The Strategic Determination of the Supply of Liquid Assets

Athanasios Geromichalos
University of California – Davis

Lucas Herrenbrueck
Simon Fraser University

Sukjoon Lee
New York University Shanghai

Revised version: June 2022, Published Version (2023): DOI 10.1016/j.red.2022.08.003

ABSTRACT ————————————————————————————————
We study asset liquidity in a model where financial assets can be liquidated for money in over-the-counter (OTC) secondary markets, in response to random liquidity needs. Traders choose to enter the market where they expect to find the best terms, understanding that their chances to trade depend on the entry decision of other investors. We find that small differences in OTC microstructure can induce very large differences in the relative liquidity of two assets. We use our model to rationalize, qualitatively and quantitatively, the superior liquidity of U.S. Treasuries over equally safe corporate debt.
——-

JEL Classification: E31, E43, E52, G12

Keywords: monetary-search models, OTC markets, endogenous liquidity, endogenous asset supply

Contact: ageromich@ucdavis.edu, herrenbrueck@sfu.ca, sukjoon.lee@nyu.edu

We would like to thank Aleksander Berentsen, Darrell Duffie, Tai-Wei Hu, Robert Jones, Oscar Jorda, Ricardo Lagos, Ed Nosal, Guillaume Rocheteau, Ina Simonovska, Alan Taylor, Dimitri Vayanos, Venky Venkateswaran, Pierre-Olivier Weill, and Randall Wright for their useful comments and suggestions, as well as participants at the 2015 St Louis FED workshop on Money, Banking, Payments, and Finance, the WEAI 90th Annual Conference, the 12th Annual Macroeconomic Workshop in Vienna, Austria, the Spring 2016 Midwest Macro Meetings, the Fall 2017 Midwest Macro Meetings, the 11th NYU Search Theory Workshop, the 2018 North American Summer Meeting of the Econometric Society, the 2022 SED Annual Meeting, and at seminars at the University of Wisconsin, Madison, the University of Saskatchewan, and the University of British Columbia. We also thank Zijian Wang for his invaluable help.

1 Introduction

Why do U.S. Treasuries sell at higher prices than corporate or municipal bonds with similar characteristics, even after controlling for safety?[1] A popular answer is "due to their *liquidity*". More precisely, the Treasury sells its bonds at a *premium* because investors expect to be able to (re)sell these bonds easily in the secondary market and are, thus, willing to pay higher prices in the primary market.[2] While this is a plausible explanation, some important questions remain. Why are the secondary markets for other types of bonds less liquid than the one for Treasuries? Is it hard(er) for sellers to find buyers due to some hardwired market friction (e.g., a poorly organized interdealer network)? Or, are there not enough buyers drawn to those markets to whom I could sell my bonds – and if so, why? Or, perhaps finding trading partners is not so hard, but there are not enough bonds to go around in the market? Finally, how do these candidate explanations (and their interaction) affect asset prices and liquidity in general equilibrium?

To answer these questions, we develop a model where liquidity depends not only on the (exogenous) characteristics of the market an asset trades in, but also on the (endogenous) decision of agents to visit that market. Our model has two main ingredients. The first is an empirically relevant concept of asset liquidity: agents can liquidate assets for money in Over-the-Counter (OTC) secondary markets which, as in Duffie, Gârleanu, and Pedersen (2005), are characterized by search and bargaining. This implies that assets are imperfect substitutes for money and have, generally, positive liquidity premia. The second ingredient is an entry decision by the agents. Each asset trades in a distinct OTC market, and agents choose to visit the market where they expect to find the best terms. Additionally, we explore the endogenous determination of the supply of liquid assets. Concretely, we focus on two issuers of assets who play a differentiated Cournot game, where, crucially, the product (asset) differentiation stems from differences in the *microstructure* of the secondary market where each asset trades.

First, we study the endogenous determination of OTC market participation, keeping asset supplies fixed. Agents receive an idiosyncratic shock that determines whether they will need, ex-post, additional liquidity in the secondary market (i.e., sell assets) or whether they will be the providers of that liquidity (i.e., buy assets). An agent who turns out to be an asset seller can only visit one OTC market at a time; since, typically, assets are costly to own due to the liquidity premium, agents choose to 'specialize' ex-ante in asset A or B. Unlike sellers, who must take into account the cost of holding a particular asset, the asset buyers make their market choice in a more 'elastic' way since their money is good to buy any asset. As a result, when one of the markets, say market A, has any kind of advantage – an exogenous matching advantage

[1] For a thorough discussion of this stylized fact, see Krishnamurthy and Vissing-Jorgensen (2012).

[2] For instance, former Assistant Secretary of the U.S. Treasury, Brian Roseboro, points precisely in this direction: "A deep, liquid, and resilient secondary market serves our goal of lowest-cost financing for the taxpayer by encouraging more aggressive bidding in the primary market." (*A Review of Treasury's Debt Management Policy*, June 3, 2002, available at http://www.treas.gov/press/releases/po3149.htm.)

or simply offering bigger surpluses because there are more A-assets to be traded – asset buyers rush into that market more eagerly than sellers. In turn, this implies that the trade probability in that market for sellers increases by far more than that for buyers. Crucially, it is the sell-probability that affects the issue price, because someone who buys an asset (in the primary market) cares about the ease of selling it later.

Through this channel, small differences in market microstructure can be *magnified* into a big endogenous liquidity advantage for one asset, even if the matching function exhibits constant returns to scale (CRS). When we consider increasing returns to scale (IRS), our channel becomes further amplified because IRS promote concentration of investors in the market with the exogenous advantage.

Thus, our model can shed some light on the superior liquidity of U.S. Treasuries over equally safe corporate or municipal bonds. One may argue that this stylized fact has an easy explanation: the secondary market for Treasuries is more well-organized (which in our model would be captured by a more efficient matching technology). However, the relative illiquidity of corporate or municipal bonds has been well-documented for many decades. If the key behind this illiquidity was just some poorly organized secondary markets, one wonders why the issuers of these bonds have not taken steps to improve the efficiency of these markets, which would lower the rate at which they can borrow. Hence, it seems unlikely that the stylized fact in question can be purely explained by differences in market efficiency. Our model can offer a deeper explanation: perhaps Treasuries have a small exogenous advantage over other types of bonds, but this is amplified into a large endogenous liquidity advantage by the fact that investors choose to concentrate their trade into the secondary market for Treasuries, rather than get exposed to the liquidity risk associated with trading other types of bonds.[3]

To quantitatively assess the importance of this amplification mechanism, we calibrate our model to basic facts about yields in US fixed income markets, and use it to estimate how large the exogenous liquidity differences must be in order to match the difference between Treasury and high quality corporate bond yields observed in the data. We find that, even if we assume CRS, our model requires the matching technology in the corporate bonds market to be just seven percent less efficient than the one in the Treasury market to perfectly match the data. And with just a small degree of IRS, the exogenous liquidity differentials that are required to match the data virtually vanish (see Section 4.1 for details).

We also perform a counterfactual exercise. The secondary corporate bonds market is known to be particularly segmented, and practitioners (BlackRock, 2014) have argued that corporate bond liquidity would benefit from moving to a more consolidated secondary market.[4] To test

[3] For instance, Helwege and Wang (2021) report that many investors choose to not participate in the corporate bonds markets altogether, because they are highly concerned about the risk of not being able to liquidate their bonds quickly and at good terms, if such a need arises.

[4] This view is supported by the empirical findings of Oehmke and Zawadowski (2016), who find that "the *fragmented* nature of the corporate bond market impedes its liquidity" (emphasis added).

this proposal, we develop a version of our model with three assets trading in three distinct secondary markets – representing Treasuries, AAA, and AA corporate bonds – and ask the model what happens when we consolidate the corporate bond markets. Indeed, we find that the liquidity premia for both AAA and AA bonds increase, while the one on Treasuries decreases.

Next, we study the duopoly game between two issuers, who realize that the demand for their assets depends on the (exogenous and endogenous) liquidity characteristics of the secondary markets where their assets trade.[5] When the matching technology exhibits CRS, asset supplies tend to be strategic substitutes. In this case, equilibrium issue sizes are low, and the prices of both assets include liquidity premia. When the matching technology exhibits IRS, asset supplies tend to be strategic complements. This promotes aggressive competition among issuers, in the sense that equilibrium issue sizes can be large, and that equilibria of the subgame tend to be in a corner in which only one of the two OTC markets operates, and therefore, only one asset ends up liquid.

We also study how changes in the exogenous market microstructure affect optimal issue sizes, and, consequently, asset prices and liquidity premia. More precisely, letting δ_i, $i = A, B$, denote the matching efficiency in the OTC market, we start with $\delta_A = \delta_B$ and study the effect of decreases in δ_B. The exogenous liquidity advantage of asset A is magnified by the entry choices of agents, which, in turn, feeds back into a rising (falling) liquidity premium on asset A (B). As δ_B declines further, there comes a point at which issuer A has an incentive to boost up her supply and drive B out of the secondary market altogether. At that point asset B becomes fully illiquid. As δ_B falls even further, the threat of competition by asset B becomes so insignificant that issuer A practically turns into a monopolist in the supply of liquid assets.

With a degree of IRS in the matching technology, this process is accelerated. We show that asset B will become completely illiquid even if the matching function in market B is almost equally efficient as the one in market A (say, $\delta_B = 0.99\delta_A$), and there is only a tiny amount of IRS in the matching function. If one were to look at these numbers, one might infer that asset B cannot be much less liquid than asset A. This conclusion would be mistaken, because it would be based only on the exogenous factors. What is more important is that agents endogenously choose to concentrate their trade in market A because they expect other agents will do the same – and, reinforcing this, because both issuers have an incentive to compete for this concentration by issuing large (enough) amounts.

Finally, our model delivers some important results regarding welfare. First, and most importantly, there exists no monotonic relationship between welfare and "liquidity" (for any measure of liquidity we could choose). Second, unlike output, social welfare tends to be maximized for small-to-intermediate quantities of liquid assets. This alone does not tell us whether

[5] In Appendix A, we provide evidence that asset issuers act strategically. That said, the duopoly game played between two issuers is not meant to be taken literally, but to highlight that allowing for an endogenous determination of asset supplies offers important economic insights that are complementary to the analysis with exogenous asset supplies.

a monopoly or a Cournot duopoly of liquid assets would be superior; each is possible, depending on parameters. However, it does tell us that aggressive competition for secondary market liquidity, where issuers issue large amounts and drive liquidity premia to zero, is suboptimal. Consequently, market segmentation and exogenous liquidity differences can be good for welfare because they tend to discourage such aggressive competition.

The present paper is related to a branch of the recent literature, often referred to as "New Monetarism" (see Lagos, Rocheteau, and Wright, 2017), that has highlighted the importance of asset liquidity for the determination of asset prices. See for example Geromichalos, Licari, and Suárez-Lledó (2007), Lagos and Rocheteau (2008), Lester, Postlewaite, and Wright (2012), Nosal and Rocheteau (2013), Andolfatto and Martin (2013), Andolfatto, Berentsen, and Waller (2014), and Hu and Rocheteau (2015). In these papers assets are 'liquid' because they serve as a medium of exchange in frictional decentralized markets.[6] In some other papers, liquidity properties stem from the fact that assets serve as collateral, as in Venkateswaran and Wright (2014) and Andolfatto, Martin, and Zhang (2017).[7] The majority of this literature has studied asset liquidity (and prices) under the simplifying assumption that asset supply is fixed. Exceptions include Rocheteau and Rodriguez-Lopez (2014) and Branch, Petrosky-Nadeau, and Rocheteau (2016). Moreover, Bethune, Sultanum, and Trachter (2019) consider an environment with asset issuance and decentralized secondary markets, but they focus on efficiency and policy rather than liquidity. Our paper is also related to Caramp (2017) who endogenizes asset creation with a focus on asset quality and asymmetric information.

A key difference of our paper with the works mentioned so far is that here asset liquidity is *indirect*. Assets never serve as media of exchange (or as collateral) to purchase consumption. Their liquidity stems from the fact that agents can sell them for money in a secondary market. This idea is exploited in a number of recent papers, including Geromichalos and Herrenbrueck (2016), Berentsen, Huber, and Marchesiani (2014, 2016), Herrenbrueck (2019a), Mattesini and Nosal (2016), and Geromichalos, Herrenbrueck, and Lee (2018). As argued earlier, we believe that this approach is empirically relevant for a large class of financial assets. A common feature of these papers is that a secondary asset market allows agents to rebalance their liquidity after an idiosyncratic expenditure need has been revealed. This idea draws upon the work of Berentsen, Camera, and Waller (2007), where the channeling of liquidity takes place through a competitive banking system. Our work is also related to Lagos and Zhang (2015), but in that paper agents use money to purchase assets (rather than goods) in an OTC financial market.

Our work is also related to the literature initiated by the seminal work of Duffie et al. (2005),

[6] Consequently, in most of these papers, assets compete with money as media of exchange. In recent work, Fernández-Villaverde and Sanches (2019) extend the Lagos and Wright (2005) framework to study the interesting question of competition among privately issued electronic currencies, such as Bitcoin and Ethereum.

[7] Some papers within this literature have shown that adopting models where assets are priced both for their role as stores of value and for their liquidity may be the key to rationalizing certain asset pricing-related puzzles. See Lagos (2010), Geromichalos, Herrenbrueck, and Salyer (2016), and Herrenbrueck (2019b).

which studies how frictions in OTC financial markets affect asset prices and trade. A non-exhaustive list of such papers includes Vayanos and Wang (2007), Weill (2007, 2008), Vayanos and Weill (2008), Lagos and Rocheteau (2009), Lagos, Rocheteau, and Weill (2011), Afonso and Lagos (2015), Chang and Zhang (2015), Üslü (2019). Our paper is uniquely distinguished from all these papers, starting with the very concept of liquidity: we have a monetary model where agents sell assets for cash after learning of a consumption opportunity, while in those papers, agents differ in the utility flow derived from holding an asset and pay for assets with transferable utility. Furthermore, we characterize the strategic incentives facing issuers of potentially liquid assets, and thereby endogenize the supply of such assets in addition to their liquidity.

Our paper is also related to a strand of the Industrial Organization literature that studies the effect of secondary markets for durable goods on the producers' pricing decisions. Examples include Rust (1985, 1986). In these papers, the existence of a secondary market, where buyers could sell the durable good in the *future*, affects the pricing decisions of sellers *now* through affecting the buyers' willingness to pay for the good.[8] In our model, if secondary markets were shut down (so that assets have to be held to maturity), agents would be only willing to buy assets at their fundamental value. The existence of secondary markets endows assets with (indirect) liquidity properties, which, in turn, allows issuers to borrow funds at lower rates (i.e., sell bonds at a price that includes a liquidity premium).

The paper is organized as follows. Section 2 describes the model. In Section 3, we study the economy with exogenous asset supplies, and in Section 4, we calibrate our model to the data. Section 5 offers microfoundations for one of the key assumptions. In Section 6, we endogenize asset supplies by characterizing the game between asset issuers, and Section 7 concludes. Appendix A discusses empirical counterparts of our modeling choices, and Appendix B contains some technical details of the model. Finally, the Web Appendix contains several extensions of our analysis – only one asset issuer being strategic, one asset issuer being a Stackelberg leader, and one asset issuer having a higher cost of creating assets than the other – and an analytical characterization of the equilibria in our model.

2 The model

Time is discrete and the horizon is infinite. Each period consists of three sub-periods where different economic activities take place. In the first sub-period, two distinct OTC financial markets open, denoted by OTC_j, $j = \{A, B\}$. Agents who hold assets of type j can sell them for money in OTC_j. One could think of asset A as Treasury bonds and asset B as high-quality corporate bonds. In the second sub-period, agents visit a decentralized goods market where trade is bilateral, and agents are anonymous and lack commitment. We refer to this market as the DM. Due

[8] Within the context of financial rather than commodity markets, this idea is also exploited by Geromichalos et al. (2016) and Arseneau, Rappoport, and Vardoulakis (2015).

to the aforementioned frictions, trade necessitates a medium of exchange in the DM, and this role can be played only by money. During the third sub-period, economic activity takes place in a centralized market, which is similar in spirit to the settlement market of Lagos and Wright (2005) (henceforth, LW). We refer to this market as the CM. There are two permanently distinct types of agents, consumers and producers, named by their role in the DM, and the measure of both is normalized to the unit. Agents live forever. There are also two agencies, $j = \{A, B\}$, that issue asset j in its respective *primary market* which opens within the third sub-period.

All agents discount the future between periods (but not sub-periods) at rate $\beta \in (0, 1)$. Consumers consume in the DM and CM sub-periods and supply labor in the CM sub-period. Their preferences within a period are given by $\mathcal{U}(X, H, q) = X - H + u(q)$, where X, H represent consumption and labor in the CM, respectively, and q consumption in the DM. Producers consume only in the CM, and they produce in both the CM and the DM. Their preferences are given by $\mathcal{V}(X, H, h) = X - H - q$, where X, H are as above, and q stands for units of production in the DM. We assume that u is twice continuously differentiable with $u' > 0$, $u'(0) = \infty$, $u'(\infty) = 0$, and $u'' < 0$. Let q^* denote the optimal level of production in a bilateral meeting in the DM, i.e., $q^* \equiv \{q : u'(q^*) = 1\}$. The issuers of assets are only present in the CM. Their preferences are given by $\mathcal{Y}(X, H) = X - H$, where X, H are as above. The issuers also discount the future at rate β. What makes them special is that they can issue assets that potentially carry liquidity premia, thus allowing them to obtain net profits out of this operation.

We now provide a detailed description of the various sub-periods. In the third sub-period, all agents consume and produce a general good or fruit. All agents (including the issuers) have access to a technology that transforms one unit of labor into one unit of the fruit. Agents can choose to hold any amount of money which they can purchase at the ongoing price φ_t (in real terms). The supply of money is controlled by the monetary authority, and it evolves according to $M_{t+1} = (1 + \mu)M_t$, with $\mu > \beta - 1$. New money is introduced, or withdrawn if $\mu < 0$, via lump-sum transfers to consumers in the CM. Money has no intrinsic value, but it possesses all the properties that make it an acceptable medium of exchange in the DM (e.g., it is portable, storable, and recognizable by agents). Agents can also purchase any amount of asset j at price p_j, $j = \{A, B\}$ (in nominal terms). These assets are one-period nominal bonds: each unit of (either) asset purchased in period t's CM pays one dollar in the CM of $t + 1$.[9] Let the supply of the assets be denoted by (A_t, B_t). In Sections 3-5, we will treat them as fixed; in Section 6, they will be chosen endogenously by the issuers. Each issuer chooses the supply of her asset as a best response to her rival's action in order to maximize profits, realizing that both her own and her rival's assets provide indirect liquidity services to an asset purchaser.

After making their portfolio decisions in the CM, consumers receive an idiosyncratic consumption shock: a measure $\ell < 1$ of consumers will have a desire to consume in the forthcoming

[9] Since the assets are nominal, in steady state their supply must grow at rate μ, too (see, for example, Berentsen and Waller, 2011).

DM. We refer to them as the C-types, and to the remaining $1-\ell$ consumers as the N-types ("not consuming"). Since consumers did not know their type when they made their portfolio choices, N-types will typically hold some cash that they will not use in the current period, while C-types may find themselves short of cash (since carrying money is costly). The OTC round of trade is placed *after* the idiosyncratic uncertainty has been resolved, but *before* the DM opens to allow a reallocation of money into the hands of those who value it most. OTC financial markets are segmented: an agent who wants to sell or purchase assets is free to enter either OTC$_A$ or OTC$_B$, but she must choose one market at a time.[10] Hence, coordination is extremely important, and agents will pick the market where they expect to find better trading conditions.

Once C-types and N-types have decided which market they wish to enter, a matching function, $f_j(C_j, N_j)$, brings together sellers (C-types) and buyers (N-types) of assets in the OTC$_j$, in bilateral matches. Throughout the paper we use the specific functional form:

$$f_j(x,y) = \delta_j \left(\frac{xy}{x+y}\right)^{1-\rho} (xy)^{\rho} \,,$$

with $\delta_j \in [0,1]$ and $\rho \in [0,1]$, and thus $f_j(x,y) \leq \min\{x,y\}$. The term δ_j captures exogenous efficiency factors in OTC$_j$, such as the density of the dealer network. The term $\rho \in [0,1]$ governs returns to scale in matching; for concreteness, notice that the elasticity of each side's matching probability with respect to scale, keeping the ratio of buyers to sellers fixed, is ρ. This functional form allows us to study both the case of CRS ($\rho = 0$) and IRS ($\rho > 0$). Within any match in either of the OTC markets, the C-type and N-type split the available surplus based on proportional bargaining (Kalai, 1977), with $\theta \in (0,1)$ denoting the C-type's bargaining power.

The second sub-period is the standard decentralized goods market of the LW model. Active consumers (C-type) meet bilaterally with producers and negotiate over the terms of trade. Exchange must take place in a *quid pro quo* fashion, and only money can serve as a medium of exchange.[11] Since all the interesting insights of the paper follow from agents' interaction in the OTC round of trade, we wish to keep the DM as simple as possible. To that end, we assume that all C-type consumers match with a producer, and that in any match the consumer makes a take-it-or-leave-it (TIOLI) offer.

Figure 1 summarizes the timing of the main actions in the model. It is important to highlight that the secondary OTC markets are completely separate from the primary markets where assets are first issued. Nevertheless, the microstructure of the secondary markets, summarized by the parameters $\delta_j, \rho,$ and θ, will determine the liquidity properties of the assets and, consequently, their selling price in the primary market.

[10] We discuss this assumption in Appendix A, and we provide microfoundations for it in Section 5. Furthermore, perfectly integrated markets are equivalent to a special case of our model, described in Proposition 1, part (e).

[11] Here we shall make this an assumption of the model. However, a number of recent papers in the monetary-search literature, such as Rocheteau (2011) and Lester et al. (2012) do not place any restrictions on which objects can serve as media of exchange and show that, under asymmetric information, fiat money will endogenously arise as a superior medium of exchange, thus, providing a micro-founded justification for our assumption.

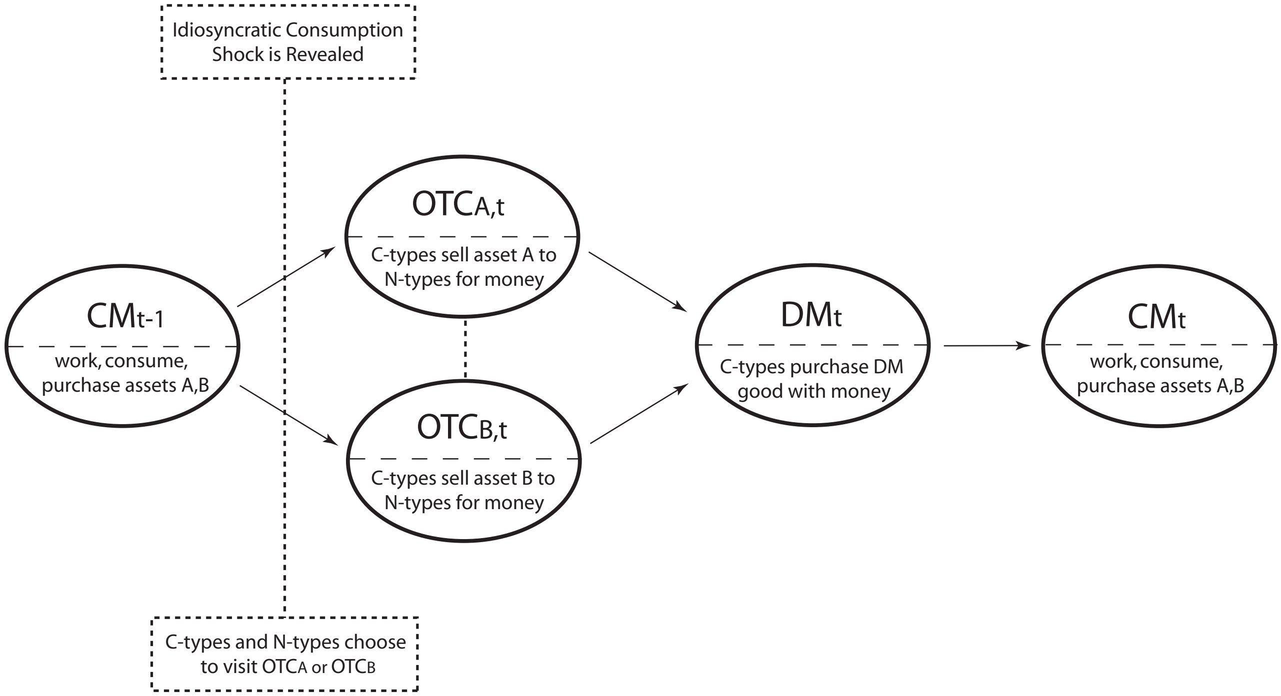


Figure 1: Timing of events.

