\frac{\nu_l}{\mu+r}$. The term $(\mu+r)\Delta_l+\lambda_p\int_{\underline{c}}^{\Delta_l}(\Delta_l-c)g(c)dc$ is strictly increasing as a function of Δ_l, and it is exactly ν_l when $\Delta_l^C = \frac{\nu_l}{\mu+r}$ because $\frac{\nu_l}{\mu+r} = \underline{c}$ so $\lambda_p \int_{\underline{c}}^{\Delta_l^C} (\Delta_l^C - c)g(c)dc$ equals zero. Therefore the only solution for the equality is $\Delta_l^C = \frac{\nu_l}{\mu+r}$.

$\boldsymbol{\Delta_l^B > \Delta_l^C}$ **:** Because of depreciation ($\mu > 0$), there is no steady state where $\phi_h = 1$, so $1 - \phi_h^B$ is strictly positive. From lemma 1 we must have $\Delta_h^B - \Delta_l^B$ strictly positive. Therefore, $\lambda_s(1-\phi_h^B)(1-\theta^B)(\Delta_h^B - \Delta_l^B) > 0$, and we can conclude that

$$(\mu+r)\Delta_l^C + \lambda_p \int_{\underline{c}}^{\Delta_l^C} (\Delta_l^C - c)g(c)dc < (r+\mu)\Delta_l^B + \lambda_p \int_{\underline{c}}^{\Delta_l^B} (\Delta_l^B - c)g(c)dc$$

from the equilibrium equation 26. Since the term $(\mu+r)\Delta_l + \lambda_p \int_{\underline{c}}^{\Delta_l} (\Delta_l - c)g(c)dc$ is strictly increasing as a function of Δ_l, the above inequality implies that $\Delta_l^B > \Delta_l^C$.

$\boldsymbol{\phi_l^C = 0}$ **:** From the equilibrium equation 27 we have

$$0 = \lambda_p(1-\phi_l^C)G(\Delta_l^C) - \mu\phi_l^C - \lambda_s\phi_l^C(1-\phi_h^C) = -\phi_l^C[\mu + \lambda_s(1-\phi_h^C)].$$

There is no issuance of assets to dealers since $G(\Delta_l^C) = G\left(\nu_l/(\mu+r)\right) = G(\underline{c}) = 0$, and $\mu + \lambda_s(1-\phi_h^C) > 0$ since ϕ_h^C is smaller than one due to depreciation. Therefore, ϕ_l^C is zero.

$\boldsymbol{\phi_l^B > \phi_l^C}$ **:** From the equilibrium equation 27 we have

$$0 = \lambda_p(1-\phi_l^B)G(\Delta_l^B) - \mu\phi_l^B - \lambda_s\phi_l^B(1-\phi_h^B) \implies \phi_l^B = \frac{\lambda_p G(\Delta_l^B)}{\lambda_p G(\Delta_l^B) + \mu + \lambda_s(1-\phi_h^B)}.$$

We know that $G(\Delta_l^B) > 0$ since $\Delta_l^B > \Delta_l^C = \underline{c}$. Therefore, $\phi_l^B > 0 = \phi_l^C$.

$\boldsymbol{\Delta_h^C > \Delta_h^B}$ **:** We showed that $\phi_l^B > 0$, lemma 1 says that $\Delta_h^B - \Delta_l^B > 0$ and, therefore, $\lambda_s\phi_l^B\theta^B(\Delta_h^B - \Delta_l^B) > 0$. We showed that $\phi_l^C = 0$ so $\lambda_s\phi_l^C\theta^C(\Delta_h^C - \Delta_l^C) = 0$. The equilibrium equation 28, $\lambda_s\phi_l^B\theta^B(\Delta_h^B - \Delta_l^B) > 0$ and $\lambda_s\phi_l^C\theta^C(\Delta_h^C - \Delta_l^C) = 0$ imply

$$(r+\mu)\Delta_h^B + \lambda_p \int_{\underline{c}}^{\Delta_h^B} (\Delta_h^B - c)g(c)dc < (r+\mu)\Delta_h^C + \lambda_p \int_{\underline{c}}^{\Delta_h^C} (\Delta_h^C - c)g(c)dc.$$

Since the term $(r+\mu)\Delta_h + \lambda_p \int_{\underline{c}}^{\Delta_h} (\Delta_h - c)g(c)dc$ is strictly increasing as a function of Δ_h, the above inequality implies that $\Delta_h^B < \Delta_h^C$.

$\boldsymbol{\phi_h^B > \phi_h^C}$**:** Define the function $F(\phi_l, \phi_h, \Delta_h; \Delta_l)$ as

$$F(\phi_l, \phi_h, \Delta_h; \Delta_l) = \begin{bmatrix} \lambda_p(1-\phi_l)G(\Delta_l) - \mu\phi_l - \lambda_s\phi_l(1-\phi_h) \\ \lambda_p(1-\phi_h)G(\Delta_h) - \mu\phi_h + \lambda_s\phi_l(1-\phi_h) \\ (r+\mu)\Delta_h + \lambda_p \int_{\underline{c}}^{\Delta_h}(\Delta_h - c)g(c)dc + \lambda_s\phi_l\theta^B(\Delta_h - \Delta_l) - \nu_h \end{bmatrix}. \tag{30}$$

Note that $F(\phi_l^B, \phi_h^B, \Delta_h^B; \Delta_l^B) = \mathbf{0}$ and $F(\phi_l^C, \phi_h^C, \Delta_h^C; \Delta_l^C) = \mathbf{0}$, where $\mathbf{0}$ is the zero column vector in $\mathbb{R}^3$. The first equality comes from the equilibrium definition, while the second comes from the equilibrium definition and $\phi_l^C = 0$. The equality $F(\phi_l, \phi_h, \Delta_h; \Delta_l) = \mathbf{0}$ implicitly defines ϕ_h as functions of Δ_l, and we can use the implicit function theorem to compute $\partial\phi_h/\partial\Delta_l$. Since $\Delta_l^B > \Delta_l^C$, if $\partial\phi_h/\partial\Delta_l$ is positive we can conclude that $\phi_h^B > \phi_h^C$.

To apply the implicit function theorem let us compute $D = det(\partial F/\partial(\phi_l,\phi_h,\Delta_h))$. We have

$$\frac{\partial F}{\partial(\phi_l, \phi_h, \Delta_h)} = \begin{bmatrix} -[\lambda_p G(\Delta_l) + \mu + \lambda_s(1-\phi_h)] & \lambda_s\phi_l & 0 \\ \lambda_s(1-\phi_h) & -[\lambda_p G(\Delta_h) + \mu + \lambda_s\phi_l] & \lambda_p(1-\phi_h)g(\Delta_h) \\ \lambda_s\theta^B(\Delta_h - \Delta_l) & 0 & r + \mu + \lambda_p G(\Delta_h) + \lambda_s\phi_l\theta^B \end{bmatrix},$$

