

# Volatility, volume, and liquidity in Over-The-Counter markets\*

Konstantin Milbradt<sup>†</sup>  
Northwestern Kellogg and NBER

December 2016

## Abstract

A cross-section of assets of heterogenous and stochastic characteristics is traded via a dealer-intermediated OTC. Liquidity shocked holders are subject to asset-specific holding costs (either assumed or induced by boundary conditions) and naturally want to sell assets to buyers who are not liquidity shocked but have holding limits. Under costly random search, market prices (and thus volatility), liquidity and volume are jointly determined in equilibrium. The equilibrium is unique, and can feature a new form of intermediation (“asset exchanges”) between holders of different assets when the natural buying capacity of investors on the sideline is too small. Considering holding costs that are linked to asset volatility via haircuts in collateralized borrowing, a feedback loop between the volatility of asset prices and holding costs arises, amplifying price volatility. Cross-margining can improve asset prices and reduce volatility by weakening this link. Finally, the model delivers a reason why firms want to issue heterogenous instead of homogenous bonds as they lead to a more efficient allocation of cross-sectional liquidity.

---

\*The author thanks Nittai Bergman, Zhiguo He, Costis Skiadas, Pierre-Olivier Weill, as well as seminar and conferences participants at CEMFI (Madrid), University of Amsterdam, Frankfurt School of Finance & Management, the FTG Summer Conference in London 2016, the 2016 UCSB Workshop for OTC markets, and the GEA conference in Berlin 2016 for comments and suggestions. All remaining errors are my own.

<sup>†</sup>Kellogg School of Management, Northwestern University, 2211 Campus Dr #4377, Evanston, IL 60208, e-mail: milbradt@northwestern.edu

# 1 Introduction

In times of crisis, some financial markets, especially over-the-counter markets, experience precipitous declines in trading or even the absence of trading, while at the same featuring large changes in prices and bid-ask spreads. The suspected cause for such market behavior is the shortage of outside capital that can absorb all the assets being offered for sale. However, not all assets within a broader market are affected equally – some asset remain relatively liquid, while others’ liquidity vanishes.

This paper derives the unique equilibrium in an over-the-counter asset market characterized by random search in a market with a cross-section of stochastic assets. This is in contrast to recent advances in the search-based OTC literature that focus on investor heterogeneity while featuring homogenous assets such as Hugonnier et al. (2014). Tractability, just as in Lagos and Rocheteau (2009), stems from the assumptions that (i) all trades have to be intermediated by dealers and (ii) a continuous inter-dealer market across all assets exists, thus linking different asset varieties. Liquidity shocked holders are subject to (asset specific) holding costs and thus want to sell assets to buyers who are currently not subject to liquidity shocks. Holding limits on investors impose an absorption capacity limit on the natural buyers, i.e., investors that are not currently subject to liquidity shocks. When there is sufficient buyers on the side-line, there is trade only between these buyers and the liquidity-shocked sellers. However, when the absorption capacity of buyers on the side-line shrinks, intermediation in the form of *asset exchanges*<sup>1</sup> arises — non-liquidity shocked investors in low effective holding costs assets offer an asset exchange with a side payment to liquidity shocked investors of high effective holding costs assets. Finally, when the process of finding a trading counter-party is costly, there may be intermediation only for some varieties of the assets with sufficient gains from trade. With costly search, *prices*, *bid-ask spreads*, and *trading volume* are jointly determined in equilibrium. We show that in general the equilibrium is unique for a wide range of holding costs functions, even those that feature endogenous elements such as a links to asset price levels and volatility.

More specifically, the model features a range of assets (varieties) within an *asset class* that has a fixed mass of possible investors. Within the *asset class*, individual assets’ type is stochastic, e.g., the distance to default of a bond can change with the individual fortunes of the firm it was issued by. Importantly, to trade, investors have to go through an OTC market run by a continuum of dealers to trade. Dealers are contacted with some intensity giving rise to a random, i.e., undirected, search framework. Meetings are short-lived, so if no transaction occurs the investor has to search again. Dealers at most meet one investor at a time, but are connected to each other via a frictionless inter-dealer market that features continuous market clearing for all traded varieties. We assume that dealers do not hold inventory to make the problem tractable.<sup>2</sup>

The equilibrium is characterized by a sufficient statistic, the surplus function  $S(x)$  for asset type  $x$ . The equilibrium traded set, i.e., all assets that feature non-zero liquidity in

---

<sup>1</sup>Here, an *exchange* refers to the exchange of one asset of type  $x$  for another asset of type  $x'$ , including a possible side-payment in cash, in a single transaction. Thus, the term is distinct from trading an asset for cash only.

<sup>2</sup>Current developments in the market for corporate bonds for example has seen dealers cut their inventory position, especially for the most illiquid bonds, and thus moving closer to the assumptions in this model.

equilibrium, is defined by all varieties  $x$  with a surplus that lies sufficiently above (or below, if asset exchanges are allowed) a benchmark surplus value  $k^*$ , reminiscent of Hopenhayn (1992). This benchmark surplus value then determines the marginal type assets, i.e., those assets that are just traded. Equilibrium  $k^*$  is derived from the market clearing condition in the inter-dealer market, matching buyer flow to seller flow. With endogenous search and linear-quadratic search costs, search-intensities are determined by the distance of the surplus of type  $x$ ,  $S(x)$ , from the marginal type assets' surplus,  $k^*$ . We find that for very steep search cost functions endogenous *intermediaries* arise in addition to the natural *buyers* and *sellers* well known in the search literature. These *intermediaries* are asset holders that offer asset exchanges to potential *sellers* to provide liquidity. Importantly, *intermediaries* recruit out of non-shocked holders with assets of low equilibrium surplus, usually associated with low holding costs, whereas asset *sellers* recruit out of liquidity shocked investors with assets of high equilibrium surplus, usually associated with high holding costs.

The assumption continuous inter-dealer market allows for a significant simplification of the link between the cross-section of the investor type distributions and the surplus function. In equilibrium, the impact of the cross-sectional distribution of asset varieties on the surplus function and the endogenous search intensities can be summarized by the benchmark surplus value  $k^*$ . The problem thus reduces to jointly solving, via market-clearing, for the benchmark surplus value and the surplus function arising from prices. We show a unique steady-state equilibrium arises and can thus rule out “liquidity begets liquidity” equilibria. As a result, liquidity is here described by two objects simultaneously, by *bid-ask spreads* and by *trading volume* (which itself arises from the endogenous *search-intensities* of the investors).

The presence of *intermediaries* that arise from asset exchanges leads to non-monotone volume in asset trades even if surplus is monotone in asset varieties. This can explain situations as observed for very safe bonds where volume concentrates on the long and short end of a specific company's outstanding bonds, with very little trading occurring for intermediate maturities. Similarly, with default risk and large holding costs for bonds in default, we can model a situation in the tradition of Miao (2005) in which trading volume mainly concentrates on bonds that are either very close to default (junk bonds) or very safe (AAA or AA), with lower volume in the middle of the rating distribution.

The model also allows for linking of market illiquidity and price volatility to margin requirements. When an investor is hit with a liquidity shock, until the investor finds a willing buying counter-party, the investor's prime broker will allow a margin loan against the investor's position (e.g. against the investor's bond holdings). As has been shown by Krishnamurthy et al. (2014), the haircut of this margin loan is often well-explained by a linear function incorporating the volatility of the bond's return process. Under the assumption that the product of haircut and price enters holding costs linearly, then the level and the slope of the pricing function w.r.t. to the characteristic  $x$  enter the holding cost function linearly. To keep the model tractable we make the simplifying assumption that the level and slope of the surplus function, instead of the pricing function, enter the holding cost function linearly. A negative feedback effect between volatility based haircuts and the price of the bond arises — riskier bonds now feature endogenously amplification of holding costs, which in turn decreases their price. The feedback effect of surplus-volatility based haircuts on surplus-volatility itself is ambiguous. When cross-margining is allowed, a diversified portfolio of bonds faces less steep haircuts, thus muting the impact of idiosyncratic volatility on holding costs, leading

to an increase in asset prices.

Lastly, the model allows a corporate finance application that addresses the observation that corporations usually have a whole cross-section of different bonds outstanding. Through the lens of the model, a cross-section of bonds, here differentiated by time-to-maturity, allocates the sparse liquidity of the OTC market more efficiently than a homogenous random-maturity (or equivalently, a sinking fund) bond would do. The key insight is that heterogeneous time-to-maturity bonds feature a cross-section of effective holding costs – the maximum time the holders might be exposed to such costs is capped by the bond’s time-to-maturity – while homogenous bonds all feature the same effective holding costs. Thus, in the heterogeneous case, the bonds with the lowest effective holding costs have zero liquidity, while the bonds with the highest effective holding costs feature maximum liquidity. In the homogenous bonds case effective holding costs are the same, so that there is rationing of trades, leading to low liquidity.

## 1.1 Literature overview

This paper builds on the literature of random matching in OTC markets. Duffie et al. (2005) is one of the first papers in this literature and features a basic setup with a homogenous assets, two types of agents and exogenous holding costs. In a follow up, Duffie et al. (2007) link the holding costs to hedging benefits via risk-aversion of agents.

The closest papers in terms of focus are Weill (2007) and He and Milbradt (2014): Weill (2007) introduces non-homogenous assets, but asset varieties are still static. Difference in expected returns in different assets arise out of liquidity properties, and not some underlying differences in cash-flows. He and Milbradt (2014) essentially incorporate a continuum of bonds that differ in time-to-maturity into a search model, and have those bonds be driven by a distance-to-default factor. However, as a sufficient supply of risk-bearing capacity on the side-line is assumed, and contact intensities are uniformly capped, assets are intermediated uniformly. This paper instead introduces a shortage of risk-bearing capacity on the side-line, which leads to liquidity concentrating on a subset of asset varieties. Further, the framework allows for asset varieties to be stochastic, whereas in He and Milbradt (2014) asset varieties (time-to-default) feature deterministic dynamics.<sup>3</sup>

Lagos and Rocheteau (2009) and Hugonnier et al. (2014) are related papers that treat heterogeneity on the investor and asset position side: Lagos and Rocheteau (2009) expand the set of possible investors (essentially, agents have different levels utility for the asset), and thus a cross-section of asset holdings in terms of positions in a homogenous asset exists. Similarly to this paper, we will also use a frictionless inter-dealer market to handle the cross-sectional distribution of asset holdings, except that our cross-section is over asset varieties, and not over holding size of the same asset. A related paper that treats the cross-sectional distribution of asset holdings in a bilateral matching model is Afonso and Lagos (2014). Hugonnier et al. (2014) expand the traditional two investor bilateral matching model to a continuum of investors of different level of desperation to sell but with homogenous contact intensities, and show that with the common asset holding restriction to  $\{0, 1\}$  that realized

---

<sup>3</sup>The model can easily handle this situation: deterministic time-to-maturity with constant holding costs  $h$  until maturity leads to relatively high trading volume (if asset exchanges are allowed) in high- and low-maturity bonds of the company, and a relatively low trading volume in medium-maturity bonds.

trades look like an intermediated network: most trades are trades from one investors to an investors just slightly less desperate than oneself, and trading volume peaks around the median investors who are effectively intermediaries. Finally, Sambalaibat (2016) introduces CDS contracts in a search market for bonds, thus expanding the asset space, to address liquidity spillovers between the two markets.

A related paper that also uses endogenous search in a bilateral matching model is Farboodi et al. (2016). In contrast to the current paper, an asymmetric equilibrium with a cross-section of search intensities arises as a response to the congestions/externality effects of the bilateral matching model, and no symmetric equilibrium exists. In our model, due to the frictionless inter-dealer market, a symmetric equilibrium arises, but there is a cross-section of search intensities driven by the heterogeneity of assets.

## 2 A heterogenous asset OTC market setup

All agents, i.e. investors and dealers, are risk-neutral, infinitely lived, with a common discount rate  $r$ . Time is continuous and indexed by  $t$ . The Appendix A provides additional proofs not provided in the text.

### 2.1 Asset type

We consider a continuum of assets indexed by a characteristic/variety/type  $x$  distributed on some finite compact set  $\mathbb{X}$  according to some steady-state density  $\mu_1(x)$  so that the total asset supply is equal to unity (i.e.  $\int_{\mathbb{X}} \mu_1(x) dx = 1$ ).<sup>4</sup> Unless specified otherwise, we take  $\mathbb{X} = [x_{min}, x_{max}] = [0, 1]$ . Type  $x$  is potentially stochastic and its dynamics can be summarized by its linear generator  $\mathcal{L}$ . Importantly, we assume that the type process for each individual asset  $x$  is continuous and i.i.d. from other assets. The dynamics of  $x$  are given by

$$dx = m(x) dt + \sigma(x) dZ. \tag{1}$$

where  $dZ$  is a standard Brownian motion, so that the linear generator is given by  $\mathcal{L}f(x) = m(x) f'(x) + \frac{\sigma^2(x)}{2} f''(x)$ .<sup>5</sup>

Assets pay a (possibly variety/type dependent) dividend / coupon / interest payment  $c(x)$  each instant, and are possibly subject to random default with zero recovery at a rate  $\delta(x)$ . Boundary conditions apply on the edges of  $\mathbb{X}$  to be specified later.

---

<sup>4</sup>Very mild technical conditions have to be imposed for the steady-state density  $\mu_1(x)$  for a compact finite set  $\mathbb{X}$ . The results of the paper go through for an infinite set  $\mathbb{X}$  with a steady-state distribution, which in turn requires either (i) sufficient mean-reversion of  $x$  or (ii) some killing of large outliers. Our example in Section 4.3 will be based on (ii).

<sup>5</sup>Very similar results hold for situations in which  $x$  is say time to maturity (so that  $\mathcal{L}f(x) = -f'(x)$ ), which we will discuss briefly in Section 4.2.

## 2.2 Investors

All investors are homogenous ex-ante in the model. They have a uniform cumulative holding limit across all varieties of the asset given by  $A$  and cannot short.<sup>6</sup> By the risk-neutrality assumption, without loss of generality, we can set  $A = 1$  and restrict investor individual holdings  $a$  to  $a \in \{0, 1\}$ .

Investors are either normal (so called high-value or H-type investors) or liquidity shocked (so called low-value or L-type investors). We will call state  $i = \{H, L\}$  the liquidity state of the investor. Investor types follow a continuous time Markov chain: Each H-type investors is subject to an i.i.d. random liquidity shock that makes him an L-type with intensity  $\xi_{HL}$ , and each L-type investor recovers back to an H-type investor with intensity  $\xi_{LH}$ , again i.i.d..

We further assume that there is a total mass  $\mu > 1$  of investors, so that the total asset supply of 1 can be held by the investors. By an appropriate law of large numbers, the steady state proportions of H and L type investors in the economy,  $\mu_H$  and  $\mu_L$ , respectively, are given by

$$\mu_H = \frac{\xi_{LH}}{\xi_{HL} + \xi_{LH}}\mu \quad \text{and} \quad \mu_L = \frac{\xi_{HL}}{\xi_{HL} + \xi_{LH}}\mu \quad (2)$$

via the flow equations.<sup>7</sup> Further, let  $\mu_{H1}(x)$  and  $\mu_{L1}(x)$  be the mass of holders of asset  $x$  of H- and L-type, respectively, so that

$$\mu_1(x) = \mu_{H1}(x) + \mu_{L1}(x).$$

The liquidity state directly impacts the investor only if he is holding the asset. In this case, the investor is subject to possibly asset specific per period holding cost of  $h(x)$ , where we assume  $h(x)$  to be continuous. For the moment, we take  $h(x)$  as exogenously specified, but will partially endogenize it in Section 5. When  $h(x) = h$ , i.e., constant, then we need the boundary conditions to generate heterogeneity across assets. Section 4.2 and 4.3 provide examples of boundary-induced heterogeneity.

It is clear that L type investors who are holding the asset, i.e.  $i = L$  and  $a = 1$  (in short denoted as L1 types), should try to sell the asset to H type investors currently not holding the asset, i.e.  $i = H$  and  $a = 0$  (in short denoted as H0 types), to save on holding costs  $h(x)$ . The literature commonly refers to L1 types as (natural) *sellers* and to H0 types as (natural) *buyers*. H type investors currently holding the asset, i.e.  $i = H$  and  $a = 1$  (in short denoted as H1 types) are the best holder of the asset if the economy is unconstrained, whereas L types investors currently not holding the asset, i.e.  $i = L$  and  $a = 0$  (in short denoted as L0 types) have no advantage in holding the asset and therefore remain inactive until they become an H0 type. Thus, traditionally, H1 types are referred to as *holders* and L0 types as *inactive* or *sidelined* investors. Later on, we will show that *holders* can become *intermediaries* by offering asset exchanges with *sellers*, breaking this traditional taxonomy of investor types.

<sup>6</sup>Here, we do not endogenize the holding limit  $A$  as in Gârleanu and Pedersen (2007) to instead concentrate on asset-side heterogeneity.

<sup>7</sup>That is,  $\dot{\mu}_H = 0 = -\xi_{HL}\mu_H + \xi_{LH}\mu_L = -\xi_{HL}\mu_H + \xi_{LH}(\mu - \mu_H)$ .

Investors can only sell through dealers that are contacted i.i.d. at an individual rate  $\rho$ . We assume that investors choose their contact rate  $\rho$  with dealers, and pay a flow cost  $\gamma(\rho)$  per unit of time with  $\gamma(0) = 0$  and  $\gamma'(\rho), \gamma''(\rho) \geq 0$ .<sup>8</sup>

Agents value the asset at some monetary value to be derived later, and when buying and selling the asset will exchange cash for the asset or vice-versa, or when exchanging an asset will exchange a side-payment of cash in addition to exchanging assets.

## 2.3 Dealers

We assume that all trade has to be intermediated by a continuum of dealers. Potential buyers and sellers search for dealers and in equilibrium find a dealer with some *endogenously* chosen intensity  $\rho_{ia}(x), i \in \{H, L\}, a \in \{0, 1\}$ .<sup>9</sup> Each dealer can only be in contact with one investor at a time. The contact is short-lived, in that if no action is taken the contact breaks down and the investor has to search again.

A central assumption is that there is a *frictionless* and *continuous* inter-dealer market that clears continuously for all traded varieties  $x$  at a price  $M(x)$  as in Lagos and Rocheteau (2007, 2009). Dealers are assumed to not hold any inventory, so that if a dealer buys an asset from a *seller* (a contacted L1 or H1 investor) the dealer will immediately sell this asset on the inter-dealer market to another dealer that is in contact with a *buyer* (a contacted H0 or L0 investor) or exchange the asset with a dealer in contact with another *seller*.

A key role in the analysis of a shortage of capital is played by what transactions are intermediated by the dealers. We consider two situations:

1. Each investor in contact with a dealer can either buy or sell an asset, but cannot do both. We will refer to this as a “*no-exchange*” situation.
2. Each investor in contact with a dealer can buy and sell up to his holding limit any asset he so desires. Thus, we specifically allow an investor to sell an asset  $x$  and simultaneously buy an asset  $x'$ . Thus, the investor exchanges his asset  $x$  for an asset  $x'$  and a side-payment in cash.

We assume that dealers and investors split the surplus that is created between them via the Nash-bargaining solution: the dealer appropriates a fraction  $(1 - \beta)$  and the investor appropriates a fraction  $\beta$  of the surplus generated between the investor and the dealer.