3 The economy with exogenous asset supply

In this section we analyze the economy, treating the supplies of assets (A, B) as given. The task of endogenizing the asset supplies is carried out in Section 6.

In order to streamline the analysis, we relegate the details of defining the value functions and characterizing the terms of trade in the OTC markets and the DM to Appendix B.1. Here we include a summary. All agents in the economy have linear preferences over labor and consumption goods in the CM, which will induce linear value functions in the CM, and make a number of economic decisions easy to characterize. First, consider a DM meeting between a producer and a C-type consumer who brings a quantity m of money. The consumer will either buy the first-best quantity q^*, or, if her money is not enough, spend all of it on the quantity $q = \varphi m < q^*$. Second, consider a meeting in the OTC market for asset $j \in \{A, B\}$, where the N-type brings a quantity $\tilde{m}$ of money, and the C-type brings a portfolio (m, d_j) of money and asset j. The N-type and C-type split the available surplus based on proportional bargaining: the N-type buys the C-type's assets and compensates him with money.[12]

[12] In OTC trade, three kinds of outcomes are possible: (a) the C-type's asset holdings could limit the trade; (b) the N-type's money holdings could limit the trade; (c) or both are so large that the pooled money is enough to purchase the first-best DM quantity ($m + \tilde{m} > q^*/\varphi$), and the C-type has enough assets to compensate the N-type. In Geromichalos and Herrenbrueck (2016), we showed that assets can only be priced (in the CM) at a determinate liquidity premium if case (a) applies in the corresponding OTC market. Case (c) is also relevant as the boundary of case (a), where an asset becomes abundant and the liquidity premium converges to zero. Case (b), however, only complicates the general equilibrium analysis. Since it does not feature a positive liquidity premium, and since our

What is the probability of matching in an OTC market for an individual agent? First, let $e_C \in [0,1]$ and $e_N \in [0,1]$ denote the fractions of C-types and N-types, respectively, who choose to enter OTC$_A$. Then, the measure of asset sellers and buyers in OTC$_A$ is given by $e_C\ell$ and $e_N(1-\ell)$, respectively, and the measure of asset sellers and buyers in OTC$_B$ is given by $(1-e_C)\ell$ and $(1-e_N)(1-\ell)$. Letting $\alpha_{ij} \in [0,1]$ denote the matching probabilities for agents of type $i = \{C, N\}$ in OTC$_j$, $j = \{A, B\}$, we have:

$$\alpha_{CA} \equiv \frac{f_A\big(e_C\ell, e_N(1-\ell)\big)}{e_C\ell}, \qquad \alpha_{CB} \equiv \frac{f_B\big((1-e_C)\ell, (1-e_N)(1-\ell)\big)}{(1-e_C)\ell}, \tag{1}$$

$$\alpha_{NA} \equiv \frac{f_A\big(e_C\ell, e_N(1-\ell)\big)}{e_N(1-\ell)}, \qquad \alpha_{NB} \equiv \frac{f_B\big((1-e_C)\ell, (1-e_N)(1-\ell)\big)}{(1-e_N)(1-\ell)}. \tag{2}$$

3.1 Optimal behavior

As shown in Appendix B.1, the producers' decisions in this model are trivial. Thus, in what follows we use the term 'agents' to refer to consumers, since these are the types who make interesting portfolio decisions. In the OTC market, these agents will take on roles as 'asset sellers' and 'asset buyers' depending on the outcome of their consumption shock (C or N, respectively). The 'producers' (who sell goods in the DM) will not come up again in the main text.

As is standard in models that build on LW, all agents choose their optimal portfolio independently of their trading histories in preceding markets. This result follows from the "no-wealth-effects" property, which, in turn, stems from the quasilinear preferences. Here, in addition to choosing an optimal portfolio of money and assets, $(\hat{m}, \hat{d}_A, \hat{d}_B)$, agents also choose which OTC market they will enter in order to sell or buy assets once their type has been revealed. The agent's choice can be analyzed with an objective function, denoted by $J(\hat{m}, \hat{d}_A, \hat{d}_B)$, which summarizes the cost and benefit from choosing portfolio $(\hat{m}, \hat{d}_A, \hat{d}_B)$. To obtain J, substitute the values of trading in the OTC markets and in the DM (Equations B.4-B.8, derived in the appendix) into the maximization operator of the CM value function (Equation B.1). After using the linearity of the value function itself (Equation B.2), we can drop all terms that do not depend on the choice variables $(\hat{m}, \hat{d}_A, \hat{d}_B)$ to obtain the objective function:

$$\begin{aligned} J(\hat{m}, \hat{d}_A, \hat{d}_B) = &-\varphi\left(\hat{m} + p_A\hat{d}_A + p_B\hat{d}_B\right) + \beta\hat{\varphi}\left(\hat{m} + \hat{d}_A + \hat{d}_B\right) \\ &+ \beta\ell\Big[u\left(\hat{\varphi}\hat{m}\right) - \hat{\varphi}\hat{m} + \max\Big\{\underbrace{\alpha_{CA}S_{CA}}_{\text{enter A}}, \underbrace{\alpha_{CB}S_{CB}}_{\text{enter B}}\Big\}\Big], \end{aligned} \tag{3}$$

so that the optimal portfolio choice is fully described by $\max J$, where the current prices of

interest is in asset issuers who seek to exploit such a premium, we exclude case (b) from our analysis. This is done by assuming that inflation is not too large, so that all agents carry at least *half* of the first-best amount of money.

money and assets, (φ, p_A, p_B), and the future price of money, $\hat{\varphi}$, are taken as given.

The interpretation of the objective function is intuitive. The first term represents the cost that the agent needs to pay in order to purchase the portfolio $(\hat{m}, \hat{d}_A, \hat{d}_B)$ in the CM, and the second term represents the benefit from selling these assets in the CM of the next period. The third term reveals that with probability ℓ the agent will be a C-type in the next period. In this case she can use her money ($\hat{m}$) to purchase consumption in the DM (generating a net surplus equal to $u(\hat{\varphi}\hat{m}) - \hat{\varphi}\hat{m}$), and she can enter OTC$_j$, $j = A, B$, in order to acquire more money by selling her assets ($\hat{d}_A$ or $\hat{d}_B$). In the last expression, the terms S_{Cj} represent the surplus for the C-type in OTC$_j$, but the agent will actually enjoy this surplus only if she gets to match in that market, an event that occurs with probability α_{Cj}.[13] Exploiting the OTC bargaining solution (i.e., Lemma 2) and Equation (B.9), one can verify that, for $j = \{A, B\}$:

$$S_{Cj} = \begin{cases} \theta[u(q^*) - u(\hat{\varphi}\hat{m}) - \hat{\varphi}(m^* - \hat{m})], \\ \qquad\qquad \text{if } \hat{\varphi}\hat{d}_j \geq (1-\theta)[u(q^*) - u(\hat{\varphi}\hat{m})] + \theta(q^* - \hat{\varphi}\hat{m}), \\ \theta[u(\hat{\varphi}(\hat{m} + \tilde{\zeta}_j)) - u(\hat{\varphi}\hat{m}) - \hat{\varphi}\tilde{\zeta}_j], \quad \text{otherwise}, \end{cases} \tag{4}$$

where $\tilde{\zeta}_j$ solves $\hat{\varphi}\hat{d}_j = (1-\theta)[u(\hat{\varphi}(\hat{m} + \tilde{\zeta}_j) - u(\hat{\varphi}\hat{m})] + \theta\hat{\varphi}\tilde{\zeta}_j$. The condition $\hat{\varphi}\hat{d}_j \geq (1-\theta)[u(q^*) - u(\hat{\varphi}\hat{m})] + \theta(q^* - \hat{\varphi}\hat{m})$ states that in this case the agent's asset holdings are "abundant", i.e., they allow her to reach the first-best amount of money, m^*, through OTC trade.

Two important observations are in order. First, while we have only imposed an exogenous segmentation assumption on the OTC markets, an *endogenous* segmentation will arise in the *primary* markets: i.e., agents will typically choose to purchase only asset A or asset B in the CM. In equilibrium, assets will trade at a premium, and agents will only pay this premium if they expect to sell the asset in the OTC. Since they can only enter one OTC (and anticipate having to choose eventually), they will choose *ex-ante* (i.e., in the CM), to "specialize" in asset A or B.[14] This, in turn, implies that an agent's portfolio choice is intertwined with the choice of which OTC market to enter in case she turns out to be a C-type. For instance, we shall see that agents who choose to trade in a less liquid OTC market will self-insure against the liquidity shock by carrying more money.

The second important observation is that the agent's choice of which market to enter if she turns out to be an N-type is unrelated with her choice of asset specialization in the CM. This is because the N-type's asset and money holdings do not affect the bargaining solution in OTC

[13] One may wonder why there is no $(1-\ell)$-term in the objective function. Does the N-type not generate value by bringing money into the OTC? Yes, this is the case, as the full value function (Equation B.1) shows. But the technical restriction (6), justified in Footnote 12, guarantees that the N-type's money is never *marginal* in OTC trade. Hence the N-branch can be dropped from the portfolio choice problem; the only decision to be made along the N-branch is which OTC market to enter.

[14] Agents may still hold the other asset if indifferent, i.e., if that asset is abundant or illiquid.

trade (see Lemma 2). As a result, regardless of her asset choice which by the time the N-type makes her OTC entry choice is sunk, this agent will enter OTC$_A$ only if:

$$\alpha_{NA} S_{NA} \geq \alpha_{NB} S_{NB}.$$

In the last expression, the terms S_{Nj} represent the surplus for the N-type in OTC$_j$. Exploiting Lemma 2 and equation (B.10), one can verify that, for $j = \{A, B\}$:

$$S_{Nj} = \begin{cases} (1-\theta)[u(q^*) - u(\hat{\varphi}\tilde{m}) - \hat{\varphi}(m^* - \tilde{m})], \\ \qquad\qquad \text{if } \hat{\varphi}\tilde{d}_j \geq (1-\theta)[u(q^*) - u(\hat{\varphi}\tilde{m})] + \theta(q^* - \hat{\varphi}\tilde{m}), \\ (1-\theta)[u(\hat{\varphi}(\tilde{m} + \tilde{\zeta}_j)) - u(\hat{\varphi}\tilde{m}) - \hat{\varphi}\tilde{\zeta}_j], \quad \text{otherwise}, \end{cases} \tag{5}$$

where $\tilde{\zeta}_j$ solves $\hat{\varphi}\tilde{d}_j = (1-\theta)[u(\hat{\varphi}(\tilde{m} + \tilde{\zeta}_j) - u(\hat{\varphi}\tilde{m})] + \theta\hat{\varphi}\tilde{\zeta}_j$ and $(\tilde{m}, \tilde{d}_j)$ stand for the N-type's expectation about the money and asset-j holdings, respectively, that her trading partner, a C-type, will carry into OTC$_j$. The condition $\hat{\varphi}\tilde{d}_j \geq (1-\theta)[u(q^*) - u(\hat{\varphi}\tilde{m})] + \theta(q^* - \hat{\varphi}\tilde{m})$ states that the asset holdings of the C-type are large enough to allow her post-OTC money balances to reach the first-best amount, m^*.

3.2 Equilibrium

In steady state, the cost of holding money can be summarized by the parameter $i \equiv (1+\mu-\beta)/\beta$; exploiting the Fisher equation, this parameter represents the nominal interest rate on an *illiquid* asset. For example, in any equilibrium it must be true that $p_j \geq 1/(1+i)$, $j = \{A, B\}$, since otherwise there would be an infinite demand for the assets; however, the inequality could be strict if the assets are liquid. The restriction $\mu > \beta - 1$ translates into $i > 0$. We also assume that:

$$i < \ell(1-\theta)\left[u'\left(q^*/2\right) - 1\right], \tag{6}$$

a technical restriction. It ensures that $q_{0j} > q^*/2$ for every agent, thus the N-type's money will never be the limiting factor in OTC trade. See our explanation in Footnote 12, and note that if we did have $q_{0j} < q^*/2$, the implied burden of the inflation tax would be enormous.

We have eleven endogenous variables.[15] First, we have the equilibrium real balances $\{z_A, z_B\}$ held by the agent who chooses to specialize in asset A or B (recall from the discussion in Section 3.1 that an agent who chooses to trade in OTC$_A$ will typically make different portfolio choices than one who chooses to trade in OTC$_B$). Next, we have the equilibrium quantities $\{q_{0A}, q_{1A}, q_{0B}, q_{1B}\}$. These represent the quantity of DM good purchased by a C-type agent who

[15] This count excludes the terms of trade in the OTC markets, since they follow directly from the main endogenous variables described in this section and Lemma 2.

either did not trade in the OTC market (indexed by 0), or who traded (indexed by 1), depending on whether they chose to specialize in asset A or asset B. Next, we have the prices of the three assets $\{\varphi, p_A, p_B\}$. Finally, we have the entry choices $\{e_C, e_N\}$, i.e., the fractions of C-types and N-types, respectively, who choose to enter OTC$_A$.

We now show that five out the eleven endogenous variables can be derived from the following six variables, $\{q_{0A}, q_{1A}, q_{0B}, q_{1B}, e_C, e_N\}$. First, we have $z_j = q_{0j}$, for $j = \{A, B\}$, since the C-type who does not trade in the OTC can only purchase the amount of DM goods that her own real money holdings, z_j, allow her to afford. Second, the price of money solves:

$$\varphi M = e_C q_{0A} + (1 - e_C) q_{0B}. \tag{7}$$

This equation is the market clearing condition in the market for money. Third, the equilibrium asset prices must satisfy the demand equations:[16]

$$p_j = \frac{1}{1+i}\left(1 + \ell\, \alpha_{Cj} \frac{\theta}{\omega(q_{1j})}[u'(q_{1j}) - 1]\right), \quad \text{for} \quad j = \{A, B\}, \tag{8}$$

where

$$\omega(q) \equiv \theta + (1 - \theta) u'(q).$$

For future reference, notice that as long as $q_{1j} < q^*$, the marginal unit of the asset allows the agent to acquire additional money which she can use in order to boost her consumption in the DM. In this case, the agent is willing to pay a *liquidity premium* in order to hold the asset. On the other hand, if $q_{1j} = q^*$, the term inside the square brackets becomes zero, and $p_j = 1/(1 + i)$, which is simply the *fundamental price* of a one-period nominal bond.

The analysis so far establishes that if one had solved for $\{q_{0A}, q_{1A}, q_{0B}, q_{1B}, e_C, e_N\}$, then the remaining five variables could also be immediately determined. Hence, hereafter we refer to these six variables as the "core" variables of the model. We now turn to the description of the equilibrium conditions that determine the core variables. Throughout this discussion, recall that the terms e_C, e_N are also implicitly affecting the arrival rates α_{Cj}.

First, the money demand equation for those specializing in asset j:

$$i = \ell\left(1 - \alpha_{Cj} \frac{\theta}{\omega(q_{1j})}\right)[u'(q_{0j}) - 1] + \ell\, \alpha_{Cj} \frac{\theta}{\omega(q_{1j})}[u'(q_{1j}) - 1], \quad \text{for} \quad j = \{A, B\}. \tag{9}$$

Note that we have defined $\alpha_{ij} = 0$ if there is no entry at all into market j. If that is the case, q_{0j} and q_{1j} are still defined as limits even though nobody actually trades at those quantities.

[16] These follow directly from obtaining the first-order conditions in the agent's objective function, i.e., Equation (3), and imposing equilibrium quantities. Notice that the asset prices do not only depend on the variables q_{1j}, but also on the equilibrium values of e_C, e_N which affect the arrival rates α_{Cj}; see Equations (1).

Next, the OTC trading protocol links q_{0j} and q_{1j}. Consider for instance market A. The bargaining solution, evaluated at equilibrium quantities, becomes:

$$q_{1j} = \min\{q^*, q_{0j} + \varphi\tilde{\zeta}_j\}, \quad \text{for} \quad j = \{A, B\},$$

where $\tilde{\zeta}_j$ solves $\varphi d_j = (1-\theta)[u(\varphi(m+\tilde{\zeta}_j) - u(\varphi m)] + \theta\varphi\tilde{\zeta}_j$ where $d_A = A/e_C$ and $d_B = B/(1-e_C)$, the amount of assets that the C-type brings into OTC$_j$.[17] Even though the aggregate supply of asset A is A, the agent under consideration holds more than the average because some agents do not hold asset A at all (they specialize in asset B), and the same argument goes also with asset B. After substituting the price of money from Equation (7) into the last expression, we obtain two equations, one for each market:

$$q_{1A} = \min\left\{q^*, q_{0A} + \frac{1}{\theta}\frac{A}{M}\frac{e_C q_{0A} + (1-e_C)q_{0B}}{e_C} - \frac{1-\theta}{\theta}[u(q_{1A}) - u(q_{0A})]\right\}, \tag{10}$$

$$q_{1B} = \min\left\{q^*, q_{0B} + \frac{1}{\theta}\frac{B}{M}\frac{e_C q_{0A} + (1-e_C)q_{0B}}{1-e_C} - \frac{1-\theta}{\theta}[u(q_{1B}) - u(q_{0B})]\right\}. \tag{11}$$

If it happens that $e_C = 1$ (no C-types enter the B-market) and $B > 0$, then we define $q_{1B} = q^*$ as a limit, because a C-type of infinitesimal size who decided to deviate and hold asset B could hold the entire stock of it, which would certainly satiate them in an OTC trade – in the hypothetical case that there was an N-type in the B-market willing to trade with them. Similarly, if $e_C = 0$ and $A > 0$, then we define $q_{1A} = q^*$.

How large can the aggregate supply of an asset be for the asset to remain scarce in OTC trades? Clearly, the asset is more likely to be scarce if its ownership is *diluted*, i.e., if many agents choose to hold that asset in the CM. So, for example, asset A is most likely to be scarce if $e_C = 1$. But in this special case, Equation (10) tells us that the asset is scarce ($q_{1A} < q^*$) only if the condition $q^* > q_{0A}[1 + A/(\theta M)] - (1-\theta)/\theta[u(q_{1A}) - u(q_{0A})]$ is satisfied. On the boundary, $q_{1A} = q^*$, so we can use the money demand equation (9) to obtain the bounds:

$$\bar{A} \equiv M\theta\left[\frac{q^*}{\bar{q}_{0A}} + \frac{1-\theta}{\theta}\frac{u(q^*) - u(\bar{q}_{0A})}{\bar{q}_{0A}} - 1\right], \text{ where } \bar{q}_{0A} \text{ solves } i = [\ell - \theta f_A(\ell, 1-\ell)][u'(\bar{q}_{0A}) - 1],$$

$$\bar{B} \equiv M\theta\left[\frac{q^*}{\bar{q}_{0B}} + \frac{1-\theta}{\theta}\frac{u(q^*) - u(\bar{q}_{0B})}{\bar{q}_{0B}} - 1\right], \text{ where } \bar{q}_{0B} \text{ solves } i = [\ell - \theta f_B(\ell, 1-\ell)][u'(\bar{q}_{0B}) - 1].$$

There are three things to notice here. First, if $A > \bar{A}$, then asset A is certain to be abundant but the reverse is not always true, because asset ownership can be *concentrated* in the hands of a few agents. Second, if we did fix $e_C = 1$ so that ownership of asset A was maximally diluted, then

[17] If the C-type's asset holdings are plentiful in the OTC, then we know that this agent will be able to purchase the first-best amount of money in the DM, hence, $q_{1j} = q^*$. On the other hand, if the asset is scarce in OTC trade, the C-type gives away all of her assets, d_j.

asset A would indeed be abundant if and only if $A \geq \bar{A}$, and conversely for asset B. Third, if the market for asset A has an exogenous liquidity advantage ($\delta_A > \delta_B$), then $\bar{A} > \bar{B}$, and vice versa. For convenience, we define the maximal upper bound on asset supply beyond which either asset is certain to be abundant:

$$\bar{D} \equiv \max\{\bar{A}, \bar{B}\}.$$

The remaining task is to characterize the OTC market entry choices. Consider first a C-type. As we have already discussed, this type at the beginning of the period has already made the choice to hold either asset A or asset B, so the choice of which market to enter has effectively been made. Evaluating equation (4) at equilibrium quantities, we find that her surplus of trading in market $j \in \{A, B\}$ equals:[18]

$$S_{Cj} = \theta[u(q_{1j}) - u(q_{0j}) - q_{1j} + q_{0j}]. \tag{12}$$

But since the agent's portfolio choice effectively determines her market choice if she turns out to be a C-type, this surplus has to be balanced not only against the probability of needing to trade and actually matching ($\ell \times \alpha_{Cj}$), but also against the cost of carrying the asset. Hence, we define the "net" surplus that the agent obtains if she chooses to specialize in asset j to be:

$$\begin{aligned}\tilde{S}_{Cj} \equiv & -iq_{0j} - [(1+i)p_j - 1][(1-\theta)(u(q_{1j}) - u(q_{0j})) + \theta(q_{1j} - q_{0j})] \quad \dots \\ & + \ell[u(q_{0j}) - q_{0j}] + \ell\, \alpha_{Cj} S_{Cj},\end{aligned} \tag{13}$$

where we can use the money and asset demand equations (8 and 9) to substitute for i and p_j. Thus, in equilibrium, the C-types' portfolio choice e_C must satisfy:

$$e_C = \begin{cases} 1, & \text{if } \tilde{S}_{CA} > \tilde{S}_{CB}, \\ 0, & \text{if } \tilde{S}_{CA} < \tilde{S}_{CB}, \\ \in [0,1], & \text{if } \tilde{S}_{CA} = \tilde{S}_{CB}. \end{cases} \tag{14}$$

Finally, we want to characterize the market choice of the N-type agents. Since these agents are asset buyers, their own asset holdings do not matter, so they can enter the market for either asset independently of which asset they chose to hold in the preceding CM. Thus, an N-type will simply enter the market in which she expects a greater surplus, accounting for the proba-

[18] This equality holds regardless of whether the asset is plentiful in the OTC meeting or not. Consider first the case of plentiful assets. For this case evaluating the relevant (i.e., the "abundant") branch of Equation (4) at equilibrium quantities yields $S_{Cj} = \theta[u(q^*) - u(q_{0j}) - q^* + q_{0j}]$, which is exactly what one would obtain if $q_{1j} = q^*$ was imposed on Equation (12). Next, consider the case of scarce assets and for simplicity focus on OTC$_j$. In this case, evaluating (4) at equilibrium quantities yields $S_{Cj} = \theta[u(q_{1j}) - u(q_{0j}) - q_{1j} + q_{0j}]$.

bility of trading. Evaluating equation (5) at equilibrium quantities implies that the surplus for the N-type who chooses to enter OTC$_j$ is given by:

$$S_{Nj} = (1-\theta)[u(q_{1j}) - u(q_{0j}) - q_{1j} + q_{0j}]. \tag{15}$$

Thus, in equilibrium, the N-types' entry choice $e_{\scriptscriptstyle N}$ must satisfy:

$$e_{\scriptscriptstyle N} = \begin{cases} 1, & \text{if } \alpha_{NA}S_{NA} > \alpha_{NB}S_{NB}, \\ 0, & \text{if } \alpha_{NA}S_{NA} < \alpha_{NB}S_{NB}, \\ \in [0,1], & \text{if } \alpha_{NA}S_{NA} = \alpha_{NB}S_{NB}. \end{cases} \tag{16}$$

We can now define a steady-state equilibrium in the model with fixed asset supplies:

Definition 1. Assume (for now) that asset supplies are fixed and equal to $(A, B) \in \mathbb{R}^2_+$. A steady-state equilibrium for the core variables of the model is a list $\{q_{0A}, q_{1A}, q_{0B}, q_{1B}, e_{\scriptscriptstyle C}, e_{\scriptscriptstyle N}\}$ such that Equations (9) for $j = \{A, B\}$, (10), and (11) hold, and agents' entry choices satisfy Equations (14) and (16).