and $D = det\left(\partial F/\partial(\phi_l,\phi_h,\Delta_h)\right)$ is

$$\begin{aligned} D &= [\lambda_p G(\Delta_l) + \mu + \lambda_s(1-\phi_h)] \times [\lambda_p G(\Delta_h) + \mu + \lambda_s\phi_l] \times [r + \mu + \lambda_p G(\Delta_h) + \lambda_s\phi_l\theta^B] \\ &\quad + \lambda_s\phi_l \times \lambda_p(1-\phi_h)g(\Delta_h) \times \lambda_s\theta^B(\Delta_h - \Delta_l) - \lambda_s\phi_l \times \lambda_s(1-\phi_h) \times [r + \mu + \lambda_p G(\Delta_h) + \lambda_s\phi_l\theta^B] \\ &= [r + \mu + \lambda_p G(\Delta_h) + \lambda_s\phi_l\theta^B]\Big\{[\lambda_p G(\Delta_l) + \mu + \lambda_s(1-\phi_h)] \times [\lambda_p G(\Delta_h) + \mu + \lambda_s\phi_l] \\ &\quad - \lambda_s\phi_l\lambda_s(1-\phi_h)\Big\} + \lambda_s\phi_l \times \lambda_p(1-\phi_h)g(\Delta_h) \times \lambda_s\theta^B(\Delta_h - \Delta_l) \\ &= [r + \mu + \lambda_p G(\Delta_h) + \lambda_s\phi_l\theta^B]\Big\{[\lambda_p G(\Delta_l) + \mu + \lambda_s(1-\phi_h)] \times [\lambda_p G(\Delta_h) + \mu] + [\lambda_p G(\Delta_l) + \mu]\lambda_s\phi_l \\ &\quad + \cancel{\lambda_s(1-\phi_h)\lambda_s\phi_l} - \cancel{\lambda_s\phi_l\lambda_s(1-\phi_h)}\Big\} + \lambda_s\phi_l \times \lambda_p(1-\phi_h)g(\Delta_h) \times \lambda_s\theta^B(\Delta_h - \Delta_l) \\ &= [(r+\mu) + \lambda_p G(\Delta_h) + \lambda_s\phi_l\theta^B]\Big\{[\lambda_p G(\Delta_l) + \mu + \lambda_s(1-\phi_h)] \times [\lambda_p G(\Delta_h) + \mu] \\ &\quad + [\lambda_p G(\Delta_l) + \mu]\lambda_s\phi_l\Big\} + \lambda_s\phi_l \times \lambda_p(1-\phi_h)g(\Delta_h) \times \lambda_s\theta^B(\Delta_h - \Delta_l) \geq (\mu + r)\mu^2 > 0. \end{aligned}$$

Since $D \geq (\mu + r)\mu^2 > 0$ is bounded away from zero, we can apply the implicit function theorem all the way from Δ_l^C to Δ_l^B. The implicit function theorem implies that

$$\underbrace{\frac{\partial F}{\partial(\phi_l, \phi_h, \Delta_h)}}_{\text{matrix A}} \begin{bmatrix} \partial\phi_l/\partial\Delta_l \\ \partial\phi_h/\partial\Delta_l \\ \partial\Delta_h/\partial\Delta_l \end{bmatrix} = -\frac{\partial F}{\partial\Delta_l} = \underbrace{\begin{bmatrix} -\lambda_p(1-\phi_l)g(\Delta_l) \\ 0 \\ \lambda_s\phi_l\theta^B \end{bmatrix}}_{\text{vector b}}.$$

We can easily solve this system using Cramer's rule. We already computed $D = det(A)$. Let us compute D_{ϕ_h}, which is the determinant of the matrix A after replacing the second column of A with the vector b.

$$
\begin{aligned}
D_{\phi_h} = & -\lambda_p(1-\phi_l)g(\Delta_l)\times\lambda_p(1-\phi_h)g(\Delta_h)\times\lambda_s\theta^B(\Delta_h-\Delta_l) \\
& +\lambda_p(1-\phi_l)g(\Delta_l)\times\lambda_s(1-\phi_h)\times[r+\mu+\lambda_pG(\Delta_h)+\lambda_s\phi_l\theta^B] \\
& +[\lambda_pG(\Delta_l)+\mu+\lambda_s(1-\phi_h)]\times\lambda_p(1-\phi_h)g(\Delta_h)\times\lambda_s\phi_l\theta^B \\
= & -\lambda_p(1-\phi_l)g(\Delta_l)\times\lambda_p(1-\phi_h)g(\Delta_h)\times\lambda_s\theta^B(-\Delta_l) \\
& +\lambda_p(1-\phi_l)g(\Delta_l)\times\lambda_s(1-\phi_h)\times[r+\mu+\lambda_pG(\Delta_h)-\lambda_p\theta^Bg(\Delta_h)\Delta_h+\lambda_s\phi_l\theta^B] \\
& +[\lambda_pG(\Delta_l)+\mu+\lambda_s(1-\phi_h)]\times\lambda_p(1-\phi_h)g(\Delta_h)\times\lambda_s\phi_l\theta^B \\
= & \lambda_p(1-\phi_l)g(\Delta_l)\times\lambda_p(1-\phi_h)g(\Delta_h)\times\lambda_s\theta^B\Delta_l \\
& +\lambda_p(1-\phi_l)g(\Delta_l)\times\lambda_s(1-\phi_h)\times[r+\mu+\lambda_p(1-\theta^B)G(\Delta_h)+\lambda_s\phi_l\theta^B] \\
& +[\lambda_pG(\Delta_l)+\mu+\lambda_s(1-\phi_h)]\times\lambda_p(1-\phi_h)g(\Delta_h)\times\lambda_s\phi_l\theta^B > 0
\end{aligned}
$$

From Cramer's rule $\partial\phi_h/\partial\Delta_l = D_{\phi_h}/D > 0$ and, therefore, $\phi_h^B > \phi_h^C$.

Proof of the seller's economy *vs* the interior economy

$\boldsymbol{\Delta_h^A > \Delta_h^B}$ **:** We showed that ϕ_l^B is strictly positive, and lemma 1 states that $\Delta_h^B - \Delta_l^B$ is strictly positive. These results imply that $\lambda_s\phi_l^B\theta^B(\Delta_h^B - \Delta_l^B)$ is strictly positive. The term $\lambda_s\phi_l^A\theta^A(\Delta_h^A - \Delta_l^A)$ is zero because θ^A is zero. From the inequality $\lambda_s\phi_l^B\theta^B(\Delta_h^B - \Delta_l^B) > 0$, the equality $\lambda_s\phi_l^A\theta^A(\Delta_h^A - \Delta_l^A) = 0$, and the equilibrium equation 28, we conclude that

$$
(r+\mu)\Delta_h^B + \lambda_p\int_{\underline{c}}^{\Delta_h^B}(\Delta_h^B - c)g(c)dc < (r+\mu)\Delta_h^A + \lambda_p\int_{\underline{c}}^{\Delta_h^A}(\Delta_h^A - c)g(c)dc.
$$

The term $(r+\mu)\Delta_h + \lambda_p\int_{\underline{c}}^{\Delta_h}(\Delta_h - c)g(c)dc$ is strictly increasing as a function of Δ_h. As a result, the above inequality implies that $\Delta_h^B < \Delta_h^A$.

$\boldsymbol{\phi_l^A > \phi_l^B}$**,** $\boldsymbol{\phi_h^A > \phi_h^B}$**, and** $\boldsymbol{\Delta_l^A > \Delta_l^B}$ **:** With abuse of notation, let us now define the function $F(\phi_l, \phi_h, \Delta_l; \theta, \Delta_h)$ as

$$
F(\phi_l,\phi_h,\Delta_l;\theta,\Delta_h) = \begin{bmatrix} \lambda_p(1-\phi_l)G(\Delta_l) - \mu\phi_l - \lambda_s\phi_l(1-\phi_h) \\ \lambda_p(1-\phi_h)G(\Delta_h) - \mu\phi_h + \lambda_s\phi_l(1-\phi_h) \\ (r+\mu)\Delta_l + \lambda_p\int_{\underline{c}}^{\Delta_l}(\Delta_l - c)g(c)dc - \lambda_s(1-\phi_h)(1-\theta)(\Delta_h-\Delta_l) - \nu_l \end{bmatrix}. \quad (31)
$$

It is easy to check that $F(\phi_l^A, \phi_h^A, \Delta_l^A; \theta^A, \Delta_h^A) = \mathbf{0}$ and $F(\phi_l^B, \phi_h^B, \Delta_l^B; \theta^B, \Delta_h^B) = \mathbf{0}$; the two equalities come from the equilibrium definition.