An equivalent interpretation for  $\beta = 1$  (i.e., the case in which the bid-ask spread is identically zero) is that investors have only random access to a Walrasian market, arriving with a chosen intensity  $\rho$ .

## 2.4 Surplus

Let  $V_{H0}$  and  $V_{L0}$  be the value functions for *buyers* and *inactive* investors that are currently not holding any assets. Let  $V_{H1}(x)$  and  $V_{L1}(x)$  be the value functions for *holders* and *sellers*

---

<sup>8</sup>To map back into the traditional framework of constant contact intensities  $\bar{\rho}$ , one could simply assume that  $\gamma(\rho) = \begin{cases} 0, & \rho \leq \bar{\rho} \\ \infty, & \rho > \bar{\rho} \end{cases}$ .

<sup>9</sup>Of course, for  $a = 0$ ,  $\rho_{i0}(x) = \rho_{i0}$ . In other words, the contact intensity cannot depend on asset type if no assets are currently held.

that are currently holding an asset of type  $x$ . Let us define the maximum surplus that can be generated by a trade of type  $x$  between two investors. This maximum surplus is generated between a *seller* of type  $x$  and a *buyer* on the side-line, and is given by

$$S(x) \equiv V_{H1}(x) - V_{H0} + V_{L0} - V_{L1}(x) \quad (3)$$

as an L1 type that sells becomes an L0 type (and therefore forgoes his outside option of remaining a potential seller with value  $V_{L1}(x)$ ) and an H0 type that buys becomes an H1 type (and therefore forgoes his outside option of remaining a potential buyer with value  $V_{H0}$ ).  $S(x)$  will be our main object of investigation for the remainder of the paper as it will pin down the trading rates  $\rho_{ia}(x)$  and the terms-of-trade/bid-ask spreads. In other words,  $S(x)$  is a sufficient statistic of the equilibrium.

Next, consider the (potential) surplus generated between a seller of type  $i$  selling an asset of type  $x$  and a dealer. The dealer has an outside option of 0, and acquiring the asset has a value of  $M(x)$  on the inter-dealer market. Thus, we have

$$S_{i \rightarrow D}(x) = M(x) + V_{i0} - V_{i1}(x). \quad (4)$$

Consider now the surplus generated between a buyer of type  $i$  buying an asset of type  $x$  and a dealer. The dealer again has an outside option of 0, but can instantaneously acquire the asset at a cost of  $M(x)$ . Thus, we have

$$S_{D \rightarrow i}(x) = V_{i1}(x) - V_{i0} - M(x). \quad (5)$$

We note that  $S(x) = S_{L \rightarrow D}(x) + S_{D \rightarrow H}(x)$ , as no additional surplus is generated by the presence of the dealers – dealers are just pass-through intermediaries. Of course, dealers collectively extract a proportion of the surplus due to bargaining power.

### 3 Dealer-Intermediated Equilibrium

Let  $\mathbb{T} \subset \mathbb{X}$  be the set of varieties available on the inter-dealer market in equilibrium, that is

$$\mathbb{T} = \{x : \rho_{L1}(x) + \rho_{H1}(x) > 0\}. \quad (6)$$

We will often refer to  $\mathbb{T}$  as the *traded set*. We can extend the traded set to all varieties via a trembling-hand equilibrium by assuming that there is a small amount of people who search at an intensity  $\varepsilon > 0$  regardless of optimality and asset or investor type.

We will now derive the conditions for equilibrium in a dealer-intermediated OTC search market. We will do this in steps. First, the following lemma establishes the surplus split that goes to buyers:

**Lemma 1** *Surplus to buyers  $S_{D \rightarrow i}(x)$  has to be constant across all traded assets  $x \in \mathbb{T}$ .*

Next, we consider market clearing. If there is a shortage of natural buyers  $\mu_{H0}$ , then (1.) under the *no-exchanges* assumption some varieties  $x$  will simply not be traded in equilibrium, whereas (2.) under the *exchanges* assumption, some assets  $x$  will be offered by H1 types in exchange for some assets  $x'$  by L1 types and a side-payment in cash.



We are now in a position to write the steady-state value functions for all 4 possible types of investors. Recall that  $\gamma(\rho) \geq 0$  is the cost of actively searching at an intensity  $\rho$ :

$$rV_{H0} = \xi_{HL}(V_{L0} - V_{H0}) + \underbrace{\max_{\rho} \left\{ \rho\beta \max_x [V_{H1}(x') - M(x') - V_{H0}] - \gamma(\rho) \right\}}_{(a)} \quad (8)$$

$$\begin{aligned} rV_{L0} &= \xi_{LH}(V_{H0} - V_{L0}) \\ [r + \delta(x)]V_{H1}(x) &= \delta(x)V_{H0} + c(x) + \xi_{HL}[V_{L1}(x) - V_{H1}(x)] + \mathcal{L}V_{H1}(x) \\ &\quad + \underbrace{\max_{\rho} \left\{ \rho\beta \left( \max_{x'} [V_{H1}(x') - M(x')] + M(x) - V_{H1}(x) \right) - \gamma(\rho) \right\}}_{(b)} \end{aligned} \quad (9)$$

$$\begin{aligned} [r + \delta(x)]V_{L1}(x) &= \delta(x)V_{L0} + c(x) - h(x) + \xi_{LH}[V_{H1}(x) - V_{L1}(x)] + \mathcal{L}V_{L1}(x) \\ &\quad + \underbrace{\max_{\rho} \left\{ \rho\beta \left( \max_{x' \in \mathbb{S}_H} [V_{L1}(x') - M(x')] + M(x) - V_{L1}(x) \right) - \gamma(\rho) \right\}}_{(c)} \end{aligned} \quad (10)$$

Here, we used the observation that  $L0$  types are at best indifferent about searching, so without loss of generality we can set  $\rho_{L0} \equiv 0$ . The terms (a), (b), and (c) summarize the benefits of trade in the model. In the *no-exchanges* case, we have (b) = 0 by assumption.

Next, let  $\mathbb{S}_H = \{x : \rho_{H1}(x) > 0\}$  and  $\mathbb{S}_L = \{x : \rho_{L1}(x) > 0\}$  be the set of varieties offered for trade by  $H1$  and  $L1$  types, respectively. Then, define marginal varieties such that  $\hat{x} \in \partial\mathbb{S}_H$  and  $\check{x} \in \partial\mathbb{S}_L$  be marginal sellers, so that

$$(b) = 0 \iff \rho_{H1}(\hat{x})\beta \left( \max_{x' \in \mathbb{S}_L} [V_{H1}(x') - M(x')] + M(\hat{x}) - V_{H1}(\hat{x}) \right) = \gamma(\rho_{H1}(\hat{x})) \quad (11)$$

$$(c) = 0 \iff \rho_{L1}(\check{x})\beta \left( \max_{x' \in \mathbb{S}_H} [V_{L1}(x') - M(x')] + M(\check{x}) - V_{L1}(\check{x}) \right) = \gamma(\rho_{L1}(\check{x})) \quad (12)$$

Of course, in case of a *no-exchanges* assumption,  $\mathbb{S}_H = \emptyset$ . In case *exchanges* are allowed, we will show below that when  $\gamma'(0) = 0$  we have  $\hat{x} = \check{x} = \bar{x}$  and thus  $\mathbb{X} = \mathbb{S}_H \cup \mathbb{S}_L$ . Next, note that by a competitive inter-dealer market and Lemma 1, we must have buyer indifference of the marginal varieties to any other type offered for sale:

$$\max_{x' \in \mathbb{S}_L} [V_{H1}(x') - M(x')] = V_{H1}(\check{x}) - M(\check{x}) \quad (13)$$

$$\max_{x' \in \mathbb{S}_H} [V_{L1}(x') - M(x')] = V_{L1}(\hat{x}) - M(\hat{x}). \quad (14)$$

That is, we must have

$$M(x) = \begin{cases} V_{H1}(x) - V_{H1}(\check{x}) + M(\check{x}), & x \in \mathbb{S}_L \\ \text{undefined (no trade)}, & x \in \text{else} \\ V_{L1}(x) - V_{L1}(\hat{x}) + M(\hat{x}), & x \in \mathbb{S}_H \end{cases} \quad (15)$$

Note that the marginally traded varieties are defined either by (i) indifference over  $\rho$  (so that  $\rho = 0$  is part of the choice set) or (ii) by the  $\rho = 0$  constraint just being binding. We note that when there is a shortage of buyers,  $H0$  types in general are not marginal.

### 3.1 Optimal search intensity

For the remainder of the paper, we will adopt the following search cost specification

$$\gamma(\rho) = \gamma_1 \rho + \frac{1}{2} \gamma_2 \rho^2 \quad (16)$$

so that  $\gamma(\rho)/\rho = \gamma_1 + \frac{1}{2} \gamma_2 \rho$ . In case of  $\gamma_2 = 0$ , we impose a uniform limit  $\rho \leq \bar{\rho}$ . Thus, we see that for the marginal agents that are just indifferent, we have

$$\beta \left( \max_{x' \in \mathbb{S}_L} [V_{H1}(x') - M(x')] + M(\hat{x}) - V_{H1}(\hat{x}) \right) = \frac{\gamma(\rho_{H1}(\hat{x}))}{\rho_{H1}(\hat{x})} = \gamma_1 \quad (17)$$

$$\beta \left( \max_{x' \in \mathbb{S}_H} [V_{L1}(x') - M(x')] + M(\check{x}) - V_{L1}(\check{x}) \right) = \frac{\gamma(\rho_{L1}(\check{x}))}{\rho_{L1}(\check{x})} = \gamma_1. \quad (18)$$

The last equality in each row follows from either  $\gamma_2 = 0$  or if  $\gamma_2 > 0$  from  $\rho_{i1}(x) = 0, i \in \{H, L\}$ . Plugging in for  $\max_{x' \in \mathbb{S}_i}$ , we have

$$V_{H1}(\check{x}) - M(\check{x}) + M(\hat{x}) - V_{H1}(\hat{x}) = \frac{\gamma_1}{\beta} \quad (19)$$

$$V_{L1}(\hat{x}) - M(\hat{x}) + M(\check{x}) - V_{L1}(\check{x}) = \frac{\gamma_1}{\beta}. \quad (20)$$

Summing the two conditions, we have

$$V_{H1}(\check{x}) + V_{L1}(\hat{x}) - V_{H1}(\hat{x}) - V_{L1}(\check{x}) = S(\check{x}) - S(\hat{x}) = 2 \frac{\gamma_1}{\beta}. \quad (21)$$

Differencing the two conditions, we have

$$M(\hat{x}) - M(\check{x}) = \frac{V_{H1}(\hat{x}) - V_{H1}(\check{x}) + [V_{L1}(\hat{x}) - V_{L1}(\check{x})]}{2}. \quad (22)$$

When we obtain the optimal search intensities below, we see that for  $\gamma_1 = 0$ , this implies  $\check{x} = \hat{x} = \bar{x}$ .

Lastly, we note the indifference condition for low types

$$V_{L0} = V_{L1}(\hat{x}) - M(\hat{x}). \quad (23)$$

Once low types (for whatever reason) are in contact with a dealer, they should be at most indifferent between buying in or staying on the side-lines.

Let us now consider the determination of the optimal  $\rho$ . First, let us introduce the benchmark value  $k$  (when  $\hat{x} = \check{x}$ , this benchmark value can be interpreted as the marginal value; when  $\hat{x} \neq \check{x}$ , it is the midpoint of the marginal values) in the *exchanges* allowed situation as defined by

$$k \equiv \frac{1}{2} [S(\hat{x}) + S(\check{x})] \iff S(\hat{x}) = k - \frac{\gamma_1}{\beta} \quad (24)$$

where we used  $S(\check{x}) - S(\hat{x}) = 2 \frac{\gamma_1}{\beta}$ . In the *no-exchanges* situation, we simply define the benchmark value by

$$k \equiv S(\hat{x}) + \frac{\gamma_1}{\beta} \iff S(\hat{x}) = k - \frac{\gamma_1}{\beta}. \quad (25)$$

Then, for any equilibrium  $k$ , we have

$$(a) = \max_{\rho} \{\rho\beta \cdot k - \gamma(\rho)\} \quad (26)$$

$$(b) = \begin{cases} \max_{\rho} \{\rho\beta [k - S(x)] - \gamma(\rho)\} & \text{exchanges allowed} \\ 0 & \text{no exchanges allowed} \end{cases} \quad (27)$$

$$(c) = \max_{\rho} \{\rho\beta [S(x) - k] - \gamma(\rho)\} \quad (28)$$

For the remainder of the paper, we will use  $\mathbb{1}_{exchange}$  as the indicator function for the *exchanges* allowed case, and

$$\hat{r} = r + \xi_{HL} + \xi_{LH}. \quad (29)$$

### 3.1.1 Linear-quadratic search costs

The optimal search intensities for  $\gamma_2 > 0$  are given by

$$\rho_{H0} = \frac{1}{\gamma_2} (\beta \cdot k - \gamma_1)^+ \quad (30)$$

$$\rho_{H1}(x) = \mathbb{1}_{exchange} \frac{1}{\gamma_2} (\beta [k - S(x)] - \gamma_1)^+ \quad (31)$$

$$\rho_{L1}(x) = \frac{1}{\gamma_2} (\beta [S(x) - k] - \gamma_1)^+ \quad (32)$$

and we see that  $\rho_{H1}(x) \rho_{L1}(x) = 0$ , i.e., varieties  $L1$  types are trying to sell  $H1$  types hold on to, and vice versa. Further, as  $S(x)$  exhibits no jump, we immediately have that for  $\gamma_1 = 0$ , we must have  $\hat{x} = \check{x} = \bar{x}$ . Plugging the search-intensities into the value-function ODEs, and then combining them via the definition of the surplus function  $S(x)$ , we are left with the following ODE:

$$\begin{aligned} [\hat{r} + \delta(x)] S(x) &= h(x) - \frac{1}{2\gamma_2} [(\beta k - \gamma_1)^+]^2 + \mathcal{L}S(x) \\ &+ \frac{1}{2\gamma_2} \mathbb{1}_{exchange} [(\beta [k - S(x)] - \gamma_1)^+]^2 - \frac{1}{2\gamma_2} [(\beta [S(x) - k] - \gamma_1)^+]^2 \end{aligned} \quad (33)$$

For  $\gamma_1 = 0$ , the surplus ODE simplifies to

$$\begin{aligned} [\hat{r} + \delta(x)] S(x) &= h(x) - \frac{1}{2\gamma_2} (\beta k)^2 + \mathcal{L}S(x) \\ &+ \frac{1}{2\gamma_2} (\beta [S(x) - k])^2 \left\{ \mathbb{1}_{exchange} \mathbb{1}_{\{S(x) \leq k\}} - \mathbb{1}_{\{S(x) \geq k\}} \right\}. \end{aligned} \quad (34)$$

This equation has a solution under mild regularity conditions (specifically, it is related to Matrix Riccati equations).

Finally, we will only consider “neutral boundary behavior/conditions”: For any finite upper boundary, the boundary condition is considered neutral if (i) in case of a boundary condition on the level of  $S$ , then  $\frac{\partial}{\partial k} [S(x_{boundary}, k) - k] \leq 0$ , and (ii) in case it is a condition on the slope of  $S$ , then  $\frac{\partial}{\partial k} S'(x_{boundary}, k) \leq 0$ . The condition for a lower boundary have the opposite signs. We can then derive the following lemma:

**Lemma 2** Define  $\hat{S}(x, k)$  as the surplus function for an arbitrary benchmark  $k$ . Then, for the linear-quadratic search costs, and neutral boundary behavior, we have

$$\frac{\partial}{\partial k} [\hat{S}(x, k) - k] = \left[ \frac{\partial \hat{S}(x, k)}{\partial k} - 1 \right] < 0.$$

### 3.1.2 Linear search costs & bounded intensities

For  $\gamma_2 = 0$ , let us introduce a uniform upper bound  $\bar{\rho}$  to keep the problem bounded. Then, for  $\gamma_1 > 0$ , the following bang-bang solutions apply:

$$\rho_{H0} = \begin{cases} \bar{\rho}, & \beta k \geq \gamma_1 \\ 0, & \text{else} \end{cases} \quad (35)$$

$$\rho_{H1}(x) = \begin{cases} \bar{\rho}, & \text{exchanges allowed \& } \beta [S(x) - k] \leq -\gamma_1 \\ 0, & \text{else} \end{cases} \quad (36)$$

$$\rho_{L1}(x) = \begin{cases} \bar{\rho} & \beta [S(x) - k] \geq \gamma_1 \\ 0, & \text{else} \end{cases}. \quad (37)$$

The surplus ODE then becomes

$$\begin{aligned} [\hat{r} + \delta(x)] S(x) &= h(x) - \bar{\rho}(\beta k - \gamma_1)^+ + \mathcal{L}S(x) \\ &\quad + \bar{\rho} \mathbb{1}_{\text{exchange}} (\beta [k - S(x)] - \gamma_1)^+ - \bar{\rho} (\beta [S(x) - k] - \gamma_1)^+. \end{aligned} \quad (38)$$

As this ODE is linear, it has a solution under very mild regularity conditions. For  $\gamma_1 = 0$ , we then have  $\rho_{ia} = \bar{\rho}$ , i.e., if search is free than wlog we have *all* agents searching at the maximum rate.<sup>10</sup>

We can again derive the same lemma as for the  $\gamma_2 > 0$  case:

**Lemma 3** Define  $\hat{S}(x, k)$  as the surplus function for an arbitrary benchmark  $k$ . Then, for the linear-capped search costs, and neutral boundary behavior, we have

$$\frac{\partial}{\partial k} [\hat{S}(x, k) - k] = \left[ \frac{\partial \hat{S}(x, k)}{\partial k} - 1 \right] < 0.$$

## 3.2 Market Clearing

For neutral boundary conditions, we have  $\left[ \frac{\partial \hat{S}(x, k)}{\partial k} - 1 \right] < 0$  both in the linear quadratic as well as in the linear capped case. The following corollary of Lemmas 2 and 3 as a building block to establish monotonicity of market-clearing w.r.t.  $k$ :

<sup>10</sup>For  $\gamma_1 = \gamma_2 = 0$ , either we assume that all investors search at their maximum rate, and some investors will be rationed, or equivalently we assume that those investors that realize they will be rationed with probability 1 do not search at all. This equivalence disappears as soon as there is any small cost to search, and most rationing disappears.

**Corollary 1** *The function  $\hat{S}(x, k) - k$  is decreasing in  $k$  for any  $x$  for our search cost specifications. Thus, the optimal contact intensity  $\hat{\rho}_{H1}(x, k)$  is increasing in  $k$  while  $\hat{\rho}_{L1}(x, k)$  is decreasing in  $k$  for any  $x$ . Further, buyers always search more than possible exchangers (which are a strict subset of holders), i.e.,  $\hat{\rho}_{H0}(k) \geq \hat{\rho}_{H1}(x, k), \forall x$ ,*

The steady-state distribution and market clearing are functions of  $\hat{\rho}_{H0}(k), \hat{\rho}_{H1}(x, k), \hat{\rho}_{L1}(x, k)$  which in turn are direct functions of the surplus  $\hat{S}(x, k)$  and  $k$  (linear in case of quadratic costs, and step functions in case of liner-capped costs). As the next step, we establish the following monotonicity of the steady-state mass equations w.r.t.  $k$ .