3.3 Characterization of equilibrium

We are now ready to characterize the equilibria of the economy, summarized by the core variables $\{q_{0A}, q_{1A}, q_{0B}, q_{1B}, e_{\scriptscriptstyle C}, e_{\scriptscriptstyle N}\}$, conditional on the asset supplies $A, B \geq 0$. Before we go to the technical details, it is helpful to gain some intuition by considering the optimal entry decision of the representative N-type, who takes as given the term $e_{\scriptscriptstyle N}$, the proportion of other N-types who enter the A-market, and best responds by entering in either market A or B. A higher value of $e_{\scriptscriptstyle N}$ implies a bigger congestion among N-types in market A, a force that *discourages* our representative N-type from entering into that market. On the other hand, a higher $e_{\scriptscriptstyle N}$ implies that a larger fraction of C-types will be drawn to market A, because C-types like a market with many N-types, and this force *encourages* our representative N-type to enter into that market. And, to make things even more interesting, a higher value of $e_{\scriptscriptstyle C}$ implies that the supply of asset A, which is fixed for now, will be *diluted* among a larger number of agents (this channel becomes more relevant if the supply of asset A is scarce). Hence, in any bilateral meeting in OTC$_A$, the surplus is more likely to be limited because the C-type is constrained by her asset holdings, yet another force that discourages our representative N-type from entering into market A.

Summing up, an increase in the term $e_{\scriptscriptstyle N}$ generates multiple and opposing forces, and may have non-monotonic effects on the optimal entry decision of the representative N-type. What one can say safely is that everything else equal, the typical N-type is more likely to enter into market A if: (i) $\delta_A > \delta_B$, because then the former market has an exogenous matching advantage; and (ii) $A > B$, because then there is a larger potential surplus when trading asset A.

Moving to the formal analysis, we construct equilibria as fixed points of e_N. To be specific: first, we fix a level of e_N; then we solve for the optimal portfolio choices through Equations (9)-(11) and (14); and finally, we define the N-types' *reply function*:

$$G(e_N) \equiv \frac{\alpha_{NA}S_{NA} - \alpha_{NB}S_{NB}}{\alpha_{NA}S_{NA} + \alpha_{NB}S_{NB}},$$

where the surplus (S) and match probability (α) terms have the optimal choices substituted. This function measures the relative benefit to an *individual* N-type from choosing the A-market over the B-market, assuming a proportion e_N of all *other* N-type agents enters the A-market, and all other decisions are conditionally optimal. To make it easier to visualize, G is scaled to lie between -1 and $+1$. A value of e_N is part of an "interior" equilibrium if $e_N \in (0, 1)$ and $G(e_N) = 0$, or a "corner" equilibrium if $e_N = 0$ and $G(0) \leq 0$ or $e_N = 1$ and $G(1) \geq 0$.

Proposition 1. *The following types of equilibria exist, and have these properties:*

(a) There exists a corner equilibrium where $e_C = e_N = 0$; only the B-market is open for trade.

(b) There exists a corner equilibrium where $e_C = e_N = 1$; only the A-market is open for trade.

(c) Assume $\rho = 0$ (CRS) and asset supplies are low enough so that assets are scarce in OTC trade. Then, $\lim_{e_N \to 0+} G(e_N) > 0 > G(0)$ and $\lim_{e_N \to 1-} G(e_N) < 0 < G(1)$; the corner equilibria are not robust to small trembles. There exists at least one interior equilibrium which is robust to small trembles.

(d) Assume $\rho > 0$ (IRS). Then, $\lim_{e_N \to 0+} G(e_N) = G(0) < 0$ and $\lim_{e_N \to 1-} G(e_N) = G(1) > 0$; the corner equilibria now are robust to small trembles. There exists at least one interior equilibrium, which may or may not be robust to small trembles.

(e) Assume $\rho = 0$ (CRS) and $\delta_A = \delta_B$ (equal market quality). Then, a symmetric equilibrium exists where $e_C = e_N = A/(A + B)$, $q_{0A} = q_{0B}$ and $q_{1A} = q_{1B}$, and $p_A = p_B$.

(f) If, in addition to the assumptions in (e), $A = B < \bar{D}/2$ (asset supplies are equal and small), $i \to 0$ (low inflation), $\theta\delta(1 - \ell) < 0.5$ (not-too-high bargaining power for the C-type), and $u''' \geq 0$ (convex marginal utility), then $G'(0.5) < 0$; that is, the symmetric equilibrium is robust to small trembles.

Proof. See Sections D.1-D.3 in the Web Appendix. □

In all cases, the C-types' entry choice e_C is optimally adjusting in the background, and it is generally an increasing function of e_N; when there are many buyers in a market, sellers would like to go to the same market. Additionally, it must be the case that $e_C = 0$ if and only if $e_N = 0$, and $e_C = 1$ if and only if $e_N = 1$ (parts (a) and (b) of the proposition). Therefore, the corners are always equilibria.

These results are depicted in Figure 2, which shows how the reply function G depends on e_N and on asset supplies, given CRS in matching. Dots at $G(0) = -1$ and $G(1) = +1$ indicate that the corners are always equilibria. In the left panel, G is shown for relatively low supplies of

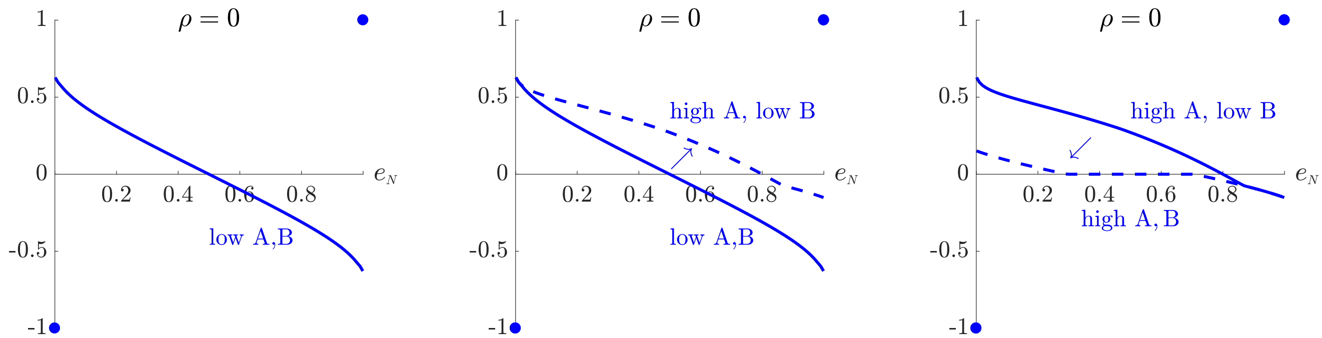


Figure 2: The reply function $G(e_N)$ for CRS ($\rho = 0$) and varying asset supplies.

A, B, and there is an interior fixed point at $e_N = 0.5$. As shown in part (c) of the Proposition, the corners are not robust to small trembles, but the interior fixed point is: if a few more N-types accidentally enter the A-market, individual N-types have an incentive to deviate back to B. In the middle panel, we show what happens for a higher supply of A: the G-function shifts up and more agents trade in the A-market, but the equilibrium is still robust.

The right panel illustrates the case where both A and B are high: the G-function shifts back down, but now it contains a flat segment for intermediate values of e_N. This is due to the fact that with high asset supplies, the aforementioned *dilution effect* disappears: if the supply of assets is high enough, each individual C-type will be able to achieve q^* in the DM after they sell their assets in the OTC (even as the fixed asset supply gets diluted among more C-types). With the dilution effect out of the picture, a higher e_N implies a higher congestion effect in market A but also a larger measure of C-types in that market (i.e., a higher equilibrium e_C). With CRS in matching these two effects completely offset each other, leading to a flat G-function; or, equivalently, a continuum of equilibria with $e_C = e_N$ when asset supplies are large enough.

We now move on to the case of IRS in the matching technology, corresponding to part (d) of Proposition 1. Figure 3 shows the reply function G under $\rho = 0.5$, an intermediate degree of IRS. In this case, a high value of e_N still implies some congestion among N-types, but this effect is dominated by the large measure of C-types drawn to market A (precisely because e_N is high). Does that mean that G will be strictly increasing? Not necessarily. Consider for instance the left panel of the figure, where both asset supplies are small, so that the dilution effect is active. If e_N is large, the typical N-type has a high probability of matching in market A because that market is flooded with C-types (as well as N-types). But each of those C-types is carrying only a tiny fraction of the supply of asset A, which was small to begin with. This force discourages the representative N-type from entering market A, and gives G the non-monotone shape seen in the left panel of Figure 3. More precisely, that picture shows that there are five equilibria: the two corners (which are both robust under small errors now), the robust interior equilibrium, and two non-robust asymmetric equilibria.

What if the supply of asset A was high but that of asset B stayed low? This case is illustrated

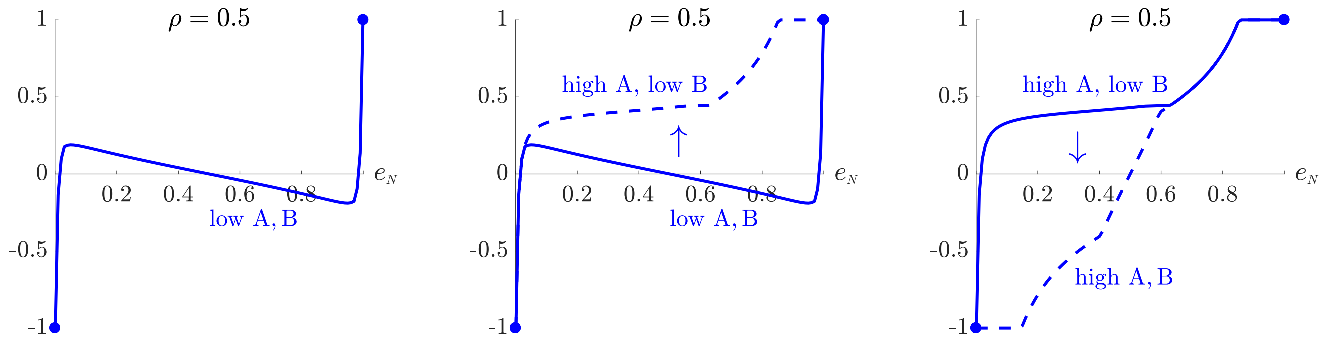


Figure 3: The reply function $G(e_N)$ for IRS ($\rho = 0.5$) and varying asset supplies.

in the middle panel of Figure 3, where one can see that the robust interior equilibrium is now *eliminated*. Unless trade was concentrated in the B-corner in the first place, N-types now have an incentive to migrate to the A-market, C-types will follow, and ultimately all trade will be in the A-corner. Finally, the right panel depicts the case where both A and B are high. In this case, the G-function shifts down (compared to the high-A, low-B case), and incentives to trade in the B-market are restored. However, when both asset supplies are large, the dilution effect vanishes and the G-function becomes increasing throughout, so the corners are the only robust equilibria. There does exist an interior equilibrium by continuity, but if it was ever played, a small shock would drive the agents into one of the corners.

Finally, as part (e) of Proposition 1 shows, the system admits a simple symmetric solution in one special case which we call "balanced CRS": there are CRS in OTC market matching ($\rho = 0$) and neither asset has an exogenous liquidity advantage ($\delta_A = \delta_B$).[19] And as part (f) of the Proposition shows, with a few more technical assumptions we can prove that $G(e_N)$ is downward-sloping in a neighborhood of the symmetric equilibrium (as depicted in the left panel of Figure 2); thus, this particular interior equilibrium is robust.

Beyond the results of Proposition 1, a general analytical characterization is not possible and most of the analysis which follows will be numerical. (The model can also not be simplified without losing essential insights.[20]) Through the rest of Section 3, to keep things simple and

[19] We use the word "balanced" to describe the assumption $\delta_A = \delta_B$. We could also call it "symmetric", but we reserve that word for equilibria where *all* variables indexed by A equal their B-counterparts (e.g., $p_A = p_B$). Even in the balanced environment, there are asymmetric equilibria: the corner equilibria for one, and additional asymmetric interior equilibria if $\rho > 0$, as shown in the left panel of Figure 3.

[20] We have a core system of six equations, and most of the endogenous variables show up in multiple equations. Moreover, the equations are non-linear and include kinks, due to the various branches that characterize the agents' market entry decisions. One may wonder whether some simplifying assumptions would allow us to achieve a complete analytical characterization. We believe that the model presented here constitutes the most parsimonious framework that can capture all the salient features of the question we are studying, hence, any further simplification would eliminate insights that we think are essential. A few examples may clarify this point. A simplifying assumption often adopted in these types of models is that the bargaining power of agents is equal to either 0 or 1. (This is precisely what we assume for the DM, because not many interesting things happen in that market.) Imposing such an assumption in the OTC would be a bad idea: it would imply that either the C-types or the N-types get no surplus from OTC trade, which would render their entry decision indeterminate. As we have explained, the agent's decision about which market to visit is one of the most important economic forces

gain some intuition about the economic forces at work, we perform some comparative statics exercises on the asset supplies and the parameters of OTC microstructure (δ_A, δ_B, ρ), maintaining the parameters $u(q) = \log(q)$, $\ell = 0.5$, $\theta = 0.5$, $i = 0.1$, and $M = 1$.

For given parameters, we guess a starting point for e_N, then iterate the function $G(e_N)$ in the direction of its sign, until convergence or until reaching a corner. Specifically, we use $e_N^0 \equiv \delta_A A/(\delta_A A + \delta_B B)$ as an efficient starting point for iteration; if a robust interior equilibrium exists, it is likely to involve more entry into the market with a higher matching probability, and/or higher trading volume. If the corners are not robust, this procedure will always find an interior equilibrium. On the other hand, a robust interior equilibrium may exist but not be found if a corner is robust and the starting point is close to it.

3.4 Comparative statics

Now that we understand the structure of possible equilibria, we want to compare asset prices in these equilibria, and interpret the comparative statics of prices with respect to quantities as the aggregate demand for these assets. These comparative statics are shown in Figure 4. In all graphs, the supply of asset A is on the horizontal axis and the supply of B is held fixed and indicated by a vertical dashed line. We show three cases: first, the simplest case of balanced CRS ($\rho = 0$ and $\delta_A = \delta_B$); second, giving an exogenous advantage to asset A ($\delta_A > \delta_B$); and third, without an advantage for either asset but with IRS in matching ($\rho > 0$). In all three examples, the graphs in the top row show the *net liquidity premia* of assets A and B, defined as:

$$L_j \equiv (1+i)p_j - 1 = \ell\, \alpha_{Cj} \frac{\theta}{\omega(q_{1j})} \left[u'(q_{1j}) - 1 \right]. \tag{17}$$

The graphs in the bottom row of the figure show the market entry choices e_C and e_N.

Notice first that some standard results are replicated in our model. First, the liquidity premium of an asset is zero if that asset is in very large supply, no matter how liquid the market for that asset is. The reason is that as the asset supply becomes large enough, $q_{1j} \to q^*$, and thus, $u'(q_{1j}) \to 1$. (One should be careful with terms here: the asset does not "lose" its liquidity properties in this case, they only become inframarginal. The asset still contributes to the overall supply of liquidity in the sense that money demand will be lower than it would be if that asset did not exist.) Furthermore, real balances decrease with inflation so the need to liquidate assets in the OTC markets becomes stronger with inflation; if the asset supplies are small enough, the liquidity premium on any liquid asset will rise with inflation, too.

In addition to these standard results, our model also delivers new insights into asset pricing

in our model. As another example, some papers (e.g., Mattesini and Nosal, 2016) gain tractability by assuming that asset trade takes place *only* in OTC markets, and the original asset holdings are given to agents in the CM as endowments, i.e., there is no primary asset market. Clearly, such an assumption here would deprive the model of one of its most important ingredients, the endogenous determination of asset supply.

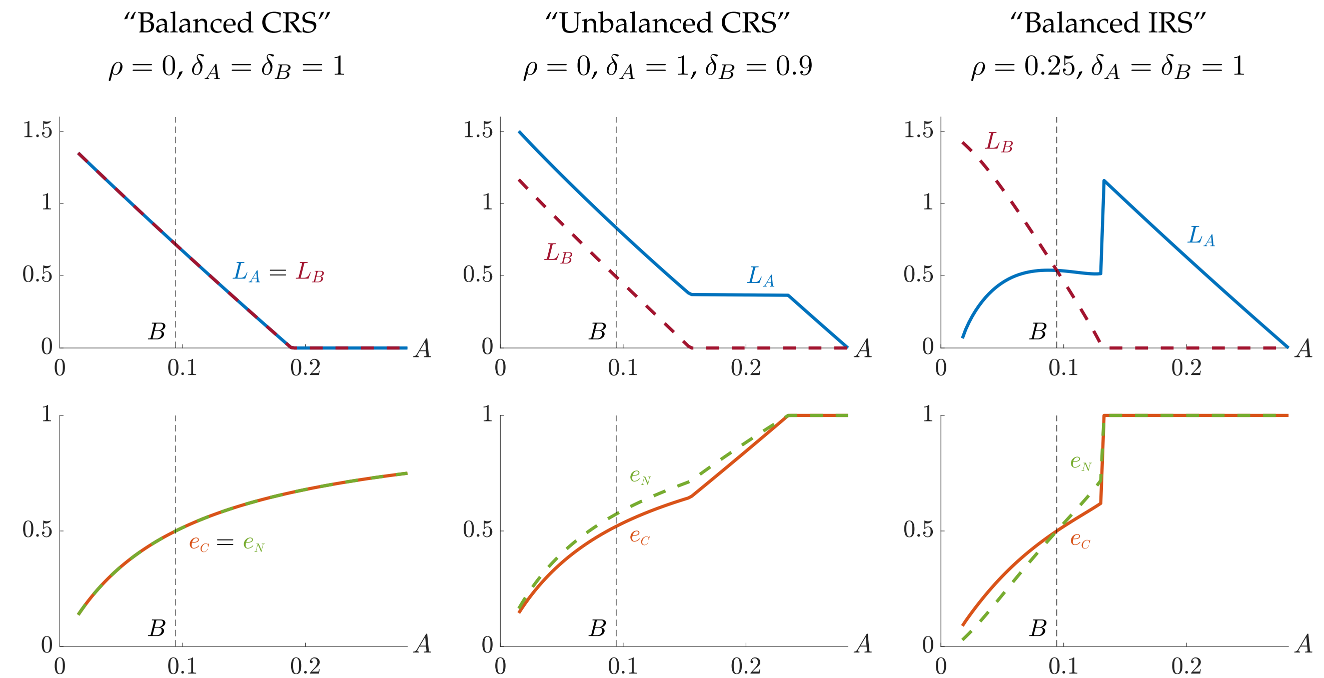


Figure 4: Net liquidity premia L_j (in %) and entry choices, varying A and holding B fixed (indicated by a vertical dashed line).

in this environment of segmented OTC markets. Three results stand out. The first is that when matching in the markets satisfies "balanced CRS" (that is, CRS and neither market having an exogenous liquidity advantage), there exists a unique interior equilibrium when the asset supplies are not too large. In this equilibrium, $e_C = e_N = A/(A+B)$, so the ratio of buyers to sellers is 1 in each market, we have $p_A = p_B$, and all the equilibrium quantities and prices only depend on the *sum* of the asset supplies, $A + B$. Thus, the assets turn out to be perfect substitutes in general equilibrium even though their secondary markets are completely segmented. Part (e) of Proposition 1 shows this formally, and the leftmost column of Figure 4 illustrates it.

The second result from this section is that exogenous liquidity differences are *amplified* by the market entry process, even with CRS. Consider a case where $\delta_A > \delta_B$, so that OTC$_A$ has an exogenous liquidity advantage. As illustrated in the middle column of Figure 4, both e_C and e_N increase, but the latter increases more. To see why, recall that an N-type's money is good in either OTC market, therefore these agents only consider the potential trading surplus in the OTC market when deciding which market to enter. In contrast, C-types can trade only the asset they have previously chosen to hold, therefore they must weigh the cost of holding either asset against the benefit of trading in the respective OTC market. Consequently, the entry decision of N-types is more sensitive to liquidity differences when choosing their market. The end result is that market tightness from the point of view of asset sellers rises in the more liquid market and falls in the less liquid one: formally, we observe that the elasticity of the endogenous ratio α_{CA}/α_{CB} with respect to the exogenous ratio δ_A/δ_B is bigger than 1. Crucially, it is the point of view of OTC asset *sellers* that matters for asset pricing at the issue stage; people who buy a newly issued asset are concerned about the conditions at which they can sell it down the road, but people who plan to buy the asset later in the secondary market have no influence on the issue price. As a consequence, even a small divergence of δ_A and δ_B will drive a larger wedge between the liquidity premia on the two assets. We view this result as Step 1 of an explanation why two assets with otherwise similar features can have big differences in their liquidity – most prominently, of course, U.S. Treasuries compared to equally safe corporate or municipal bonds.

The third result from this section is that IRS in matching encourage market concentration, i.e., corner equilibria. This is illustrated in the rightmost column of Figure 4. Near the origin, we have a case of $A \ll B$, so asset A is barely traded in OTC markets (though not entirely absent due to the fact that ownership of asset B is much more diluted). As the supply of A increases, more agents are willing to trade it in the OTC market because of the increase in potential trading surplus; and crucially, N-types are more sensitive to this increase, so the ratio e_N/e_C rises as A increases. This is important because again, it means that asset A becomes rapidly more attractive to C-types through *two* channels (market tightness and IRS).[21] As asset demand in the CM by future C-types determines the issue price, the resulting increase in liquidity is so strong that it makes the price of asset A upward sloping in its supply – at least, until that supply is so large that the force of diminishing marginal utility takes over.[22] But we are not done. When the supply of A becomes even larger, all OTC trade becomes concentrated in the market for A, and asset B ceases to be liquid at all. As this happens, the price of asset A jumps upward discontinuously; later, we will see that this effect of increasing returns provides a powerful incentive to the issuer of an asset to issue up to the point where competing assets are driven out of secondary markets.

This result is Step 2 of our explanation why two assets with otherwise similar features can have big differences in their liquidity. Even with a modest degree of IRS in matching, an asset in smaller supply is likely to be significantly less liquid than one which is in larger supply, as agents prefer to enter the market where gains from trade are larger, and through their own entry help to make this market "thick". And consider how this would interact with the first step described above: even with a small exogenous difference in market efficiency, the disadvantaged market is likely to see significantly less entry, and thereby becomes very "thin" indeed.[23] In

[21] To be precise: with IRS and $\delta_A = \delta_B$, we observe $e_C < e_N$ in the interior if and only if $A < B$. The more plentiful asset is more liquid.

[22] Weill (2008) has a result of similar flavor: he studies an extension of Duffie et al. (2005) with multiple assets, keeping the aggregate supply of tradable assets constant but allowing some assets to be in larger supply than others. He finds that the more plentiful assets are easier to find and have a higher price.

[23] Interpreting market A as the market for U.S. Treasuries, there is an additional element that may add to this market's liquidity: the Federal Reserve (Fed) often participates in this market by selling or buying large quantities of assets. For instance, in the period between November 2008 and September 2011, the Fed purchased $1.19 Trillion of Treasury debt, as part of a program now known as quantitative easing (QE). While our paper does not explicitly model interventions of the Fed in the financial markets, in the form of open market operations or QE, it is reasonable to expect that the presence of a big player such as the Fed in that market will be a pole

the next section, we calibrate our model to quantitatively assess the power of this amplification mechanism.

4 Quantitative analysis

4.1 Calibration

We calibrate the model to the Treasury bonds and AAA/AA corporate bonds markets and test whether the calibrated model can quantitatively match the observed spread between Treasuries and (the weighted average of) AAA/AA corporate bond yields.[24] More precisely, we feed the model with data on Treasury and corporate bond supplies, and examine the exogenous liquidity differential in OTC market efficiency that would allow the model to perfectly match the data.

For the utility function, we use $u(q) = q^{1-\sigma}/(1-\sigma)$. Thus, we need to calibrate ten parameters: the supplies of the Treasury bonds, the AAA/AA corporate bonds, and the money stock (A, B, and M), the nominal interest rate on an illiquid bond ($i = (1+\mu-\beta)/\beta$, which subsumes time preference and expected inflation), the elasticity of marginal utility (σ), the fraction of C-type agents (ℓ), the relative bargaining power of C-types (θ), the degree of returns to scale in the OTC matching function (ρ), and the matching efficiency in the two OTC markets (δ_A and δ_B).

Some of these parameters have straightforward empirical targets, while others do not. For the latter, we show what combinations of parameters are required in order to exactly match the yield spreads observed in the data.