The equality $F(\phi_l, \phi_h, \Delta_l; \theta, \Delta_h) = \mathbf{0}$ implicitly defines ϕ_l, ϕ_h, and Δ_l as functions of θ and Δ_h. So we can use the implicit function theorem to compute $\partial\phi_l/\partial\Delta_h$, $\partial\phi_h/\partial\Delta_h$, $\partial\Delta_l/\partial\Delta_h$,

$\partial\phi_l/\partial\theta$, $\partial\phi_h/\partial\theta$ and $\partial\Delta_l/\partial\theta$. We know that $\Delta_l^A > \Delta_l^B$ and $\theta^A < \theta^B$. Therefore, if $\partial\phi_l/\partial\Delta_h$, $\partial\phi_h/\partial\Delta_h$, and $\partial\Delta_l/\partial\Delta_h$ are positive, and $\partial\phi_l/\partial\theta$, $\partial\phi_h/\partial\theta$, and $\partial\Delta_l/\partial\theta$ are negative, we can conclude that $\phi_l^A > \phi_l^B$, $\phi_h^A > \phi_h^B$, and $\Delta_l^A > \Delta_l^B$.

To apply the implicit function theorem let us compute $D = det(\partial F/\partial(\phi_l,\phi_h,\Delta_l))$. We have

$$\frac{\partial F}{\partial(\phi_l,\phi_h,\Delta_l)} = \begin{bmatrix} -[\lambda_p G(\Delta_l)+\mu+\lambda_s(1-\phi_h)] & \lambda_s\phi_l & \lambda_p(1-\phi_l)g(\Delta_l) \\ \lambda_s(1-\phi_h) & -[\lambda_p G(\Delta_h)+\mu+\lambda_s\phi_l] & 0 \\ 0 & \lambda_s(1-\theta)(\Delta_h-\Delta_l) & r+\mu+\lambda_p G(\Delta_l)+\lambda_s(1-\phi_h)(1-\theta) \end{bmatrix},$$

and $D = det\left(\partial F/\partial(\phi_l,\phi_h,\Delta_l)\right)$ is

$$\begin{aligned} D = {} & [\lambda_p G(\Delta_l)+\mu+\lambda_s(1-\phi_h)] \times [\lambda_p G(\Delta_h)+\mu+\lambda_s\phi_l] \times [r+\mu+\lambda_p G(\Delta_l)+\lambda_s(1-\phi_h)(1-\theta)] \\ & + \lambda_p(1-\phi_l)g(\Delta_l) \times \lambda_s(1-\phi_h) \times \lambda_s(1-\theta)(\Delta_h-\Delta_l) \\ & - \lambda_s\phi_l \times \lambda_s(1-\phi_h) \times [r+\mu+\lambda_p G(\Delta_l)+\lambda_s(1-\phi_h)(1-\theta)] \\ = {} & [\lambda_p G(\Delta_l)+\mu] \times [\lambda_p G(\Delta_h)+\mu+\lambda_s\phi_l] \times [r+\mu+\lambda_p G(\Delta_l)+\lambda_s(1-\phi_h)(1-\theta)] \\ & + \lambda_p(1-\phi_l)g(\Delta_l) \times \lambda_s(1-\phi_h) \times \lambda_s(1-\theta)(\Delta_h-\Delta_l) \\ & - \cancel{\lambda_s\phi_l \times \lambda_s(1-\phi_h)} \times [r+\mu+\lambda_p G(\Delta_l)+\lambda_s(1-\phi_h)(1-\theta)] \\ & + \cancel{\lambda_s(1-\phi_h) \times \lambda_s\phi_l} \times [r+\mu+\lambda_p G(\Delta_l)+\lambda_s(1-\phi_h)(1-\theta)] \\ & + \lambda_s(1-\phi_h) \times [\lambda_p G(\Delta_h)+\mu] \times [r+\mu+\lambda_p G(\Delta_l)+\lambda_s(1-\phi_h)(1-\theta)] \\ = {} & [\lambda_p G(\Delta_l)+\mu] \times [\lambda_p G(\Delta_h)+\mu+\lambda_s\phi_l] \times [r+\mu+\lambda_p G(\Delta_l)+\lambda_s(1-\phi_h)(1-\theta)] \\ & + \lambda_p(1-\phi_l)g(\Delta_l) \times \lambda_s(1-\phi_h) \times \lambda_s(1-\theta)(\Delta_h-\Delta_l) \\ & + \lambda_s(1-\phi_h) \times [\lambda_p G(\Delta_h)+\mu] \times [r+\mu+\lambda_p G(\Delta_l)+\lambda_s(1-\phi_h)(1-\theta)] \geq \mu^2(r+\mu) > 0 \end{aligned}$$

Since $D \geq (\mu+r)\mu^2 > 0$ is bounded away from zero we can apply the implicit function theorem all the way from Δ_h^A to Δ_h^B and θ^A to θ^B.

The implicit function theorem implies that

$$\underbrace{\frac{\partial F}{\partial(\phi_l,\phi_h,\Delta_l)}}_{\text{matrix A}} \begin{bmatrix} \partial\phi_l/\partial\Delta_h \\ \partial\phi_h/\partial\Delta_h \\ \partial\Delta_l/\partial\Delta_h \end{bmatrix} = -\frac{\partial F}{\partial\Delta_h} = \underbrace{\begin{bmatrix} 0 \\ -\lambda_p(1-\phi_h)g(\Delta_h) \\ \lambda_s(1-\phi_h)(1-\theta) \end{bmatrix}}_{\text{vector } \mathrm{b}_h}.$$

and

$$\underbrace{\frac{\partial F}{\partial(\phi_l,\phi_h,\Delta_l)}}_{\text{matrix A}} \begin{bmatrix} \partial\phi_l/\partial\theta \\ \partial\phi_h/\partial\theta \\ \partial\Delta_l/\partial\theta \end{bmatrix} = -\frac{\partial F}{\partial\theta} = \underbrace{\begin{bmatrix} 0 \\ 0 \\ -\lambda_s(1-\phi_h)(\Delta_h-\Delta_l) \end{bmatrix}}_{\text{vector } \mathrm{b}_\theta}.$$

We can solve the systems using Cramer's rule.

- Label $D_{\phi_l}^{\Delta_h}$ the determinant of A after replacing the first column of A with b_h.

$$\begin{aligned}D_{\phi_l}^{\Delta_h} = & -\lambda_p(1-\phi_l)g(\Delta_l) \times \lambda_p(1-\phi_h)g(\Delta_h) \times \lambda_s(1-\theta)(\Delta_h-\Delta_l) \\ & + \lambda_p(1-\phi_l)g(\Delta_l) \times [\lambda_p G(\Delta_h) + \mu + \lambda_s\phi_l] \times \lambda_s(1-\phi_h)(1-\theta) > 0\end{aligned}$$

From Cramer's rule $\partial\phi_l/\partial\Delta_h = D_{\phi_l}^{\Delta_h}/D > 0$. Label $D_{\phi_l}^{\theta}$ the determinant of A after replacing the first column of A with b_θ.

$$D_{\phi_l}^{\theta} = -\lambda_p(1-\phi_l)g(\Delta_l) \times [\lambda_p G(\Delta_h) + \mu + \lambda_s\phi_l] \times \lambda_s(1-\phi_h)(\Delta_h-\Delta_l) < 0$$

From Cramer's rule $\partial\phi_l/\partial\theta = D_{\phi_l}^{\theta}/D < 0$. Since $\partial\phi_l/\partial\Delta_h$ is positive and $\partial\phi_l/\partial\theta$ is negative, we can conclude that $\phi_l^A > \phi_l^B$.