**Lemma 4** *The steady-state mass  $\hat{\mu}_{H1}(x, k)$  is decreasing in  $k$ , the steady-state mass  $\hat{\mu}_{L1}(x, k)$  is increasing in  $k$ , and the steady state mass  $\hat{\mu}_{H0}(k)$  is increasing in  $k$ .*

Market clearing is given by

$$\underbrace{\hat{\rho}_{H0}(k) \hat{\mu}_{H0}(k)}_{\text{Buyers}} + \underbrace{\int_{\mathbb{X}} \hat{\rho}_{H1}(x, k) \hat{\mu}_{H1}(x, k) dx}_{\text{Exchangers}} = \underbrace{\int_{\mathbb{X}} \hat{\rho}_{L1}(x, k) \hat{\mu}_{L1}(x, k) dx}_{\text{Sellers \& Exchangers}}. \quad (39)$$

Of course, if no *exchanges* are allowed, then the second term is simply identically zero as  $\hat{\rho}_{H1}(x, k) = 0$ . To generalize, suppose there is default at a rate  $\delta(x)$ , and suppose further that all assets are reborn and allocated to H0 types on a competitive primary market according to the distribution  $\mu_1(x)$ . Then, the steady-state mass equation for  $\hat{\mu}_{H1}(x, k)$  solves

$$0 = -\xi_{HL} \hat{\mu}_{H1}(x, k) + \xi_{LH} \hat{\mu}_{L1}(x, k) + \mathcal{L}^* \hat{\mu}_{H1}(x, k) + \hat{\rho}_{L1}(x, k) \hat{\mu}_{L1}(x, k) - \hat{\rho}_{H1}(x, k) \hat{\mu}_{H1}(x, k) + \delta(x) \hat{\mu}_{L1}(x, k) \quad (40)$$

where  $\mathcal{L}^*$  is the adjoint operator of  $\mathcal{L}$ .<sup>11</sup> Here, the last term  $\delta(x) \hat{\mu}_{L1}(x, k)$  is the inflow of defaulting assets that are reborn and then allocated to only H types. Rearranging, we have

$$[\xi_{HL} + \xi_{LH} + \hat{\rho}_{L1}(x, k) + \hat{\rho}_{H1}(x, k) + \delta(x)] \hat{\mu}_{H1}(x, k) = [\xi_{LH} + \hat{\rho}_{L1}(x, k) + \delta(x)] \mu_1(x) + \mathcal{L}^* \hat{\mu}_{H1}(x, k). \quad (41)$$

Taking integrals w.r.t.  $x$  over  $\mathbb{X}$ , and using  $\mu_1(x) = \mu_{H1}(x) + \mu_{L1}(x)$ , we have

$$\int_{\mathbb{X}} \hat{\rho}_{L1}(x, k) \hat{\mu}_{L1}(x, k) dx - \int_{\mathbb{X}} \hat{\rho}_{H1}(x, k) \hat{\mu}_{H1}(x, k) dx = \int_{\mathbb{X}} [\xi_{HL} + \xi_{LH} + \delta(x)] \hat{\mu}_{H1}(x, k) dx - \int_{\mathbb{X}} \delta(x) \mu_1(x) dx - \xi_{LH} - bc(k) \quad (42)$$

<sup>11</sup>Any atomistic rebirth to a single point  $x_0$  does not affect the differential equation, but only boundary conditions at  $x_0$ . As long as  $x_0$  is independent of  $k$  the argument below is unchanged. The argument becomes more subtle when  $x_0(k)$ , that is, when the resetting point is a function of  $k$ .

where we define the boundary conditions

$$\begin{aligned} bc(k) &\equiv \int_{\mathbb{X}} \mathcal{L}^* \hat{\mu}_{H1}(x, k) dx \\ &= \left[ \frac{\sigma^2}{2} \hat{\mu}'_{H1}(x_{max}, k) - m \cdot \hat{\mu}_{H1}(x_{max}, k) \right] - \left[ \frac{\sigma^2}{2} \hat{\mu}'_{H1}(x_{min}, k) - m \cdot \hat{\mu}_{H1}(x_{min}, k) \right]. \end{aligned} \quad (43)$$

Here, the integral  $\int_{\mathbb{X}} \mathcal{L}^* \hat{\mu}_{H1}(x, k) dx$  imposes the boundary conditions at  $x_{min}$  and  $x_{max}$ . The term  $bc(k)$  is identically zero in the reflection case.<sup>12</sup> Finally, let us define the market imbalance function by

$$\begin{aligned} ib(k) &\equiv \int_{\mathbb{X}} \hat{\rho}_{L1}(x, k) \hat{\mu}_{L1}(x, k) dx - \int_{\mathbb{X}} \hat{\rho}_{H1}(x, k) \hat{\mu}_{H1}(x, k) dx - \hat{\rho}_{H0}(k) \hat{\mu}_{H0}(k) \\ &= \int_{\mathbb{X}} [\xi_{LH} + \xi_{HL} + \delta(x) + \hat{\rho}_{H0}(k)] \hat{\mu}_{H1}(x, k) dx - \int_{\mathbb{X}} \delta(x) \mu_1(x) dx - \hat{\rho}_{H0}(k) \mu_H \\ &\quad - \xi_{LH} - bc(k). \end{aligned} \quad (44)$$

Then  $ib(k^*) = 0$  defines an equilibrium benchmark value  $k^*$ .

Let us define two quantities,  $\bar{k}$  and  $\underline{k}$ , by the fixed point relations

$$\bar{k} = \max_x \hat{S}(x, \bar{k}) \quad (45)$$

$$\underline{k} = \min_x \hat{S}(x, \underline{k}) \quad (46)$$

and we naturally have  $0 \leq \underline{k} \leq \bar{k}$ . Then, from the derivation of the optimal search intensities, it immediately follows that  $ib(0) > 0 > ib(\bar{k})$ . In the Appendix, we additionally show that for any  $bc'(k) \geq 0$ , we have  $ib'(k) < 0$ , which establishes the following proposition:

**Proposition 1** *Assume  $bc'(k) \geq 0$ . Then there exists a unique dealer-intermediated equilibrium characterized by the benchmark value  $k^* \in [0, \bar{k}]$ .*

### 3.3 Properties of the equilibrium

The equilibrium is defined by the joint determination of the surplus function  $S(x)$  and the equilibrium search intensities  $\rho_{H0}, \rho_{H1}(x), \rho_{L1}(x)$  via the benchmark value  $k^*$ . Importantly, the inter-dealer market allows us to reduce the cross-sectional search problem to a problem of finding trading cutoffs (via the benchmark value) instead of having to account for the whole cross-sectional distribution of asset holdings. This is reminiscent of Hopenhayn (1992): the benchmark surplus value  $k^*$  defines the equilibrium surplus function  $S(x) \equiv \hat{S}(x; k^*)$ , which in turn defines the equilibrium search rates  $\rho_{H0}, \rho_{H1}(x), \rho_{L1}(x)$ , the inter-dealer market price and thus the traded set  $\mathbb{T}$ . Unlike in Hopenhayn (1992), however, we do not use monotonicity of  $S(x)$ , as we are solving via the benchmark surplus value  $k^*$ , instead of the marginal type (which might be non-unique if  $S(x)$  is non-monotone). With equilibrium surplus  $S(x) = \hat{S}(x, k^*)$  and benchmark surplus value  $k^*$  in hand, the differential equations for individual valuations  $V_{H1}(x), V_{L1}(x)$  are now easily solved as they are simple linear ODEs with the particular parts (b), (c) completely characterized by  $S(x)$  and  $k^*$ .

<sup>12</sup>Reflection at the boundaries implies  $\left[ \frac{\sigma^2}{2} \hat{\mu}'_{H1}(x, k) - m \cdot \hat{\mu}_{H1}(x, k) \right]_{x \in \partial \mathbb{X}} = 0$ , so that  $\int_{\mathbb{X}} \mathcal{L}^* \hat{\mu}_{H1}(x, k) dx = 0$ .

**Exchanges provide intermediation services.** Let us now consider why exchanges arise. Assume  $\gamma_1 = 0$  for simplicity. The intuition for exchanges is as follows: if there is a shortage of natural buyers  $\mu_{H0}$ , without *exchanges* some L1 types would be rationed from selling, essentially stuck holding some medium surplus assets while only high surplus assets are traded. Consider H1 types holding a low surplus asset, which most likely will have low holding costs. Even though the asset holding capacity of this investor is tied up, there is benefits to applying the investor's holding capacity instead to asset with high holding costs as the investor is currently not subject to any holding costs. To do this, the H1 type exchanges the asset with the afore-rationed L1 type, thereby resulting in the L1 type holding an asset with low current holding costs. Thus, an asset *exchange* provides additional liquidity to parts of the market that would otherwise be rationed, lowering overall surplus and improving liquidity. The impact on volume is ambiguous as long as search is costly – when surplus decreases, so can the contact intensity of the investors.

Next, let us investigate when exchanges occur. By definition, exchanges do not occur when  $\rho_{H1}(x, k) = 0$  for all  $x \in \mathbb{X}$ :

**Corollary 2** *Let  $k^*$  be the market clearing marginal surplus. Then, we have*

$$\begin{aligned} k^* \in [0, \underline{k}] &\iff \rho_{H1}(x, k^*) = 0, \forall x \in \mathbb{X} \\ k^* \in (\underline{k}, \bar{k}] &\iff \rho_{H1}(x, k^*) > 0, \exists x \in \mathbb{X} \end{aligned}$$

*so that there are no asset exchanges / intermediation for  $k^* \in [0, \underline{k}]$  and there are at least some asset exchanges / intermediation for  $k^* \in (\underline{k}, \bar{k}]$ .*

Note that for  $k^* \in [0, \underline{k}]$  no true marginal type exists as all sellers are of L1 type, and all buyers are of the H0 type. Nevertheless, sellers are able to extract some surplus  $S(x) - k^* > 0$ , and buyers receive surplus  $k^* \geq 0$  conditional on trade. In general, when we ban asset exchanges, a very similar equilibrium holds: only L1 types become sellers, and only H0 are buyers, and there exists a unique benchmark surplus  $k_{NS}$  where *NS* stands for no-exchanges.

**Cross-sectional Volume.** First, we note that  $\rho_{H1}(x) \cdot \rho_{L1}(x) = 0$  so H1 and L1 types never try to offer the same assets for trade. Further, we note that either for the case of sufficient outside buyers ( $k \leq \underline{k}$ ) or for insufficient outside buyers ( $k \geq \bar{k}$ ) but with zero marginal cost of search at  $\rho = 0$  ( $\gamma_1 = 0$ ), we see trading across the whole range of assets, i.e.,  $\mathbb{T} = \mathbb{X}$ . Regardless, as discussed above, the equilibrium trading structure features *differentiated* market volume by type  $x$  — there is more trade for assets with either very high surplus (and also with very low surplus in case of *exchanges*) than for assets close (or below in case of *no-exchanges*) to the benchmark surplus value  $k$ . Thus, in general even with *monotone* holding costs we can see *non-monotone* volume patterns when asset *exchanges* are allowed. When *exchanges* are allowed but do not occur in equilibrium, i.e.,  $k^* \leq \underline{k}$ , then the equilibrium trading intensities follow the shape of the surplus function.

Trading volume for variety  $x$  in our model is given by

$$V(x) \equiv \rho_{H1}(x) \mu_{H1}(x) + \rho_{L1}(x) \mu_{L1}(x) \tag{47}$$

which is the trading volume (single counted) of type  $x$  every instant. Further, the (dollar) bid-ask spread is given by  $BA(x) = (1 - \beta)S(x)$  for  $x \in \mathbb{S}_L$  and is thus just a constant multiple of the surplus function.<sup>13</sup> We note that for  $k \in (\underline{k}, \bar{k})$  we have a non-monotone volume function  $V(x)$  as there exists at least one point at which volume is 0. Finally, define total trading volume (or *integrated* trading volume) in the market (by contracts traded, not by pricing volume) as

$$TV \equiv \int_0^1 V(x) dx. \quad (48)$$

**When does the cross-sectional distribution of assets not matter for volume?** The cross-sectional distribution of assets matters when it influences investors trading decision. First, consider the situation in which  $\gamma_1 = \gamma_2 = 0$  and  $\rho \leq \bar{\rho}$ . Then, if  $k < \underline{k}$ , the exchanges and no-exchanges equilibrium coincide and assets are uniformly intermediated, as is the case for example in He and Milbradt (2014). This implies that the existence of say some very high holding-costs assets does not affect the trading of some very low holding-costs assets. This equilibrium only comes about when  $\mu_{H0}$  is sufficiently high. Next, consider instead the situation in which

$$\bar{\rho} \cdot \mu_{H0} < \bar{\rho} \int_{\mathbb{X}} \mu_{L1}(x) dx \iff \mu_H < 1. \quad (49)$$

Then, there is a shortage of natural buyers. However, even though this will be reflected in how the surplus is split between buyers and sellers, without endogenous search there will be no impact on the actual traded set:

1. In the case of *no-exchanges*, the L1 types with the lowest surplus, i.e., with assets  $x$  such that  $S(x) \leq k$ , are rationed from selling. Consider now a situation with monotone holding costs (wlog increasing), so that there exists  $\bar{x}$  such that  $S(\bar{x}) = k$  and  $S(x) > k$  if and only if  $x > \bar{x}$ . Now, consider a positive shock to holding costs that does not destroy the monotonicity of the holding cost function. Then, by constant search intensities, market clearing immediately implies that the marginal type  $\bar{x}$  cannot change, even though  $k$  increases. In other words, even as  $k$  changes to account for the higher holding costs and thus the higher surplus throughout,  $\bar{x}$  stays constant. Consequently, volumes stays constant, while bid-ask spreads widen.
2. Similarly, even when we allow *exchanges*, again assuming monotone holding costs, there will be exchanger of the H1 type with  $S(x) \leq k$  and sellers of the L1 type with  $S(x) \geq k$ . But by constant search intensities, market clearing immediately implies that the marginal type  $\bar{x}$  cannot change after a monotonicity-preserving shock to holding costs. In other words, even as  $k$  changes to account for the higher holding costs and thus the higher surplus throughout,  $\bar{x}$  and thus volume stays constant even though bid-ask spreads widen.

**When does the cross-sectional distribution of assets matter for volume?** Thus, for the cross-sectional distribution to matter, we need endogenous search, i.e., costly search,

---

<sup>13</sup>The bid-ask spread when *exchanges* are allowed, i.e., on  $x \in \mathbb{S}_H$ , is given in the appendix. The bid-ask spread is ill-defined as it is just one leg of a two leg transaction intermediated by the same dealer.



or we need to assume sufficient changes in the monotonicity of the holding-cost function. The intuition is simple: the surplus function indexes how attractive a trading opportunity  $x$  is to a buyer, or how desperately a seller is trying to exit an asset position  $x$ . Thus, when  $S(x)$  is very high, we would expect both the seller or the prospective buyer to try harder to contact a dealer, resulting in high  $\rho_{L1}(x)$  and high  $\rho_{H0}(x)$ . In the previous example, as contact rates were fixed, this could not happen. With flexible contact rates, greater opportunities in the form of higher  $S(x)$  attract more search activity. However, some opportunities  $x$  becoming more attractive (high  $S(x)$ ) can concentrate volume on these opportunities to the detriment of some intermediate opportunities  $x'$  (low  $S(x')$ ).

Essentially, once more extreme assets exists, and intermediation capacity is limited by costly search, more extreme assets raise the surplus that a prospective buyer can expect, thereby making it harder for intermediate surplus assets to attract any trades.

**Free entry, buyer's and seller's markets.** Next, we will link the equilibrium characterization to notions of buyer's market, seller's market, and free entry.

To be more precise, we say a market features *free-entry* if

$$V_{H0} = V_{L0} = 0 \tag{50}$$

in equilibrium. We call a market a *seller's market* if seller's are able to extract all surplus, that is  $S(x)$  so that  $k = 0$ . We call a market a *buyer's market* if buyer's are able to extract all surplus. The equilibrium outcome of our model is usually something between these two extremes: seller's extract a surplus  $[S(x) - k]^+$  and buyer's extract a surplus  $k > 0$ .

First, consider the situation in which  $V_{H0} = V_{L0} = 0$ . Imposing the free-entry condition, we see that we must have  $(a) = \max_{\rho} \{\rho\beta \cdot k - \gamma(\rho)\} = 0$ . This can only happen in our linear-quadratic search specification if  $k = \frac{\gamma_1}{\beta}$ . Note that  $\rho_{H0} = \frac{1}{\gamma_2}(\beta \cdot k - \gamma_1)^+$ , so that if  $k = \frac{\gamma_1}{\beta}$  and  $\gamma_2 > 0$  we immediately have  $\rho_{H0} = 0$ , i.e., no trade at all. Thus, free-entry with trade is only possible under linear search costs, i.e.  $\gamma_2 = 0$ , so that indifference leads to search of  $\rho_{H0} = \bar{\rho}$  for at least some agents.<sup>14</sup>

Next, we observe that  $V_{H0} > 0$  implies that  $k > \frac{\gamma_1}{\beta}$ , so that some positive fraction of the surplus is appropriated by the buyers.<sup>15</sup> Additionally,  $V_{H0} > 0$  immediately implies positive trading volume, i.e.,  $\rho_{H0} > 0$ , regardless of setup. Thus, outside of a model with no search costs and a capped rate  $\bar{\rho}$ , the model can never feature a true free-entry situation due to its lack of congestion.

---

<sup>14</sup>Let us walk through a simple example: Consider  $\gamma_1 = \gamma_2 = 0$  and  $\rho \leq \bar{\rho}$ . Then everyone searches. If there is too many natural *buyers*, then some of these *buyers* must be rationed, and we therefore must have  $k = 0$  (they do not make anything in equilibrium), a result established in DGP. Why?  $\beta k$  is the benefit per unit of searching  $\rho$ . If there is no cost, then there will be an oversupply of *buyers* if  $\bar{\rho}$  is high enough. Then, we need rationing, but rationing cannot at the same time deliver any equilibrium surplus to the realized non-rationed *buyers*, as otherwise the realized rationed *buyers* would undercut everyone by accepting just a bit less surplus. This is equivalent to free-entry in the model.

<sup>15</sup>For the quadratic search specification, we have

$$V_{H0} \propto \frac{1}{2} \frac{1}{\gamma_2} [(\beta k - \gamma_1)^+]^2$$

so that  $V_{H0} > 0 \iff k > \frac{\gamma_1}{\beta}$ .

**Local Volatility.** Due to our stochastic setup, prices and liquidity (in terms of the bid-ask measure) exhibit local volatility via the Brownian component as in Gârleanu and Pedersen (2007), but in contrast to Gârleanu and Pedersen (2007) our model features a cross-section of assets that are differentially exposed to  $x$  and thus a cross-section of asset volatilities. If  $P(x)$  is the mid-point trading price (or the bid-, or the ask-price), we have local volatility at  $x$  of

$$\sigma(x) P'(x) dB(x) \neq 0. \quad (51)$$

As we will see below, having a tractable expressions for price volatility will allow introduction of haircuts that are linked to asset price volatilities. Volatility disappears only when heterogeneity does not influence price, something that would happen only for very specific symmetric boundary conditions.

## 4 Applications

We will now present a series of applications and extensions to the framework. We assume  $\gamma_1 = 0 < \gamma_2$  for tractability and ease of presentation.