First, for the Treasury bond supply (A) and the AAA/AA corporate bond supply (B), we use the data from the report by S&P Global on the U.S. corporate debt market as of January 2019 (S&P Global, 2019); this is the only data on the outstanding amount of corporate bonds we could find. For M, we use the MZM monetary aggregate (from FRED). We divide the Treasury and corporate bond supplies by the money stock to obtain 0.1390 of Treasury supply and 0.0313 of corporate bond supply, relative to a normalized $M = 1$. For i, we cannot use any observed interest rate since no traded asset is perfectly illiquid; instead, we use an estimate of 7%/year based on time preference, expected real growth, and expected inflation (Herrenbrueck, 2019b). We set $\ell = 0.5$ for symmetry (equal numbers of potential buyers and sellers in the secondary

of attraction for other investors, too: if I want to sell assets in the secondary market (like the C-types in our model) and I know that someone is purchasing billions worth of asset A in OTC$_A$, why would I go anywhere else? Readers who are interested in how one could model direct interventions of the Fed in financial markets in a similar framework are referred to Herrenbrueck (2019a) and Geromichalos and Herrenbrueck (2022). For a careful empirical characterization of the effects of QE, see Song and Zhu (2018).

[24] One could argue that only AAA corporate bonds are default-free and, therefore, comparable to Treasuries with respect to their risk component. However, the supply of AAA corporate bonds has decreased so dramatically in the last decade that any reasonable calibration including only those assets in the role of "asset B" would imply that no agent ever chooses to visit OTC$_B$. To avoid this trivial result, we choose to interpret asset B as the class of "very safe" corporate bonds, and this certainly includes AA as well.

	Description	Value
A	supply of Treasury bonds	0.1390
B	supply of AAA/AA corporate bonds	0.0313
M	money supply	1
i	nominal interest rate on an illiquid bond (yearly)	7%
ℓ	fraction of C-type agents	0.5
σ	elasticity of marginal utility	0.34
θ	relative bargaining power of C-types	See Table 2
ρ	elasticity of the OTC matching function	See Table 2
δ_A	matching efficiency in the OTC market for Treasuries	1
δ_B	matching efficiency in the OTC market for corporate bonds	See Table 2

Table 1: Key parameter values.

asset markets). Next, we follow the procedure of Rocheteau, Wright, and Zhang (2018) (where to be consistent with the rest of our calibration, we use MZM as the monetary aggregate) and set $\sigma = 0.34$ to match the slope of the empirical U.S. money demand function. For the matching efficiency in the OTC market for Treasuries (δ_A), we normalize it to 1, so that we can interpret values of δ_B smaller than 1 as the exogenous liquidity disadvantage for corporate bonds. A summary of the calibrated parameter values can be found in Table 1.

This leaves us with three parameters – the bargaining power (θ), the scale elasticity (ρ), and the matching efficiency in the OTC market for corporate bonds (δ_B) – which have no direct counterparts in the data. We consider a number of combinations of θ, ρ, and for each combination, we ask what δ_B should be equal to so that our model can perfectly match the yield differentials observed in the data. For ρ, as we will see soon, some amount of IRS will lower the burden on other parameters for the model to match the data, but ρ does not have to be particularly high. Indeed, if ρ is too high, then one of the assets will attract all secondary asset market trade, and the other one will be completely illiquid, a result which would be counterfactual. Thus, all the values of ρ we consider are in the *neighborhood* of the CRS case (between 0 and 0.03).

For the target, the yield differential between Treasuries and AAA/AA corporate bonds, we restrict attention to the period from 2018 to 2020.[25] We calculate this target as the difference between the average of the market yields on U.S. Treasury securities at 20- and 30-year constant maturities from FRED (GS20 and GS30) and the weight average of AAA and AA corporate bond yields from the ICE BofA AAA and AA US Corporate Index (C0A1 and C0A2).

[25] While the data on Treasury bond supply is available for every year, we only have data for the corporate bond supply in 2019. Since this data limitation forces us to focus on a period around 2019, we choose to calculate the yield differential for the period 2018 to 2020. However, our results are robust to expanding the period over which this yield differential is calculated.

	$\rho = 0$	$\rho = 0.01$	$\rho = 0.02$	$\rho = 0.03$
$\theta = 0.1$	—	—	—	—
$\theta = 0.2$	0.5251	0.5367	0.5485	0.5606
$\theta = 0.3$	0.7122	0.7254	0.7388	0.7525
$\theta = 0.4$	0.7995	0.8133	0.8274	0.8417
$\theta = 0.5$	0.8507	0.8649	0.8793	0.8939
$\theta = 0.6$	0.8848	0.8992	0.9138	0.9287
$\theta = 0.7$	0.9094	0.9239	0.9387	0.9537
$\theta = 0.8$	0.9282	0.9428	0.9577	0.9728
$\theta = 0.9$	0.9434	0.9581	0.9731	0.9883

Table 2: Values for δ_B, given θ and ρ, in an exact match of our model to the data.

Table 2 presents the results. Each cell of the table shows the value of δ_B that is needed to exactly match the observed spread in the data, given the values of θ and ρ. For a given θ, a higher ρ implies a higher δ_B; that is, with a higher degree of IRS, a smaller amount of exogenous liquidity disadvantage is needed to match the data. Similarly, for a given ρ, a higher θ implies a higher δ_B; that is, with a higher bargaining power for asset sellers, a smaller amount of exogenous liquidity disadvantage is needed to match the data. However, notice that if θ is very small, e.g., 0.1, there is no value of δ_B that would allow the model to match the data. This is because in our model agents are willing to pay a (high) liquidity premium, if they expect to sell the asset "down the road". The price at which they can sell depends positively on θ, thus, the channels highlighted by our model are stronger for greater values of that parameter.[26] Of course, we exclude $\theta = 1$ so that asset buyers have a meaningful market entry decision.

Overall, the results show that our model needs minimal degrees of scale elasticity and a liquidity disadvantage for corporate bonds to match the data. For example, when the bargaining power is 0.8, a model with the scale elasticity of 0 (which corresponds the constant-returns-to-scale case) implies 0.93 for the matching efficiency in the corporate bonds secondary market. Thus, even if we completely shut down the IRS channel, our model requires the matching technology in the corporate bonds market to be just 7% less efficient than the one in the Treasury market to perfectly match the data. As expected, if we allow for a positive value of the scale elasticity, ρ, the exogenous liquidity differentials that are required to match the data decrease even further. For instance, if $\rho = 0.01$, the implied value for the matching efficiency for the corporate bonds secondary market becomes 0.94. For a scale elasticity of 0.02, that number becomes 0.96. Thus, our model does not need an unreasonably high value of scale elasticity to explain the data; in fact, it is virtually in the neighborhood of CRS.[27]

[26] Importantly, a high value of θ does not give one of the assets a relative advantage, since a higher θ scales liquidity premia of both assets proportionally (see Equation 17).

[27] For context, most theoretical finance papers use a congestion-free matching function with scale elasticity

4.2 Evaluating the contribution of each channel of the model

The discussion so far reveals that there are three channels delivering the main results of the analysis: endogenous entry in the OTC markets, exogenous liquidity differences (δ_B), and increasing returns to scale (ρ). The goal of this section is to quantitatively determine the individual contribution of each of those channels. To that end, we build on the calibration from Section 4.1, and we ask how much explanatory power we would lose if we were to shut down each of those channels one at a time. The results of this exercise are illustrated in Tables 3, 4, and 5. All of those tables remind the reader what δ_B should be, for any given combination of (θ, ρ), for our model to perfectly match the data.

Table 3 reports how much explanatory power our model loses, for the various parameter specifications, if instead of allowing agents to enter OTC markets optimally, we assumed that the measure of agents entering each market is held in constant proportion to the supply of the respective asset. Specifically, we assume $A/e_C = B/(1 - e_C)$ and $A/e_N = B/(1 - e_N)$, so that $e_C = e_N = A/(A + B)$. To illustrate the results of the exercise, suppose that $\theta = 0.7$ and $\rho = 0.02$ (so that with $\delta_B = 0.9387$ our model perfectly matches the data). If one was to shut down the endogenous entry decision of the agents, our model would not able to explain 48% of the yield differential in the data, which is to say that this channel is responsible for 48% of the model's explanatory power (for the given set of parameters). Note that the importance of this channel is quite stable across the different combinations of (θ, ρ, δ_B) that fit the data and roughly equal to 50%.

Table 4 illustrates the importance of exogenous liquidity differences. The reported numbers illustrate the loss of explanatory power if one assumes that δ_B equals 1, as opposed to the value that would perfectly match the data. As an example, consider again $\theta = 0.7$ and $\rho = 0.02$. Our result suggests that eliminating exogenous liquidity differences (i.e., assuming $\delta_B = 1$ rather than 0.9387) would result in losing 66% of the model's explanatory power. Notice that with $\rho = 0$ (and $\delta_B = 1$) we are in the "balanced CRS" case, where our theory predicts that both assets should be priced equally, in other words, our model cannot explain any of the yield differentials in the data. Also, note that for a given θ, increasing ρ results in a smaller loss of explanatory power. This is because, as discussed Section 4.1, higher exogenous liquidity differences and higher scale elasticity work as substitutes.

Table 5 highlights the contribution of the IRS channel of our analysis. The percentages reported in this table represent the loss of explanatory power, if one assumes that ρ equals 0 (CRS), as opposed to the value that would perfectly match the data. For instance, suppose that $\theta = 0.7$. The analysis in Section 4.1 reveals that with $\rho = 0.02$ and $\delta_B = 0.9387$ our model perfectly fits the data. The table shows that removing IRS, i.e., setting $\rho = 0$, while leaving the other parameters unchanged, would imply a 31% loss of the model's ability to match the data. Naturally,

$\rho = 1$; see for example Duffie et al. (2005) and Vayanos and Wang (2007).

	$\rho = 0$	$\rho = 0.01$	$\rho = 0.02$	$\rho = 0.03$
$\theta = 0.3$	$\delta_B = 0.7122$ 42%	$\delta_B = 0.7254$ 43%	$\delta_B = 0.7388$ 44%	$\delta_B = 0.7525$ 44%
$\theta = 0.5$	$\delta_B = 0.8507$ 44%	$\delta_B = 0.8649$ 45%	$\delta_B = 0.8793$ 46%	$\delta_B = 0.8939$ 46%
$\theta = 0.7$	$\delta_B = 0.9094$ 47%	$\delta_B = 0.9239$ 48%	$\delta_B = 0.9387$ 48%	$\delta_B = 0.9537$ 49%
$\theta = 0.9$	$\delta_B = 0.9434$ 52%	$\delta_B = 0.9581$ 53%	$\delta_B = 0.9731$ 53%	$\delta_B = 0.9883$ 54%

Table 3: Percentage unexplainable without endogenous market participation ($e_C = e_N = A/(A+B)$ versus optimal e_C, e_N).

	$\rho = 0$	$\rho = 0.01$	$\rho = 0.02$	$\rho = 0.03$
$\theta = 0.3$	$\delta_B = 0.7122$ 100%	$\delta_B = 0.7254$ 94%	$\delta_B = 0.7388$ 89%	$\delta_B = 0.7525$ 83%
$\theta = 0.5$	$\delta_B = 0.8507$ 100%	$\delta_B = 0.8649$ 90%	$\delta_B = 0.8793$ 79%	$\delta_B = 0.8939$ 69%
$\theta = 0.7$	$\delta_B = 0.9094$ 100%	$\delta_B = 0.9239$ 83%	$\delta_B = 0.9387$ 66%	$\delta_B = 0.9537$ 49%
$\theta = 0.9$	$\delta_B = 0.9434$ 100%	$\delta_B = 0.9581$ 73%	$\delta_B = 0.9731$ 46%	$\delta_B = 0.9883$ 20%

Table 4: Percentage unexplainable without exogenous liquidity difference ($\delta_B = 1$ versus δ_B as shown in the table).

	$\rho = 0$	$\rho = 0.01$	$\rho = 0.02$	$\rho = 0.03$
$\theta = 0.3$	$\delta_B = 0.7122$ 0%	$\delta_B = 0.7254$ 4%	$\delta_B = 0.7388$ 9%	$\delta_B = 0.7525$ 13%
$\theta = 0.5$	$\delta_B = 0.8507$ 0%	$\delta_B = 0.8649$ 9%	$\delta_B = 0.8793$ 18%	$\delta_B = 0.8939$ 28%
$\theta = 0.7$	$\delta_B = 0.9094$ 0%	$\delta_B = 0.9239$ 15%	$\delta_B = 0.9387$ 31%	$\delta_B = 0.9537$ 48%
$\theta = 0.9$	$\delta_B = 0.9434$ 0%	$\delta_B = 0.9581$ 25%	$\delta_B = 0.9731$ 51%	$\delta_B = 0.9883$ 78%

Table 5: Percentage unexplainable without IRS ($\rho = 0$ versus ρ as shown in the table).

all the numbers in the first column of the table are 0's since that calibration was based on the assumption that $\rho = 0$. Another thing to note is that in the first row of the table (for $\theta = 0.3$), shutting down the IRS channel does not take away too much explanatory power. This is because for such a low value of θ, most of the model's explanatory power stems from the low values of δ_B, which in our current exercise remain unchanged.

4.3 Implications of consolidating secondary markets

The secondary marketplace for corporate bonds is often believed to be quite segmented. An interesting report by BlackRock (2014) provides evidence in support of this belief, and suggests that consolidating the secondary asset markets for different classes of corporate bonds would be beneficial for the issuers. Our model confirms this conjecture, because it predicts that consolidating the segmented secondary markets will increase secondary market liquidity, and, thus, lower the rate at which these corporations can borrow funds. More importantly, our model can be used to quantitatively assess the size of these benefits. To that end, we use the calibrated model in Section 4.1, but we adjust it to include three assets, asset A, asset B_1, and asset B_2, interpreting asset B_1 as AAA corporate bonds and asset B_2 as AA corporate bonds. (Per usual, asset A is interpreted as Treasury bonds). We then perform the following counterfactual exercise: we ask what the gains in terms of higher liquidity premia (or, equivalently, lower borrowing rates) would be from consolidating the secondary asset markets for AAA and AA corporate bonds. The model with three assets and three secondary markets is described in detail in Appendix B.2.

The main results of our quantitative exercise are summarized in Table 6.[28] We can see that, based on our calibration, before the secondary market consolidation AAA corporate bonds (asset B_1) do not enjoy any liquidity premium. This is because their asset supply is so small, and the benefit of trading that asset in a segmented market is so limited, that no agents choose to visit OTC_{B_1}. AA corporate bonds (asset B_2) enjoy a liquidity premium, but, of course, that premium is lower than the one for Treasury bonds (asset A). Consolidating OTC_{B_1} and OTC_{B_2} implies that both of these assets will now enjoy the same, and higher liquidity premium, equal to 1.4014%. However, the lion's share of the benefit obtained by the market consolidation is enjoyed by the issuers of AAA corporate bonds, who suffered disproportionately from the market segmentation, due to the extremely low supply of this asset (see Footnote 28). Overall, the

[28] Three clarifications are in order regarding this quantitative exercise. First, in the baseline exercise we saw that the supply of asset B was 0.0313, and that included both AAA and AA corporate bonds. Here we break this into the supply of asset $B_1 = 0.0051$ (AAA) and $B_2 = 0.0262$ (AA). Note that the supply of AA bonds is roughly five times greater than the supply of AAA. Second, even though we know from section 4.1 that our model can match the data for various values of θ and ρ, for this exercise we focus on $\theta = 0.7$ and $\rho = 0.03$. Regarding our choice of θ, we have already argued that our model performs better for higher values of that parameter. Regarding the choice of ρ, we choose a relatively high value because our exercise, i.e., evaluate the gains from consolidating the secondary asset markets, is more meaningful in the presence of some IRS. Third, even when the secondary markets for assets B_1 and B_2 are segmented, the matching efficiency in both markets is common, and given by δ_B.

Parameters	$\rho = 0.03$	$\theta = 0.7$	$\delta_B = 0.9537$
Liquidity premia,	Asset A	Asset B_1	Asset B_2
before consolidation	1.7576%	0%	1.4001%
after consolidation	1.7414%	1.4014%	1.4014%
Changes in liquidity premia	-0.0162%	1.4014%	0.0013%

Table 6: Impact of consolidating secondary asset markets on liquidity premia.

consolidation of the corporate bonds secondary markets attracts more agents into that market, leading to an endogenous decrease in the liquidity (premium) of Treasury bonds.

5 Microfoundations: market segmentation

One of the crucial assumptions of our model is that the secondary markets for assets A and B are segmented, and that agents can visit only one market per period. The first assumption is certainly empirically relevant, as we discuss in Appendix A. The second assumption is not meant to be taken literally, and only intends to capture the idea that trading a particular type of assets is costly, and agents will visit more frequently the secondary market where they expect to find the best trading terms. While we choose to maintain this assumption in the baseline model for tractability purposes, in this section we show that the limited participation assumption can arise endogenously as a result of a more general model, where agents have the option to trade both assets in the secondary market.

To that end, we consider an environment with two distinct secondary "marketplaces".[29] In one of the marketplaces, there is a consolidated secondary market where agents can trade both assets, A and B. Agents can choose to pay a fee κ_2 and access this market. In the other marketplace, there are two segmented sub-markets, just like in the original version of our paper. Agents pay a fee κ_1 to access this marketplace, and, once they enter, they must choose which of the two sub-markets to go to (either the one for asset A or the one for asset B). Within any market or submarket, matching takes place according to the usual matching function described in the original version of the paper. This version of the model captures the idea that if agents want to, they should be able to avoid the limited participation imposed in the baseline model. (This exercise is similar in spirit to Geromichalos et al. (2018).)

The details and solution of the model can be found in Appendix B.3, and the findings are presented in Figure 5. Our analysis implies that there exist two robust corner equilibria; one where all agents concentrate in the consolidated marketplace, and one where all agents con-

[29] In this extended model, there is still only a single primary asset market (the CM) where all agents and assets trade. In particular, this means that the allocation of money or assets to either of the two secondary "marketplaces" is completely unrestricted.

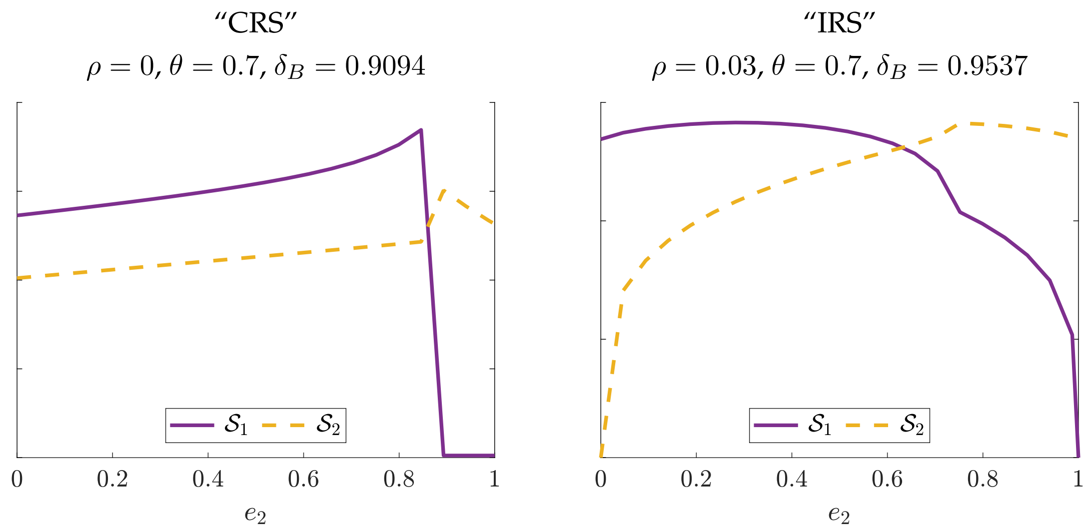


Figure 5: The typical agent's surplus from visiting the segmented marketplace (S_1, defined in (B.40)) and the consolidated marketplace (S_2, defined in (B.41)) as a function of the fraction of agents visiting the consolidated marketplace (e_2).

centrate in the segmented marketplace. Of course, the latter coincides with the equilibrium described in our baseline model, where segmentation was imposed as an assumption. One should notice that this analysis has abstracted from the entry fees κ_1, κ_2. In other words, we have established that even if $\kappa_1 = \kappa_2$, having all agents concentrate in the segmented marketplace is a robust equilibrium of the model. However, if we interpret the κ terms as costs of evaluating assets traded in the various secondary markets, it is reasonable to assume that κ_2 should be larger than κ_1. If that was the case, it is theoretically possible that concentrating in the segmented marketplace becomes the *unique* robust equilibrium of the model.

6 The economy with strategically chosen asset supply

So far the analysis has assumed that asset supplies are exogenously given. The goal of this section is to highlight that allowing for an endogenous determination of asset supplies offers important economic insights that are complementary to the analysis with exogenous asset supplies. To that end, we study the non-cooperative duopoly game between issuers of asset A and B, who realize that the OTC market microstructure (δ_A, δ_B, ρ) and the entry decisions of agents (i.e., the equilibrium of the 'subgame' described in Section 3.1) determine the demand for their assets. In Appendix A, we provide some justification for treating the issuers of these two assets as strategic players, whose decisions can affect market outcomes. However, this particular market structure is clearly not meant to be taken literally; in the real world there are multiple asset issuers, and each of them has a different degree of market power. Rather than developing an intractable model that attempts to incorporate all these details, we present the solution to a simpler model, which allows the reader to extrapolate what kind of outcomes one might obtain under their preferred market structure.[30]

6.1 The game between the asset issuers

We look at the non-cooperative game between two issuers who seek to maximize their utility. They live only in the CM, where they can work, consume, and issue assets. Their utility within the period is $\mathcal{Y}(X, H) = X - H$, where X, H denote consumption and work effort, and they discount the future by the same factor β as all agents. They take into account that the real price at which they can sell their asset, φp_j, depends on the supplies of both assets. For example, the problem of issuer A who has issued A^- assets in the previous period can be described by the following Bellman equation:

$$W^A(A^-) = \max_{X,H,A} \left\{X - H + \beta W^A(A)\right\}$$
$$\text{s.t. } X + \varphi A^- = H + \varphi p_A A,$$

which we can simplify to yield:

$$W^A(A^-) = -\varphi A^- + \max_{A} \left\{\varphi p_A A + \beta W^A(A)\right\}.$$

Just like for private agents, the issuer's choice of A does not depend on their previous choices. We can use this, plus the fact that in steady state $\varphi/\hat{\varphi} = (1+\mu)$, to solve for issuer A's objective:

$$\begin{aligned} J^A &= \frac{\varphi}{1+i}\left[(1+i)p_A - 1\right] A \\ &= \frac{\varphi}{1+i}\left(\ell\, \alpha_{CA} \frac{\theta}{\omega(q_{1A})}\left[u'(q_{1A}) - 1\right]\right) A. \end{aligned} \tag{18}$$

With an analogous derivation, issuer B's objective is:

$$J^B = \frac{\varphi}{1+i}\left(\ell\, \alpha_{CB} \frac{\theta}{\omega(q_{1B})}\left[u'(q_{1B}) - 1\right]\right) B. \tag{19}$$

Simply put, each issuer seeks to maximize the product of the net liquidity premium L_j and the supply of their asset, taking into account that their choice of asset supply affects the general equilibrium choices of the agents.

[30] As one example, one may argue that political agents for whom liquidity rent is not the only consideration, such as the U.S. Treasury, are not Nash players but are able to precommit. We do explore this possibility in the Web Appendix: first, a "semi-strategic" case where the supply of A is set non-strategically, and issuer B best-responds to it; second, a Stackelberg duopoly where issuer A moves first and precommits to a (typically, large) issue size before B best-responds. And if we take the repeated interaction between the issuers seriously, there are even more possibilities, but they go beyond the scope of our paper.