- Label $D_{\phi_h}^{\Delta_h}$ the determinant of A after replacing the second column of A with b_h.

$$\begin{aligned}D_{\phi_h}^{\Delta_h} = & [\lambda_p G(\Delta_l) + \mu + \lambda_s(1-\phi_h)] \times \lambda_p(1-\phi_h)g(\Delta_h) \times [r + \mu + \lambda_p G(\Delta_l) + \lambda_s(1-\phi_h)(1-\theta)] \\ & + \lambda_p(1-\phi_l)g(\Delta_l) \times \lambda_s(1-\phi_h) \times \lambda_s(1-\phi_h)(1-\theta) > 0\end{aligned}$$

From Cramer's rule $\partial\phi_h/\partial\Delta_h = D_{\phi_h}^{\Delta_h}/D > 0$. Label $D_{\phi_h}^{\theta}$ the determinant of A after replacing the second column of A with b_θ.

$$D_{\phi_l}^{\theta} = -\lambda_p(1-\phi_l)g(\Delta_l) \times \lambda_s(1-\phi_h) \times \lambda_s(1-\phi_h)(\Delta_h-\Delta_l) < 0$$

From Cramer's rule $\partial\phi_h/\partial\theta = D_{\phi_l}^{\theta}/D < 0$. Since $\partial\phi_h/\partial\Delta_h$ is positive and $\partial\phi_h/\partial\theta$ is negative, we can conclude that $\phi_h^A > \phi_h^B$.

- Label $D_{\Delta_l}^{\Delta_h}$ the determinant of A after replacing the third column of A with b_h.

$$\begin{aligned}D_{\Delta_l}^{\Delta_h} = & [\lambda_p G(\Delta_l) + \mu + \lambda_s(1-\phi_h)] \times [\lambda_p G(\Delta_h) + \mu + \lambda_s\phi_l] \times \lambda_s(1-\phi_h)(1-\theta) \\ & - \lambda_s\phi_l \times \lambda_s(1-\phi_h) \times \lambda_s(1-\phi_h)(1-\theta) \\ & - [\lambda_p G(\Delta_l) + \mu + \lambda_s(1-\phi_h)] \times \lambda_p(1-\phi_h)g(\Delta_h) \times \lambda_s(1-\theta)(\Delta_h-\Delta_l) \\ = & [\lambda_p G(\Delta_l) + \mu + \lambda_s(1-\phi_h)] \times [\mu + \lambda_s\phi_l] \times \lambda_s(1-\phi_h)(1-\theta) \\ & - \lambda_s\phi_l \times \lambda_s(1-\phi_h) \times \lambda_s(1-\phi_h)(1-\theta) \\ & + [\lambda_p G(\Delta_l) + \mu + \lambda_s(1-\phi_h)] \times \lambda_p G(\Delta_h) \times \lambda_s(1-\phi_h)(1-\theta) \\ & - [\lambda_p G(\Delta_l) + \mu + \lambda_s(1-\phi_h)] \times \lambda_p(1-\phi_h) \times \lambda_s(1-\theta)g(\Delta_h)\Delta_h \\ & + [\lambda_p G(\Delta_l) + \mu + \lambda_s(1-\phi_h)] \times \lambda_p(1-\phi_h)g(\Delta_h) \times \lambda_s(1-\theta)\Delta_l \\ = & [\lambda_p G(\Delta_l) + \mu + \lambda_s(1-\phi_h)] \times [\mu + \lambda_s\phi_l] \times \lambda_s(1-\phi_h)(1-\theta)\end{aligned}$$

$$
\begin{aligned}
&- \lambda_s\phi_l \times \lambda_s(1-\phi_h) \times \lambda_s(1-\phi_h)(1-\theta) \\
&+ [\lambda_p G(\Delta_l) + \mu + \lambda_s(1-\phi_h)] \times \cancel{\lambda_p\lambda_s(1-\theta)(1-\phi_h) \times G(\Delta_h)} \\
&- [\lambda_p G(\Delta_l) + \mu + \lambda_s(1-\phi_h)] \times \cancel{\lambda_p\lambda_s(1-\theta)(1-\phi_h) \times G(\Delta_h)} \\
&+ [\lambda_p G(\Delta_l) + \mu + \lambda_s(1-\phi_h)] \times \lambda_p(1-\phi_h)g(\Delta_h) \times \lambda_s(1-\theta)\Delta_l \\
= {} & [\lambda_p G(\Delta_l) + \mu + \lambda_s(1-\phi_h)] \times [\mu + \lambda_s\phi_l] \times \lambda_s(1-\phi_h)(1-\theta) \\
&- \lambda_s\phi_l \times \lambda_s(1-\phi_h) \times \lambda_s(1-\phi_h)(1-\theta) \\
&+ [\lambda_p G(\Delta_l) + \mu + \lambda_s(1-\phi_h)] \times \lambda_p(1-\phi_h)g(\Delta_h) \times \lambda_s(1-\theta)\Delta_l \\
= {} & [\lambda_p G(\Delta_l) + \mu + \lambda_s(1-\phi_h)] \times \mu \times \lambda_s(1-\phi_h)(1-\theta) \\
&+ [\lambda_p G(\Delta_l) + \mu] \times \lambda_s\phi_l \times \lambda_s(1-\phi_h)(1-\theta) \\
&+ \lambda_s\phi_l \times \cancel{\lambda_s(1-\phi_h)} \times \lambda_s(1-\phi_h)(1-\theta) \\
&- \lambda_s\phi_l \times \cancel{\lambda_s(1-\phi_h)} \times \lambda_s(1-\phi_h)(1-\theta) \\
&+ [\lambda_p G(\Delta_l) + \mu + \lambda_s(1-\phi_h)] \times \lambda_p(1-\phi_h)g(\Delta_h) \times \lambda_s(1-\theta)\Delta_l \\
= {} & [\lambda_p G(\Delta_l) + \mu + \lambda_s(1-\phi_h)] \times \mu \times \lambda_s(1-\phi_h)(1-\theta) \\
&+ [\lambda_p G(\Delta_l) + \mu] \times \lambda_s\phi_l \times \lambda_s(1-\phi_h)(1-\theta) \\
&+ [\lambda_p G(\Delta_l) + \mu + \lambda_s(1-\phi_h)] \times \lambda_p(1-\phi_h)g(\Delta_h) \times \lambda_s(1-\theta)\Delta_l > 0
\end{aligned}
$$

From Cramer's rule ${}^{\partial\Delta_l}/_{\partial\Delta_h} = {}^{D^{\Delta_h}_{\Delta_l}}/_{D} > 0$. Label $D^{\theta}_{\Delta_l}$ the determinant of the matrix A after replacing the third column of A with the vector b_θ.

$$
\begin{aligned}
D^{\theta}_{\Delta_l} = {} & - [\lambda_p G(\Delta_l) + \mu + \lambda_s(1-\phi_h)] \times [\lambda_p G(\Delta_h) + \mu + \lambda_s\phi_l] \times \lambda_s(1-\phi_h)(\Delta_h - \Delta_l) \\
&+ \lambda_s\phi_l \times \lambda_s(1-\phi_h) \times \lambda_s(1-\phi_h)(\Delta_h - \Delta_l) \\
= {} & - [\lambda_p G(\Delta_l) + \mu] \times [\lambda_p G(\Delta_h) + \mu + \lambda_s\phi_l] \times \lambda_s(1-\phi_h)(\Delta_h - \Delta_l) \\
&- \lambda_s(1-\phi_h) \times [\lambda_p G(\Delta_h) + \mu + \lambda_s\phi_l] \times \lambda_s(1-\phi_h)(\Delta_h - \Delta_l) \\
&+ \lambda_s\phi_l \times \lambda_s(1-\phi_h) \times \lambda_s(1-\phi_h)(\Delta_h - \Delta_l) \\
= {} & - [\lambda_p G(\Delta_l) + \mu] \times [\lambda_p G(\Delta_h) + \mu + \lambda_s\phi_l] \times \lambda_s(1-\phi_h)(\Delta_h - \Delta_l) \\
&- \lambda_s(1-\phi_h) \times [\lambda_p G(\Delta_h) + \mu] \times \lambda_s(1-\phi_h)(\Delta_h - \Delta_l) \\
&- \lambda_s(1-\phi_h) \times \lambda_s\phi_l \times \cancel{\lambda_s(1-\phi_h)(\Delta_h - \Delta_l)} \\
&+ \lambda_s\phi_l \times \lambda_s(1-\phi_h) \times \cancel{\lambda_s(1-\phi_h)(\Delta_h - \Delta_l)} \\
= {} & - [\lambda_p G(\Delta_l) + \mu] \times [\lambda_p G(\Delta_h) + \mu + \lambda_s\phi_l] \times \lambda_s(1-\phi_h)(\Delta_h - \Delta_l) \\
&- \lambda_s(1-\phi_h) \times [\lambda_p G(\Delta_h) + \mu] \times \lambda_s(1-\phi_h)(\Delta_h - \Delta_l) < 0
\end{aligned}
$$