### 4.1 Static asset heterogeneity

First, let us consider the situation in which asset heterogeneity does not change, that is  $\mathcal{L}f(x) = 0$ . Then, for heterogeneity to matter, we need  $h(x)$  to be non-constant. We change this assumption in Section 4.2 and 4.3, where the boundary conditions induce asset heterogeneity even though holding costs are constant, i.e.,  $h(x) = h_0$ .

Let us assume that exchanges are allowed. Then, we have  $\mathcal{L}^x f(x) = 0$ , and we have

$$\hat{r} \cdot S(x) = h(x) - \frac{1}{2\gamma_2} (\beta k)^2 + \frac{1}{2\gamma_2} (\beta [S(x) - k])^2 \left\{ \mathbb{1}_{\{S(x) \leq k\}} - \mathbb{1}_{\{S(x) \geq k\}} \right\}. \quad (52)$$

So for a given arbitrary  $k$  and  $x$ , we can solve  $\hat{S}(x, k)$   $x$ -by- $x$  to get

$$\hat{S}(x, k) = \begin{cases} \frac{-\left[\hat{r} - \frac{\beta^2}{\gamma_2} k\right] + \sqrt{\left[\hat{r} - \frac{\beta^2}{\gamma_2} k\right]^2 + 2\frac{\beta^2}{\gamma_2} \left(h(x) - \frac{\beta^2}{\gamma_2} k^2\right)}}{\frac{\beta^2}{\gamma_2}} & k < \underline{k}(x) \\ \frac{\left[\hat{r} + \frac{\beta^2}{\gamma_2} k\right] - \sqrt{\left[\hat{r} + \frac{\beta^2}{\gamma_2} k\right]^2 - 2\frac{\beta^2}{\gamma_2} h(x)}}{\frac{\beta^2}{\gamma_2}}, & k > \underline{k}(x) \end{cases}. \quad (53)$$

where  $\underline{k}(x) \equiv \frac{-\hat{r} + \sqrt{(\hat{r})^2 + 2\frac{\beta^2}{\gamma_2} h(x)}}{\frac{\beta^2}{\gamma_2}}$ . As long as  $h(x)$  is measurable, volume and thus market

imbalance are clearly defined  $x$ -by- $x$ . Differentiating  $\hat{S}(x, k)$  w.r.t.  $h(x)$ , we see that high  $h(x)$  assets feature higher surplus, and thus are assets that L1 types try to sell out of, whereas low  $h(x)$  assets feature lower surplus, and thus are assets that H1 types will offer to exchange for high  $h(x)$  assets. Search intensities vary in line with holding costs  $h(x)$ , and translate into volume via the assume steady-state distribution  $\mu_1(x)$  and the induced steady-state distribution  $\mu_{H1}(x)$ . Aggregate effects are driven by the common valuation benchmark  $k^*$ . Note that with  $h(x) = cst$ , the model collapses to an endogenous search-version of Duffie et al. (2005).

## 4.2 Deterministic heterogeneity from time-to-maturity in bonds

Next, let us consider risk-free bonds: First, let us interpret  $c(x) = c$  as the bond's coupon, and  $\tau = x$  as its time-to-maturity, so that  $\mathcal{L}f(\tau) = -f'(\tau)$ . Second, following Leland and Toft (1996) we assume that bonds' maturities are evenly distributed on  $[0, T]$ , i.e.,  $\mu_1(\tau) = \frac{1}{T}$ : Whenever a bond matures, i.e. at  $\tau = 0$ , it is re-issued as a  $\tau = T$  bond. Third, we have the boundary conditions  $S(0) = 0$  and  $\mu_{H1}(T) = \frac{1}{T}$ : as a bond pays out cash at maturity, there is no surplus to be gained from trading it at  $\tau = 0$ , and when it is reissued, it is sold to only  $H0$  types on a perfectly liquid primary market.<sup>16</sup> Fourth, we assume that  $h(x) = h_0$ , so that heterogeneity amongst bonds is purely driven by time-to-maturity  $\tau$  via the boundary condition at  $\tau = 0$ .

Then, we have the (auxiliary) surplus being defined by

$$\hat{r} \cdot \hat{S}(\tau, k) = h_0 - \frac{1}{2\gamma_2} (\beta k)^2 - \hat{S}'(\tau, k) + \frac{1}{2\gamma_2} (\beta [\hat{S}(\tau, k) - k])^2 \left\{ \mathbb{1}_{\{\hat{S}(\tau, k) \leq k\}} - \mathbb{1}_{\{\hat{S}(\tau, k) \geq k\}} \right\} \quad (54)$$

with boundary condition  $\hat{S}(0, k) = 0$ . Next, we note that the (auxiliary) steady-state mass equation for  $\hat{\mu}_{H1}(\tau, k)$  is given by

$$0 = -\xi_{HL} \hat{\mu}_{H1}(\tau, k) + \xi_{LH} \hat{\mu}_{L1}(\tau, k) + \hat{\rho}_{L1}(\tau, k) \hat{\mu}_{L1}(\tau, k) - \hat{\rho}_{H1}(\tau, k) \hat{\mu}_{H1}(\tau, k) + \hat{\mu}'_{H1}(\tau, k) \quad (55)$$

with boundary condition  $\hat{\mu}_{H1}(T, k) = \mu_1(T) = \frac{1}{T}$ . Next, note that  $bc(k) = \hat{\mu}_{H1}(T, k) - \hat{\mu}_{H1}(0, k)$  so that

$$ib(k) = [\xi_{LH} + \xi_{HL} + \hat{\rho}_{H0}(k)] \int_0^T \hat{\mu}_{H1}(\tau, k) d\tau - \xi_{LH} - \hat{\rho}_{H0}(k) \mu_H - [\hat{\mu}_{H1}(T, k) - \hat{\mu}_{H1}(0, k)]. \quad (56)$$

We have  $ib(0) > 0 > ib(\bar{k})$  with  $ib'(k) < 0$  where  $\bar{k} < \frac{h_0}{\bar{r}}$ . A closed-form solution exists to the system of (non-linear) first-order ODEs, but is cumbersome. Importantly, the boundary condition introduces strong monotonicity properties, in that  $\hat{S}'(\tau, k) > 0$  and  $\hat{\mu}'_{H1}(\tau, k) < 0$  in equilibrium.

The figures provide further intuition. The left panel of Figure 1 shows the surplus function and the benchmark surplus. We see that surplus is monotonically increasing in time-to-maturity, as was conjectured. The right panel shows the volume for each time to maturity. As there is a shortage of buyers, exchanges arise for low time-to-maturity bonds. Volume is generated by search intensities and the mass of agents searching, shown in the left and right panels of Figure 2 respectively. Given the monotonicity properties, as  $\mu_{L1}(T) = 0$ , we have zero volume at  $\tau = T$ , but positive L1 driven volume on  $(\bar{\tau}, T)$ . Further, we have H1 (exchange) driven volume on  $(0, \bar{\tau})$ , with a maximum at  $\tau = 0$  due to the fact that  $\mu'_{L1}(\tau) = -\mu'_{H1}(\tau) < 0$  as shown in the right panel of Figure 2. Further, not shown in the graph, we have a spike volume at  $\tau = T$  on the primary market. Thus, to sum up, we have primary market volume at  $\tau = T$ , some secondary market volume close to issuance, and then again secondary market volume close to maturity.

<sup>16</sup>Without perfect liquidity, an additional condition arises, as now primary sales are competing with secondary market sales. This does not however change the liquidity mechanism we are highlighting here.

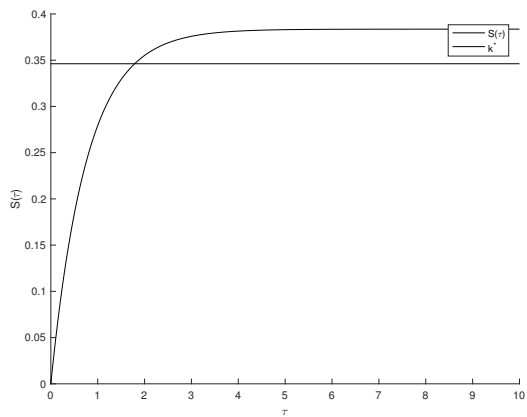


Figure 1: **Left:** Surplus; **Right:** Volume

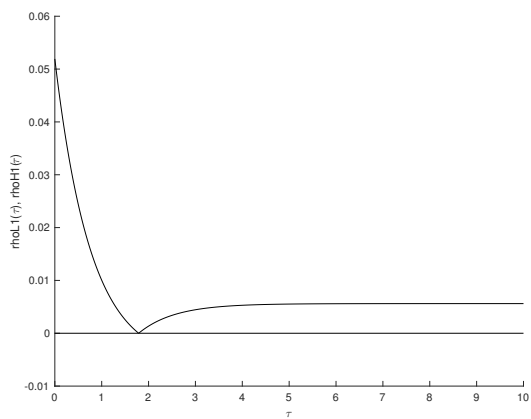


Figure 2: **Left:** Search Intensities; **Right:** Steady-State Masses

### 4.3 Stochastic heterogeneity from distance-to-default in bonds

Next, let us consider a setting involving bonds subject to default risk, such as in Leland (1994) in which firms issue infinite maturity bonds at firm inception. To this end, let us interpret  $x$  as log cash-flows, and suppose that bonds default at some fixed exogenously given boundary  $x_{min}$ . Then, we have  $\mathbb{X} = [x_{min}, \infty)$ . We further assume  $m(x) = m$ ,  $\sigma(x) = \sigma$ , and that the exogenous holding costs are constant at  $h(x) = h_0$ . Finally, suppose that as in He and Milbradt (2014), default occurs at  $x = x_{min}$  with a boundary surplus condition of  $S(x_{min}) = \bar{S}$ , and we concentrate on situations for which  $S'(x) < 0$  so that  $\bar{S} > \lim_{x \rightarrow \infty} S(x) = \underline{S}$  – as we will see below,  $\underline{S}$  is determined in equilibrium due to the endogenous search assumption. Implicitly, we are assuming that holding costs in default are higher than outside of default.

To have a tractable cross-sectional distribution, let us assume a constant “success” rate of firms at an intensity  $\delta$  that leads to immediate cash-payouts of the bond. To close the model, we assume that each departing firms is reborn at a random  $x$  according to a rebirth rate  $b(x)$  and issues one unit of debt to high-value investors only. In a more sophisticated model, such as for example Miao (2005), new firms would optimally choose their initial leverage at birth by issuing bonds of a certain type  $x_0$ . We interpret the variation in initial leverage represented by  $b(x)$  as an outcome of un-modeled firm and issuance costs heterogeneity.<sup>17</sup>

**Steady-state Distributions.** The steady-state distribution  $\mu_1(x)$  is easily calculated via the FKE

$$0 = \frac{\sigma^2}{2} \mu_1''(x) - m \cdot \mu_1'(x) + \overbrace{\frac{\sigma^2}{2} \mu_1'(x_{min}) b(x)}^{\text{Default \& Rebirth}} + \overbrace{b(x) \int_{\mathbb{X}} \delta \cdot \mu_1(x) dx - \delta \mu_1(x)}^{\text{Random death \& Rebirth}} \quad (57)$$

with boundary conditions

$$\mu_1(x_{min}) = 0 \quad \text{and} \quad cst = \int_{\mathbb{X}} \mu_1(x) dx = 1. \quad (58)$$

Let us assume that the rebirth distribution  $b(x) = \lambda e^{-\lambda(x-x_{min})}$ , i.e., an exponential distribution. Then, the steady state bond distribution is given by

$$\mu_1(x) = \frac{\lambda}{\lambda + \eta_1} (-\eta_1) e^{\eta_1(x-x_{min})} + \frac{\eta_1}{\lambda + \eta_1} \lambda e^{-\lambda(x-x_{min})}, \quad (59)$$

i.e., a mixture of two exponential functions with  $\eta_1 \equiv \frac{m - \sqrt{m^2 + 2\delta\sigma^2}}{\sigma^2} < 0$ .

Next, the type specific steady-state distribution  $\mu_{H1}(x)$  is given by the FKE

$$0 = [\xi_{LH} + \rho_{L1}(x)] \mu_1(x) + \mathcal{L}^* \mu_{H1}(x) + \left[ \frac{\sigma^2}{2} \mu_1'(x_{min}) + \delta \right] b(x) - [\xi_{HL} + \xi_{LH} + \rho_{L1}(x) + \rho_{H1}(x) + \delta] \mu_{H1}(x). \quad (60)$$

---

<sup>17</sup>Having  $b(x)$  as a continuous function of  $x$  makes the numerical analysis easier to handle, whereas a singular inflow point  $x_0$  requires the pasting of two differential equations. Importantly, the choice of  $b(x)$  does not alter any of the qualitative liquidity results.

where  $\mathcal{L}^* f(x) = -m \cdot f'(x) + \frac{\sigma^2}{2} f''(x)$  and  $\frac{\sigma^2}{2} \mu_1'(x_{min}) b(x)$  is the inflow of replaced defaulted firms. The boundary condition is  $\mu_{H1}(x_{min}) = 0$  and the integrability condition is given by  $\int_{\mathbb{X}} \mu_{H1}(x) dx < 1$ . Note that the last term, the inflow from default and random resetting, is exogenous. Unlike in the  $\mu_1(x)$  case, the particular parts anchor the function, so that the integrability condition is equivalent to a non-exploding condition.

**Market clearing.** The boundary conditions for the integrated FKE are given by

$$\begin{aligned} bc(k) &= \lim_{x_{max} \rightarrow \infty} \left[ \frac{\sigma^2}{2} \hat{\mu}'_{H1}(x_{max}, k) - m \cdot \hat{\mu}_{H1}(x_{max}, k) \right] - \left[ \frac{\sigma^2}{2} \hat{\mu}'_{H1}(x_{min}, k) - m \cdot \hat{\mu}_{H1}(x_{min}, k) \right] \\ &= 0 - \left[ \frac{\sigma^2}{2} \hat{\mu}'_{H1}(x_{min}, k) \right] \end{aligned} \quad (61)$$

as we must have

$$\lim_{x_{max} \rightarrow \infty} \hat{\mu}_{H1}(x_{max}, k) = \lim_{x_{max} \rightarrow \infty} \hat{\mu}'_{H1}(x_{max}, k) = 0 \quad (62)$$

for any steady-state density that exists. Thus, we have

$$\begin{aligned} ib(k) &= [\xi_{HL} + \xi_{LH} + \delta + \hat{\rho}_{H0}(k)] \int_{\mathbb{X}} \hat{\mu}_{H1}(x) dx - \xi_{LH} - \delta - \hat{\rho}_{H0}(k) \mu_H \\ &\quad - \frac{\sigma^2}{2} [\mu_1'(x_{min}) - \hat{\mu}'_{H1}(x_{min}, k)] \end{aligned} \quad (63)$$

where we substituted in from the integrated FKE. Note that  $\mu_1'(x_{min}) = -\lambda\eta_1 > 0$  is a constant, and that  $\hat{\mu}'_{L1}(x_{min}, k) = \mu_1'(x_{min}) - \hat{\mu}'_{H1}(x_{min}, k)$ . We again have  $ib(0) > 0 > ib(\bar{k})$  with  $ib'(k) < 0$  where  $\bar{k} = \bar{S}$ .

**Surplus.** Next, let us compute the surplus function when  $x \rightarrow \infty$ . We assume  $\gamma_1 = 0$  throughout for ease of exposition. First, the auxiliary surplus function  $\hat{S}(x, k)$  is given by the non-linear ODE

$$\begin{aligned} (\hat{r} + \delta) \hat{S}(x; k) &= h_0 - \frac{1}{2\gamma_2} (\beta k)^2 + m \cdot \hat{S}'(x; k) + \frac{1}{2} \sigma^2 \hat{S}''(x; k) \\ &\quad + \frac{1}{2\gamma_2} \left( \beta [\hat{S}(x; k) - k] \right)^2 \left\{ \mathbf{1}_{\{\hat{S}(x; k) \leq k\}} - \mathbf{1}_{\{\hat{S}(x; k) \geq k\}} \right\} \end{aligned} \quad (64)$$

with boundary conditions

$$\hat{S}(x_{min}; k) = \bar{S} \quad \text{and} \quad \lim_{x \rightarrow \infty} |\hat{S}(x; k)| = \hat{\underline{S}}(k) < \infty. \quad (65)$$

Here,  $\hat{\underline{S}}(k)$  is an endogenous function: As  $x \rightarrow \infty$ , risk disappears, and  $\hat{S}(x; k)$  becomes perfectly flat (i.e.,  $\lim_{x \rightarrow \infty} \mathcal{L}^x \hat{S}(x, k) = 0$ ) at its limit value

$$\hat{\underline{S}}(k) = \begin{cases} \frac{-\left[(\hat{r} + \delta) - \frac{\beta^2}{\gamma_2} k\right] + \sqrt{\left[(\hat{r} + \delta) - \frac{\beta^2}{\gamma_2} k\right]^2 + 2 \frac{\beta^2}{\gamma_2} \left(h_0 - \frac{\beta^2}{\gamma_2} k^2\right)}}{\frac{\beta^2}{\gamma_2}} & k < \underline{k} \\ \frac{\left[(\hat{r} + \delta) + \frac{\beta^2}{\gamma_2} k\right] - \sqrt{\left[(\hat{r} + \delta) + \frac{\beta^2}{\gamma_2} k\right]^2 - 2 \frac{\beta^2}{\gamma_2} h_0}}{\frac{\beta^2}{\gamma_2}}, & k > \underline{k} \end{cases} \quad (66)$$

where

$$\underline{k} = \hat{\underline{S}}(\underline{k}) = \frac{-(\hat{r} + \delta) + \sqrt{(\hat{r} + \delta)^2 + 2\frac{\beta^2}{\gamma_2}h_0}}{\frac{\beta^2}{\gamma_2}}. \quad (67)$$

Finally, for the assumption  $\hat{S}'(x; k) < 0$  to hold we need  $\bar{S} > \hat{\underline{S}}(k^*)$ . Therefore, let us concentrate on the case where  $\bar{S} > \max_{k \in [0, \bar{S}]} \hat{\underline{S}}(k)$ . To this end, note that the (unconstrained) maximum of  $\hat{\underline{S}}(k)$  is given by  $\hat{\underline{S}}(k_{max}) = 2k_{max}$  where

$$k_{max} = \frac{-(\hat{r} + \delta) + \sqrt{(\hat{r} + \delta)^2 + \frac{\beta^2}{\gamma_2}h_0}}{\frac{\beta^2}{\gamma_2}}. \quad (68)$$

Thus, let us assume  $\bar{S} > 2k_{max}$  throughout. Further, note that even though  $\hat{\underline{S}}(k)$  is single-peaked at  $k_{max}$  and thus non-monotone, we still have  $\frac{\partial}{\partial k} [\hat{\underline{S}}(k) - k] < 0$ .