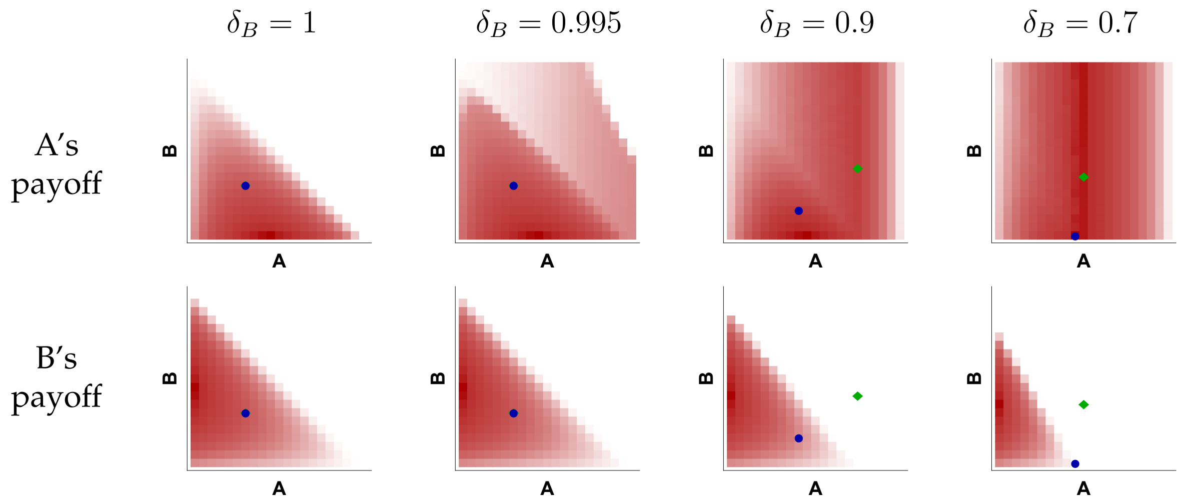


Figure 6: Payoffs as functions of asset supplies, with CRS ($\rho = 0$) and asset A having an exogenous liquidity advantage over asset B ($\delta_B \leq \delta_A = 1$). Darker-shade cells indicate larger payoffs, and white indicates zero. The circle and diamond points indicate particular Nash equilibria.

6.2 Strategic structure of the game

As before, we begin with the simplest case: "balanced CRS" in financial markets ($\rho = 0$ and $\delta_A = \delta_B$). As we saw in Proposition 1 above, the two corner equilibria are not robust to small errors; consequently, the interior equilibrium defined in part (e) of the proposition is the interesting one to study here. In this equilibrium, liquidity premia are positive, equal ($L_A = L_B > 0$), and depend only on the sum $A+B$: the assets are perfect substitutes and are priced along a common demand curve. And because the assets are perfect substitutes, the only Nash equilibrium of the game between the issuers is the symmetric Cournot equilibrium where both assets are issued in the same quantity, each approximately one-third of the quantity $\bar{D}$ that would drive the liquidity premium to zero.

Next, we are interested in the effects of exogenous liquidity differences. Specifically, we set δ_A equal to 1 and let δ_B vary, while maintaining CRS. For this discussion, we set $\theta = 0.7$; the remaining parameters are equal to the ones discussed in the calibration of the model (Section 4.1). Figure 6 illustrates the results: the leftmost column shows the balanced CRS case, and the rightmost column confirms that if B has too much of a disadvantage, the interior equilibrium ceases to exist and all OTC trade is in the A-market. Issuer A gets to issue the monopoly quantity, approximately one-half of $\bar{D}$, and issuer B issues an arbitrary amount because asset B is illiquid in any case.

The intermediate values of δ_B, where the B-market is only a little bit worse than the A-market, show the transition. As we had already seen in Figure 4 (middle column), the demand curve for asset A has a kink whenever δ_B is less than δ_A. As long as δ_B is close enough, the Cournot-style interior equilibrium survives. When δ_B becomes too small, however, A prefers

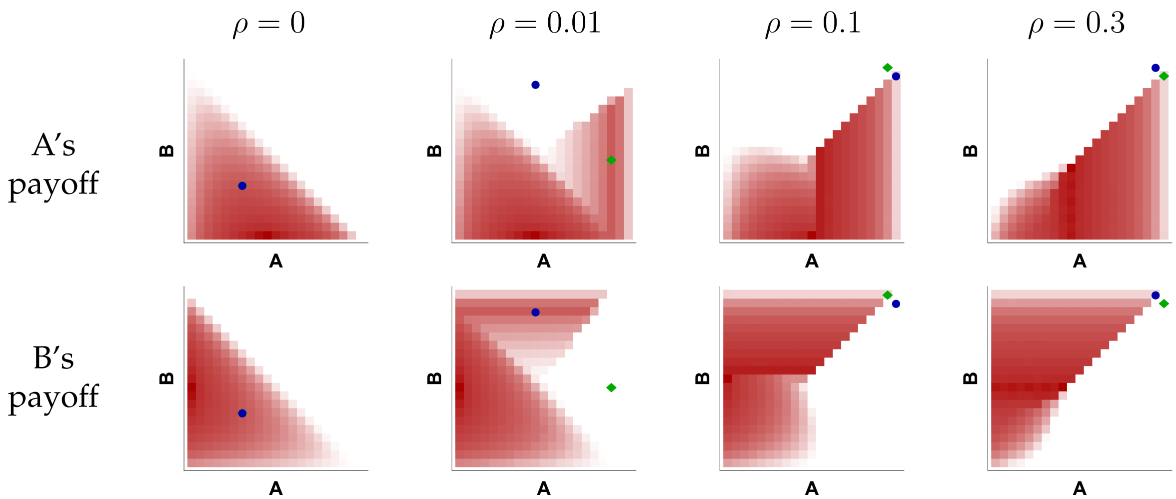


Figure 7: Payoffs as functions of asset supplies, with IRS but no exogenous liquidity advantage ($\delta_A = \delta_B = 1$). Darker-shade cells indicate larger payoffs, and white indicates zero. The circle and diamond points indicate particular Nash equilibria.

to jump to a very large quantity that concentrates OTC trade in the A-market and drives B's liquidity premium to zero – even at the expense of a low liquidity premium for A itself. We will analyze the consequences for the economy in more detail later, but we can already see that the total supply of liquid assets is largest if B is somewhat illiquid, smallest if B is very illiquid, and in between if both A and B are very liquid.

To summarize: with CRS in financial markets, the structure of the game resembles Cournot competition. If not too unbalanced, CRS promotes the interior equilibrium in OTC markets where every asset is somewhat liquid. As a result, we see relatively small issue sizes.

Finally, we look at how the issuers' incentives are affected by IRS in financial markets. Specifically, we set $\delta_A = \delta_B = 1$ and let ρ vary. These results are illustrated in Figure 7; the leftmost column repeats the balanced CRS case from the previous figure, and the rightmost column illustrates how a strong degree of IRS makes the symmetric interior entry equilibrium so unstable that it is never reached as the subgame of the issuers' game. Why? Let us go back to Figure 3. Suppose that asset supplies are small and the interior entry equilibrium is played – i.e., both OTC markets are active. Issuer A has a strong incentive to supply more: yes, this moves her down her own demand curve (reducing her profits), but at the same time, the bigger surpluses in the A-market attract so many traders that the B-market shuts down (increasing A's profits). Of course, B has a symmetric incentive. With IRS, traders prefer to concentrate in one market, so the reward to issuers of offering a bigger trading surplus than their competitor becomes enormous.[31] Consequently, there is a Nash equilibrium (of the discretized model) where quantity A is so close to $\bar{D}$ that issuer B does not find it profitable to issue any more, because

[31] Recall: when computing corner equilibria, we made the tie-breaking assumption that traders are more likely to pick the corner of the asset of which there is a larger supply. See the discussion at the end of Section 3.3.

their asset would trade at a zero liquidity premium, either due to being illiquid or due to being plentiful. And there is a mirror Nash equilibrium with A and B's roles reversed, which is what the two dots in the top right corner of Figure 7 indicate.

For intermediate values of ρ, we see a smooth transformation of the playing field. For low ρ, assets tend to be strategic substitutes where issuers prefer to issue neither too little nor too much, but for high ρ, assets become strategic complements where issuers strongly prefer to issue *more* than the other. Crucially, the fact that the playing field changes smoothly does not mean that the Nash equilibria change smoothly. On the contrary: as ρ increases, we get a jump transition from Cournot-type equilibria of low issue sizes and both assets being liquid to asymmetric equilibria of high issue sizes and only one asset being liquid. Note that the critical amount of IRS is approximately $\rho = 0.01$ – not particularly large – because it is the competition between issuers that gives a small amount of IRS a big endogenous 'kick'.

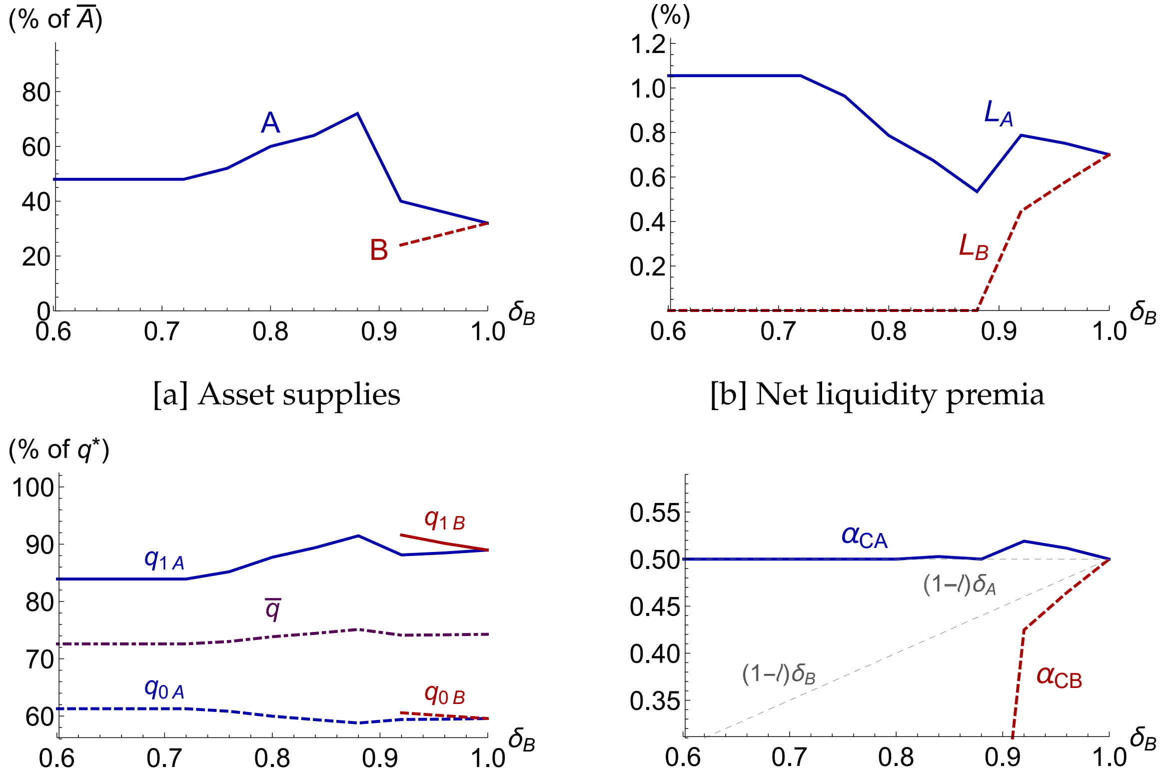
We can say that with enough IRS in financial markets, despite being a game in quantities rather than prices, the strategic structure of the game *resembles* Bertrand competition rather than Cournot. This promotes corner equilibria, where one asset ends up being very liquid and the other one not liquid at all. As long as there are no exogenous differences in market quality ($\delta_A \approx \delta_B$), in such equilibria the 'winning', liquid asset must be in large supply (close to $\bar{D}$).

6.3 Comparative statics

In this section, we analyze the comparative statics with respect to δ_B of the static Nash equilibria of the issuers' game. We consider both constant returns in matching and small amounts of increasing returns. Our goal is to understand what happens if one of the assets (A for concreteness) has an exogenous matching advantage, and how the answer to this question interacts with the returns to scale in matching. Throughout this section, we hold $\delta_A = 1$ fixed.[32]

The case of CRS is illustrated in Figure 8. As δ_B declines slightly from 1 (the balanced case), A begins to issue more and B begins to issue less (panel [a]), but the strategic pattern of a Cournot game is maintained. The exogenous liquidity advantage of asset A is magnified by the entry choices of agents (panel [d]), which feeds back into a rising liquidity premium on asset A and a falling liquidity premium on asset B (panel [b]). Outputs diverge: C-types who hold asset A end up purchasing smaller quantities q_{0A} and q_{1A}, but the probability that they will obtain the larger one of the two, q_{1A}, increases. Conversely, C-types who still hold asset B despite its liquidity disadvantage are compensated with higher quantities q_{0B} and q_{1B} (panel [c]).

[32] One detail to be aware of is how we compute the Nash equilibria. We iterate best responses of the two issuers on a finite grid of possible asset supplies which excludes asset supplies which we know can never give positive payoffs: zero and supplies exceeding $\bar{D}$. The starting point is the smallest positive asset supply on the grid (e.g., the point $(0.05\bar{D}, 0.05\bar{D})$ on a 20×20-grid). The remaining choice is whether we let A or B move first. In this section, all equilibria are computed with A moving first; the equilibria where B moves first are usually identical, payoff-identical, or mirror images. In Figures 6 and 7, Nash Equilibria where A moves first are indicated with a circle point, and those where B moves first are indicated with a diamond point.



[a] Asset supplies

[b] Net liquidity premia

[c] Quantities traded in DM, and average

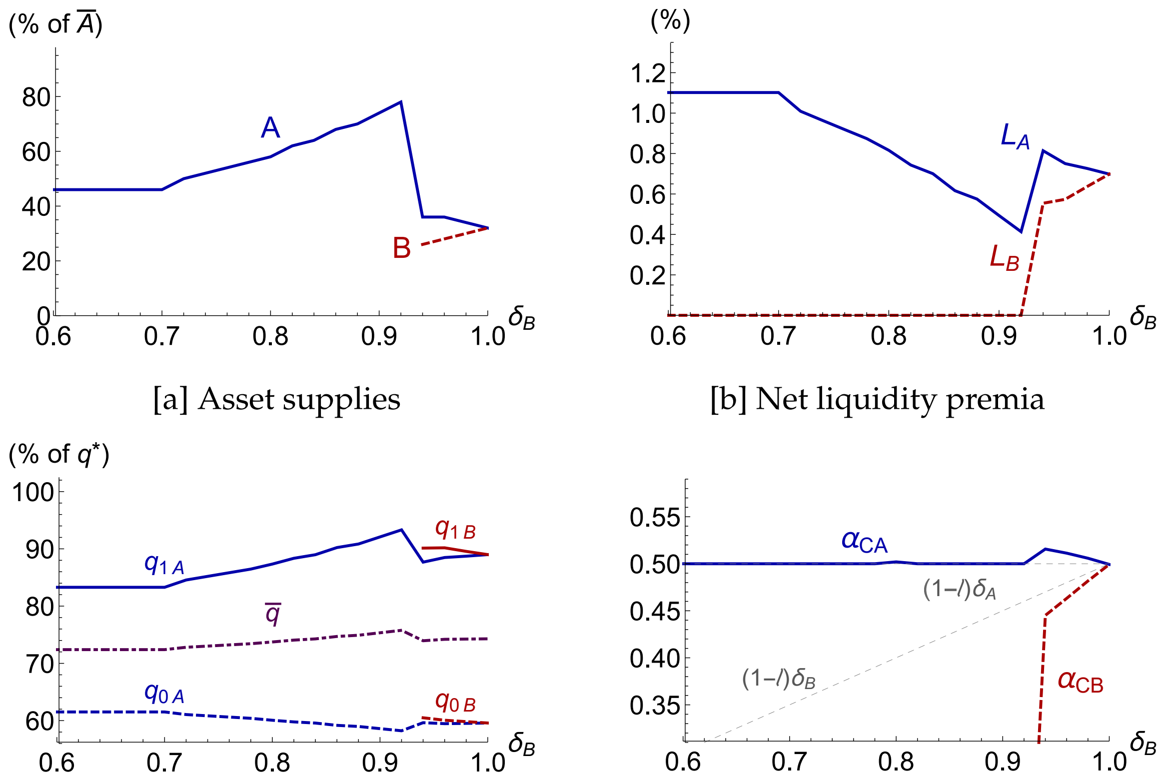
[d] OTC matching probabilities

Figure 8: Comparative statics of the strategic equilibria with respect to δ_B, with CRS ($\rho = 0$).

As δ_B declines further, we observe a discontinuity. At some point, the benefit to A from ramping up the issue size all the way to drive out B from the financial markets becomes too strong, so this is what A does. Asset B becomes fully illiquid, and therefore its issue size and the quantities q_{0B} and q_{1B} become indeterminate. As a result of this aggressive competition, average output of DM goods is highest at the discontinuity. If δ_B declines even more, the threat of trading asset B gradually diminishes; eventually, A becomes a monopolist who issues an intermediate quantity of asset A and average output declines to its lowest value. (Welfare is a more complicated story, as we explain in Section 6.4 below.)

When we allow for a very small degree of IRS in matching, $\rho = 0.002$ (illustrated in Figure 9), the results are almost identical to those with CRS, as one might expect given that ρ is so close to zero. Even so, we can see that the transition from the interior equilibrium to the A-corner where asset B is illiquid happens 'sooner', i.e., for a higher value of δ_B, than under CRS. Increasing returns make it slightly easier for A to drive B out of the market: in the example, A will do so for $\delta_B = 0.92$ under $\rho = 0.002$ but not under $\rho = 0$.

Based on Figure 7, one would guess that when increasing returns are strong enough, the Cournot-style equilibrium is eliminated in favor of aggressive competition for secondary market liquidity. But *how* strong do they need to be? Our perhaps surprising answer is: not very.



[a] Asset supplies

[b] Net liquidity premia

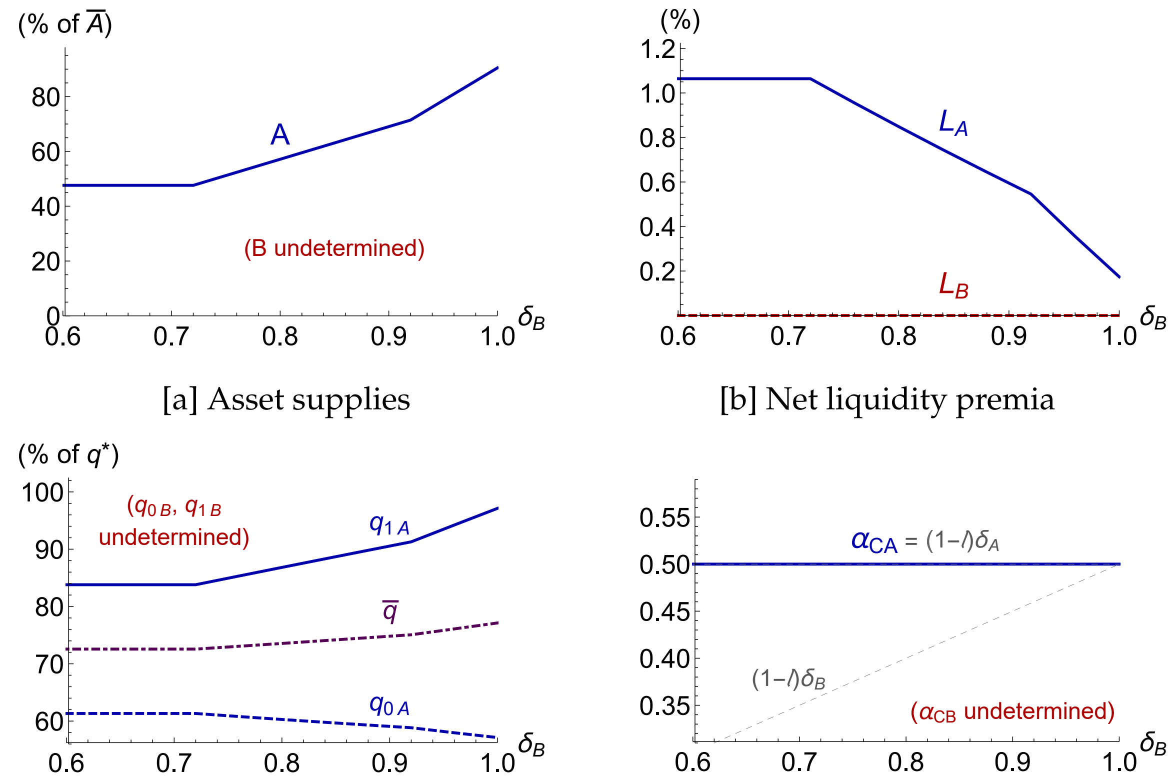
[c] Quantities traded in DM, and average

[d] OTC matching probabilities

Figure 9: Comparative statics with respect to δ_B, with just the tiniest bit of IRS ($\rho = 0.002$).

As Figure 10 illustrates, the transition occurs already for $\rho < 0.01$. Even with $\rho = 0.01$, a tiny degree of IRS, issuer competition is fierce and for any value of δ_B, only one asset ends up being liquid. However, this does not mean that δ_B stops mattering. When $\delta_B = \delta_A = 1$, an issuer who wishes to capture the secondary market must issue the quantity $\bar{D}$, which also drives her own payoff to zero. But as δ_B declines, so does the threat of B's competition, and therefore A's issuance is *negatively* related to her strategic advantage δ_A/δ_B.

While our model abstracts from a number of factors that are certainly influencing the borrowing decisions of the real-world issuers (the U.S. Treasury, large corporations, etc.), our theory generates solutions that resemble patterns in real-world asset markets. For instance, Figures 8 and 9 illustrate how even a small disadvantage of market B manifests itself as a higher matching probability for sellers of asset A (panel [d]), hence a larger liquidity premium for asset A (panel [b]), and how this mechanism is reinforced by issuer B's decision to scale back their issue size (panel [a]). The question whether the Treasury should be considered a strategic agent is interesting but not dispositive. In the Web Appendix, we consider a "semi-strategic" model where issuer B is strategic but issuer A is not. We show that the implications – at least, as far as issuer B is concerned – are broadly the same.



[a] Asset supplies

[b] Net liquidity premia

[c] Quantities traded in DM, and average

[d] OTC matching probabilities

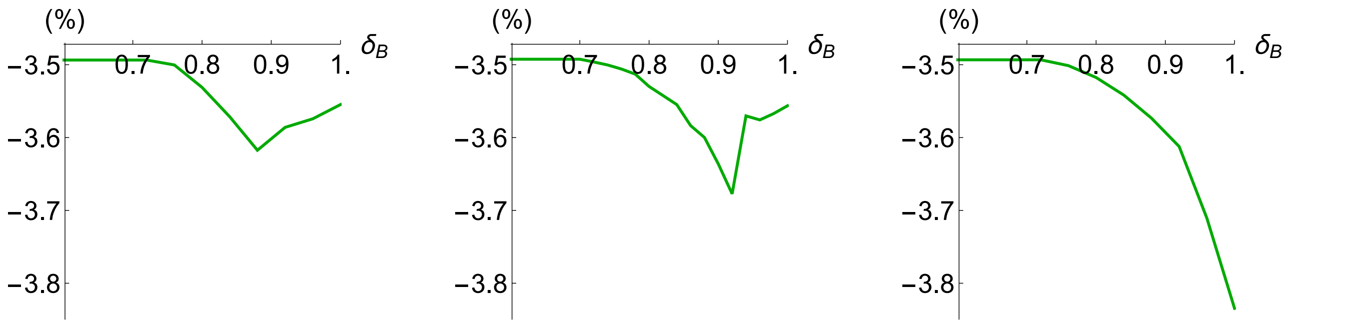
Figure 10: Comparative statics with respect to δ_B, with a little bit more IRS ($\rho = 0.01$).

6.4 The relationship between asset supplies, output, and welfare

We define social welfare $\mathcal{W}$ to be the total surplus across all DM trades, as follows:

$$\begin{aligned} \mathcal{W} \equiv\; & \ell e_C \Big((1-\alpha_{CA}) \left[u(q_{0A}) - q_{0A} \right] + \alpha_{CA} \left[u(q_{1A}) - q_{1A} \right] \Big) \quad \ldots \\ & + \ell (1-e_C) \Big((1-\alpha_{CB}) \left[u(q_{0B}) - q_{0B} \right] + \alpha_{CB} \left[u(q_{1B}) - q_{1B} \right] \Big). \end{aligned} \tag{20}$$

In monetary models of this kind, there is no general relationship between the supply of liquid assets and output or welfare. For example, consider the corner equilibrium where only the OTC market for asset A is open (or assume for a moment that A is the only asset). Applying a recent result by Herrenbrueck and Geromichalos (2017) and Huber and Kim (2017), it can be shown that welfare is a *decreasing* function of the asset supply in a neighborhood $(\bar{A} - \epsilon, \bar{A})$. Why? First, note that as A increases (but is still below $\bar{A}$), q_{0A} falls and $(q_{1A}, \tilde{q}_{1A})$ rise, so the effect on average output is ambiguous and depends on parameters. However, the welfare impacts of these changes are weighted by the marginal utility term $u'(q) - 1$. If A is close to $\bar{A}$, then $u'(q_{1A})$ is close to $u'(q^*) = 1$; thus, the welfare gain which successful traders receive from higher A vanishes, but the welfare loss of unsuccessful traders does not, and the overall welfare effect is negative. This is confirmed by combining panels [a] and [c] of Figure 10, showing increasing



[a] $\rho = 0$ [b] $\rho = 0.002$ [c] $\rho = 0.01$

Figure 11: Welfare as a function of δ_B, measured as equivalent CM consumption, in percent deviations from the first-best.

asset supply and output near $\bar{A}$, with panel [c] of Figure 11, which shows the drop in welfare.