From Cramer's rule ${}^{\partial\Delta_l}/_{\partial\theta} = {}^{D^{\theta}_{\Delta_l}}/_{D} < 0$. Since ${}^{\partial\Delta_l}/_{\partial\Delta_h}$ is positive and ${}^{\partial\Delta_l}/_{\partial\theta}$ is negative, we can conclude that $\Delta_l^A > \Delta_l^B$.

A.3 Proof of Lemma 2

The Hamiltonian of the planner's problem is

$$H = \phi_l v_l + \phi_h v_h - \lambda_p \left[(\pi_l - \phi_l) \int_0^{c_l} c g(c) dc + (\pi_h - \phi_h) \int_0^{c_h} c g(c) dc \right]$$
$$+ \gamma_l \Big\{ \lambda_p G(c_l)(\pi_l - \phi_l) + \lambda_s \phi_h p_{hl} (\pi_l - \phi_l) - \mu \phi_l - \lambda_s (\pi_h - \phi_h) p_{lh} \phi_l \Big\}$$
$$+ \gamma_h \Big\{ \lambda_p G(c_h)(\pi_h - \phi_h) + \lambda_s \phi_l p_{lh} (\pi_h - \phi_h) - \mu \phi_h - \lambda_s (\pi_l - \phi_l) p_{hl} \phi_h \Big\},$$

where γ_l and γ_h are the co-state variables associated with the constraints on the law of motion of the distribution of investors (1) and (2). First, consider the first-order conditions with respect to c_l and c_h;

$$0 = \frac{\partial H}{\partial c_l} = -\lambda_p (\pi_l - \phi_l) g(c_l) c_l + \lambda_p (\pi_l - \phi_l) g(c_l) \gamma_l \quad \text{and}$$
$$0 = \frac{\partial H}{\partial c_h} = -\lambda_p (\pi_h - \phi_h) g(c_h) c_h + \lambda_p (\pi_h - \phi_h) g(c_h) \gamma_h.$$

It is immediate from these conditions that $\gamma_l = c_l$ and $\gamma_h = c_h$. The planner equalizes the marginal cost of issuing assets to either low or high types to the shadow prices or marginal social gains associated with the Lagrange multipliers of the feasibility constraints (1)-(2). The terms coinciding with the mass of agents affected cancel out in both the gain and cost. Now, consider the first-order conditions with respect to ϕ_l and ϕ_h;

$$r\gamma_l = \dot{\gamma}_l + \frac{\partial H}{\partial \phi_l} = \dot{\gamma}_l + v_l - \mu \gamma_l$$
$$- \lambda_p \int_0^{\gamma_l} (\gamma_l - c) g(c) dc + (\gamma_h - \gamma_l) \lambda_s \left[\phi_h p_{hl} + (\pi_h - \phi_h) p_{lh} \right],$$
$$r\gamma_h = \dot{\gamma}_h + \frac{\partial H}{\partial \phi_h} = \dot{\gamma}_h + v_h - \mu \gamma_h$$
$$- \lambda_p \int_0^{\gamma_h} (\gamma_h - c) g(c) dc - (\gamma_h - \gamma_l) \lambda_s \left[\phi_l p_{lh} + (\pi_l - \phi_l) p_{hl} \right],$$

where we have used the fact that $c_i = \gamma_i$ for $i \in \{l, h\}$. The above two equations imply that $\gamma_l < \gamma_h$. To see that this is the case, note that we can write the difference $\gamma_h - \gamma_l$ as

$$r(\gamma_h - \gamma_l) = (\dot{\gamma}_h - \dot{\gamma}_l) + v_h - v_l - \mu(\gamma_h - \gamma_l)$$

$$
\begin{aligned}
&- \lambda_p \left[\int_0^{\gamma_h} (\gamma_h - c) g(c) dc - \int_0^{\gamma_l} (\gamma_l - c) g(c) dc \right] \\
&- (\gamma_h - \gamma_l) \lambda_s \left[(\pi_l - \phi_l) p_{hl} + \phi_l p_{lh} \right] - (\gamma_h - \gamma_l) \lambda_s \left[\phi_h p_{hl} + (\pi_h - \phi_h) p_{lh} \right].
\end{aligned}
$$

Suppose to the contrary that $\gamma_l \geq \gamma_h$. Then the left hand side of the above equation would be smaller or equal to zero. However, in order to have the right hand side smaller or equal to zero, $\dot{\gamma}_h - \dot{\gamma}_l$ would have to be strictly negative. Since $\gamma_h \geq 0$, this would imply that γ_l needs to converge to infinity at a rate higher than r, which would imply an explosive path for γ_l (that is, $\lim e^{-rt} \gamma_l = \infty$), a violation of the transversality condition. Hence, $\gamma_l < \gamma_h$. Finally, consider the first-order conditions with respect to the trade probabilities, p_{hl} and p_{lh};

$$
\begin{aligned}
\frac{\partial H}{\partial p_{lh}} &= -\gamma_l \lambda_s \phi_l (\pi_h - \phi_h) + \gamma_h \lambda_s \phi_l (\pi_h - \phi_h) \geq 0 \\
\frac{\partial H}{\partial p_{hl}} &= \gamma_l \lambda_s (\pi_l - \phi_l) \phi_h - \gamma_h \lambda_s (\pi_l - \phi_l) \phi_h \leq 0.
\end{aligned}
$$

with complementary slackness with multipliers of constraints that $p_{ij} \in [0, 1]$ for $i, j = l, h$. Notice, since $\gamma_h > \gamma_l$, that the inequalities must hold with a strict inequality implying that $p_{lh} = 1$ and $p_{hl} = 0$. In words, it is never optimal for the planner to reallocate assets from high-type investors to low-type investors since the shadow value of high types is always strictly larger than that of low types.

A.4 Proof of Proposition 2

To see this is the case, let an asset allocation $\{\boldsymbol{\phi}, \boldsymbol{c}, \boldsymbol{p}\}$ combined with reservation values Δ_l and Δ_h be a decentralized equilibrium and assume, by way of contradiction, that this asset allocation solves the planner's problem. By Proposition 2, there exist co-state variables γ_l and γ_h such that γ_l, γ_h, ϕ_l, and ϕ_h solve the system of differential equations (14)-(17). By Definition 1, Δ_l, Δ_h, ϕ_l, and ϕ_h solve the system of differential equations (9)-(12). Equations (15) and (10) imply that

$$
\begin{aligned}
\dot{\phi}_l &= \lambda_p (\pi_l - \phi_l) G(\gamma_l) - \mu \phi_l - \lambda_s \phi_l (\pi_h - \phi_h) && (32) \\
&= \lambda_p (\pi_l - \phi_l) G(\Delta_l) - \mu \phi_l - \lambda_s \phi_l (\pi_h - \phi_h), && (33)
\end{aligned}
$$