**Numerical Results.** The proofs of Section 3 can be easily adapted to show there exists a unique equilibrium. To show some of the implications of our setup for bond-trading, we now solve the model for a set of parameters. We pick the parameters as follows:

Parameter	Value	Description	Parameter	Value	Description
$r$	5%	Interest rate	$x_{min}$	0	default boundary
$c$	5%	Coupon rate	$h_0$	.1	exogenous holding costs
$\xi_{HL}$	1	Liquidity shock	$\bar{S}$	4	default surplus
$\xi_{LH}$	.25	Liquidity shock recovery	$\gamma_1$	0	linear search costs
$\delta$	1%	Random success	$\gamma_2$	1	quadratic search costs
$\beta$	15%	Bargaining power	$\mu$	1.1	mass agents
$m$	0	drift	$\lambda$	0.1%	rebirth parameter
$\sigma$	40%	volatility	$k^*$	0.0935	equilibrium benchmark

Figure 3 and Figure 4 illustrate the outcomes. First, the left panel of Figure 3 depicts as the solid black line the surplus function  $S(x)$  and the cutoff value  $k^*$ . As  $\underline{S} < k^* < \bar{S}$ , asset exchanges occur in equilibrium between H1 types holding highly rated bonds (large distance to default, i.e., high  $x$ ) and L1 types holding lowly rated bonds (small distance to default, i.e.,  $x$  close to  $x_{min}$ ). We further note that surplus is decreasing in distance to default, i.e.,  $S'(x) < 0$ . The right panel depicts the trading volume for the exogenous holding cost model as a solid line. Note that volume is non-monotone, with volume peaking in lowly rated bonds (small  $x$ ) and (at a much lower peak) for highly rated bonds.<sup>18</sup> In real life, we often see flurries of trading when bonds are at imminent risk of default.

Figure 4 depicts the search intensities of the different agents as solid black lines. We see that H1 types with high  $x$  bonds search, but at a low intensity, whereas L1 types with low  $x$  bonds search with increasingly higher intensities. Unlike volume, search intensities are monotonically increasing for  $x > \bar{x}$  where  $S(\bar{x}) = k^*$  and monotonically decreasing for

<sup>18</sup>As for highly rated bonds, there is a long tail of very safe bonds. Depending on where the AAA category starts, it is the integral of the volume that matters.

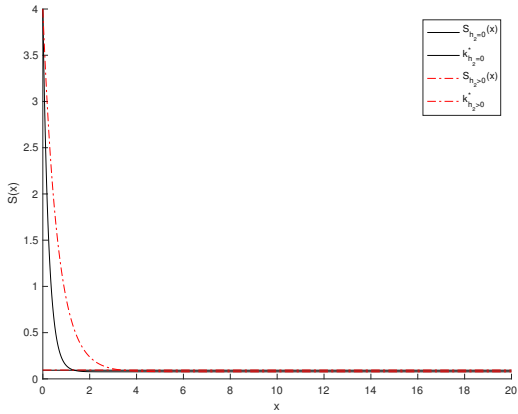


Figure 3: **Left:** Surplus; **Right:** Volume

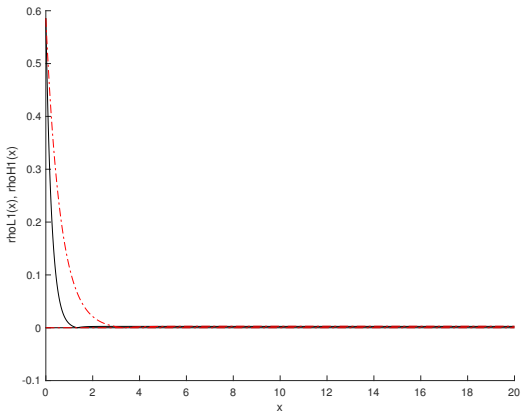


Figure 4: **Left:** Search intensities; **Right:** Holding costs

$x < \bar{x}$ . The difference to volume comes from the steady state distribution  $\mu_{H1}(x), \mu_{L1}(x)$ . In this example, a large proportion of asset holders are of the low type, as shocks die out at a low rate and counter-party search is relatively expensive. Finally, the right panel depicts the exogenous holding  $h_0(x)$  of the agents as a horizontal black line at  $h_0 = 1$ .

## 5 Volatility, Margins and Holding Costs

Next, suppose holding costs, via say internal Value-at-Risk (VaR) weights or haircuts, are linked to asset price dynamics, specifically price volatility. Consider the following situation: after a liquidity shock, the agent is subject to a large cash shortfall. To make up this shortfall, the agent can either borrow unsecuredly at a rate  $r + \chi$  or securedly at a rate  $r$ . Holding a unit of asset  $x$  with current market price  $P(x)$  allows the agent to pledge the asset in a margin loan to receive the secured rate  $r$ . However, this loan will be subject to a haircut  $hair$  so that only  $[1 - hair] P(x)$  can be borrowed at the rate  $r$  for a saving of  $\chi$  per unit of



cash borrowed.<sup>19</sup> As Krishnamurthy et al. (2014) show, haircuts are well described as linear functions of return volatility in the data. Applying Ito's lemma, we thus have

$$Volatility(\log P(x)) \propto \sigma \frac{|P'(x)|}{P(x)}. \quad (69)$$

Supposing that the haircut is a linear function of return volatility, and in addition there is some baseline exogenous holding costs  $h_0(x)$ , the total effective holding costs can then be written as

$$\tilde{h}_0(x) - \chi[1 - hair] P(x) = h_0(x) + h_1 P(x) + h_2 |P'(x)|. \quad (70)$$

The last term arises by multiplying the return volatility with the price itself, cancelling out the price term in the denominator. In a single asset model with stochastic distance to default, the linear specification of the holding costs still allows the model to be solved in closed form. However, in our setup of many heterogenous assets, this specification breaks the role of  $S(x)$  as a sufficient statistic for calculations of search intensities and holding costs.

To keep the model tractable, we make the simplifying assumption that holding costs are a linear function of the surplus function  $S(x)$  (and  $S'(x)$ ) instead of being a linear function of the the price function  $P(x)$  (and  $P'(x)$ ). Thus, holding costs are a function of both the level of liquidity  $S(x)$  and of the volatility/slope of liquidity  $S'(x)$ :<sup>20</sup>

$$h(x) = h_0(x) + h_1 S(x) + h_2 |S'(x)|. \quad (71)$$

where  $h_0(x)$  denotes the *exogenous* part of the holding costs as specified above, and  $h_1 S(x) + h_2 |S'(x)|$  denotes the *endogenous* part of the holding costs. This defines a functional fixed point in that the equilibrium holding costs function is now an equilibrium object as level  $S(x)$  and volatility of liquidity  $|S'(x)|$  are equilibrium objects themselves. This specification changes the pricing equation for  $S(x)$  directly. Plugging in, we see that that the equilibrium drift of the pricing measure,

$$\tilde{m}(x) \equiv m(x) + h_2 \left( \mathbb{1}_{\{S'(x) > 0\}} - \mathbb{1}_{\{S'(x) < 0\}} \right) \quad (72)$$

diverges from the drift of the physical measure we use for the steady-state mass equations. Similarly, we now have a distorted discount rate

$$\tilde{r} \equiv \hat{r} - h_1. \quad (73)$$

Thus, haircuts linked to price dynamics introduce a distortion in the dynamics of the fundamental  $x$  that is akin to the risk-aversion adjustment when using the risk-neutral pricing approach.

---

<sup>19</sup>It is unimportant which price  $P(x)$  is being used, as long as it is a linear combination of the ask-price  $A(x)$  and the bid-price  $B(x)$ .

<sup>20</sup>The argument here is that with monotone prices and monotone surplus, price volatility and surplus volatility are roughly proportional, i.e.,  $\sigma |P'(x)| \approx cst \times \sigma |S'(x)|$  for some constant  $cst$ . If we were not interested in the joint cross-sectional properties of liquidity and volume (i.e.  $k^*$ ), and instead assumed uniform intermediation of all  $x$  at a rate  $\bar{p}$  as in say He and Milbradt (2014) brought about by a sufficient mass of H0 buyers on the sideline then we could solve the haircut to pricing volatility feedback loop without approximation.

## 5.1 Revisiting the Leland example

Let us now revisit the Leland example of Section 4.2, but under partially endogenous holding costs as just discussed. Assuming  $h(x) = h_0(x) + h_1 \hat{S}(x; k) + h_2 |\hat{S}'(x; k)|$ , then the surplus function is given by the non-linear ODE

$$\begin{aligned} (\hat{r} + \delta - h_1) \hat{S}(x; k) &= h_0 - \frac{1}{2\gamma_2} (\beta k)^2 + (m - h_2) \hat{S}'(x; k) + \frac{1}{2} \sigma^2 \hat{S}''(x; k) \\ &\quad + \frac{1}{2\gamma_2} \left( \beta [\hat{S}(x; k) - k] \right)^2 \left\{ \mathbb{1}_{\{\hat{S}(x; k) \leq k\}} - \mathbb{1}_{\{\hat{S}(x; k) \geq k\}} \right\} \end{aligned} \quad (74)$$

where we made the assumption  $\hat{S}'(x; k) \leq 0$  so that  $|\hat{S}'(x; k)| = -\hat{S}'(x; k)$ . The boundary conditions are again given by

$$\hat{S}(x_{min}; k) = \bar{S} \quad \text{and} \quad \lim_{x \rightarrow \infty} |\hat{S}(x; k)| = \underline{\hat{S}}(k) < \infty \quad (75)$$

for a slight generalization of  $\underline{\hat{S}}(k)$  that now includes  $h_1 \neq 0$ : As  $x \rightarrow \infty$ , risk disappears, and  $\hat{S}(x; k)$  becomes perfectly flat at the limit value

$$\underline{\hat{S}}(k) = \begin{cases} \frac{-\left[(\hat{r} + \delta - h_1) - \frac{\beta^2}{\gamma_2} k\right] + \sqrt{\left[(\hat{r} + \delta - h_1) - \frac{\beta^2}{\gamma_2} k\right]^2 + 2\frac{\beta^2}{\gamma_2} \left(h_0 - \frac{\beta^2}{\gamma_2} k^2\right)}}{\frac{\beta^2}{\gamma_2}} & k < \underline{k} \\ \frac{\left[(\hat{r} + \delta - h_1) + \frac{\beta^2}{\gamma_2} k\right] - \sqrt{\left[(\hat{r} + \delta - h_1) + \frac{\beta^2}{\gamma_2} k\right]^2 - 2\frac{\beta^2}{\gamma_2} h_0}}{\frac{\beta^2}{\gamma_2}}, & k > \underline{k} \end{cases} \quad (76)$$

where

$$\underline{k} = \underline{\hat{S}}(\underline{k}) = \frac{-(\hat{r} + \delta - h_1) + \sqrt{(\hat{r} + \delta - h_1)^2 + 2\frac{\beta^2}{\gamma_2} h_0}}{\frac{\beta^2}{\gamma_2}}. \quad (77)$$

The maximum occurs at

$$k_{max} = \frac{-(\hat{r} + \delta - h_1) + \sqrt{(\hat{r} + \delta - h_1)^2 + \frac{\beta^2}{\gamma_2} h_0}}{\frac{\beta^2}{\gamma_2}}. \quad (78)$$

and the (unconstrained) maximum is given by  $\underline{\hat{S}}(k_{max}) = 2k_{max}$ . Thus, we assume  $\bar{S} > 2k_{max}$ .

**Liquidity feedback.** Note that by the differentiability of  $S(x)$ , we can write

$$S(x_{min}) - \lim_{x \rightarrow \infty} S(x) = \bar{S} - \underline{S} = - \int_{x_{min}}^{\infty} S'(x) dx. \quad (79)$$

To understand the outcome of the volatility feedback better, consider first a limit value  $\underline{S}$  that is independent of  $k$ . Then, the local volatility of the surplus,  $\sigma |S'(x)|$ , integrates to a constant regardless of  $k$ , as  $S(x_{min}) - \lim_{x \rightarrow \infty} S(x) = \bar{S} - \underline{S}$ . Consequently, when we change the volatility loading of holding costs  $h_2 > 0$ , it will simply shift around volatilities  $\sigma |S'(x)|$

while integrating to the same constant. This, however, does not mean that aggregate asset volatility is constant: aggregate volatility is an expression

$$\sigma \int_{x_{min}}^{\infty} |S'(x)| \mu_1(x) dx \quad (80)$$

and as long as  $\mu_1(x)$  is not flat, changes in the slope  $S'(x)$  change aggregate surplus volatility.

Next, consider the additional impact of the equilibrium benchmark  $k^*$  on the limit value  $\underline{S} = \hat{S}(k^*)$ . As  $h_2$  shifts,  $k^*$  shifts, and we additionally have volatility affecting the boundary condition. Thus, even though we are mechanically excluding a feedback effect of volatility on the endogenously chosen default boundary  $x_{min}$  (as is common in Leland-type models), we nevertheless find *aggregate* volatility effects.

**Numerical Results.** We can now solve the model for a set of parameters. We pick the parameters in addition to the ones in Section 4.2, and thus have a new  $k^*$ :

Parameter	Value	Description
$h_1$	0	surplus-level holding costs
$h_2$	.7	surplus-vol holding costs
$k^*$	0.0954	equilibrium benchmark

The dashed red lines in Figure 3 and Figure 4 illustrate the outcomes. First, left panel of in Figure 3, we see that the new surplus function under endogenous holding costs, depicted by the dashed red line, lies above the surplus function induced by the exogenous holding costs function, depicted by the solid black line. Again, asset exchanges are occurring in equilibrium. Second, the right panel depicts the changes in trading volume, again depicted as dashed red line: as agents for low  $x$  now face larger endogenous holding costs due to increases surplus volatility, the volume peak has shifted up substantially. Similarly, volume for very safe bonds (generated by asset exchanges) has also increased, albeit at a smaller scaled to take advantage of a market awash in desperate sellers.

Next, the left panel of Figure 4 shows as dashed red lines the new (usually higher) search intensities induced by the endogenous higher holding costs. Finally, the right panel shows the amplification in holding costs by depicting  $h(x)$  as a dashed-red line. We see that the amplifications of the exogenous holding costs (which we fixed at  $h_0 = 1$ ) is on the order of a factor of 4.5.

As holding costs are amplified, surplus increases uniformly compared to the previous  $h_2 = 0.1$  example. By the logic of optimally chosen search intensities, the previously marginal investors with  $\rho_{L1}(x) = 0$  are now actively searching. Most of the previously very active investors, that is with high  $S(x)$ , are now searching even more, but as  $\hat{S}(x, k^*) - k^*$  has increased as  $k^*$  increased. If  $k^*$  had remained constant, we see that there is an increase in sellers (extensive margin), as well as the already active sellers would increase their search intensity (intensive margin). To clear the market, the benchmark value  $k^*$  (and thus the surplus appropriated by buyers) had to increase, thereby attracting more buyers. By the quadratic cost specification, some of the already desperate sellers (with high  $S(x)$ ) do not increase their search intensities as much as the outside buyers, thus leaving an opportunity for previously inactive natural sellers with varieties around the old cutoff  $\bar{x}$ .

## 5.2 Cross-margining, volatility and asset prices

Let us now consider the introduction of cross-margining in this model. Thus far, by an appropriate assumption of divisibility of search for each  $x$ , the model applied equally to portfolios of bonds (with the aggregate amount of bonds held equal to  $a = 1$ ) or single positions of  $a = 1$  in a specific bond. This was without loss of generality as we had risk-neutral agents that were exposed to an individual (but not asset) level liquidity shock. Thus, there were no diversification benefits from diversified portfolio holdings, as holding costs were specific to the asset, but only applied when the asset holder was exposed to a shock.

With the introduction of endogenous holding costs arising out of the volatility of the surplus, there can now be a role for diversified portfolios. Consider first the situation in which any broker dealer does not allow any cross-margining. Then, holding costs are applied  $x$ -by- $x$ , and investors are indifferent between all portfolios.

Next, consider the situation in which a broker dealer allows for cross-margining. In practice, this means that the risk of the margin-loan is calculated on the volatility of the total pledged portfolio, and not  $x$ -by- $x$  (i.e., asset by asset). Then, a diversified portfolio of all assets in the proportions  $\mu_1(x)$  would lead to zero volatility in the surplus and in the prices of the portfolio. Consequently, the applicable holding costs asset by asset would be only the residual holding costs  $h_0(x)$  after the volatility components  $h_2$  are set to zero. Thus, surplus decreases everywhere, total volatility decreases, and asset prices increase. However, as the optimal search intensities of the investors adjust to the decreases surplus, trading volume and the traded set tend to decrease as well. In contrast, if  $\gamma_1 = \gamma_2 = 0$  and  $\rho \leq \bar{\rho}$ , then unless there is some non-monotonicity in the surplus function, only bid-ask spreads adjust while volume stays constant.

## 6 Optimal heterogeneity of bonds

Here, we present a corporate finance application of the trading environment. We will show that introducing asset heterogeneity via type  $\tau$  can be beneficial if it allocates liquidity risk in a way that interacts with optimal search incentives that leads to an improvement in the allocation of the liquidity risk vis-a-vis a homogenous asset world. The intuition is as follows: In a homogenous asset world, all investors are equally exposed to future effective holding costs. If there is a shortage of buyers on the sideline, liquidity shocked investors will be uniformly rationed. In a heterogenous asset world, endogenous search focuses volume on those varieties that feature the highest surplus, which are most likely those asset that feature the highest effective holding costs. Thus, if there is a shortage of buyers on the sideline, liquidity concentrates on providing liquidity to the liquidity shocked investors with the highest effective holding costs, while ignoring those liquidity shocked investors with low effective holding costs, leading to a decrease in overall surplus and higher asset prices. Importantly, our setup does not feature thick market externalities that would naturally lead to concentrations in liquidity. Our example, however, does feature an interaction with the primary market placement technology that leads to

Consider a firm that continuously issues finite-maturity risk-free debt, while keeping

the outstanding face-value of bonds (wlog  $F = 1$ ) constant. The market for bonds has a trading technology characterized by  $\gamma_1 = \gamma_2 = 0$  and  $\rho \leq \bar{\rho}$ , and it features constant holding costs  $h$  for each liquidity shocked investor holding the bond. We assume that large debt issues are prohibitively expensive, and instead consider two debt maturity structures featuring stationary rollover: (1) each outstanding bond has a deterministic time-to-maturity  $\tau \in [0, T]$ , and bonds are evenly spread out between  $[0, T]$  so that  $\mu_1(\tau) = \frac{1}{T}$  as in Section 4.2, or (2) each bond has a stochastic maturity that arrives with intensity  $\delta$ , so that all outstanding bonds are homogenous. Bonds are issued for  $D_{primary}(\tau) = V_{H1}(\tau) - V_{H0}$  on the primary market if the marginal buyer is H0, and for  $D_{primary}(\tau) = V_{L1}(\tau) - V_{L0}$  if the marginal buyer at  $\tau$  is L0.