Next, we are interested in the relationship between market microstructure and welfare. First, using the fact that an asset supply close to $\bar{D}$ is always 'too much' from a welfare perspective, we argue that any condition that leads to aggressive competition among the asset issuers is best avoided. In particular, the general intuition that the more competition, the better for social welfare, is not valid when it comes to liquid assets. The same reasoning would apply when matching is CRS and we compare a Cournot oligopoly of few versus many competitors.

Second, there is less clarity when we are far from the aggressive "everyone issues $\bar{D}$" case. For our paremetrization, a Cournot duopoly is better for welfare than a monopoly, so it is also possible to have too little competition. But the exact turning point will depend on details.

Third, and perhaps surprisingly, the effect of the exogenous 'market quality' parameter δ_B on welfare is not monotonic. In fact, for IRS and $\delta_B \approx \delta_A$, the effect is negative, because similarity promotes aggressive competition. For CRS, we have shown that $\delta_B \ll \delta_A$ promotes a monopoly and $\delta_B \approx \delta_A$ promotes a duopoly, but it is *intermediate* values of δ_B that promote the most aggressive competition, the largest supply of liquid assets, and a dip in welfare. It is also important to recognize that little of the welfare results can be ascribed to the direct effect on the extensive margin of OTC trade, as the [d]-panels of Figures 8-10 show: asset B is endogenously illiquid for $\delta_B < 0.9$, no OTC trade in that market actually takes place, but the threat that it might still affects the equilibrium.

7 Conclusion

We develop a model in which an asset's liquidity and, hence, its equilibrium price depend on:

1. The microstructure of the secondary market where that asset trades;
2. The microstructure of the secondary market(s) where "competing" assets trade;

3. The decision of agents to visit these secondary markets; (which in turn depends on the microstructure of the various markets), and;
4. The endogenous supply of the various assets.

Our model delivers a number of new insights. Even with small amounts of increasing returns, asset demand curves can be upward sloping because IRS encourages market concentration and agents are more likely to concentrate in market of an asset with plentiful supply. We also show that small differences in the microstructure of an OTC market can be magnified into a big endogenous liquidity advantage for one asset, because traders would prefer to be in the thick market, and through their own entry help make it even thicker. Our model predicts that for a reasonable set of parameters, a big and well-established borrower such as the Treasury can enjoy a significant liquidity advantage, to the point where they may be the *only* issuer of assets that trade at a liquidity premium. But in our model, whether the Treasury will be a monopolist in the issuance of liquid assets or not is endogenous.

Appendix

A Discussion of key modeling choices

We adopt a matching function that admits both CRS and IRS as subcases, and we present results for each case, but one may say that some of the most interesting results of the paper are derived using IRS. So how realistic are increasing returns to scale in financial markets? Quite realistic, in fact, which is well-established both at the theoretical and the empirical level. Duffie et al. (2005), and the vast majority of papers that follow their seminal work, adopt an IRS matching technology.[33] Furthermore, a number of empirical finance papers seem to confirm the relevance of IRS in OTC markets: for example, there is strong evidence that markets with higher trading volumes have lower bid-ask spreads. (See the discussion on page 54 of Vayanos and Wang (2013).) If higher spreads are associated with longer search times, this would be an argument suggesting IRS, since it would imply that markets with higher trading volume have more traders who are searching, and have lower bid ask spreads because trading delays are shorter. Are higher bid-ask spreads indeed associated with longer search times? Any theoretical model of OTC trade we are aware of would predict so.[34] As for the data, Amihud and Mendelson (1986) provide support in favor of this empirical regularity. Also, a quick glance at some of the main OTC markets suggests that this relationship is indeed true: markets that are characterized by long trading delays, e.g., municipal bonds, are also typically characterized by large spreads. Crucially, our own analysis in Section 4 shows that only a few percentage points of IRS are needed in order to explain big divergences in market outcomes.

A key assumption in our analysis is that secondary markets are segmented and agents can visit only one per period. The first part of this assumption is certainly realistic: Treasuries and municipal (or corporate) bonds do trade in secondary markets that are completely distinct. The second part of the assumption, according to which agents can visit only *one* market, is stronger – but, clearly, it is not meant to be taken literally. It does not imply a significant loss of generality since it is a qualitative rather than quantitative ingredient for our model's central mechanism: it is just a stark way to capture the idea that even if some investors do visit multiple markets, they will visit the market where they expect to find better trading conditions more frequently. Further, in the absence of market segmentation, the assets would be perfect substitutes and their prices would always be identical. But the empirical finance literature abounds with examples of assets that have pretty much identical characteristics, yet they trade at significantly different

[33] In that paper, the total number of matches between buyers and sellers of assets is given by $2\lambda\mu_B\mu_S$, where μ_B, μ_S are the respective measures of buyers and sellers, equivalent to $\rho = 1$ in our model. Hence, the arrival rate of a buyer to a seller is $2\lambda\mu_B$, which does not depend of the number of sellers. This process is therefore not just IRS, but completely congestion-free.

[34] For example, in Duffie et al. (2005), the bid-ask spread is strictly decreasing in the arrival rate of trading opportunities. Faster arrival rates imply a better outside option for the investor, thus a better bargaining position.

prices (and in secondary markets with very different levels of liquidity, as measured by bid-ask spreads, trading volumes, etc). It is also a fact that most fixed income dealers simply do not intermediate multiple kinds of securities. Clearly there is some cost to becoming an expert in a specific security, even for such similar ones as Treasury and AAA corporate bonds. Thus, market segmentation is not only essential for our results, but also the empirically relevant case. For a more detailed discussion of this assumption, see Geromichalos et al. (2018).

In our model, we study a differentiated Cournot game played by two bond issuers. One question that arises is whether in reality bond issuers are strategic. ('Strategic' has two relevant meanings: whether the issuers' objective includes profit/rent maximization, and whether they have market power.) First, the quote of the Assistant Secretary of the Treasury (presented in Footnote 2) clearly indicates that the Treasury is interested in maximizing its rent from debt issuance (although they call it "minimizing borrowing costs"). Similar evidence can be found for debt issuing corporations. Greenwood, Hanson, and Stein (2010) document that debt issuing corporations pay close attention to the actions taken by the Treasury and respond to these moves by filling in the supply gaps created by changes in government financing patterns. For another example, Robert Tipp, the Managing Director and Chief Investment Strategist of Prudential Investment Management, highlights that chief financial officers in big corporations are paying close attention to the market conditions and especially to the demand for bonds issued by the biggest player in the market: the Treasury.[35]

Given the discussion so far, we think that writing down a model where issuers play a differentiated Cournot game is a reasonable choice. In the baseline model, we focus on the case of a duopoly, but one can always generalize the model to include a general number of issuers, $N > 2$. Then, how much market power each issuer has is a (decreasing) function of N. Overall, we think that a model where issuers play a Cournot game is closer to reality than a model where issuers behave competitively.[36] Crucially, our baseline model can be easily extended in order to study a variety of alternative market structures. The Web Appendix contains three (out of many possible) such extensions: C.1 studies equilibrium in the presence of a non-strategic issuer with fixed supply; C.2 studies Stackelberg equilibria where issuer A is the first mover; C.3 allows for issuer B to have a positive marginal cost of issuing assets.

[35] Source: http://www.marketwatch.com/story/treasury-yields-edge-higher-apple-expected-to-issue-bonds-2016-02-16.

[36] For instance, in September 2013, Verizon issued bonds worth 49 billion dollars; in January 2016, Anheuser-Busch InBev issued bonds worth 46 billion dollars; in March 2018, CVS issued bonds worth 40 billion dollars (the list goes on). It would be hard to argue that when these corporations issue debt of this size they behave as measure zero agents whose actions have no effect on market prices.

B Details of the baseline model and its extensions

B.1 Baseline model

B.1.1 Value functions

We begin with the description of the value functions in the CM. Consider first a buyer who enters this market with m units of fiat money and d_j units of asset $j = \{A, B\}$. The Bellman equation of the buyer is given by:

$$W(m, d_A, d_B) = \max_{\substack{X,H,\hat{m},\\ \hat{d}_A,\hat{d}_B}} \left\{ X - H + \beta\, \mathbb{E}_i \left\{ \max \left\{ \Omega^i_A\left(\hat{m}, \hat{d}_A, \hat{d}_B\right), \Omega^i_B\left(\hat{m}, \hat{d}_A, \hat{d}_B\right) \right\} \right\} \right\}$$
$$\text{s.t. } X + \varphi(\hat{m} + p_A\hat{d}_A + p_B\hat{d}_B) = H + \varphi(m + \mu M + d_A + d_B),$$

where variables with hats denote portfolio choices for the next period, and $\mathbb{E}$ denotes the expectations operator. The price of money is expressed in terms of the general good but the price of bonds is expressed in nominal terms. The function Ω^i_j represents the value function in the OTC market for asset $j \in \{A, B\}$ for a buyer of type $i \in \{C, N\}$, to be described in more detail below. At the optimum, X and H are indeterminate but their difference is not. Using this fact and substituting $X - H$ from the budget constraint into W yields:

$$\begin{aligned} W(m, d_A, d_B) = \varphi(m + \mu M + d_A + d_B) \quad \ldots \\ + \max_{\hat{m},\hat{d}_A,\hat{d}_B} \Big\{ & - \varphi(\hat{m} + p_A\hat{d}_A + p_B\hat{d}_B) \\ & + \beta\ell \max\left\{\Omega^C_A\left(\hat{m}, \hat{d}_A, \hat{d}_B\right), \Omega^C_B\left(\hat{m}, \hat{d}_A, \hat{d}_B\right)\right\} \\ & + \beta(1-\ell) \max\left\{\Omega^N_A\left(\hat{m}, \hat{d}_A, \hat{d}_B\right), \Omega^N_B\left(\hat{m}, \hat{d}_A, \hat{d}_B\right)\right\} \Big\}. \end{aligned} \tag{B.1}$$

In the last expression, we have also used the fact that the representative buyer will be a C-type with probability ℓ in order to replace the expectations operator. As is standard in models that build on LW, the optimal choice of the agent does not depend on the current state (due to the quasi-linearity of $\mathcal{U}$), and the CM value function is linear. We write:

$$W(m, d_A, d_B) = \varphi(m + d_A + d_B) + \Upsilon, \tag{B.2}$$

where the constant Υ collects the remaining terms that do not depend on the state variables m, d_A, d_B.

As is well-known, a seller will not wish to leave the CM with positive amounts of money and bond holdings. Therefore, when entering the CM a seller will only hold money that she received as payment in the preceding DM, and her CM value function is given by:

$$W^S(m) = \max_{X,H} \left\{X - H + V^S\right\}$$
$$\text{s.t. } X = H + \varphi m,$$

where V^S denotes the seller's value function in the forthcoming DM. We can again use the budget constraint to substitute $X - H$ and show that W^S will be linear:

$$W^S(m) = \varphi m + V^S \equiv \Upsilon^S + \varphi m. \tag{B.3}$$

We now turn to the description of the OTC value functions. Recall that $e_C \in [0,1]$ and $e_N \in [0,1]$ denote the fraction of C-types and N-types, respectively, who are entering OTC$_A$. Using the matching probabilities α_{ij} defined in Equations (1)-(2), we can now define the value function for an agent of type $i = \{C, N\}$ who decides to enter OTC$_j$, $j = \{A, B\}$. Let ζ_j, denote the amount of money that gets transferred to the C-type, and χ_j the amount of assets (of type j) that gets transferred to the N-type in a typical match in OTC$_j$, $j = \{A, B\}$. These terms are described in detail in Lemma 2 below. We have:

$$\Omega_A^C(m, d_A, d_B) = \alpha_{CA} V\left(m + \zeta_A, d_A - \chi_A, d_B\right) + \left(1 - \alpha_{CA}\right) V(m, d_A, d_B), \tag{B.4}$$

$$\Omega_B^C(m, d_A, d_B) = \alpha_{CB} V\left(m + \zeta_B, d_A, d_B - \chi_B\right) + \left(1 - \alpha_{CB}\right) V(m, d_A, d_B), \tag{B.5}$$

$$\Omega_A^N(m, d_A, d_B) = \alpha_{NA} W\left(m - \zeta_A, d_A + \chi_A, d_B\right) + \left(1 - \alpha_{NA}\right) W(m, d_A, d_B), \tag{B.6}$$

$$\Omega_B^N(m, d_A, d_B) = \alpha_{NB} W\left(m - \zeta_B, d_A, d_B + \chi_B\right) + \left(1 - \alpha_{NB}\right) W(m, d_A, d_B), \tag{B.7}$$

where V denotes a buyer's value function in the DM. Notice that N-type agents proceed directly to next period's CM.

Lastly, consider the value functions in the DM. Let q denote the quantity of goods traded, and τ the total payment in units of fiat money. These terms are described in detail in Lemma 1 below. The DM value function for a buyer who enters that market with portfolio (m, d_A, d_B) is given by:

$$V(m, d_A, d_B) = u(q) + W(m - \tau, d_A, d_B), \tag{B.8}$$

and the DM value function for a seller (who enters with no money or assets) is given by:

$$V^S = -q + \beta W^S(\tau).$$

B.1.2 The terms of trade in the OTC markets and the DM

Consider a meeting between a C-type consumer with portfolio (m, d_A, d_B) and a producer who, in the beginning of the DM sub-period, holds no money or assets. The two parties bargain over a quantity q to be produced by the producer and a cash payment τ, to be made by the consumer. The consumer makes a TIOLI offer maximizing her surplus subject to the producer's participation constraint and the cash constraint. The bargaining problem can be described by:

$$\max_{\tau,q} \{u(q) + W(m - \tau, d_A, d_B) - W(m, d_A, d_B)\}$$
$$\text{s.t. } -q + W^S(\tau) - W^S(0) = 0,$$

and the cash constraint $\tau \leq m$. Substituting the value functions W, W^S from (B.2) and (B.3) into the expressions above, allows us to simplify this problem to:

$$\max_{\tau,q} \{u(q) - \varphi\tau\}$$
$$\text{s.t. } q = \varphi\tau,$$

and $\tau \leq m$. The solution to the bargaining problem is described in the following lemma.

Lemma 1. *Let m^* denote the amount of money that, given the CM value of money, φ, allows the consumer to purchase the first-best quantity q^*, i.e., let $m^* = q^*/\varphi$. Then, the solution to the bargaining problem is given by $\tau(m) = \min\{m, m^*\}$ and $q(m) = \varphi \min\{m, m^*\}$.*

Proof. The proof is standard and it is, therefore, omitted. □

The solution to the bargaining problem is straightforward. The only variable that affects the solution is the consumer's money holdings. As long as the buyer carries m^* or more, the first-best quantity q^* will always be produced. If, on the other hand, $m < m^*$, the consumer does not have enough cash to induce the seller to produce q^*. The cash constrained buyer will give up all her money, $\tau(m) = m$, and the producer will produce the quantity of good that satisfies her participation constraint under $\tau(m) = m$, namely, $q = \varphi m$.

While Lemma 1 describes the bargaining solution for all possible money holdings by the C-type consumer, we know that, since $\mu > \beta - 1$, the cost of carrying money is strictly positive and a consumer will never choose to hold $m > m^*$.[37] Hence, from now on we will focus on the binding branch of the bargaining solution, i.e., we will set $\tau(m) = m$ and $q(m) = \varphi m$.

We now describe the terms of trade in the OTC markets. Consider a meeting in OTC$_j$,

[37] Even if the consumer in question matches with an N-type in the preceding OTC round and acquires some extra liquidity, she will never choose to adjust her post-OTC money balances in a way that these exceed m^*. This would be unnecessary since carrying m^* is already enough to buy her the first-best quantity in the forthcoming DM.

$j = \{A, B\}$, between a C-type carrying the portfolio (m, d_A, d_B) and an N-type with portfolio $(\tilde{m}, \tilde{d}_A, \tilde{d}_B)$. These agents negotiate over an amount of money, ζ_j, to be transferred to the C-type, and an amount of type-j assets, χ_j, to be transferred to the N-type. Recall that the C-type and N-type split the available surplus based on proportional bargaining, with $\theta \in (0, 1)$ denoting the C-type's bargaining power. In the match under consideration, the surpluses for the C-type and the N-type agents are given by:

$$\begin{aligned} S_{Cj} &= V(m + \zeta_j, d_A - \mathbb{I}\{j = A\}\, \chi_A, d_B - \mathbb{I}\{j = B\}\, \chi_B) - V(m, d_A, d_B) \qquad \text{(B.9)} \\ &= u(\varphi(m + \zeta_j)) - u(\varphi m) - \varphi \chi_j, \\ S_{Nj} &= W(\tilde{m} - \zeta_j, \tilde{d}_A + \mathbb{I}\{j = A\}\, \chi_A, \tilde{d}_B + \mathbb{I}\{j = B\}\, \chi_B) - W(\tilde{m}, \tilde{d}_A, \tilde{d}_B) = \varphi(\chi_j - \zeta_j), \qquad \text{(B.10)} \end{aligned}$$

where $\mathbb{I}$ denotes the identity function, and the second equalities in the equations above exploit the definitions of the functions V, W (i.e., Equations (B.8) and (B.2), respectively). The bargaining problem is described by:

$$\max_{\zeta_j, \chi_j} S_{Cj} \quad \text{s.t.} \quad S_{Cj} = \frac{\theta}{1 - \theta} S_{Nj}, \quad \chi_j \leq d_j.$$

We restrict attention to equilibria where the N-type's money holdings never limit the trade, hence the corresponding constraint $\zeta_j \leq \tilde{m}$ is slack. A sufficient condition that guarantees this in equilibrium is given by inequality (6): inflation rates must be low enough that C-types (who carry m units of money) and N-types (who carry $\tilde{m}$) can always obtain the first-best m^* if they were to pool their money ($m + \tilde{m} \geq m^*$). Actual trade may achieve m^* or not, depending on whether the C-type carries enough assets to compensate the N-type for her money. Excluding the scarce-money branch of the bargaining solution is convenient: that branch ultimately generates a kink in the value function, which gives rise to an asset pricing indeterminacy, as we extensively analyzed in Geromichalos and Herrenbrueck (2016). It is also innocent for the purposes of our present paper: assets can only be priced (in the CM) at a determinate liquidity premium if $\chi_j \leq d_j$ binds (in the OTC) but $\zeta_j \leq \tilde{m}$ does not. Since our interest is in asset issuers who seek to exploit a positive premium, we think the restriction is acceptable.

The solution to the bargaining problem is described in the following lemma.

Lemma 2. *The bargaining solution is given by:*

$$\zeta_j(m, d_j) = \min\{\tilde{\zeta}_j(m, d_j), m^* - m\}$$

and

$$\chi_j(m, d_j) = \frac{z(\zeta_j(m, d_j))}{\varphi} = \min\{d_j, \bar{d}\,\},$$

where we have defined:

$$\tilde{\zeta}_j(m, d_j) \equiv \{\zeta : \varphi d_j = z(\zeta)\},$$

$$z(\zeta) \equiv (1-\theta)[u(\varphi(m+\zeta)) - u(\varphi m)] + \theta\varphi\zeta,$$

$$\bar{d} \equiv \frac{z(m^* - m)}{\varphi}.$$

Proof. It is straightforward to check that the suggested answer satisfies the necessary and sufficient conditions for maximization in each case. □

The OTC bargaining solution is intuitive. Agents' objective is to maximize the available total surplus of the match. This surplus is generated by transferring more money to the C-type, and it is maximized when the C-type's post-OTC money holdings are $m + \zeta_j = m^*$. However, in order to "afford" this transfer of liquidity, the C-type needs to have enough assets, and the critical level of asset holdings that allows her to acquire $\zeta_j = m^* - m$ is given by $\bar{d}$.

Summing up, if the C-type carries a sufficient amount of assets (defined as $\bar{d}$), then the money transfer will be optimal, i.e., $\zeta_j = m^* - m$, and the asset transfer will satisfy $\chi_j = \bar{d}$. On the other hand, if the C-type is constrained by her asset holdings (i.e., if $d_j < \bar{d}$), then the C-type will give up all her assets, $\chi_j = d_j$, and she will receive a money transfer which is smaller than $m^* - m$; more precisely, it satisfies $\zeta_j = \tilde{\zeta}_j$ ($< m^* - m$).

B.2 Model with three assets and three secondary asset markets

B.2.1 Analysis of value functions and terms of trade

First, we analyze the value functions in the CM. The value function of a buyer who enters the CM with m unit of money and d_j units of asset j, $j = \{A, B_1, B_2\}$, is given by:

$$W(m, d_A, d_{B_1}, d_{B_2}) = \max_{\substack{X,H,\hat{m},\\ \hat{d}_A,\hat{d}_{B_1},\hat{d}_{B_2}}} \Bigg\{ X - H \quad \dots$$

$$+\beta\,\mathbb{E}_i \left\{ \max\left\{ \Omega^i_A\Big(\hat{m}, \hat{d}_A, \hat{d}_{B_1}, \hat{d}_{B_2}\Big), \Omega^i_{B_1}\Big(\hat{m}, \hat{d}_A, \hat{d}_{B_1}, \hat{d}_{B_2}\Big), \Omega^i_{B_2}\Big(\hat{m}, \hat{d}_A, \hat{d}_{B_1}, \hat{d}_{B_2}\Big) \right\}\right\} \Bigg\}$$

$$\text{s.t. } X + \varphi(\hat{m} + p_A\hat{d}_A + p_{B_1}\hat{d}_{B_1} + p_{B_2}\hat{d}_{B_2}) = H + \varphi(m + \mu M + d_A + d_{B_1} + d_{B_2}),$$

where Ω^i_j denotes the value function of an i-type buyer, $i = \{C, N\}$, who enters the OTC market for asset j.

Next, the value functions in the OTC markets are given by:

$$\Omega^C_A(m, d_A, d_{B_1}, d_{B_2}) = \alpha_{CA} V\left(m + \zeta_A, d_A - \chi_A, d_{B_1}, d_{B_2}\right) + (1 - \alpha_{CA})\, V(m, d_A, d_{B_1}, d_{B_2}),$$

$$\Omega^C_{B_1}(m, d_A, d_{B_1}, d_{B_2}) = \alpha_{CB_1} V(m + \zeta_{B_1}, d_A, d_{B_1} - \chi_{B_1}, d_{B_2}) + (1 - \alpha_{CB_1}) V(m, d_A, d_{B_1}, d_{B_2}),$$

$$\Omega^C_{B_2}(m, d_A, d_{B_1}, d_{B_2}) = \alpha_{CB_2} V(m + \zeta_{B_2}, d_A, d_{B_1}, d_{B_2} - \chi_{B_2}) + (1 - \alpha_{CB_2}) V(m, d_A, d_{B_1}, d_{B_2}),$$

$$\Omega^N_{A}(m, d_A, d_{B_1}, d_{B_2}) = \alpha_{NA} W(m - \zeta_{A}, d_A + \chi_A, d_{B_1}, d_{B_2}) + (1 - \alpha_{NA}) W(m, d_A, d_{B_1}, d_{B_2}),$$

$$\Omega^N_{B_1}(m, d_A, d_{B_1}, d_{B_2}) = \alpha_{NB_1} W(m - \zeta_{B_1}, d_A, d_{B_1} + \chi_{B_1}, d_{B_2}) + (1 - \alpha_{NB_1}) W(m, d_A, d_{B_1}, d_{B_2}),$$

$$\Omega^N_{B_2}(m, d_A, d_{B_1}, d_{B_2}) = \alpha_{NB_2} W(m - \zeta_{B_2}, d_A, d_{B_1}, d_{B_2} + \chi_{B_2}) + (1 - \alpha_{NB_2}) W(m, d_A, d_{B_1}, d_{B_2}),$$

where ζ_j is the amount of money that gets transferred to a C-type, and χ_j is the amount of asset j that gets transferred to an N-type in a typical match in OTC$_j$.