or $\gamma_l = \Delta_l$. Analogously, equations (17) and (12) imply that

$$
\dot{\phi}_h = \lambda_p (\pi_h - \phi_h) G(\gamma_h) - \mu \phi_h + \lambda_s \phi_l (\pi_h - \phi_h) \quad (34)
$$

$$= \lambda_p(\pi_h - \phi_h)G(\Delta_h) - \mu\phi_h + \lambda_s\phi_l(\pi_h - \phi_h), \tag{35}$$

or $\gamma_h = \Delta_h$. The above two results, together with (14)-(16) and (9)-(11), imply that

$$\begin{aligned}
r\gamma_l &= \dot{\gamma}_l + \nu_l - \mu\gamma_l - \lambda_p \int_0^{\Delta_l} (\Delta_l - c)g(c)dc + \lambda_s(\pi_h - \phi_h)(\gamma_h - \gamma_l) \\
&= \dot{\gamma}_l + \nu_l - \mu\gamma_l - \lambda_p\theta_a \int_0^{\Delta_l} (\Delta_l - c)g(c)dc + \lambda_s(\pi_h - \phi_h)(1-\theta)(\gamma_h - \gamma_l), \quad \text{and} \\
r\gamma_h &= \dot{\gamma}_h + \nu_h - \mu\gamma_h - \lambda_p \int_0^{\Delta_h} (\Delta_h - c)g(c)dc - \lambda_s\phi_l(\gamma_h - \gamma_l) \\
&= \dot{\gamma}_h + \nu_h - \mu\gamma_h - \lambda_p\theta_a \int_0^{\Delta_h} (\Delta_h - c)g(c)dc - \lambda_s\phi_l\theta(\gamma_h - \gamma_l).
\end{aligned}$$

If our candidate asset allocation coincides with a decentralized equilibrium, then we must be able to find constants θ_a and θ that solve (32)-(34). We can write the system as

$$\begin{aligned}
&\lambda_p \int_0^{\Delta_l} (\Delta_l - c)g(c)dc\theta_a + \lambda_s(\pi_h - \phi_h)(\gamma_h - \gamma_l)\theta = \lambda_p \int_0^{\Delta_l} (\Delta_l - c)g(c)dc \quad \text{and} \\
&\lambda_p\theta_a \int_0^{\Delta_h} (\Delta_h - c)g(c)dc\theta_a + \lambda_s\phi_l(\gamma_h - \gamma_l)\theta = \lambda_p \int_0^{\Delta_h} (\Delta_h - c)g(c)dc + \lambda_s\phi_l(\gamma_h - \gamma_l).
\end{aligned}$$

Additionally, since θ_a and θ are bargaining powers, we must have that $\theta_a, \theta \in [0,1]$. Moreover, it is easy to show that a solution of the differential equations discussed above cannot have either $\gamma_l = 0$, $\gamma_h = 0$, $\phi_h = \pi_h$, or $\phi_l = 0$ for all but a measure zero of period t's. Therefore, $\lambda_p \int_0^{\Delta_h} (\Delta_h - c)g(c)dc$ and $\lambda_s\phi_l(\gamma_h - \gamma_l)$ most both be strictly positive and the only $\theta_a, \theta \in [0,1]$ that satisfy the second equation in the above system are $\theta_a = \theta = 1$. But note that $\theta_a = \theta = 1$ does not solve the first equation in this system. Which contradicts that the asset allocation solves the planner's problem.

A.5 Proof of Proposition 3

The proof of proposition 3 is analogous to the proof of proposition 1, and we omit it here.

A.6 Proof of Proposition 4

We know by Lemma 2 that $\{\phi_l^*, \phi_h^*, \gamma_l, \gamma_h\}$ solves the differential equations (14)-(17). Then, after replacing $\boldsymbol{\tau}$, we can see that $\{\phi_l^*, \phi_h^*, \gamma_l, \gamma_h\}$ also solves the decentralized equilibrium with intervention equations (19)-(22). And $\boldsymbol{\tau}$ is feasible because $2\bar{\tau} = \phi_l\tau_l - \phi_h\tau_h$.

B A model with a continuum of investor types

In this section we augment the economy to allow for a continuum of investor types, as studied in Hugonnier et al. (2014). Unlike the simple model with two types of investors, here all investors buy and sell assets in the secondary market, thus all of them act as intermediaries. Because investors serve on both sides of the market, all of them fail to internalize the full gains from trade when buying and selling, adding a new layer of complexity with respect to the simple model with two types where each type of investor failed to internalize the full gains from trade of either selling or buying assets. We show that the decentralized equilibrium is always inefficient, provided that the gains from a particular trade in the secondary market have to be fully split by the meeting participants. We also study an extension of the model where we allow the bargaining power to be fully dependent on the identities of trade participants in a given meeting, and we show that the decentralized equilibrium continues to remain inefficient.

We call an investor not holding an asset a non-owner investor (subscript n in the equations below), and we call an investor holding an asset an owner investor (subscript o in the equations below). There is a measure two of investors, and let $F(\nu)$ denote the cumulative distribution of investor types, with support $[\underline{\nu}, \bar{\nu}]$. Likewise, we use $\Phi(\nu)$ to denote the cumulative distribution of owner investors, with $\Phi(\nu) \leq F(\nu)$ for all ν, with $\phi(\nu) \equiv \partial\Phi(\nu)/\partial\nu$. Further, let $p^a(c, \nu_n)$ denote the probability that an issuer with issuance cost c trades with a non owner investor of type ν_n at the primary market, and let $p^b(\nu_o, \nu_n)$ denote the probability that an owner investor of type ν_o trades with a non owner investor of type ν_n at the secondary market.

The evolution of the distribution $\Phi(\nu)$ is given by

$$\begin{aligned}\dot{\Phi}(\nu) = & -\mu\Phi(\nu) - \lambda_s \int_{\underline{\nu}}^{\nu} \int_{\nu}^{\bar{\nu}} p^b(\nu_o, \nu_n)[f(\nu_n) - \phi(\nu_n)]\phi(\nu_o) d\nu_o d\nu_n \\ & + \lambda_s \int_{\nu}^{\bar{\nu}} \int_{\underline{\nu}}^{\nu} p^b(\nu_o, \nu_n)[f(\nu_n) - \phi(\nu_n)]\phi(\nu_o) d\nu_o d\nu_n \\ & + \lambda_p \int_{\underline{c}}^{\bar{c}} p^a(c, \nu_n)[f(\nu_n) - \phi(\nu_n)]g(c) dc d\nu_n \ .\end{aligned}$$

The equation states that the fraction of owner investors of type less or equal to ν holding assets suffers an outflow in two ways. First, assets can mature. Second, owner investors with type $\nu_o \leq \nu$ can sell it to a non-owner investor satisfying $\nu_n > \nu$. Likewise, the fraction of owner investors of type less or equal to ν holding assets gets an inflow also in two ways. First, a non-owner investor of type $\nu_n \leq \nu$ can buy the asset in the primary market from an issuer. Second, a non-owner investor of type $\nu_n \leq \nu$ buys the asset from an owner investor of type $\nu_o > \nu$. Notice that this last case will not be a feature of the decentralized equilibrium nor of the efficient allocation. It proves useful to provide the evolution of the density function $\phi(\nu)$,

$$\begin{aligned}\dot{\phi}(\nu) = & -\mu\phi(\nu) - \lambda_s \int_{\nu}^{\bar{\nu}} [f(\nu_n) - \phi(\nu_n)]\phi(\nu) d\nu_n + \lambda_s \int_{\underline{\nu}}^{\nu} p^b(\nu_o, \nu)[f(\nu) - \phi(\nu)]\phi(\nu_o) d\nu_o \\ & - \lambda_s \int_{\underline{\nu}}^{\nu} p^b(\nu, \nu_n)[f(\nu_n) - \phi(\nu_n)]\phi(\nu) d\nu_n + \lambda_s \int_{\nu}^{\bar{\nu}} p^b(\nu_o, \nu)[f(\nu) - \phi(\nu)]\phi(\nu_o) d\nu_o \\ & + \lambda_p \int_{\underline{c}}^{\bar{c}} p^a(c, \nu)[f(\nu) - \phi(\nu)] g(c) dc \ . \end{aligned} \tag{36}$$