Note that with uniform intermediation intensities, there is  $\bar{\rho}\mu_{H0}$  natural buyers arriving in the market, and  $\bar{\rho} \cdot \bar{\mu}_{L1}$  natural sellers. Note then that

$$\mu_{H0} > \bar{\mu}_{L1} \iff \mu_H - \bar{\mu}_{H1} > \bar{\mu}_{L1} \iff \mu_H > \bar{\mu}_{H1} + \bar{\mu}_{L1} = 1. \quad (81)$$

**Homogenous bond.** First, note that the homogenous bond can never feature exchanges. Second, it follows the Duffie et al. (2005) pricing formula without bilateral meetings that<sup>21</sup>

$$(\hat{r} + \delta)S = h - \bar{\rho}\beta S \iff S(\delta) \equiv \frac{h}{\hat{r} + \bar{\rho}\beta + \delta} \quad (82)$$

regardless of  $\mu_{H0} > \mu_{L1}$  or vice-versa. For  $\mu_{H0} > \mu_{L1}$ , we have

$$0 = \xi_{HL}\mu_{H1} - (\xi_{LH} + \bar{\rho} + \delta)\mu_{L1} \iff \mu_{L1}(\delta) = \frac{\xi_{HL}}{\xi_{HL} + \xi_{LH} + \bar{\rho} + \delta}. \quad (83)$$

**Heterogenous bond.** The heterogenous bond case follows the liquidity and pricing in Section 4.2. Surplus is given by

$$\hat{r}S(\tau) = h - \bar{\rho}\beta k - S'(\tau) + \bar{\rho}\mathbb{1}_{exchange}\beta[k - S(\tau)]^+ - \bar{\rho}\beta[S(\tau) - k]^+ \quad (84)$$

with  $S'(\tau) > 0$  and  $S(0) = 0$ . For  $\mu_{H0} > \mu_{L1}$ , we have  $k = 0$  and thus

$$\hat{r}S(\tau) = h - \bar{\rho}\beta S(\tau) - S'(\tau) \iff S(\tau) = \frac{h}{\hat{r} + \bar{\rho}\beta} \left[1 - e^{-(\hat{r} + \bar{\rho}\beta)\tau}\right]. \quad (85)$$

For  $\mu_{H0} < \mu_{L1}$ , however, we either have (i) rationing of bonds  $\tau \in [0, \bar{\tau}_{NS}]$  if exchanges are not allowed, and (ii) we have exchanges for  $\tau \in [0, \bar{\tau}_{exchange}]$  where  $0 < \bar{\tau}_{exchange} < \bar{\tau}_{NS}$ . Concentrating on case (ii), we have  $\mathbb{1}_{exchange} = 1$  and the surplus function is as derived in case  $\mu_{H0} > \mu_{L1}$ .

The density  $\mu_{L1}(\tau)$  solves

$$0 = \xi_{HL}\mu_{H1}(\tau) - \xi_{LH}\mu_{L1}(\tau) - \bar{\rho}\mu_{L1}(\tau)\mathbb{1}_{\{\tau > \bar{\tau}\}} + \bar{\rho}\mu_{H1}(\tau)\mathbb{1}_{\{\tau < \bar{\tau}\}}\mathbb{1}_{exchange} + \mu'_{L1}(\tau) \quad (86)$$

with  $\mu_{L1}(T) = 0$  and where  $\bar{\tau} \geq 0$  solves the market clearing condition

$$\bar{\rho}\mu_{H0} + \mathbb{1}_{exchange}\bar{\rho} \int_0^{\bar{\tau}} \mu_{H1}(\tau) d\tau \geq \bar{\rho} \int_{\bar{\tau}}^T \mu_{L1}(\tau) d\tau \quad (87)$$

<sup>21</sup>This can be easily derived from the formula in the heterogenous equation below by setting  $S'(\tau) = \delta S$  and noting that  $k \in \{0, S\}$  depending on which side of the market is tighter.

with equality when  $\bar{\tau} > 0$ . When  $\bar{\tau} = 0$ , then we have

$$\mu_{L1}(\tau) = \frac{1}{T} \frac{\xi_{HL}}{\xi_{HL} + \xi_{LH} + \bar{\rho}} \left[ 1 - e^{-(\xi_{HL} + \xi_{LH} + \bar{\rho})(\tau - T)} \right]. \quad (88)$$

**Average surplus.** Next, we want to calculate the average surplus in the market. To this end, we need to calculate the chance that a randomly selected  $L1$  type holds a maturity  $\tau$  asset, which is given by  $\frac{\mu_{L1}(\tau)}{\int_0^T \mu_{L1}(\tau) d\tau}$ . Integrating, we have

$$\begin{aligned} \bar{\mu}_{L1}(T) &= \frac{\xi_{HL}}{\xi_{HL} + \xi_{LH} + \bar{\rho}} - \frac{\xi_{HL} \left[ 1 - e^{-(\xi_{HL} + \xi_{LH} + \bar{\rho})T} \right]}{T (\xi_{HL} + \xi_{LH} + \bar{\rho})^2} \\ &= \frac{\xi_{HL}}{\xi_{HL} + \xi_{LH} + \bar{\rho}} \left\{ 1 - \frac{\left[ 1 - e^{-(\xi_{HL} + \xi_{LH} + \bar{\rho})T} \right]}{T (\xi_{HL} + \xi_{LH} + \bar{\rho})} \right\}. \end{aligned} \quad (89)$$

Define  $\delta_\mu(T)$  as the random maturity that gives exactly the same amount of liquidity shocked holders as the heterogenous case with deterministic maturity  $T$ , i.e.,  $\bar{\mu}_{L1}(T) = \mu_{L1}(\delta_\mu(T))$ . Then

$$\delta_\mu(T) \equiv f(T(\xi_{HL} + \xi_{LH} + \bar{\rho})) \frac{1}{T} \quad \text{where} \quad f(x) \equiv \frac{x(1 - e^{-x})}{x - (1 - e^{-x})}. \quad (90)$$

Note that  $f(x) \in (1, 2)$  and  $f'(x) < 0$ . We note that  $\mu_{L1}\left(\frac{2}{T}\right) < \bar{\mu}_{L1}(T) < \mu_{L1}\left(\frac{1}{T}\right)$ , so matching expected bond maturity *at issuance*, i.e.,  $\delta = \frac{1}{T}$ , we have the heterogenous market is more efficient in terms of reducing the total amount of liquidity shocked holders. However, matching average maturity of *outstanding* bonds, i.e.,  $\delta = \frac{2}{T}$ , we have that the homogenous market is more efficient in terms of reducing the total amount of liquidity shocked holders (as it also implies faster rollover, which uses the primary market resetting mechanisms more often). Next, integrating surplus  $S(\tau)$  w.r.t.  $\frac{\mu_{L1}(\tau)}{\int_0^T \mu_{L1}(\tau) d\tau}$ , after tedious algebra we can show that

$$\bar{S}(T) \equiv \frac{\int_0^T S(\tau) \mu_{L1}(\tau) d\tau}{\int_0^T \mu_{L1}(\tau) d\tau} < S\left(\frac{1}{T}\right) \quad (91)$$

but  $S\left(\frac{2}{T}\right) \leq \bar{S}(T)$ .<sup>22</sup> Let us impose that  $\delta = \delta_\mu(T)$  from above so that we are comparing cases with the same total amount of liquidity shocked holders. Then, we have  $\bar{S}(T) < S(\delta_\mu(T))$  – the average surplus is higher in the homogenous case when we equate the total amount of liquidity shocked holders. Further, this difference becomes larger as soon as  $\mu_{H0} < \bar{\mu}_{L1}$ , as then liquidity provision in the heterogenous case is on the assets that have

<sup>22</sup>Integrating  $S(\tau)$  w.r.t.  $\mu_1(\tau)$  does not give the average surplus of a seller, as sellers are *not* equally distributed on  $[0, T]$ . We would then have

$$\bar{S}(T) \equiv \int_0^T \frac{S(\tau)}{T} d\tau = \frac{h}{\hat{r} + \bar{\rho}\beta} \left[ 1 - \frac{1 - e^{-(\hat{r} + \bar{\rho}\beta)T}}{(\hat{r} + \bar{\rho}\beta)T} \right]$$

and we can show that  $S\left(\delta = \frac{1}{T}\right) > \bar{S}(T) > S\left(\delta = \frac{2}{T}\right)$ .

the largest surplus, whereas in the homogenous case it is still uniformly across assets. Let  $\delta_S(T)$  be the random maturity that equates the average surplus

$$\bar{S}(T) = S(\delta_S(T)). \quad (92)$$

Consider starting with parameters such that  $\mu_H = 1$ , and then perturbing parameters such that  $\mu_H = 1 - \varepsilon < 1$ , so that we are just entering rationing after the perturbation. Then, liquidity will be withdrawn uniformly in the homogenous asset case, raising (average) surplus. However, in the heterogenous asset case, liquidity will be withdrawn from very low  $\tau$ , say for  $\tau \in [0, l]$  with  $l$  small. But the assets that are losing liquidity have almost zero surplus  $S(\tau) \approx 0$  afterward as they are such short maturity, leading to almost no increase in the average surplus, and with no effect on  $\mu_{L1}(\tau), \tau \geq l$ .<sup>23</sup> Thus, average surplus increases significantly less due to asset heterogeneity, interacted with the primary market technology, to concentrate liquidity on the most needed investors.

## 7 Conclusion

We considered a cross-section of assets of heterogenous and stochastic characteristics traded via an OTC search market with dealer-intermediated trade. Trade takes place as liquidity shocked holders are subject to asset-specific holding costs and thus want to sell assets to buyers who are not liquidity shocked. The buyer's absorption capacity is limited due to a holding limits assumption common in the search literature. Under endogenous contact intensities, market prices (and their volatility), liquidity and volume are jointly determined in equilibrium. The equilibrium is unique under mild conditions. When the natural buying capacity of investors on the sideline is too small, either some asset varieties are rationed or additional intermediation ("asset exchanges") between holders of different assets arises, providing additional liquidity. We partially endogenize holding costs by linking them to haircuts in collateralized borrowing, which yielded a positive linear relation between volatility of the surplus and the holding costs of an asset. This introduced a feedback loop leading to an amplification mechanism between the volatility of the surplus and holding costs of an asset. As investors are risk-neutral and subject to insurable liquidity shocks, there is no role for traditional diversification. However, in the case of volatility linked holding costs, there is a role for diversification if cross-margining on haircuts is allowed. In this case, a diversified bond portfolio can reduce the exposure to the volatility-linked portion of the holding costs, leading to lower illiquidity, lower asset price volatility, and higher asset prices.

---

<sup>23</sup>The reason is that we have a first order ODE that is pinned down by a boundary condition at  $T$ . Thus, contagion can only flow from high to low  $\tau$ , not the other way around.

## References

- Afonso, Gara and Ricardo Lagos**, “Trade Dynamics in the Market for Federal Funds,” *Econometrica*, 2014, *forthcomin*.
- Duffie, Darrell, Nicolae Gârleanu, and Lasse H Pedersen**, “Over-the-Counter Markets,” *Econometrica*, 2005, *73* (6), 1815–1847.
- , – , and – , “Valuation in Over-the-Counter Markets,” *Review of Financial Studies*, nov 2007, *20* (6), 1865–1900.
- Farboodi, Maryam, Gregor Jarosch, and Robert Shimer**, “Meeting technologies in decentralized asset markets,” *Working Paper*, 2016.
- Gârleanu, Nicolae and Lasse H Pedersen**, “Liquidity and risk management,” *American Economic Review*, 2007, *97* (2), 193–197.
- He, Zhiguo and Konstantin Milbradt**, “Endogenous liquidity and defaultable bonds,” *Econometrica*, 2014, *82* (4), 1443–1508.
- Hopenhayn, Hugo A**, “Entry, Exit, and Firm Dynamics in Long Run Equilibrium,” *Econometrica*, 1992, *60* (5), 1127–1150.
- Hugonnier, Julien, Benjamin Lester, and Pierre-Olivier Weill**, “Heterogeneity in Decentralized Asset Markets,” *Working Paper*, 2014.
- Krishnamurthy, Arvind, Stefan Nagel, and Dmitry Orlov**, “Sizing Up Repo,” *Journal of Finance*, 2014.
- Lagos, Ricardo and Guillaume Rocheteau**, “Search in asset markets: Market structure, liquidity, and welfare,” *American Economic Review*, 2007, *97* (2), 198–202.
- and – , “Liquidity in asset markets with search frictions,” *Econometrica*, 2009, *77* (2), 403–426.
- Leland, Hayne**, “Corporate debt value, bond covenants, and optimal capital structure,” *Journal of Finance*, sep 1994, *49* (4), 1213–1252.
- and **Klaus Bjerre Toft**, “Optimal capital structure, endogenous bankruptcy, and the term structure of credit spreads,” *Journal of Finance*, jul 1996, *51* (3), 987–1019.
- Miao, J**, “Optimal capital structure and industry dynamics,” *Journal of Finance*, dec 2005, *60* (6), 2621–2659.
- Sambalaibat, Batchimeg**, “A theory of liquidity spillover between bond and cds markets,” *Working Paper*, 2016.
- Weill, Pierre-Olivier**, “Leaning against the wind,” *Review of Economic Studies*, 2007, *74* (4), 1329–1354.



# A Appendix

## A.1 Proofs Surplus Function (Lemma 1-4, Proposition 1)

**Proof of Lemma 1.** To resolve the trade selection issue, we make three observations regarding buyer-dealer and seller-dealer pairs, i.e. for investors that get to act this instant because they actually matched with a dealer:

Observation 1: Any investor of type  $i$  that buys an asset of type  $x$  via a dealer receives a surplus  $\beta S_{D \rightarrow i}(x)$  from this leg of the trade. For the moment, let us assume that all relevant assets are on offer so that maximizing over  $\mathbb{T}$  is the same as maximizing over  $\mathbb{X}$ . Taking the inter-dealer prices  $M(x)$  as given for the moment, we note that a buyer-dealer pair can maximize their individual surpluses by selecting varieties  $x$  such that  $x$  is in the set

$$\arg \max_{x' \in \mathbb{X}} S_{D \rightarrow i}(x') = \arg \max_{x' \in \mathbb{X}} \{V_{i1}(x') - V_{i0} - M(x')\}. \quad (\text{A.1})$$

Observation 2: Observation 1 then implies that surplus to buyers has to be constant across all traded assets. Suppose it were not. Then there exists a type  $x \in \mathbb{T}$  that has strictly higher surplus than type  $x' \in \mathbb{T}$ , i.e.  $S_{D \rightarrow i}(x) > S_{D \rightarrow i}(x')$ . But this cannot be, as then all buyers would try to buy type  $x$ , driving up its inter-dealer price  $M(x)$  to clear the market (note that the supply of the assets is fixed by contact intensities in the inter-dealer market), in the process lowering  $S_{D \rightarrow i}(x)$ . ■

Plugging in for  $k$ , we have

$$\begin{aligned} & (a) \\ &= \max_{\rho} \{ \rho \beta [V_{H1}(\check{x}) - M(\check{x}) - V_{H0}] - \gamma(\rho) \} \\ &= \max_{\rho} \{ \rho \beta [(V_{H1}(\check{x}) - M(\check{x})) - V_{H0} + (V_{L0} - V_{L1}(\hat{x}) + M(\hat{x}))] - \gamma(\rho) \} \\ &= \max_{\rho} \left\{ \rho \beta \left[ \left( V_{H1}(\hat{x}) - M(\hat{x}) + \frac{\gamma(\rho_{H1}(\hat{x}))}{\beta \rho_{H1}(\hat{x})} \right) - V_{H0} + (V_{L0} - V_{L1}(\hat{x}) + M(\hat{x})) \right] - \gamma(\rho) \right\} \\ &= \max_{\rho} \left\{ \rho \beta \left[ S(\hat{x}) + \frac{\gamma_1}{\beta} \right] - \gamma(\rho) \right\} \\ &= \max_{\rho} \{ \rho \beta \cdot k - \gamma(\rho) \} \end{aligned} \quad (\text{A.2})$$

as well as

$$\begin{aligned} & (b) \\ &= \max_{\rho} \{ \rho \beta [(V_{H1}(\check{x}) - M(\check{x})) + M(x) - V_{H1}(x)] - \gamma(\rho) \} \\ &= \max_{\rho} \{ \rho \beta [V_{H1}(\check{x}) - M(\check{x}) + V_{L1}(x) - V_{L1}(\hat{x}) + M(\hat{x}) - V_{H1}(x)] - \gamma(\rho) \} \\ &= \max_{\rho} \left\{ \rho \beta \left[ V_{H1}(\check{x}) + V_{L1}(x) - V_{L1}(\hat{x}) + \frac{V_{H1}(\hat{x}) - V_{H1}(\check{x}) + [V_{L1}(\hat{x}) - V_{L1}(\check{x})]}{2} - V_{H1}(x) \right] - \gamma(\rho) \right\} \\ &= \max_{\rho} \left\{ \rho \beta \left[ \frac{1}{2} [S(\hat{x}) + S(\check{x})] - S(x) \right] - \gamma(\rho) \right\} \\ &= \max_{\rho} \{ \rho \beta [k - S(x)] - \gamma(\rho) \} \end{aligned} \quad (\text{A.3})$$

and

$$\begin{aligned} (c) &= \max_{\rho} \{ \rho \beta [(V_{L1}(\hat{x}) - M(\hat{x})) + M(x) - V_{L1}(x)] - \gamma(\rho) \} \\ &= \max_{\rho} \{ \rho \beta [S(x) - k] - \gamma(\rho) \}. \end{aligned} \quad (\text{A.4})$$

**Proof of Lemma 2.** For the optimal search intensity with  $\gamma_2 > 0$ , we have

$$\gamma'(\rho_{H0}) = \beta \cdot k \quad (\text{A.5})$$

$$\gamma'(\rho_{H1}(x)) = \beta [k - S(x)] \quad (\text{A.6})$$

$$\gamma'(\rho_{L1}(x)) = \beta [S(x) - k] \quad (\text{A.7})$$

so that

$$\rho_{H0} = \frac{1}{\gamma_2} [\beta \cdot k - \gamma_1]^+ \quad (\text{A.8})$$

$$\rho_{H1}(x) = \frac{1}{\gamma_2} [\beta [k - S(x)] - \gamma_1]^+ \quad (\text{A.9})$$

$$\rho_{L1}(x) = \frac{1}{\gamma_2} [\beta [S(x) - k] - \gamma_1]^+ \quad (\text{A.10})$$

and we see that  $\rho_{H1}(x) \rho_{L1}(x) = 0$ , i.e., varieties  $L1$  types are trying to sell  $H1$  types hold on to, and vice versa.