Finally, the value function in the DM is given by:

$$V(m, d_A, d_{B_1}, d_{B_2}) = u(q) + W(m - \tau, d_A, d_{B_1}, d_{B_2}).$$

We now turn to the description of the terms of trade in the various markets, starting with the DM. Consider a meeting between a C-type consumer with m units of money and a producer. Given that the C-type consumer makes a take-it-or-leave-it offer and that the liquidity constraint will always bind due to the cost of carrying money, the bargaining solution is given by:

$$q(m) = \varphi m \qquad \text{and} \qquad \tau(m) = m.$$

Next, we turn to the terms of trade in the OTC markets. Consider a meeting in OTC$_j$ between a C-type with portfolio (m, d_A, d_B) and an N-type with portfolio $(\tilde{m}, \tilde{d}_A, \tilde{d}_B)$. The bargaining surpluses of an i-type buyer from an OTC$_j$ trading, S_{ij}, are given by:

$$S_{Cj} = V(m + \zeta_j, d_A - \mathbb{I}\{j = A\}\,\chi_A, d_{B_1} - \mathbb{I}\{j = B_1\}\,\chi_{B_1}, d_{B_2} - \mathbb{I}\{j = B_2\}\,\chi_{B_2}) - V(m, d_A, d_{B_1}, d_{B_2})$$
$$= u(\varphi(m + \zeta_j)) - u(\varphi m) - \varphi\chi_j,$$

$$S_{Nj} = W(\tilde{m} - \zeta_j, \tilde{d}_A + \mathbb{I}\{j = A\}\,\chi_A, \tilde{d}_{B_1} + \mathbb{I}\{j = B_1\}\,\chi_{B_1}, \tilde{d}_{B_2} + \mathbb{I}\{j = B_2\}\,\chi_{B_2}) - W(\tilde{m}, \tilde{d}_A, \tilde{d}_{B_1}, \tilde{d}_{B_2})$$
$$= \varphi(\chi_j - \zeta_j).$$

The bargaining solution (ζ_j, χ_j) solves:

$$\varphi\chi_j = (1 - \theta)[u(\varphi(m + \zeta_j)) - u(\varphi m)] + \theta\varphi\zeta_j,$$
$$\zeta_j = \min\{\tilde{\zeta}_j, m^* - m\},$$
$$\tilde{\zeta}_j = \{\zeta : \varphi d_j = (1 - \theta)[u(\varphi(m + \zeta)) - u(\varphi m)] + \theta\varphi\zeta\}.$$

We can now derive the objective function of a buyer in the CM, which is given by:

$$J(\hat{m}, \hat{d}_A, \hat{d}_{B_1}, \hat{d}_{B_2}) = -\varphi\big(\hat{m} + p_A\hat{d}_A + p_{B_1}\hat{d}_{B_1} + p_{B_2}\hat{d}_{B_2}\big) + \beta\hat{\varphi}\big(\hat{m} + \hat{d}_A + \hat{d}_{B_1} + \hat{d}_{B_2}\big) \quad \dots$$
$$+ \beta\ell\Big[u(\hat{\varphi}\hat{m}) - \hat{\varphi}\hat{m} + \max\Big\{\alpha_{CA}S_{CA},\, \alpha_{CB_1}S_{CB_1},\, \alpha_{CB_2}S_{CB_2}\Big\}\Big].$$

B.2.2 Matching probabilities

Let $e_{Cj} \in [0,1]$ and $e_{Nj} \in [0,1]$ denote the fractions of C-types and N-types, respectively, who choose to enter OTC$_j$, $j = \{A, B_1, B_2\}$. Then, the measure of asset sellers and buyers in OTC$_j$ is given by $e_{Cj}\ell$ and $e_{Nj}(1-\ell)$, respectively, and the measure of asset sellers and buyers in OTC$_B$ is given by $(1-e_{Cj})\ell$ and $(1-e_{Nj})(1-\ell)$. Letting $\alpha_{ij} \in [0,1]$ denote the matching probabilities for agents of type $i = \{C, N\}$ in OTC$_j$, $j = \{A, B_1, B_2\}$, we have:

$$\alpha_{CA} \equiv \frac{f_A(e_{CA}\ell, e_{NA}(1-\ell))}{e_{CA}\ell}, \qquad \alpha_{NA} \equiv \frac{f_A(e_{CA}\ell, e_{NA}(1-\ell))}{e_{NA}(1-\ell)},$$
$$\alpha_{CB_1} \equiv \frac{f_B(e_{CB_1}\ell, e_{NB_1}(1-\ell))}{e_{CB_1}\ell}, \qquad \alpha_{NB_1} \equiv \frac{f_B(e_{CB_1}\ell, e_{NB_1}(1-\ell))}{e_{NB_1}(1-\ell)},$$
$$\alpha_{CB_2} \equiv \frac{f_B(e_{CB_2}\ell, e_{NB_2}(1-\ell))}{e_{CB_2}\ell}, \qquad \alpha_{NB_2} \equiv \frac{f_B(e_{CB_2}\ell, e_{NB_2}(1-\ell))}{e_{NB_2}(1-\ell)},$$

where:

$$e_{CA} + e_{CB_1} + e_{CB_2} = 1,$$
$$e_{NA} + e_{NB_1} + e_{NB_2} = 1.$$

B.2.3 Equilibrium

We now describe the steady state equilibrium of the model with three assets and three secondary asset markets. The core variables are $\{q_{0A}, q_{1A}, q_{0B_1}, q_{1B_1}, q_{0B_2}, q_{1B_2}, e_{CA}, e_{NA}, e_{CB_1}, e_{NB_1}, e_{CB_2}, e_{NB_2}\}$, and we will describe the derivation of equilibrium following the same methodology as in the baseline model. For this analysis recall that we have defined $\omega(q) \equiv \theta + (1-\theta)u'(q)$.

First, the money demand equations are given by:

$$i = \ell\left(1 - \alpha_{Cj}\frac{\theta}{\omega(q_{1j})}\right)[u'(q_{0j}) - 1] + \ell\alpha_{Cj}\frac{\theta}{\omega(q_{1j})}[u'(q_{1j}) - 1]. \tag{B.11}$$

The OTC trading protocols:

$$q_{1j} = \min\{q^*, q_{0j} + \varphi\tilde{\zeta}_j\},$$

combined with the OTC bargaining solutions:

$$\varphi d_j = (1-\theta)[u(\varphi(m+\tilde{\zeta}_j) - u(\varphi m)] + \theta\varphi\tilde{\zeta}_j,$$

and the market clearing conditions:

$$A = e_{CA}d_A, \quad B_1 = e_{CB_1}d_{B_1}, \quad B_2 = e_{CB_2}d_{B_2}, \quad \varphi M = e_{CA}q_{0A} + e_{CB_1}q_{0B_1} + e_{CB_2}q_{0B_2},$$

yield:

$$q_{1A} = \min\left\{q^*, q_{0A} + \frac{1}{\theta}\frac{A}{M}\frac{e_{CA}q_{0A} + e_{CB_1}q_{0B_1} + e_{CB_2}q_{0B_2}}{e_{CA}} - \frac{1-\theta}{\theta}[u(q_{1A}) - u(q_{0A})]\right\}, \tag{B.12}$$

$$q_{1B_1} = \min\left\{q^*, q_{0B_1} + \frac{1}{\theta}\frac{B_1}{M}\frac{e_{CA}q_{0A} + e_{CB_1}q_{0B_1} + e_{CB_2}q_{0B_2}}{e_{CB_1}} - \frac{1-\theta}{\theta}[u(q_{1B_1}) - u(q_{0B_1})]\right\}, \tag{B.13}$$

$$q_{1B_2} = \min\left\{q^*, q_{0B_2} + \frac{1}{\theta}\frac{B_2}{M}\frac{e_{CA}q_{0A} + e_{CB_1}q_{0B_1} + e_{CB_2}q_{0B_2}}{e_{CB_2}} - \frac{1-\theta}{\theta}[u(q_{1B_2}) - u(q_{0B_2})]\right\}. \tag{B.14}$$

The next important task is to describe the agents' entry decisions. For that, it is useful to start by describing the liquidity premia of the various assets. The liquidity premium of asset j, denoted by L_j, is given by the percentage difference between an asset's price and its fundamental value. In other words, L_j solves:

$$p_j = \frac{1}{1+i}\Big(1 + L_j\Big),$$

where:

$$L_j = \ell\alpha_{Cj}\frac{\theta}{\omega(q_{1j})}[u'(q_{1j}) - 1].$$

For the optimal entry of C-type and N-type agents, there are seven possible cases:

Case 1: $e_{CA} \in (0,1)$, $e_{CB_1} \in (0,1)$, $e_{CB_2} \in (0,1)$ if:

$$\begin{aligned} \tilde{S}_{CA} &= \tilde{S}_{CB_1} = \tilde{S}_{CB_2}, \\ \alpha_{NA}S_{NA} &= \alpha_{NB_1}S_{NB_1} = \alpha_{NB_2}S_{NB_2}. \end{aligned} \tag{B.15}$$

Case 2: $e_{CA} \in (0,1)$, $e_{CB_1} \in (0,1)$, $e_{CB_2} = 0$ if:

$$\begin{aligned} \tilde{S}_{CA} &= \tilde{S}_{CB_1} > \tilde{S}_{CB_2}, \\ \alpha_{NA}S_{NA} &= \alpha_{NB_1}S_{NB_1} > \alpha_{NB_2}S_{NB_2}. \end{aligned} \tag{B.16}$$

Case 3: $e_{CA} \in (0,1)$, $e_{CB_1} = 0$, $e_{CB_2} \in (0,1)$ if:

$$\begin{aligned} \tilde{S}_{CA} &= \tilde{S}_{CB_2} > \tilde{S}_{CB_1}, \\ \alpha_{NA} S_{NA} &= \alpha_{NB_2} S_{NB_2} > \alpha_{NB_1} S_{NB_1}. \end{aligned} \tag{B.17}$$

Case 4: $e_{CA} = 0$, $e_{CB_1} \in (0,1)$, $e_{CB_2} \in (0,1)$ if:

$$\begin{aligned} \tilde{S}_{CB_1} &= \tilde{S}_{CB_2} > \tilde{S}_{CA}, \\ \alpha_{NB_1} S_{NB_1} &= \alpha_{NB_2} S_{NB_2} > \alpha_{NA} S_{NA}. \end{aligned} \tag{B.18}$$

Case 5: $e_{CA} = 1$, $e_{CB_1} = 0$, $e_{CB_2} = 0$ if:

$$\begin{aligned} \tilde{S}_{CA} &> \max\{\tilde{S}_{CB_1}, \tilde{S}_{CB_2}\}, \\ \alpha_{NA} S_{NA} &> \max\{\alpha_{NB_1} S_{NB_1}, \alpha_{NB_2} S_{NB_2}\}. \end{aligned} \tag{B.19}$$

Case 6: $e_{CA} = 0$, $e_{CB_1} = 1$, $e_{CB_2} = 0$ if:

$$\begin{aligned} \tilde{S}_{CB_1} &> \max\{\tilde{S}_{CA}, \tilde{S}_{CB_2}\}, \\ \alpha_{NB_1} S_{NB_1} &> \max\{\alpha_{NA} S_{NA}, \alpha_{NB_2} S_{NB_2}\}. \end{aligned} \tag{B.20}$$

Case 7: $e_{CA} = 0$, $e_{CB_1} = 0$, $e_{CB_2} = 1$ if:

$$\begin{aligned} \tilde{S}_{CB_2} &> \max\{\tilde{S}_{CA}, \tilde{S}_{CB_1}\}, \\ \alpha_{NB_2} S_{NB_2} &> \max\{\alpha_{NA} S_{NA}, \alpha_{NB_1} S_{NB_1}\}. \end{aligned} \tag{B.21}$$

where:

$$\begin{aligned} \tilde{S}_{Cj} &= -iq_{0j} - L_j[(1-\theta)(u(q_{1j}) - u(q_{0j})) + \theta(q_{1j} - q_{0j})] + \ell[u(q_{0j}) - q_{0j}] + \ell \alpha_{Cj} S_{Cj}, \\ S_{Cj} &= \theta[u(q_{1j}) - u(q_{0j}) - q_{1j} + q_{0j}], \\ S_{Nj} &= (1-\theta)[u(q_{1j}) - u(q_{0j}) - q_{1j} + q_{0j}]. \end{aligned}$$

Definition of equilibrium

Definition 2. For given asset supplies $\{A, B_1, B_2\}$, the steady-state equilibrium for the core variables of the model consists of the equilibrium quantities and entry choices, $\{q_{0A}$, q_{1A}, q_{0B_1}, q_{1B_1}, q_{0B_2}, q_{1B_2}, e_{CA}, e_{NA}, e_{CB_1}, e_{NB_1}, e_{CB_2}, $e_{NB_2}\}$, such that (B.11), (B.12), (B.13), (B.14), and one of (B.15), (B.16), (B.17), (B.18), (B.19), (B.20), (B.21) hold. The remaining variables follow directly from the core variables as in the baseline model.

B.3 Model with two marketplaces

B.3.1 Analysis of value functions and terms of trade

First, we analyze the value functions in the CM. The representative agent has five state variables. As always m represents money holdings, and d_A, d_B denote the amounts of asset A and B, respectively, carried by the agent for trade in the segmented marketplace. The new states d_{A2}, d_{B2} represent the amounts of asset A and B, receptively, carried by the agent for trade in the consolidated marketplace.[38] The value function of an agent who enters the CM is given by:

$$\begin{aligned} W(m, d_A, d_B, d_{A2}, d_{B2}) = \max_{\substack{X,H,\hat{m}, \\ \hat{d}_A,\hat{d}_B,\hat{d}_{A2},\hat{d}_{B2}}} &\Big\{ X - H + \beta \max \Big\{ \mathcal{M}_0(\hat{m}, \hat{d}_A, \hat{d}_B, \hat{d}_{A2}, \hat{d}_{B2}), \\ &\mathcal{M}_1(\hat{m}, \hat{d}_A, \hat{d}_B, \hat{d}_{A2}, \hat{d}_{B2}) - \kappa_1,\ \mathcal{M}_2(\hat{m}, \hat{d}_A, \hat{d}_B, \hat{d}_{A2}, \hat{d}_{B2}) - \kappa_2 \Big\}\Big\} \\ \text{s.t. } X + \varphi(\hat{m} + p_A\hat{d}_A + p_B\hat{d}_B + p_A\hat{d}_{A2} + p_B\hat{d}_{B2}) & \\ = H + \varphi(m + \mu M + d_A + d_B + d_{A2} + d_{B2}), & \end{aligned}$$

where:

$$\begin{aligned} \mathcal{M}_0(\hat{m}, \hat{d}_A, \hat{d}_B, \hat{d}_{A2}, \hat{d}_{B2}) &= V(\hat{m}, \hat{d}_A, \hat{d}_B, \hat{d}_{A2}, \hat{d}_{B2}), \\ \mathcal{M}_1(\hat{m}, \hat{d}_A, \hat{d}_B, \hat{d}_{A2}, \hat{d}_{B2}) &= \mathbb{E}_i\Big[\max \Big\{ \Omega_{iA}(\hat{m}, \hat{d}_A, \hat{d}_B, \hat{d}_{A2}, \hat{d}_{B2}), \Omega_{iB}(\hat{m}, \hat{d}_A, \hat{d}_B, \hat{d}_{A2}, \hat{d}_{B2}) \Big\}\Big], \\ \mathcal{M}_2(\hat{m}, \hat{d}_A, \hat{d}_B, \hat{d}_{A2}, \hat{d}_{B2}) &= \mathbb{E}_i\Big[\Omega_{i2}(\hat{m}, \hat{d}_A, \hat{d}_B, \hat{d}_{A2}, \hat{d}_{B2}) \Big]. \end{aligned}$$

Notice that $\mathcal{M}_0$ is the value function of not participating in any marketplace. (Since entering either of the two marketplaces entails a cost, the agent should always the option to not participate.) $\mathcal{M}_1$ is the value function of participating in the segmented marketplace, and κ_1 is the associated entry cost. $\mathcal{M}_2$ is the value function of participating in the consolidated marketplace, and κ_2 is the associated entry cost. (A mnemonic rule similar to the one described in Footnote 38 also applies here.)

Next, we move to the value functions in the OTC markets. We start with the segmented marketplace. Let Ω_{ij} denote the value function of an i-type agent, $i = \{C, N\}$, who enters OTC$_j$, $j = \{A, B\}$. These value functions are given by:

$$\begin{aligned} \Omega_{CA}(m, d_A, d_B, d_{A2}, d_{B2}) = \ &\alpha_{CA} V(m + \zeta_A, d_A - \chi_A, d_B, d_{A2}, d_{B2}) \\ &+ (1 - \alpha_{CA}) V(m, d_A, d_B, d_{A2}, d_{B2}), \end{aligned}$$

[38] As a mnemonic rule, the “2” is meant to remind the reader that these variables pertain to the market where *two* assets can be traded, namely, the consolidated market.

$$\Omega_{CB}(m, d_A, d_B, d_{A2}, d_{B2}) = \alpha_{CB} V(m + \zeta_B, d_A, d_B - \chi_B, d_{A2}, d_{B2}) + (1 - \alpha_{CB}) V(m, d_A, d_B, d_{A2}, d_{B2}),$$

$$\Omega_{NA}(m, d_A, d_B, d_{A2}, d_{B2}) = \alpha_{NA} W(m - \zeta_A, d_A + \chi_A, d_B, d_{A2}, d_{B2}) + (1 - \alpha_{NA}) W(m, d_A, d_B, d_{A2}, d_{B2}),$$

$$\Omega_{NB}(m, d_A, d_B, d_{A2}, d_{B2}) = \alpha_{NB} W(m - \zeta_B, d_A, d_B + \chi_B, d_{A2}, d_{B2}) + (1 - \alpha_{NB}) W(m, d_A, d_B, d_{A2}, d_{B2}),$$

where ζ_j is the amount of money transferred to a C-type, and χ_j is the amount of asset j transferred to an N-type in a typical match in OTC$_j$.

Now we turn to the value functions in the consolidated marketplace. Let Ω_{i2} denote the value function of an i-type agent who enters this marketplace. These value functions are given by:

$$\Omega_{C2}(m, d_A, d_B, d_{A2}, d_{B2}) = \alpha_{C2} V(m + \zeta_2, d_A, d_B, d_{A2} - \chi_{A2}, d_{B2} - \chi_{B2}) + (1 - \alpha_{C2}) V(m, d_A, d_B, d_{A2}, d_{B2}),$$

$$\Omega_{N2}(m, d_A, d_B, d_{A2}, d_{B2}) = \alpha_{N2} W(m - \zeta_2, d_A, d_B, d_{A2} + \chi_{A2}, d_{B2} + \chi_{B2}) + (1 - \alpha_{N2}) W(m, d_A, d_B, d_{A2}, d_{B2}),$$

where ζ_2 is the amount of money transferred to a C-type, and χ_{j2} is the amount of asset j transferred to an N-type in a typical match.

Finally, the value function in the DM is given by:

$$V(m, d_A, d_B, d_{A2}, d_{B2}) = u(q) + W(m - \tau, d_A, d_B, d_{A2}, d_{B2}).$$

We now turn to the description of the terms of trade in the various markets, starting with the DM. Since this version of the model has no differences regarding the bargaining protocol in the DM, the bargaining solution is still given by:

$$q(m) = \varphi m \quad \text{and} \quad \tau(m) = m.$$

Next, we turn to the terms of trade in the OTC markets. First consider a meeting in OTC$_j$ in the segmented the marketplace between a C-type with portfolio $(m, d_A, d_B, d_{A2}, d_{B2})$ and an N-type with portfolio $(\tilde{m}, \tilde{d}_A, \tilde{d}_B, \tilde{d}_{A2}, \tilde{d}_{B2})$. By definition, a C-type who enters this marketplace has $d_{A2} = d_{B2} = 0$, and either $d_A = 0$ or $d_B = 0$ (because an agent who goes to the segmented marketplace specializes only in one asset). Let S_{ij} denote the surplus of an i-type in OTC$_j$.

These surpluses are given by:

$$
\begin{aligned}
S_{CA} &= V(m+\zeta_A, d_A-\chi_A, d_B, d_{A2}, d_{B2}) - V(m, d_A, d_B, d_{A2}, d_{B2}) \\
&= u(\varphi(m+\zeta_A)) - u(\varphi m) - \varphi\chi_A, \\
S_{NA} &= W(\tilde{m}-\zeta_A, \tilde{d}_A+\chi_A, \tilde{d}_B, \tilde{d}_{A2}, \tilde{d}_{B2}) - W(\tilde{m}, \tilde{d}_A, \tilde{d}_B, \tilde{d}_{A2}, \tilde{d}_{B2}) \\
&= -\varphi\zeta_A + \varphi\chi_A, \\
S_{CB} &= V(m+\zeta_B, d_A, d_B-\chi_B, d_{A2}, d_{B2}) - V(m, d_A, d_B, d_{A2}, d_{B2}) \\
&= u(\varphi(m+\zeta_B)) - u(\varphi m) - \varphi\chi_B, \\
S_{NB} &= W(\tilde{m}-\zeta_B, \tilde{d}_A, \tilde{d}_B+\chi_B, \tilde{d}_{A2}, \tilde{d}_{B2}) - W(\tilde{m}, \tilde{d}_A, \tilde{d}_B, \tilde{d}_{A2}, \tilde{d}_{B2}) \\
&= -\varphi\zeta_B + \varphi\chi_B.
\end{aligned}
$$

We now describe the bargaining solutions. For (ζ_A, χ_A), we have:

$$
\begin{aligned}
\varphi\chi_A &= (1-\theta)[u(\varphi(m+\zeta_A)) - u(\varphi m)] + \theta\varphi\zeta_A, \\
\zeta_A &= \min\{m^* - m, \tilde{\zeta}_A\}, \\
\tilde{\zeta}_A &= \{\zeta : \varphi d_A = (1-\theta)[u(\varphi(m+\zeta)) - u(\varphi m)] + \theta\varphi\zeta\}.
\end{aligned}
$$

For (ζ_B, χ_B), we have:

$$
\begin{aligned}
\varphi\chi_B &= (1-\theta)[u(\varphi(m+\zeta_B)) - u(\varphi m)] + \theta\varphi\zeta_B, \\
\zeta_B &= \min\{m^* - m, \tilde{\zeta}_B\}, \\
\tilde{\zeta}_B &= \{\zeta : \varphi d_B = (1-\theta)[u(\varphi(m+\zeta)) - u(\varphi m)] + \theta\varphi\zeta\}.
\end{aligned}
$$

Next consider a meeting in the consolidated marketplace between a C-type with portfolio $(m, d_A, d_B, d_{A2}, d_{B2})$ and an N-type with portfolio $(\tilde{m}, \tilde{d}_A, \tilde{d}_B, \tilde{d}_{A2}, \tilde{d}_{B2})$. By definition, a C-type agent who enters this marketplace has $d_A = d_B = 0$. Let S_{i2} denote the surplus of an i-type in this marketplace. These surpluses are given by:

$$
\begin{aligned}
S_{C2} &= V_n(m+\zeta_2, d_A, d_B, d_A-\chi_{A2}, d_B-\chi_{B2}) - V_n(m, d_A, d_B, d_{A2}, d_{B2}) \\
&= u(\varphi(m+\zeta_2)) - u(\varphi m) - \varphi\chi_{A2} - \varphi\chi_{B2}, \\
S_{N2} &= W_n(\tilde{m}-\zeta_2, \tilde{d}_A, \tilde{d}_B, \tilde{d}_A+\chi_{A2}, \tilde{d}_B+\chi_{B2}) - W_n(\tilde{m}, \tilde{d}_A, \tilde{d}_B, \tilde{d}_{A2}, \tilde{d}_{B2}) \\
&= -\varphi\zeta_2 + \varphi\chi_{A2} + \varphi\chi_{B2}.
\end{aligned}
$$

We now describe the bargaining solutions. For $(\zeta_2, \chi_{A2}, \chi_{B2})$:

$$\varphi\chi_{A2} + \varphi\chi_{B2} = (1-\theta)[u(\varphi(m+\zeta_2)) - u(\varphi m)] + \theta\varphi\zeta_2,$$
$$\zeta_2 = \min\{m^* - m, \tilde{\zeta}_2\},$$
$$\tilde{\zeta}_2 = \{\zeta : \varphi d_A + \varphi d_B = (1-\theta)[u(\varphi(m+\zeta)) - u(\varphi m)] + \theta\varphi\zeta\}.$$

We can now derive the objective function of an agent in the CM, which is given by:

$$\begin{aligned} &J(\hat{m}, \hat{d}_A, \hat{d}_B, \hat{d}_{A2}, \hat{d}_{B2}) \\ &\quad = -\varphi(\hat{m} + p_A\hat{d}_A + p_B\hat{d}_B + p_A\hat{d}_{A2} + p_B\hat{d}_{B2}) \quad \ldots \\ &\qquad + \beta\hat{\varphi}[\hat{m} + \hat{d}_A + \hat{d}_B + \hat{d}_{A2} + \hat{d}_{B2}] + \beta\ell[u(\hat{\varphi}\hat{m}) - \hat{\varphi}\hat{m}] \quad \ldots \\ &\qquad + \beta\max\Big\{0, \max\Big\{\max\{\ell\alpha_{CA}S_{CA}, \ell\alpha_{CB}S_{CB}\Big\} + (1-\ell)\max\Big\{\alpha_{NA}S_{NA}, \alpha_{NB}S_{NB}\Big\} - \kappa_1, \\ &\qquad\qquad \ell\alpha_{C2}S_{C2} + (1-\ell)\alpha_{N2}S_{N2} - \kappa_2\Big\}\Big\}. \end{aligned}$$

This objective function is quite different and more complicated compared to the baseline model, therefore a couple of comments are in order. Notice that the agent's objective has three layers of maximum operators. First, the agent chooses whether to participate in any marketplace at all, or walk away and obtain a zero surplus from OTC trade. Second, conditional on participating, then the agent chooses whether she will go to the segmented or the consolidated marketplace. Third, if the agent chose to visit the segmented marketplace, she must then decide whether to visit OTC$_A$ or OTC$_B$. Another, more subtle difference is that, unlike the baseline model, here the surplus terms of an N-type appear in the objective function. This is because agents choose their marketplace before the idiosyncratic liquidity shock has been revealed. As a result, the agent's optimal decision must take into account the surplus she will make as an N-type.