We now solve for the decentralized equilibrium under the following two assumptions: (i) $p^a(c, \nu_n) = 1$ if $c \leq \Delta(\nu_n)$ and $p^a(c, \nu_n) = 0$ otherwise, (ii) $p^b(\nu_o, \nu_n) = 1$ if $\nu_o \leq \nu_n$ and $p^b(\nu_o, \nu_n) = 0$ otherwise. Later we verify that these two assumptions are satisfied in the decentralized equilibrium. Under these assumptions, the equilibrium exhibits a simple pattern of trade in the both primary and secondary markets. In the primary market, issuers issue whenever they find a non owner investor with reservation value $\Delta(\nu_n)$ above their cost of issuance c. In the secondary market, owner investors of of type ν_n sell to any non-owner investor they encounter in the secondary market, as long as $\nu_o > \nu_n$. These trade patterns are the analogous ones to those obtained in the simple model with only two types of investors. Using this observation, the value functions for a non-owner and owner investor of type ν are given by

$$\begin{aligned} rV_n(\nu) &= \dot{V}_n(\nu) + \lambda_p \int_{\underline{c}}^{\Delta(\nu)} [\Delta(\nu) - c] g(c) dc + \lambda_s \theta \int_{\underline{\nu}}^{\nu} [\Delta(\nu) - \Delta(\nu_o)]\phi(\nu_o) d\nu_o \ , \\ rV_o(\nu) &= \dot{V}_o(\nu) + \nu - \mu\Delta(\nu) + \lambda_s(1-\theta) \int_{\nu}^{\bar{\nu}} [\Delta(\nu_n) - \Delta(\nu)]\{f(\nu_n) - \phi(\nu_n)\} d\nu_n \ , \end{aligned}$$

where $\Delta(\nu) \equiv V_o(\nu) - V_n(\nu)$. Using these expressions we can derive an expression for $\Delta(\nu)$,

$$\begin{aligned} (r+\mu)\Delta(\nu) = \dot{\Delta}(\nu) + \nu - \lambda_p \int_{\underline{c}}^{\Delta(\nu)} [\Delta(\nu) - c] g(c) dc \\ + \lambda_s(1-\theta) \int_{\nu}^{\bar{\nu}} [\Delta(\nu_n) - \Delta(\nu)]\{f(\nu_n) - \phi(\nu_n)\} d\nu_n \\ - \lambda_s \theta \int_{\underline{\nu}}^{\nu} [\Delta(\nu) - \Delta(\nu_o)]\phi(\nu_o) d\nu_o \ . \end{aligned} \tag{37}$$

It is easy to show that $\Delta(\nu)$ is increasing in ν, which validates the assumption made above regarding the trade patterns in the economy.[15] The expression for the reservation value is analogous to that one obtained for the model with two investor types, as presented in equations (9) and (11). The only difference is that while in the simple model low-valuation investors were sellers in the secondary market and high-valuation investors were buyers in the secondary market, in the model with continuum of types all investors are both buyers and sellers in the secondary market. Thus, all of them serve a role as intermediators, while that role was only assigned to the low-valuation investors in the model with two investor types. We next define a decentralized equilibrium in the economy with a continuum of investor types,

Definition 4. *A decentralized equilibrium is a set of trading protocols $p^a(c, \nu_n), p^b(\nu_o, \nu_n)$ for all $c \in [\underline{c}, \bar{c}]$, $\nu_n \in [\underline{\nu}, \bar{\nu}]$ and $\nu_o \in [\underline{\nu}, \bar{\nu}]$, an asset allocation and bounded reservation values for investors $\{\phi(v), \Delta(\nu)\}$ for all $\nu \in [\underline{\nu}, \bar{\nu}]$, that solve the system of differential equations given by equations (36) and (37), with initial conditions $\phi(\nu) = \phi^0(\nu)$, for all $\nu \in [\underline{\nu}, \bar{\nu}]$. The trading protocols satisfy (i) $p^a(c, \nu_n) = 1$ if $c \leq \Delta(\nu_n)$ and $p^a(c, \nu_n) = 0$ otherwise, and (ii) $p^b(\nu_o, \nu_n) = 1$ if $\nu_o \leq \nu_n$ and $p^b(\nu_o, \nu_n) = 0$ otherwise.*

We now solve for the efficient allocation. The efficient allocation solves the following problem,

$$\max_{\phi, p^a, p^b} \int_0^\infty e^{-rt} \left\{ \int_{\underline{\nu}}^{\bar{\nu}} \nu\phi(\nu)d\nu - \lambda_p \int_{\underline{\nu}}^{\bar{\nu}} \int_{\underline{c}}^{\bar{c}} p^a(c, \nu)c\{f(\nu) - \phi(\nu)\}g(c)dcd\nu \right\} dt ,$$

subject to equation (36) for all $\nu \in [\underline{\nu}, \bar{\nu}]$ and $\int_{\underline{\nu}}^{\bar{\nu}} \phi(\nu)d\nu = 1$. We solve for the efficient allocation by forming the Hamiltonian of the problem, $\mathscr{H}$. We use $\gamma(\nu)$ as the co-state variable for equation (36), and $\bar{\gamma}$ as the multiplier of the second restriction (i.e. the asset density must add up to one).

We begin by studying the optimal choices for the controls p^a and p^b. Differentiating the Hamiltonian with respect to $p^a(c, \nu)$ provides

$$\frac{\partial \mathscr{H}}{\partial p^a(c, \nu)} = \lambda_p[f(\nu) - \phi(\nu)]\{\gamma(\nu) - c\}g(c)dcd\nu .$$

This expression is positive if $\gamma(\nu) \geq c$ and negative otherwise. This shows that the efficient allocation requires that an issuer issues whenever they encounter a non-owner investor with co-state variable $\gamma(\nu)$ above his cost c. That is, $p^a(c, \nu) = 1$ if $\gamma(\nu) \geq c$, and $p^a(c, \nu) = 0$

[15] The proof is analogous to the proof of Lemma 1, and thus it is not provided.

if $\gamma(\nu) < c$. Likewise, differentiating the Hamiltonian with respect to $p^b(\nu_0, \nu_n)$ provides

$$\frac{\partial \mathcal{H}}{\partial p^b(\nu_o, \nu_n)} = \lambda_s [f(\nu_n) - \phi(\nu_n)] \phi(\nu_o) \{\gamma(\nu_n) - \gamma(\nu_o)\} d\nu_o \nu_n \ ,$$

which is positive if $\gamma(\nu_n) \geq \gamma(\nu_o)$. This shows that the efficient allocation requires that an owner investor of type ν_o sells whenever he encounters a non-owner investor with co-state variable $\gamma(\nu_n)$ above his co-state variable $\gamma(\nu_n)$.

For the co-state variable $\gamma(\nu)$ we use that optimal control requires $\partial \mathcal{H} / \partial \gamma(\nu) = r\gamma(\nu) - \dot{\gamma}(\nu)$. After operating with this expression, we obtain the following expression,

$$\begin{aligned}(r + \mu)\gamma(\nu) =& \dot{\gamma}(\nu) + \nu - \lambda_p \int_{\underline{c}}^{\gamma(\nu)} [\gamma(\nu) - c] g(c) dc \\ &+ \lambda_s \int_{\nu}^{\bar{\nu}} [\gamma(\nu_n) - \gamma(\nu)] \{f(\nu_n) - \phi(\nu_n)\} d\nu_n \\ &- \lambda_s \int_{\underline{\nu}}^{\nu} [\gamma(\nu) - \gamma(\nu_o)] \phi(\nu_o) d\nu_o \ . \end{aligned} \tag{38}$$

The next definition describes an efficient allocation.