Let us plug in the equilibrium search intensities to derive the surplus function. We have

$$\begin{aligned} (a) &= \rho_{H0} \beta k - \gamma(\rho_{H0}) \\ &= \rho_{H0} \left( \beta k - \gamma_1 - \frac{1}{2} \gamma_2 \rho_{H0} \right) \\ &= \frac{1}{\gamma_2} [\beta k - \gamma_1]^+ \left( \beta k - \gamma_1 - \frac{1}{2} \gamma_2 \frac{1}{\gamma_2} [\beta \cdot k - \gamma_1]^+ \right) \\ &= \frac{1}{2\gamma_2} \left( [\beta k - \gamma_1]^+ \right)^2 \end{aligned} \quad (\text{A.11})$$

as well as

$$\begin{aligned} (b) &= \rho_{H1}(x) \beta [k - S(x)] - \gamma(\rho_{H1}(x)) \\ &= \rho_{H1}(x) \left[ \beta [k - S(x)] - \gamma_1 - \frac{1}{2} \gamma_2 \rho_{H1}(x) \right] \\ &= \frac{1}{\gamma_2} [\beta [k - S(x)] - \gamma_1]^+ \left[ [\beta [k - S(x)] - \gamma_1]^+ - \frac{1}{2} \gamma_2 \frac{1}{\gamma_2} [\beta [k - S(x)] - \gamma_1]^+ \right] \\ &= \frac{1}{2\gamma_2} \left( [\beta [k - S(x)] - \gamma_1]^+ \right)^2 \end{aligned} \quad (\text{A.12})$$

and finally

$$(c) = \rho_{L1}(x) \beta [S(x) - k] - \gamma(\rho_{L1}(x)) \quad (\text{A.13})$$

$$= \rho_{L1}(x) \left[ \beta [S(x) - k] - \gamma_1 - \frac{1}{2} \gamma_2 \rho_{L1}(x) \right] \quad (\text{A.14})$$

$$= \frac{1}{2\gamma_2} \left( [\beta [S(x) - k] - \gamma_1]^+ \right)^2. \quad (\text{A.15})$$

Next, we derive the monotonicity of the surplus function w.r.t. an arbitrary  $k$ . To this end, let us subtract  $[\hat{r} + \delta(x)]k$  from both sides of the equation to get

$$\begin{aligned} [\hat{r} + \delta(x)] [S(x) - k] &= h(x) - [\hat{r} + \delta(x)]k - \frac{1}{2\gamma_2} \left( [\beta k - \gamma_1]^+ \right)^2 + \mathcal{L}[S(x) - k] \\ &\quad + \frac{1}{2\gamma_2} \left( [\beta [k - S(x)] - \gamma_1]^+ \right)^2 - \frac{1}{2\gamma_2} \left( [\beta [S(x) - k] - \gamma_1]^+ \right)^2. \end{aligned} \quad (\text{A.16})$$

Let us define  $\hat{S}(x, k)$  as the *surplus function for an arbitrary (possibly non-equilibrium)  $k$*  which induces trading rates  $\hat{\rho}_{H0}(k), \hat{\rho}_{H1}(x, k), \hat{\rho}_{L1}(x, k)$ . Differentiating w.r.t.  $k$ , we have

$$\begin{aligned}
[\hat{r} + \delta(x)] \frac{\partial}{\partial k} [\hat{S}(x, k) - k] &= -[\hat{r} + \delta(x)] + \mathcal{L} \frac{\partial}{\partial k} [\hat{S}(x, k) - k] \\
&\quad - \frac{1}{2\gamma_2} 2[\beta k - \gamma_1]^+ \beta \\
&\quad + \frac{1}{2\gamma_2} 2[-\beta [\hat{S}(x, k) - k] - \gamma_1]^+ (-\beta) \frac{\partial}{\partial k} [\hat{S}(x, k) - k] \\
&\quad - \frac{1}{2\gamma_2} 2[\beta [\hat{S}(x, k) - k] - \gamma_1]^+ \beta \frac{\partial}{\partial k} [\hat{S}(x, k) - k]. \tag{A.17}
\end{aligned}$$

Rewriting, we have

$$\begin{aligned}
\left[ \hat{r} + \delta(x) + \frac{\beta}{\gamma_2} \left\{ \begin{array}{l} [\beta [\hat{S}(x, k) - k] - \gamma_1]^+ \\ + [-\beta [\hat{S}(x, k) - k] - \gamma_1]^+ \end{array} \right\} \right] \frac{\partial}{\partial k} [\hat{S}(x, k) - k] &= - \left[ \hat{r} + \delta(x) + \frac{1}{2\gamma_2} 2[\beta k - \gamma_1]^+ \beta \right] \\
&\quad + \mathcal{L} \frac{\partial}{\partial k} [\hat{S}(x, k) - k]. \tag{A.18}
\end{aligned}$$

which we can summarize as

$$[\hat{r} + \delta(x) + (\text{positive function of } x)] \left[ \frac{\partial \hat{S}(x, k)}{\partial k} - 1 \right] = -(\text{positive function of } x) + \mathcal{L} \left[ \frac{\partial \hat{S}(x, k)}{\partial k} - 1 \right]. \tag{A.19}$$

Using Feynman-Kac, and imposing neutral boundary behavior, we immediately have

$$\left[ \frac{\partial \hat{S}(x, k)}{\partial k} - 1 \right] < 0. \tag{A.20}$$

■

**Proof of Lemma 3.** Equilibrium search intensities are then

$$\begin{aligned}
(a) = \rho_{H0} \beta k - \gamma(\rho_{H0}) &= \rho_{H0} (\beta k - \gamma_1) \\
&= \bar{\rho} [\beta k - \gamma_1]^+ \tag{A.21}
\end{aligned}$$

as well as

$$\begin{aligned}
(b) = \rho_{H1}(x) \beta [k - S(x)] - \gamma(\rho_{H1}(x)) &= \rho_{H1}(x) [\beta [k - S(x)] - \gamma_1] \\
&= \bar{\rho} [\beta [k - S(x)] - \gamma_1]^+ \tag{A.22}
\end{aligned}$$

and finally

$$\begin{aligned}
(c) = \rho_{L1}(x) \beta [S(x) - k] - \gamma(\rho_{L1}(x)) &= \rho_{L1}(x) [\beta [S(x) - k] - \gamma_1] \\
&= \bar{\rho} [\beta [S(x) - k] - \gamma_1]^+. \tag{A.23}
\end{aligned}$$

Plugging this in, we have

$$\begin{aligned}
[\hat{r} + \delta(x)] S(x) &= h(x) + \mathcal{L} S(x) + (b) - (c) - (a) \\
&= h(x) + \mathcal{L} S(x) \\
&\quad + \bar{\rho} [\beta [k - S(x)] - \gamma_1]^+ - \bar{\rho} [\beta [S(x) - k] - \gamma_1]^+ - \bar{\rho} [\beta k - \gamma_1]^+. \tag{A.24}
\end{aligned}$$

Next, we derive the monotonicity of the surplus function w.r.t. an arbitrary  $k$ . Again, subtracting

$[\hat{r} + \delta(x)]k$  from both sides, we have

$$\begin{aligned} [\hat{r} + \delta(x)][S(x) - k] &= h(x) - [\hat{r} + \delta(x)]k - \bar{\rho}[\beta k - \gamma_1]^+ + \mathcal{L}[S(x) - k] \\ &\quad + \bar{\rho}[\beta[k - S(x)] - \gamma_1]^+ - \bar{\rho}[\beta[S(x) - k] - \gamma_1]^+. \end{aligned} \quad (\text{A.25})$$

Again, let us define  $\hat{S}(x, k)$  as the *surplus function for an arbitrary (possibly non-equilibrium)  $k$*  which induces trading rates  $\hat{\rho}_{H0}(k)$ ,  $\hat{\rho}_{H1}(x, k)$ ,  $\hat{\rho}_{L1}(x, k)$ . Then differentiating w.r.t.  $k$ , we see that

$$\begin{aligned} [\hat{r} + \delta(x)] \frac{\partial}{\partial k} [\hat{S}(x, k) - k] &= -[\hat{r} + \delta(x)] - \bar{\rho} \mathbf{1}_{\{\beta k > \gamma_1\}} \beta + \mathcal{L} \frac{\partial}{\partial k} [\hat{S}(x, k) - k] \\ &\quad + \bar{\rho} \mathbf{1}_{\{\beta[k - \hat{S}(x, k)] > \gamma_1\}} (-\beta) \frac{\partial}{\partial k} [\hat{S}(x, k) - k] \end{aligned} \quad (\text{A.26})$$

$$- \bar{\rho} \mathbf{1}_{\{\beta[\hat{S}(x, k) - k] > \gamma_1\}} \beta \frac{\partial}{\partial k} [\hat{S}(x, k) - k] \quad (\text{A.27})$$

which we can write as

$$\left[ \hat{r} + \delta(x) + \bar{\rho} \beta \left\{ \begin{array}{l} \mathbf{1}_{\{\beta[k - \hat{S}(x, k)] > \gamma_1\}} \\ + \mathbf{1}_{\{\beta[\hat{S}(x, k) - k] > \gamma_1\}} \end{array} \right\} \right] \frac{\partial}{\partial k} [\hat{S}(x, k) - k] = -[\hat{r} + \delta(x) + \bar{\rho} \mathbf{1}_{\{\beta k > \gamma_1\}} \beta] + \mathcal{L} \frac{\partial}{\partial k} [\hat{S}(x, k) - k] \quad (\text{A.28})$$

so that again

$$[\hat{r} + \delta(x) + (\text{positive function of } x)] \left[ \frac{\partial \hat{S}(x, k)}{\partial k} - 1 \right] = -(\text{positive function of } x) + \mathcal{L} \left[ \frac{\partial \hat{S}(x, k)}{\partial k} - 1 \right] \quad (\text{A.29})$$

so that

$$\left[ \frac{\partial \hat{S}(x, k)}{\partial k} - 1 \right] < 0. \quad (\text{A.30})$$

■

**Proof of Lemma 4.** Pick a particle in state  $(x, H)$ . Then, for a given  $k$ , it has a chance of being transported to  $L$  of  $\xi_{LH} + \hat{\rho}_{H1}(x, k)$  which is increasing in  $k$ . Furthermore, for a given  $k$ , a particle in state  $(x, L)$  has a chance of being transported to  $H$  of  $\xi_{HL} + \hat{\rho}_{L1}(x, k)$ , which is decreasing in  $k$ . Then, by  $\mu_1(x) = \hat{\mu}_{H1}(x, k) + \hat{\mu}_{L1}(x, k)$ , it must be that as we increase  $k$ ,  $\frac{d\hat{\mu}_{H1}(x, k)}{dk} < 0 < \frac{d\hat{\mu}_{L1}(x, k)}{dk}$ . Next, note that

$$\hat{\mu}_{H0}(k) = \mu_H - \int_{\mathbb{X}} \hat{\mu}_{H1}(x, k) dx, \quad (\text{A.31})$$

so that

$$\hat{\mu}'_{H0}(k) = - \int_{\mathbb{X}} \frac{d\hat{\mu}_{H1}(x, k)}{dk} dx > 0. \quad (\text{A.32})$$

Here, we implicitly assume that default  $\delta(x)$  is at most type neutral or tilted towards H types. ■

**Proof of Proposition 1.** Write the market-clearing equation in imbalance form, that is

$$\begin{aligned} ib(k) &\equiv (\xi_{LH} + \xi_{HL}) \int_{\mathbb{X}} \hat{\mu}_{H1}(x, k) dx + \int_{\mathbb{X}} \delta(x) [\hat{\mu}_{H1}(x, k) - \mu_1(x)] dx - \xi_{LH} - \hat{\rho}_{H0}(k) \hat{\mu}_{H0}(k) \\ &\quad - bc(k) \end{aligned} \quad (\text{A.33})$$

so that  $ib(k^*) = 0$  defines an equilibrium  $k$ . Taking derivatives w.r.t.  $k$ , we have

$$\begin{aligned} ib'(k) &= (\xi_{LH} + \xi_{HL}) \int_{\mathbb{X}} \partial_k \hat{\mu}_{H1}(x, k) dx + \int_{\mathbb{X}} \delta(x) \partial_k \hat{\mu}_{H1}(x, k) dx \\ &\quad - \hat{\rho}'_{H0}(k) \hat{\mu}_{H0}(k) - \hat{\rho}_{H0}(k) \hat{\mu}'_{H0}(k) - bc(k) < 0. \end{aligned} \quad (\text{A.34})$$

Note that by Lemma 4, we have both integrals being negative as well as  $\hat{\mu}'_{H0}(k) > 0$ . By Lemma 2 and 3,

we have  $\hat{\rho}'_{H0}(k) > 0$ , and by assumption  $bc'(k) \geq 0$ , so that we have  $ib(k)$  decreasing in  $k$ . This establishes a unique  $k^*$  that characterizes the equilibrium. ■

## A.2 Relation to free entry

For example, take the linear-quadratic cost specification. Then, we have

$$\rho_{H0} = \frac{1}{\gamma_2} (\beta k - \gamma_1)^+. \quad (\text{A.35})$$

Then, plugging this in, free-entry requires

$$\begin{aligned} 0 &= \rho_{H0}\beta k - \left( \gamma_1 \rho_{H0} + \frac{1}{2} \gamma_2 \rho_{H0}^2 \right) \\ &= \frac{1}{\gamma_2} (\beta k - \gamma_1) \beta k - \gamma_1 \frac{1}{\gamma_2} (\beta k - \gamma_1) - \frac{1}{2} \gamma_2 \left[ \frac{1}{\gamma_2} (\beta k - \gamma_1) \right]^2 \\ &= \frac{1}{\gamma_2} \left[ (\beta k)^2 - 2\gamma_1 \beta k + \gamma_1^2 \right] - \frac{1}{2} \frac{1}{\gamma_2} (\beta k - \gamma_1)^2 \\ &= \frac{1}{2} \frac{1}{\gamma_2} (\beta k - \gamma_1)^2. \end{aligned} \quad (\text{A.36})$$

## A.3 Leland implementation

**Limit values.** Under the assumption  $S'(x) \leq 0$  we have the fixed point relation  $\underline{k} = \min_x \hat{S}(x, \underline{k}) = \lim_{x \rightarrow \infty} \hat{S}(x, \underline{k}) = \hat{S}(\underline{k})$ . Then, ignoring  $S'(x)$  and  $S''(x)$  in the ODE, we have

$$(\hat{r} + \delta - h_1) \underline{S} = h_0 - \frac{1}{2\gamma_2} (\beta k)^2 + \frac{1}{2\gamma_2} (\beta [\underline{S} - k])^2 \{ \mathbf{1}_{\{\underline{S} \leq k\}} - \mathbf{1}_{\{\underline{S} \geq k\}} \}. \quad (\text{A.37})$$

First, suppose that  $k = \underline{S}$ , so that we are solving for  $\underline{k}$ . Then we have

$$(\hat{r} + \delta - h_1) k = h_0 - \frac{1}{2\gamma_2} (\beta k)^2 \iff \frac{1}{2} \frac{\beta^2}{\gamma_2} k^2 + (\hat{r} + \delta - h_1) k - h_0 = 0 \quad (\text{A.38})$$

which has one positive and one negative root. We take the positive root

$$\underline{S}(\underline{k}) = \underline{k} \equiv \frac{-(\hat{r} + \delta - h_1) + \sqrt{(\hat{r} + \delta - h_1)^2 + 2 \frac{\beta^2}{\gamma_2} h_0}}{\frac{\beta^2}{\gamma_2}}. \quad (\text{A.39})$$

Next, suppose that  $\underline{S} > k$  so that we are considering  $k \in [0, \underline{k})$ . Then, we have

$$\begin{aligned} (\hat{r} + \delta - h_1) \underline{S} &= h_0 - \frac{1}{2\gamma_2} (\beta k)^2 - \frac{1}{2\gamma_2} \beta^2 (\underline{S}^2 - 2k\underline{S} + k^2) \\ &= h_0 - \frac{1}{2\gamma_2} \beta^2 (\underline{S}^2 - 2k\underline{S} + 2k^2) \end{aligned} \quad (\text{A.40})$$

so that

$$\begin{aligned} 0 &= \frac{1}{2} \frac{\beta^2}{\gamma_2} (\underline{S}^2 - 2k\underline{S} + 2k^2) + (\hat{r} + \delta - h_1) \underline{S} - h_0 \\ &= \frac{1}{2} \frac{\beta^2}{\gamma_2} \underline{S}^2 + \left[ (\hat{r} + \delta - h_1) - \frac{\beta^2}{\gamma_2} k \right] \underline{S} - \left( h_0 - \frac{\beta^2}{\gamma_2} k^2 \right). \end{aligned} \quad (\text{A.41})$$

We select the higher root, as it converges to  $\underline{k}$ :

$$\hat{\underline{S}}(k) = \frac{-\left[(\hat{r} + \delta - h_1) - \frac{\beta^2}{\gamma_2}k\right] + \sqrt{\left[(\hat{r} + \delta - h_1) - \frac{\beta^2}{\gamma_2}k\right]^2 + 2\frac{\beta^2}{\gamma_2}\left(h_0 - \frac{\beta^2}{\gamma_2}k^2\right)}}{\frac{\beta^2}{\gamma_2}}. \quad (\text{A.42})$$

Lastly, suppose that  $\underline{S} < k$ . Then, we have

$$\begin{aligned} (\hat{r} + \delta - h_1)\underline{S} &= h_0 - \frac{1}{2\gamma_2}(\beta k)^2 + \frac{1}{2\gamma_2}\beta^2(\underline{S}^2 - 2k\underline{S} + k^2) \\ &= h_0 + \frac{1}{2\gamma_2}\beta^2(\underline{S}^2 - 2k\underline{S}) \end{aligned} \quad (\text{A.43})$$

so that

$$\begin{aligned} 0 &= \frac{1}{2}\frac{\beta^2}{\gamma_2}(\underline{S}^2 - 2k\underline{S}) - (\hat{r} + \delta - h_1)\underline{S} + h_0 \\ &= \frac{1}{2}\frac{\beta^2}{\gamma_2}\underline{S}^2 - \left[(\hat{r} + \delta - h_1) + \frac{\beta^2}{\gamma_2}k\right]\underline{S} + h_0. \end{aligned} \quad (\text{A.44})$$

This quadratic equation has two positive roots (as  $\frac{c}{a} > 0$  and  $-\frac{b}{a} > 0$ ). We pick the smaller root, as it converges to  $\underline{k}$ :

$$\hat{\underline{S}}(k) = \frac{\left[(\hat{r} + \delta - h_1) + \frac{\beta^2}{\gamma_2}k\right] - \sqrt{\left[(\hat{r} + \delta - h_1) + \frac{\beta^2}{\gamma_2}k\right]^2 - 2\frac{\beta^2}{\gamma_2}h_0}}{\frac{\beta^2}{\gamma_2}}. \quad (\text{A.45})$$

This is only defined for

$$\left[(\hat{r} + \delta - h_1) + \frac{\beta^2}{\gamma_2}k\right]^2 - 2\frac{\beta^2}{\gamma_2}h_0 > 0. \quad (\text{A.46})$$

For  $k \in (\underline{k}, \bar{S})$ , the LHS is minimized at  $k = \underline{k}$ . Plugging in, and simplifying, we have

$$\left[(\hat{r} + \delta - h_1) + \frac{\beta^2}{\gamma_2}\underline{k}\right]^2 - 2\frac{\beta^2}{\gamma_2}h_0 = (\hat{r} + \delta - h_1) > 0 \quad (\text{A.47})$$

and thus the root is defined everywhere.

Lastly, we want  $\bar{S} \geq \max_{k \in [0, \bar{S}]} \hat{\underline{S}}(k)$ . It is easy to show that  $\hat{\underline{S}}'(k) < 0$  for  $k \in (\underline{k}, \infty)$ . We thus find the maximum for  $k \in (0, \underline{k})$ .<sup>24</sup> It occurs at

$$k_{max} = \frac{-(\hat{r} + \delta - h_1) + \sqrt{(\hat{r} + \delta - h_1)^2 + \frac{\beta^2}{\gamma_2}h_0}}{\frac{\beta^2}{\gamma_2}}. \quad (\text{A.48})$$

We assume that

$$\bar{S} > \max_{k \in [0, \bar{S}]} \hat{\underline{S}}(k) = \hat{\underline{S}}(k_{max}) = 2k_{max}. \quad (\text{A.49})$$

After tedious algebra, one can show that

$$\frac{\partial}{\partial k} \left[ \hat{\underline{S}}(k) - k \right] < 0 \quad (\text{A.50})$$

---

<sup>24</sup>Even though  $\hat{\underline{S}}(k)$  is not monotone on  $(0, \underline{k})$ ,  $\left[\hat{\underline{S}}(k) - k\right]$  is, and our uniqueness proofs go through with slight modifications.

so that the boundary condition still satisfies the neutral requirement, and the proofs of Lemma 2 and 3 go through.