B.3.2 Matching probabilities

With agents' choosing between two marketplaces, matching probabilities change drastically compared to the baseline model. Denote the *measure* of agents who do not participate in any marketplace by e_0. Also, let e_2 denote the *fraction* of marketplace participants who choose to go to the consolidated marketplace.[39] Let e_C denote the fraction of C-types who go to OTC$_A$ among the segmented marketplace participants. Similarly, let e_N denote the fraction of N-types who visit OTC$_A$, conditional on having chosen the segmented marketplace.

For the segmented marketplace, α_{ij} is the matching probability of an i-type who enters

[39] Notice that while the term e_0 is a measure, the term e_2 is a fraction. We do not explicitly define the measure of the segmented marketplace participants, since we can always write it as $(1-e_0)(1-e_2)$.

OTC$_j$. These matching probabilities are given by:

$$\alpha_{CA} = \frac{f_A[(1-e_0)(1-e_2)e_C\ell, (1-e_0)(1-e_2)e_N(1-\ell)]}{(1-e_0)(1-e_2)e_C\ell},$$

$$\alpha_{CB} = \frac{f_B[(1-e_0)(1-e_2)(1-e_C)\ell, (1-e_0)(1-e_2)(1-e_N)(1-\ell)]}{(1-e_0)(1-e_2)(1-e_C)\ell},$$

$$\alpha_{NA} = \frac{f_A[(1-e_0)(1-e_2)e_C\ell, (1-e_0)(1-e_2)e_N(1-\ell)]}{(1-e_0)(1-e_2)e_N(1-\ell)},$$

$$\alpha_{NB} = \frac{f_B[(1-e_0)(1-e_2)(1-e_C)\ell, (1-e_0)(1-e_2)(1-e_N)(1-\ell)]}{(1-e_0)(1-e_2)(1-e_N)(1-\ell)},$$

which are equal to:

$$\alpha_{CA} = [(1-e_0)(1-e_2)]^\rho \, \frac{f_A[e_C\ell, e_N(1-\ell)]}{e_C\ell},$$

$$\alpha_{CB} = [(1-e_0)(1-e_2)]^\rho \, \frac{f_B[(1-e_C)\ell, (1-e_N)(1-\ell)]}{(1-e_C)\ell},$$

$$\alpha_{NA} = [(1-e_0)(1-e_2)]^\rho \, \frac{f_A[e_C\ell, e_N(1-\ell)]}{e_N(1-\ell)},$$

$$\alpha_{NB} = [(1-e_0)(1-e_2)]^\rho \, \frac{f_B[(1-e_C)\ell, (1-e_N)(1-\ell)]}{(1-e_N)(1-\ell)}.$$

For the consolidated marketplace, α_{i2} is the matching probability of an i-type. These matching probabilities are given by:

$$\alpha_{C2} = \frac{f_2[(1-e_0)e_2\ell, (1-e_0)e_2(1-\ell)]}{(1-e_0)e_2\ell}, \qquad \alpha_{N2} = \frac{f_2[(1-e_0)e_2\ell, (1-e_0)e_2(1-\ell)]}{(1-e_0)e_2(1-\ell)},$$

which are equal to:

$$\alpha_{C2} = [(1-e_0)e_2]^\rho \, \delta_2(1-\ell), \qquad \alpha_{N2} = [(1-e_0)e_2]^\rho \, \delta_2\ell,$$

where δ_2 is the matching efficiency in the OTC market in the consolidated marketplace.

B.3.3 Equilibrium

We now describe the steady state equilibrium of the model with two marketplaces. We proceed as follows. First, we describe the equilibrium conditions (importantly, the demand for the various assets) implied by agents who (i) do not participate in any marketplaces; (ii) participate in the segmented marketplace; and (iii) participate in the consolidated marketplace. We do so taking as given the measures of agents in the various marketplaces (including the non-participants). Second, we endogenize these measures by studying agents' optimal entry decisions in the various marketplaces.

Equilibrium conditions implied by non-participants Here there is only one core variable, q. Non-participants rely only on their own money holdings, and their money demand equation is given by:

$$i = \ell(u'(q) - 1). \tag{B.22}$$

Since in the last step of equilibrium characterization we will describe the agents' entry decisions, it is useful to define the non-participant agent's surplus, which is given by:

$$S_0 = -iq + \ell[u(q) - q].$$

Equilibrium conditions implied by segmented marketplace participants Here the core variables are $\{q_{0A}, q_{1A}, q_{0B}, q_{1B}, e_C, e_N\}$, as in the baseline model. Recall that we have defined $\omega(q) \equiv \theta + (1-\theta)u'(q) \geq 1$.

First, the money demand equations are given by:

$$i = \ell\left[1 - \alpha_{CA}\frac{\theta}{\omega(q_{1A})}\right][u'(q_{0A}) - 1] + \ell\alpha_{CA}\frac{\theta}{\omega(q_{1A})}[u'(q_{1A}) - 1], \tag{B.23}$$

$$i = \ell\left[1 - \alpha_{CB}\frac{\theta}{\omega(q_{1B})}\right][u'(q_{0B}) - 1] + \ell\alpha_{CB}\frac{\theta}{\omega(q_{1B})}[u'(q_{1B}) - 1]. \tag{B.24}$$

The OTC trading protocols:

$$q_{1A} = \min\{q^*, q_{0A} + \varphi\tilde{\zeta}_A\},$$
$$q_{1B} = \min\{q^*, q_{0B} + \varphi\tilde{\zeta}_B\},$$

combined with the OTC bargaining solutions:

$$\varphi d_A = (1-\theta)[u(q_{1A}) - u(q_{0A})] + \theta\varphi\tilde{\zeta}_A,$$
$$\varphi d_B = (1-\theta)[u(q_{1B}) - u(q_{0B})] + \theta\varphi\tilde{\zeta}_B,$$

yield:

$$q_{1A} = \min\left\{q^*, q_{0A} + \frac{\varphi d_A - (1-\theta)[u(q_{1A}) - u(q_{0A})]}{\theta}\right\}, \tag{B.25}$$

$$q_{1B} = \min\left\{q^*, q_{0B} + \frac{\varphi d_B - (1-\theta)[u(q_{1B}) - u(q_{0B})]}{\theta}\right\}. \tag{B.26}$$

Next, we describe the liquidity premia implied by the demand of segmented marketplace participants, denoted by L_j. We define these premia as the percentage difference between an asset's price and its fundamental value. Assuming that the measure of agents visiting the segmented marketplace is positive, then L_j solves:

$$p_A = \frac{1}{1+i}(1+L_A), \qquad p_B = \frac{1}{1+i}(1+L_B),$$

where:

$$L_A = \ell \alpha_{CA} \frac{\theta}{\omega(q_{1A})}[u'(q_{1A}) - 1], \tag{B.27}$$

$$L_B = \ell \alpha_{CB} \frac{\theta}{\omega(q_{1B})}[u'(q_{1B}) - 1]. \tag{B.28}$$

If the measure of agents visiting this marketplace is zero, the liquidity premia will be exclusively determined by the demand of the consolidated marketplace participants.

The optimal entry of C-types (who have chosen the segmented marketplace) is characterized by:

$$e_C = \begin{cases} 1, & \tilde{S}_{CA} > \tilde{S}_{CB}, \\ 0, & \tilde{S}_{CA} < \tilde{S}_{CB}, \\ \in [0,1], & \tilde{S}_{CA} = \tilde{S}_{CB}, \end{cases} \tag{B.29}$$

where:

$$\tilde{S}_{CA} = - i q_{0A} - L_A \varphi d_A + \ell \Big[u(q_{0A}) - q_{0A} + \alpha_{CA} S_{CA} \Big],$$

$$S_{CA} = \theta \Big(u(q_{1A}) - u(q_{0A}) - q_{1A} + q_{0A} \Big),$$

and

$$\tilde{S}_{CB} = - i q_{0B} - L_B \varphi d_B + \ell \Big[u(q_{0B}) - q_{0B} + \alpha_{CB} S_{CB} \Big],$$

$$S_{CB} = \theta \Big(u(q_{1B}) - u(q_{0B}) - q_{1B} + q_{0B} \Big).$$

Similarly, for the N-types' decision, we have:

$$e_N = \begin{cases} 1, & \alpha_{NA} S_{NA} > \alpha_{NB} S_{NB}, \\ 0, & \alpha_{NA} S_{NA} < \alpha_{NB} S_{NB}, \\ \in [0,1], & \alpha_{NA} S_{NA} = \alpha_{NB} S_{NB}, \end{cases} \tag{B.30}$$

where:

$$S_{NA} = (1-\theta)\Big(u(q_{1A}) - u(q_{0A}) - q_{1A} + q_{0A}\Big),$$
$$S_{NB} = (1-\theta)\Big(u(q_{1B}) - u(q_{0B}) - q_{1B} + q_{0B}\Big).$$

Equilibrium conditions implied by consolidated marketplace participants Here the core variables are $\{q_0, q_1\}$. Recall that we have defined $\omega(q) \equiv \theta + (1-\theta)u'(q) \geq 1$.

First, the money demand equation is given by:

$$i = \ell\left[1 - \alpha_{C2}\frac{\theta}{\omega(q_1)}\right][u'(q_0) - 1] + \ell\alpha_{C2}\frac{\theta}{\omega(q_1)}[u'(q_1) - 1]. \tag{B.31}$$

The OTC trading protocols:

$$q_1 = \min\{q^*, q_0 + \varphi\tilde{\zeta}_2\},$$

combined with the OTC bargaining solutions:

$$\varphi d_{A2} + \varphi d_{B2} = (1-\theta)[u(q_1) - u(q_0)] + \theta\varphi\tilde{\zeta}_2,$$

yield:

$$q_1 = \min\left\{q^*, q_0 + \frac{\varphi d_{A2} + \varphi d_{B2} - (1-\theta)[u(q_1) - u(q_0)]}{\theta}\right\}. \tag{B.32}$$

Next, we describe the liquidity premia implied by the demand of consolidated marketplace participants, denoted by L_{j2}. As usual, we define these premia as the percentage difference between an asset's price and its fundamental value. Assuming that the measure of agents visiting the consolidated marketplace is positive, then L_{j2} solves:

$$p_{A2} = \frac{1}{1+i}(1 + L_{A2}), \qquad p_{B2} = \frac{1}{1+i}(1 + L_{B2}),$$

where:

$$L_{A2} = \ell\alpha_{C2}\frac{\theta}{\omega(q_1)}[u'(q_1) - 1], \tag{B.33}$$
$$L_{B2} = \ell\alpha_{C2}\frac{\theta}{\omega(q_1)}[u'(q_1) - 1]. \tag{B.34}$$

If the measure of agents visiting this marketplace is zero, the liquidity premia will be exclusively determined by the demand of the segmented marketplace participants.

Since in the last step of equilibrium characterization we will describe the agents' entry decisions, it is useful to define the consolidated marketplace participants' surpluses. For the C-types, we have:

$$\tilde{S}_{C2} = -iq_0 - L_A\varphi d_{A2} - L_B\varphi d_{B2} + \ell\Big[u(q_0) - q_0 + \alpha_{C2}S_{C2}\Big],$$
$$S_{C2} = \theta\Big(u(q_1) - u(q_0) - q_1 + q_0\Big).$$

For the N-types, we have:

$$S_{N2} = (1-\theta)\Big(u(q_1) - u(q_0) - q_1 + q_0\Big).$$

Market clearing and no-arbitrage conditions The money market clearing condition is:

$$\varphi M = (1-e_0)[(1-e_2)[e_C q_{0A} + (1-e_C)q_{0B}] + e_2 q_0] + e_0 q. \tag{B.35}$$

The asset market clearing conditions are:

$$S_A = (1-e_0)[(1-e_2)e_C d_A + e_2 d_{A2}], \tag{B.36}$$

$$S_B = (1-e_0)[(1-e_2)(1-e_C)d_B + e_2 d_{B2}]. \tag{B.37}$$

Finally, assuming positive measures of agents in both marketplaces, no-arbitrage requires the liquidity premia implied by both types of marketplace participants to be equal:

$$L_A = L_{A2}, \tag{B.38}$$

$$L_B = L_{B2}. \tag{B.39}$$

Entry choices of marketplaces Recall that the surplus of a non-participant agent was defined in section B.3.3 and denoted by S_0. The surplus of an agent in the segmented marketplace is given by:

$$S_1 = \tilde{S}_{CA}\cdot\mathbb{I}\{e_C > 0\} + \tilde{S}_{CB}\cdot\mathbb{I}\{e_C = 0\} \quad \ldots \tag{B.40}$$
$$+ (1-\ell)\Big[\alpha_{NA}S_{NA}\cdot\mathbb{I}\{e_N > 0\} + \alpha_{NB}S_{NB}\cdot\mathbb{I}\{e_N = 0\}\Big].$$

Finally, the surplus of an agent choosing the consolidated marketplace is given by:

$$S_2 = \tilde{S}_{C2} + (1-\ell)\alpha_{N2}S_{N2}. \tag{B.41}$$

The optimal marketplace choice is then characterized by:

$$e_0 = \begin{cases} 1, & S_0 > \max\{S_1 - \kappa_1, S_2 - \kappa_2\}, \\ 0, & S_0 < \max\{S_1 - \kappa_1, S_2 - \kappa_2\}, \\ \in [0,1], & S_0 = \max\{S_1 - \kappa_1, S_2 - \kappa_2\}, \end{cases} \tag{B.42}$$

and, conditional on $e_0 < 1$:

$$e_2 = \begin{cases} 1, & S_2 - \kappa_2 > S_1 - \kappa_1, \\ 0, & S_2 - \kappa_2 < S_1 - \kappa_1, \\ \in [0,1], & S_2 - \kappa_2 = S_1 - \kappa_1. \end{cases} \tag{B.43}$$

Definition of equilibrium

Definition 3. For given asset supplies $\{A, B\}$, the steady-state equilibrium of the model consists of 21 variables: the equilibrium quantity $\{q\}$ implied by the agents not participating in any marketplace; the equilibrium quantities $\{q_{0A}, q_{1A}, q_{0B}, q_{1B}\}$, the entry choices $\{e_C, e_N\}$, the liquidity premia $\{L_A, L_B\}$, and the asset holdings $\{d_A, d_B\}$, implied by the agents participating in the segmented marketplace; the equilibrium quantities $\{q_0, q_1\}$, the liquidity premia $\{L_{A2}, L_{B2}\}$, and the asset holdings $\{d_{A2}, d_{B2}\}$, implied by agents participating in the consolidated marketplace; the price of money $\{\varphi\}$; and the entry decisions of marketplaces $\{e_0, e_2\}$. These equilibrium variables are determined by (B.22), (B.23), (B.24), (B.25), (B.26), (B.27), (B.28), (B.29), (B.30), (B.31) (B.32), (B.33), (B.34), (B.35), (B.36), (B.37), (B.38), (B.39), (B.42), and (B.43).

References

Afonso, G. and R. Lagos (2015). Trade dynamics in the market for federal funds. *Econometrica 83*(1), 263–313.

Amihud, Y. and H. Mendelson (1986). Asset pricing and the bid-ask spread. *Journal of financial Economics 17*(2), 223–249.

Andolfatto, D., A. Berentsen, and C. Waller (2014). Optimal disclosure policy and undue diligence. *Journal of Economic Theory 149*, 128–152.

Andolfatto, D. and F. M. Martin (2013). Information disclosure and exchange media. *Review of Economic Dynamics 16*(3), 527–539.

Andolfatto, D., F. M. Martin, and S. Zhang (2017). Rehypothecation and liquidity. *European Economic Review 100*, 488–505.

Arseneau, D. M., D. Rappoport, and A. Vardoulakis (2015). Secondary market liquidity and the optimal capital structure. *Available at SSRN 2594558*.

Berentsen, A., G. Camera, and C. Waller (2007). Money, credit and banking. *Journal of Economic Theory 135*(1), 171–195.

Berentsen, A., S. Huber, and A. Marchesiani (2014). Degreasing the wheels of finance. *International economic review 55*(3), 735–763.

Berentsen, A., S. Huber, and A. Marchesiani (2016). The societal benefit of a financial transaction tax. *European Economic Review 89*, 303–323.

Berentsen, A. and C. Waller (2011). Outside versus inside bonds: A modigliani–miller type result for liquidity constrained economies. *Journal of Economic Theory 146*(5), 1852–1887.

Bethune, Z., B. Sultanum, and N. Trachter (2019). Asset issuance in over-the-counter markets. *Review of Economic Dynamics 33*, 4–29.

BlackRock (2014). Corporate bond market structure: The time for reform is now. Technical report.

Branch, W. A., N. Petrosky-Nadeau, and G. Rocheteau (2016). Financial frictions, the housing market, and unemployment. *Journal of Economic Theory 164*, 101–135.

Caramp, N. E. (2017). Sowing the seeds of financial crises: Endogenous asset creation and adverse selection. Working paper, MIT.

Chang, B. and S. Zhang (2015). Endogenous market making and network formation. *Available at SSRN 2600242*.

Duffie, D., N. Gârleanu, and L. H. Pedersen (2005, November). Over-the-counter markets. *Econometrica 73*(6), 1815–1847.

Fernández-Villaverde, J. and D. Sanches (2019). Can currency competition work? *Journal of Monetary Economics 106*, 1–15.

Geromichalos, A. and L. Herrenbrueck (2016). Monetary policy, asset prices, and liquidity in over-the-counter markets. *Journal of Money, Credit, and Banking 48*(1), 35–79.

Geromichalos, A. and L. Herrenbrueck (2022). The liquidity-augmented model of macroeconomic aggregates. *Review of Economic Dynamics 45*, 134–167.

Geromichalos, A., L. Herrenbrueck, and S. Lee (2018). Asset safety versus asset liquidity. Working paper, UC Davis.

Geromichalos, A., L. Herrenbrueck, and K. Salyer (2016). A search-theoretic model of the term premium. *Theoretical Economics 11*(3), 897–935.

Geromichalos, A., J. M. Licari, and J. Suárez-Lledó (2007, October). Monetary policy and asset prices. *Review of Economic Dynamics 10*(4), 761–779.

Greenwood, R., S. Hanson, and J. C. Stein (2010). A gap-filling theory of corporate debt maturity choice. *The Journal of Finance 65*(3), 993–1028.

Helwege, J. and L. Wang (2021). Liquidity and price pressure in the corporate bond market: Evidence from mega-bonds. *Journal of Financial Intermediation 48*(100922).

Herrenbrueck, L. (2019a). Frictional asset markets and the liquidity channel of monetary policy. *Journal of Economic Theory 181*, 82–120.

Herrenbrueck, L. (2019b). Interest rates, moneyness, and the fisher equation. Working paper, Simon Fraser University.

Herrenbrueck, L. and A. Geromichalos (2017). A tractable model of indirect asset liquidity. *Journal of Economic Theory 168*, 252 – 260.

Hu, T.-W. and G. Rocheteau (2015). Monetary policy and asset prices: A mechanism design approach. *Journal of Money, Credit and Banking 47*(S2), 39–76.

Huber, S. and J. Kim (2017). On the optimal quantity of liquid bonds. *Journal of Economic Dynamics and Control 79*, 184–200.

Kalai, E. (1977, October). Proportional solutions to bargaining situations: Interpersonal utility comparisons. *Econometrica 45*(7), 1623–30.

Krishnamurthy, A. and A. Vissing-Jorgensen (2012). The aggregate demand for treasury debt. *Journal of Political Economy 120*(2), 233–267.

Lagos, R. (2010, November). Asset prices and liquidity in an exchange economy. *Journal of Monetary Economics 57*(8), 913–930.

Lagos, R. and G. Rocheteau (2008, September). Money and capital as competing media of exchange. *Journal of Economic Theory 142*(1), 247–258.

Lagos, R. and G. Rocheteau (2009). Liquidity in asset markets with search frictions. *Econometrica 77*(2), 403–426.

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

Lagos, R., G. Rocheteau, and R. Wright (2017). Liquidity: A new monetarist perspective. *Journal of Economic Literature 55*(2), 371–440.

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

Lagos, R. and S. Zhang (2015). Monetary exchange in over-the-counter markets: A theory of speculative bubbles, the fed model, and self-fulfilling liquidity crises. Technical report, National Bureau of Economic Research.

Lester, B., A. Postlewaite, and R. Wright (2012). Information, liquidity, asset prices, and monetary policy. *The Review of Economic Studies 79*(3), 1209–1238.

Mattesini, F. and E. Nosal (2016). Liquidity and asset prices in a monetary model with otc asset markets. *Journal of Economic Theory 164*, 187–217.

Nosal, E. and G. Rocheteau (2013). Pairwise trade, asset prices, and monetary policy. *Journal of Economic Dynamics and Control 37*(1), 1–17.

Oehmke, M. and A. Zawadowski (2016). The anatomy of the cds market. *The Review of Financial Studies 30*(1), 80–119.

Rocheteau, G. (2011). Payments and liquidity under adverse selection. *Journal of Monetary Economics 58*(3), 191–205.

Rocheteau, G. and A. Rodriguez-Lopez (2014). Liquidity provision, interest rates, and unemployment. *Journal of Monetary Economics 65*, 80–101.

Rocheteau, G., R. Wright, and C. Zhang (2018). Corporate finance and monetary policy. *American Economic Review 108*(4–5), 1147–1186.

Rust, J. (1985). Stationary equilibrium in a market for durable assets. *Econometrica 53*(4), 783–805.

Rust, J. (1986). When is it optimal to kill off the market for used durable goods? *Econometrica 54*(1), 65–86.

Song, Z. and H. Zhu (2018). Quantitative easing auctions of treasury bonds. *Journal of Financial Economics 128*(1), 103–124.

S&P Global (2019). U.S. corporate debt market: The state of play in 2019.

Üslü, S. (2019). Pricing and liquidity in decentralized asset markets. *Econometrica 87*(6), 2079–2140.

Vayanos, D. and J. Wang (2013). Market liquidity–theory and empirical evidence. *Handbook of the Economics of Finance 2*(B), 1289–1361.

Vayanos, D. and T. Wang (2007). Search and endogenous concentration of liquidity in asset markets. *Journal of Economic Theory 136*(1), 66–104.

Vayanos, D. and P.-O. Weill (2008). A search-based theory of the on-the-run phenomenon. *The Journal of Finance 63*(3), 1361–1398.

Venkateswaran, V. and R. Wright (2014). Pledgability and liquidity: a new monetarist model of financial and macroeconomic activity. Technical report.

Weill, P.-O. (2007). Leaning against the wind. *The Review of Economic Studies 74*(4), 1329–1354.

Weill, P.-O. (2008). Liquidity premia in dynamic bargaining markets. *Journal of Economic Theory 140*(1), 66–96.