Definition 5. *An efficient allocation is a set of trading protocols $p^a(c, \nu_n), p^b(\nu_o, \nu_n)$ for all $c \in [\underline{c}, \bar{c}]$, $\nu_n \in [\underline{\nu}, \bar{\nu}]$, and $\nu_o \in [\underline{\nu}, \bar{\nu}]$, an asset allocation and co-state variables for investors $\{\phi(v), \gamma(\nu)\}$ for all $\nu \in [\underline{\nu}, \bar{\nu}]$, that solve the system of differential equations given by equations (36) and (38), with initial conditions $\phi(\nu) = \phi^0(\nu)$, for all $\nu \in [\underline{\nu}, \bar{\nu}]$. The trading protocols satisfy (i) $p^a(c, \nu_n) = 1$ if $c \leq \Delta(\nu_n)$ and $p^a(c, \nu_n) = 0$ otherwise, and (ii) $p^b(\nu_o, \nu_n) = 1$ if $\nu_o \leq \nu_n$ and $p^b(\nu_o, \nu_n) = 0$ otherwise.*

Since in a model with a continuum of types, all investors resale assets, the inefficiency appears twice for each investor, rather than once as in the model with two types. In the simpler model, low-valuation investors fail to internalize the full gains from trade when selling, while the high-valuation investors failed to internalize the full gains from trade when buying. In the model with a continuum of types, each type of investor fails to internalize both sources of gains from trade, as all of them buy and sell assets in the secondary market.

Lemma 3. *The decentralized-equilibrium is never efficient.*

Lemma 3 extends the inefficiency result obtained in the simple model to the model with a continuum of types. As in the simple model, the nature of the lack of efficiency stems from the failure of the decentralized equilibrium to internalize the full gains from trade when investors are buying and selling assets in the secondary market. At a superficial level, there seems to be a way to split the trade surplus that allows investors of type ν to fully internalize the gains from trade. Under Nash bargaining the split of trade surplus is governed by the bargaining power θ. This choice of bargaining power must be such that $\theta = \theta(\nu)$, so that different investor types have different bargaining powers, and thus internalize different shares of the trade surplus. However, when a seller of type ν and buyer of type $\tilde{\nu} \geq \nu$ trade, their bargaining powers must add up to one–in other words, the sum of surplus that each party internalizes must add up to the total trade surplus. That is, the surplus generated by the trade $\Delta(\tilde{\nu}) - \Delta(\nu)$ is fully divided by the trade participants. This restriction guarantees that it is not possible to choose a set of bargaining powers that depend on the investor type to make the decentralized equilibrium is efficient.

B.1 Trading partners

A natural conjecture that follows from the previous finding is that if the bargaining power were to be allowed to depend on the identity of both parties in a trade meeting, $\theta(\nu_o, \nu_n)$, the decentralized equilibrium could be made efficient for the appropriate choice of bargaining weights. With this in mind we augment the model to allow for the bargaining power to depend on the identity of trade participants. In particular, at the secondary market, for any seller of type ν_o and any buyer of type $\nu_n \geq \nu_o$ for all $\nu_o \in [\underline{\nu}, \bar{\nu}]$, let $\theta = \theta(\nu_o, \nu_n) \in [0, 1]$ denote the bargaining power of the buyer in a meeting between these two investors.

In terms of the decentralized equilibrium, the differential equation for the reservation value presented in equation (37) is now given by

$$\begin{aligned}(r+\mu)\Delta(\nu) = \dot{\Delta}(\nu) + \nu - \lambda_p \int_{\underline{c}}^{\Delta(\nu)} [\Delta(\nu) - c] g(c) dc \\ + \lambda_s \int_{\nu}^{\bar{\nu}} \{1 - \theta(\nu, \nu_n)\}[\Delta(\nu_n) - \Delta(\nu)]\{f(\nu_n) - \phi(\nu_n)\} d\nu_n \\ - \lambda_s \int_{\underline{\nu}}^{\nu} \theta(\nu_o, \nu)[\Delta(\nu) - \Delta(\nu_o)]\phi(\nu_o) d\nu_o \, . \end{aligned} \tag{39}$$

All the other equations for the decentralized equilibrium remain unchanged.

Efficiency of the decentralized equilibrium requires that $\{\phi(\nu), \Delta(\nu)\}$ is also a solution to the efficient allocation problem. As before, this reduces to checking whether there is a way to choose $\theta(\nu_o, \nu_n)$ that makes equation (39) identical to (38). This reduces to

$$\begin{aligned}&\int_{\nu}^{\bar{\nu}} \{1 - \theta(\nu, \nu_n)\}[\Delta(\nu_n) - \Delta(\nu)]\{f(\nu_n) - \phi(\nu_n)\} d\nu_n - \int_{\underline{\nu}}^{\nu} \theta(\nu_o, \nu)[\Delta(\nu) - \Delta(\nu_o)]\phi(\nu_o) d\nu_o \\ &= \int_{\nu}^{\bar{\nu}} [\Delta(\nu_n) - \Delta(\nu)]\{f(\nu_n) - \phi(\nu_n)\} d\nu_n - \int_{\underline{\nu}}^{\nu} [\Delta(\nu) - \Delta(\nu_o)]\phi(\nu_o) d\nu_o \, . \end{aligned} \tag{40}$$

This condition has to be satisfied for every $\nu \in [\underline{\nu}, \bar{\nu}]$. In particular, it has to be satisfied for $\nu = \underline{\nu}$ and $\nu = \bar{\nu}$. For investors with the lowest valuation $\nu = \underline{\nu}$, the condition reduces to

$$\int_{\underline{\nu}}^{\bar{\nu}} \theta(\underline{\nu}, \nu_n)[\Delta(\nu_n) - \Delta(\underline{\nu})]\{f(\nu_n) - \phi(\nu_n)\}d\nu_n = 0 \,. \tag{41}$$

Thus, $\theta(\underline{\nu}, \nu_n) = 0$ for all $\nu_n \in [\underline{\nu}, \bar{\nu}]$ as $[\Delta(\nu_n) - \Delta(\underline{\nu})]\{f(\nu_n) - \phi(\nu_n)\} > 0$ for all $\nu_n \in [\underline{\nu}, \bar{\nu}]$. That is, every trade in the secondary market that includes the lowest-valuation investor must assign all the gains from trade to the seller. This occurs because for an investor of type $\underline{\nu}$, trading in the secondary market only involves selling an asset, and thus the investor fails to internalize the full gains from trade when selling, which can only be corrected by giving investors of type $\underline{\nu}$ the full bargaining power.

For investors with the highest valuation $\nu = \bar{\nu}$, the previous condition is given by

$$\int_{\underline{\nu}}^{\bar{\nu}} \{1 - \theta(\nu_o, \nu)\}[\Delta(\bar{\nu}) - \Delta(\nu_o)]\phi(\nu_o)d\nu_o = 0 \,. \tag{42}$$

Using the same logic as before, this condition is only satisfied if $\theta(\nu_o, \bar{\nu}) = 1$ for all $\nu_o \in [\underline{\nu}, \bar{\nu}]$, so that whenever a trade includes the highest-valuation investors as buyers, the full gains from trade must go to the buyer. This is an immediate consequence of the fact that these investors only buy in the secondary market and thus fail to internalize the full gains from trade only when buying assets. This can only be corrected by giving investors of type $\bar{\nu}$ the full bargaining power.

A natural implication of these two-limit cases is that the decentralized equilibrium with trade-specific bargaining powers is also inefficient. Whenever an owner investor of type $\underline{\nu}$ meets a non owner investor of type $\bar{\nu}$, efficiency would require that we provide full bargaining power to both buyer and seller, violating the restriction that investors can at most share the surplus generated in the trade at hand.