**Steady-state distribution.** Let us assume for tractability that  $b(x)$  follows an exponential distribution, i.e.,

$$b(x) = \lambda e^{-\lambda(x-x_{min})} \quad (\text{A.51})$$

so that the particular part is given by  $\left[\frac{\sigma^2}{2}\mu'_1(x_{min}) + \delta \times cst\right] e^{-\lambda(x-x_{min})}$ . Then, we have the following solution:

$$\mu_1(x) = C_1 e^{\eta_1(x-x_{min})} + C_2 e^{\eta_2(x-x_{min})} + C_\lambda \lambda e^{-\lambda(x-x_{min})} \quad (\text{A.52})$$

where

$$\eta_{1/2} = \frac{m \pm \sqrt{m + 2\delta\sigma^2}}{\sigma^2} \quad (\text{A.53})$$

$$C_\lambda = \frac{\frac{\sigma^2}{2}\mu'_1(x_{min}) + \delta \times cst}{\delta - \frac{\sigma^2}{2}\lambda^2 - m\lambda} \lambda \quad (\text{A.54})$$

with  $\eta_1 < 0 < \eta_2$ . Then, we see that  $C_2 = 0$  as otherwise the integral does not converge. Thus, we write

$$\mu_1(x) = C_1 e^{\eta_1(x-x_{min})} + C_\lambda e^{-\lambda(x-x_{min})}. \quad (\text{A.55})$$

Let us treat  $C_\lambda$  as a unknown, and impose two boundary conditions:

$$\mu_1(x_{min}) = C_1 + C_\lambda = 0 \quad (\text{A.56})$$

$$\int_{\mathbb{X}} \mu_1(x) dx = -\frac{C_1}{\eta_1} + \frac{C_\lambda}{\lambda} = 1. \quad (\text{A.57})$$

We thus ignore the possible third boundary condition linking  $C_\lambda$  to  $\mu'_1(x_{min})$ . We will see that this boundary condition is redundant. This is solved by

$$C_1 = -\frac{\eta_1 \lambda}{\lambda + \eta_1} \quad (\text{A.58})$$

$$C_\lambda = \frac{\eta_1 \lambda}{\lambda + \eta_1}. \quad (\text{A.59})$$

**Market clearing convergence.** Convergence of  $k$  to market clearing  $k^*$  is well behaved and can be inspected in Figure 5.

## A.4 Trading prices

This model features bilateral bargaining and thus has “private” prices. We need to make an assumption that trades are restricted to 2 trades per dealer contact. Let  $\gamma_1 = 0$  so that  $\hat{x} = \check{x} = \bar{x}$ .

A dealer buys at  $B(x)$  and sells at  $A(x)$ . We will maintain this interpretation for the moment, but it will become clear that for exchanges only looking at one leg of the transaction can lead to non-sensical results.

First, note that

$$M(\bar{x}) = V_{L1}(\bar{x}) - V_{L0} \quad (\text{A.60})$$

from the indifference condition.

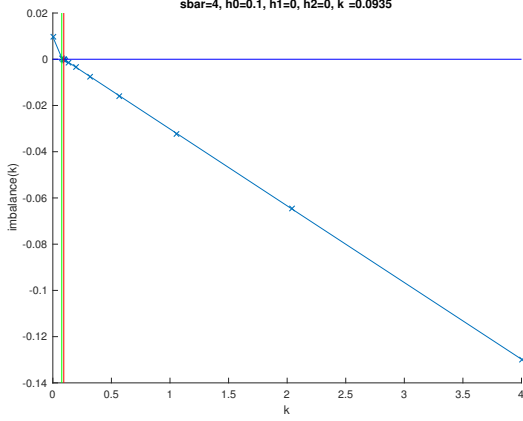


Figure 5: **Left:** Convergence for  $h_2 = 0$ ; **Right:** Convergence for  $h_2 = 0.7$ .

**No-exchange.** Second, consider  $x \in \{x : S(x) \geq k\}$ . On this set, the dealer buys from L1 types and sells to H0 types, appropriating surpluses of

$$\text{Buy: } (1 - \beta) [V_{L0} - V_{L1}(x) + M(x)] = M(x) - B(x) \quad (\text{A.61})$$

$$\text{Sell: } (1 - \beta) [V_{H1}(x) - V_{H0} - M(x)] = A(x) - M(x). \quad (\text{A.62})$$

Plugging in for  $M(x)$ , we have

$$\begin{aligned} B(x) &= \beta [V_{H1}(x) - V_{H1}(\bar{x}) + M(\bar{x})] + (1 - \beta) [V_{L1}(x) - V_{L0}] \\ &= \beta [V_{H1}(x) - V_{H1}(\bar{x}) + V_{L1}(\bar{x}) - V_{L0}] + (1 - \beta) [V_{L1}(x) - V_{L0}] \\ &= \beta [V_{H1}(x) - (V_{H1}(\bar{x}) - V_{L1}(\bar{x}) + V_{L0} - V_{H0}) - V_{H0}] + (1 - \beta) [V_{L1}(x) - V_{L0}] \\ &= \beta [V_{H1}(x) - S(\bar{x}) - V_{H0}] + (1 - \beta) [V_{L1}(x) - V_{L0}] \\ &= \beta [V_{H1}(x) - k - V_{H0}] + (1 - \beta) [V_{L1}(x) - V_{L0}] \\ &= \beta [V_{H1}(x) - k - V_{H0} - (V_{L1}(x) - V_{L0})] + [V_{L1}(x) - V_{L0}] \\ &= [V_{L1}(x) - V_{L0}] + \beta [S(x) - k] \end{aligned} \quad (\text{A.63})$$

$$\begin{aligned} A(x) &= \beta [V_{H1}(x) - V_{H1}(\bar{x}) + M(\bar{x})] + (1 - \beta) [V_{H1}(x) - V_{H0}] \\ &= \beta [V_{H1}(x) - V_{H1}(\bar{x}) + V_{L1}(\bar{x}) - V_{L0}] + (1 - \beta) [V_{H1}(x) - V_{H0}] \\ &= \beta [V_{H1}(x) - (V_{H1}(\bar{x}) - V_{L1}(\bar{x}) + V_{L0} - V_{H0}) - V_{H0}] + (1 - \beta) [V_{H1}(x) - V_{H0}] \\ &= \beta [V_{H1}(x) - k - V_{H0}] + (1 - \beta) [V_{H1}(x) - V_{H0}] \\ &= [V_{H1}(x) - V_{H0}] - \beta k \end{aligned} \quad (\text{A.64})$$

for a bid-ask spread of

$$A(x) - B(x) = (1 - \beta) [V_{H1}(x) - V_{H0} - V_{L1}(x) + V_{L0}] = (1 - \beta) S(x). \quad (\text{A.65})$$

**Exchange.** Third, consider  $x \in \{x : S(x) \leq k\}$ . On this set, the dealer buys from H1 types and sells to L0 types. However, these transactions are just one leg of a two legged trade, something that is important to bear in mind when interpreting the results. Here, we have the dealer buying  $x$  and selling  $x' \in \{x : S(x) \geq k\}$ . Wlog assume  $x' = \bar{x}$ . Then, the dealer buys  $x$  from an H1 type, and then sells the resulting H0 type  $x$ . Suppose that the dealer still offers ask price  $A(\bar{x})$  in this deal. Then, the dealer appropriates surplus

$$\text{H0 exchange: } (1 - \beta) [V_{H1}(\bar{x}) - M(\bar{x}) + M(x) - V_{H1}(x)] = M(x) - B(x) + A(\bar{x}) - M(\bar{x}) \quad (\text{A.66})$$



Plugging in for  $M(x)$ , we have

$$\begin{aligned}
B(x) &= \beta [V_{L1}(x) - V_{L1}(\bar{x}) + M(\bar{x})] + (1 - \beta) [V_{H1}(x) - V_{H0}] \\
&= \beta [V_{L1}(x) - V_{L1}(\bar{x}) + V_{L1}(\bar{x}) - V_{L0}] + (1 - \beta) [V_{H1}(x) - V_{H0}] \\
&= \beta [V_{L1}(x) - V_{L0}] + (1 - \beta) [V_{H1}(x) - V_{H0}] \\
&= [V_{H1}(x) - V_{H0}] - \beta S(x).
\end{aligned} \tag{A.67}$$

Lastly, note that the L1 type does not appropriate any surplus in the buying leg of the transaction, so that

$$\begin{aligned}
A(x) &= M(x) \\
&= V_{L1}(x) - V_{L0} \\
&= \beta [V_{L1}(x) - V_{L1}(\bar{x}) + M(\bar{x})] + (1 - \beta) [V_{L1}(x) - V_{L0}] \\
&= \beta [V_{L1}(x) - V_{L1}(\bar{x}) + V_{L1}(\bar{x}) - V_{L0}] + (1 - \beta) [V_{L1}(x) - V_{L0}] \\
&= \beta [V_{L1}(x) - V_{L0}] + (1 - \beta) [V_{L1}(x) - V_{L0}] \\
&= [V_{L1}(x) - V_{L0}].
\end{aligned} \tag{A.68}$$

for a bid-ask spread of

$$A(x) - B(x) = (1 - \beta) [V_{L1}(x) - V_{L0} - V_{H1}(x) + V_{H0}] = -(1 - \beta) S(x) < 0.$$

Thus, if we were to impose the prices we see agents trading on  $\{x : S(x) \geq k\}$  as captured by  $A(\bar{x})$  and  $B(\bar{x})$ , we get the result that agents are trading at a negative bid-ask spread. This, however, does not violate the incentive condition of the agents involved — they are negotiating a private deal that is composed of two legs, and neither leg on its own has a pinned down price. Thus, when projecting on leg onto the pricing derived above, we see that the other leg has strange properties. But this is to be expected — agents engage in a *seemingly unprofitable* leg of a trade to get access to the other *profitable* leg of the trade, and thus only the difference between both surpluses matters: Suppose  $x$  is exchanged for  $x'$  (thus  $S(x') - S(x) \geq 0$ ) by an H0 agent, then the total intermediation surplus appropriated by the dealers is given by

$$[A(x') - B(x')] + [A(x) - B(x)] = (1 - \beta) [S(x') - S(x)] \geq 0.$$

**Primary market.** We assume here that there exists a competitive friction-less primary market for the asset. The price is given by  $P(x)$ . This is the price that unconstrained H0 types are willing to pay. As we always have a positive mass of H0 types waiting on the side-line, this does not invalidate any of our results above. Any H0 types waiting on the side-line who are not currently in contact with a dealer are willing to pay

$$P(x) = V_{H1}(x) - V_{H0} \tag{A.69}$$

for the asset.

**Valuations as functions of  $k^*$  and  $S(x)$  when  $\delta(x) = 0$ .** Note that the valuations can be written as

$$rV_{H0} = \xi_{HL} (V_{L0} - V_{H0}) + \rho_{H0} \beta \cdot k^* - \gamma(\rho_{H0}) \tag{A.70}$$

$$rV_{L0} = \xi_{LH} (V_{H0} - V_{L0}) \tag{A.71}$$

$$\begin{aligned}
rV_{H1}(x) &= c(x) + \xi_{HL} [V_{L1}(x) - V_{H1}(x)] + \mathcal{L}V_{H1}(x) \\
&\quad + \{\rho_{H1}(x) [k^* - S(x)] - \gamma(\rho_{H1}(x))\}
\end{aligned} \tag{A.72}$$

$$\begin{aligned}
rV_{L1}(x) &= c(x) - h(x) + \xi_{LH} [V_{H1}(x) - V_{L1}(x)] + \mathcal{L}V_{L1}(x) \\
&\quad + \{\rho_{L1}(x) [S(x) - k^*] - \gamma(\rho_{L1}(x))\}.
\end{aligned} \tag{A.73}$$

Suppose there is no delay at default  $x = x_{min}$ . Then, we would have  $V_{H1}(x_{min}) = \alpha L(x_{min}) + V_{H0}$  and  $V_{L1}(x_{min}) = \alpha L(x_{min}) + V_{L0}$  so that  $S(x_{min}) = 0$ . Thus, we need some delay at default to introduce some post default surplus. To keep this out of the model, we simply assume that  $S(x_{min}) = \bar{S}$  sufficiently larger than  $h_0$  by introducing  $\alpha_H$  and  $\alpha_L$  so that  $V_{H1}(x_{min}) = \alpha_H L(x_{min}) + V_{H0}$  and  $V_{L1}(x_{min}) = \alpha_L L(x_{min}) + V_{L0}$  and thus  $S(x_{min}) = (\alpha_H - \alpha_L) L(x_{min})$ .

**Valuations for  $\gamma_2 > 0$ .** Let us assume  $\gamma_2 > 0$ . Then, plugging in for  $\rho_{ia}(x)$ , we have

$$rV_{H0} = \xi_{HL}(V_{L0} - V_{H0}) + \frac{1}{2\gamma_2} \left( [\beta k - \gamma_1]^+ \right)^2 \quad (\text{A.74})$$

$$rV_{L0} = \xi_{LH}(V_{H0} - V_{L0}) \quad (\text{A.75})$$

so that

$$V_{H0} - V_{L0} = \frac{1}{2\gamma_2} \frac{\left( [\beta k - \gamma_1]^+ \right)^2}{r + \xi_{HL} + \xi_{LH}}$$

and thus

$$V_{H0} = \left( \frac{r + \xi_{LH}}{r} \right) \frac{1}{2\gamma_2} \frac{\left( [\beta k - \gamma_1]^+ \right)^2}{r + \xi_{HL} + \xi_{LH}} \quad (\text{A.76})$$

$$V_{L0} = \frac{\xi_{LH}}{r} \frac{1}{2\gamma_2} \frac{\left( [\beta k - \gamma_1]^+ \right)^2}{r + \xi_{HL} + \xi_{LH}}. \quad (\text{A.77})$$

Then, we can write the valuations  $V_{H1}(x)$  and  $V_{L1}(x)$  as ODEs with exogenous functions  $S(x)$  and  $k$ :

$$\begin{aligned} rV_{H1}(x) &= c(x) + \xi_{HL}[V_{L1}(x) - V_{H1}(x)] + \mathcal{L}V_{H1}(x) \\ &\quad + \frac{1}{2\gamma_2} \left( [\beta[k - S(x)] - \gamma_1]^+ \right)^2 \\ &= c(x) - \xi_{HL}[S(x) + V_{H0} - V_{L0}] + \mathcal{L}V_{H1}(x) \\ &\quad + \frac{1}{2\gamma_2} \left( [\beta[k - S(x)] - \gamma_1]^+ \right)^2 \\ &= c(x) - \xi_{HL} \left[ S(x) + \frac{1}{2\gamma_2} \frac{\left( [\beta k - \gamma_1]^+ \right)^2}{r + \xi_{HL} + \xi_{LH}} \right] + \mathcal{L}V_{H1}(x) \\ &\quad + \frac{1}{2\gamma_2} \left( [\beta[k - S(x)] - \gamma_1]^+ \right)^2 \end{aligned} \quad (\text{A.78})$$

and

$$\begin{aligned}
rV_{L1}(x) &= c(x) - h(x) + \xi_{LH} [V_{H1}(x) - V_{L1}(x)] + \mathcal{L}V_{L1}(x) \\
&\quad + \frac{1}{2\gamma_2} \left( [\beta[S(x) - k] - \gamma_1]^+ \right)^2 \\
&= c(x) - h(x) + \xi_{LH} [S(x) + V_{H0} - V_{L0}] + \mathcal{L}V_{L1}(x) \\
&\quad + \frac{1}{2\gamma_2} \left( [\beta[S(x) - k] - \gamma_1]^+ \right)^2 \\
&= c(x) - h(x) + \xi_{LH} \left[ S(x) + \frac{1}{2\gamma_2} \frac{\left( [\beta k - \gamma_1]^+ \right)^2}{r + \xi_{HL} + \xi_{LH}} \right] + \mathcal{L}V_{L1}(x) \\
&\quad + \frac{1}{2\gamma_2} \left( [\beta[S(x) - k] - \gamma_1]^+ \right)^2. \tag{A.79}
\end{aligned}$$

## A.5 Solutions to $S(x)$ with $\gamma_1 = 0$ and relation to the Matrix Riccati equation

For this appendix, let us assume  $\gamma_1 = 0$  and a monotone surplus function, wlog,  $S'(x) < 0$ . Then, the surplus equation becomes

$$\begin{aligned}
[\hat{r} + \delta(x) - h_1] S(x) &= h_0(x) - \frac{1}{2\gamma_2} (\beta k)^2 + (m - h_2) S'(x) + \frac{\sigma^2}{2} S''(x) \\
&\quad + \frac{1}{2\gamma_2} (\beta [S(x) - k])^2 \{ \mathbb{1}_{\{S(x) \leq k\}} - \mathbb{1}_{\{S(x) \geq k\}} \}. \tag{A.80}
\end{aligned}$$

First, let us assume that  $k \in [0, \underline{k}]$ . Then, we know that  $S(x) \geq k, \forall x$ , and the surplus ODE becomes

$$\begin{aligned}
[\hat{r} + \delta(x) - h_1] S(x) &= h_0(x) - \frac{1}{2\gamma_2} (\beta k)^2 + (m - h_2) S'(x) + \frac{\sigma^2}{2} S''(x) \\
&\quad - \frac{1}{2\gamma_2} (\beta [S(x) - k])^2. \tag{A.81}
\end{aligned}$$

Next, let us define  $y_1(x) \equiv S(x)$  and  $y_2(x) \equiv S'(x)$ . Then, we have

$$\begin{aligned}
y_1'(x) &= y_2(x) \\
[\hat{r} + \delta(x) - h_1] y_1(x) &= h_0(x) - \frac{1}{2\gamma_2} (\beta k)^2 + (m - h_2) y_2(x) + \frac{\sigma^2}{2} y_2'(x) \\
&\quad - \frac{1}{2\gamma_2} \beta^2 \left( y_1(x)^2 - 2k y_1(x) + k^2 \right). \tag{A.82}
\end{aligned}$$

Rewriting in Matrix form, we have

$$\begin{aligned}
\begin{bmatrix} y_1'(x) \\ y_2'(x) \end{bmatrix} &= \begin{bmatrix} 0 \\ -\frac{2}{\sigma^2} \left( h_0(x) - \frac{1}{\gamma_2} (\beta k)^2 \right) \end{bmatrix} + \begin{bmatrix} 0 & 1 \\ \frac{2}{\sigma} \left( \hat{r} + \delta(x) - h_1 - \frac{\beta^2 k}{\gamma_2} \right) & -\frac{2}{\sigma} (m - h_2) \end{bmatrix} \times \begin{bmatrix} y_1(x) \\ y_2(x) \end{bmatrix} \\
&\quad + \begin{bmatrix} y_1(x) \\ y_2(x) \end{bmatrix}^\top \times \begin{bmatrix} \frac{1}{2\gamma_2} \beta^2 & 0 \\ 0 & 0 \end{bmatrix} \times \begin{bmatrix} y_1(x) \\ y_2(x) \end{bmatrix} \tag{A.83}
\end{aligned}$$

which we can write as

$$\mathbf{y}'(x) = \mathbf{a}(x) + \mathbf{B}(x) \times \mathbf{y}(x) + \mathbf{y}(x)^\top \times \mathbf{C} \times \mathbf{y}(x). \tag{A.84}$$

This is clearly a Matrix Riccati Equation. We have boundary conditions

$$\lim_{x \rightarrow \infty} \mathbf{y}(x) = \underline{\mathbf{y}} = \begin{bmatrix} \hat{S}(k^*) \\ 0 \end{bmatrix}. \quad (\text{A.85})$$

Even as this Riccati Equation has no closed form solution, strong existence results apply.