Asset Safety versus Asset Liquidity

Athanasios Geromichalos
University of California - Davis

Lucas Herrenbrueck
Simon Fraser University

Sukjoon Lee
New York University Shanghai

This Version: July 2022
First Version: October 2018

ABSTRACT

Many economists assume that safer assets are more liquid, and some have practically used "safe" and "liquid" as synonyms. But these terms are not synonyms, and mixing them up can lead to confusion and wrong policy recommendations. We build a multi-asset model where an asset's safety and liquidity are well-defined and distinct, and examine their relationship in general equilibrium. We show that the common belief that "safety implies liquidity" is generally justified, but also identify conditions under which this relationship can be reversed. We use our model to rationalize, qualitatively and quantitatively, a prominent safety-liquidity reversal observed in the data.

JEL Classification: E31, E43, E52, G12

Keywords: asset safety, asset liquidity, over-the-counter markets, search-and-matching, liquidity premium

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

We are grateful to David Andolfatto, Nicolas Caramp, Darrell Duffie, Janet Jiang, Ricardo Lagos, Jose Lopez, Florian Madison, Monika Piazzesi, Guillaume Rocheteau, Martin Schneider, Sanjay Singh, Alan Taylor, Alberto Trejos, Venky Venkateswaran, Randall Wright, four anonymous referees, and the editor for their comments and suggestions, as well as participants at the 2018 St Louis Fed Workshop on Money, Banking, Payments, and Finance, the 2018 Wisconsin School of Business Conference on Money, Banking, and Asset Markets, the 2020 AEA Annual Meeting, the HKBU-NTU-NYUSH-Sinica-Yonsei Macro/Finance Seminar Series, the SENA Workshop on Financial Networks, and seminars at Bank of Canada, National University of Singapore, Sogang University, Stanford University, UC Davis, UC Irvine, and UCSC. Edited by Greg Kaplan.

1 Introduction

Recently, there has been a lot of attention on the role of safe assets and liquid assets in the macroeconomy. Many economists, both academics and practitioners, seem to believe that safer assets are also more liquid, and some go a step further by practically using the two terms as synonyms or by merging them into the single term "safe and liquid assets".[1] However, the terms are not synonyms: Safety refers to the probability that the (issuer of the) asset will pay the promised cash flow, at maturity, and liquidity refers to the ease with which an investor can sell the asset if needed, before maturity.[2] Mixing up an asset's safety and liquidity is not just semantics; it can lead to false conclusions and misguided policy recommendations.

For instance, when a credit rating agency characterizes a certain bond as AAA, should investors think of this as an assessment (only) of its safety or also of its liquidity? And, if the answer is affirmative, how can one explain the fact that (the virtually default free) AAA corporate bonds are considered less liquid than their riskier AA counterparts? Moreover, a recent literature in empirical macro-finance measures the so-called safety premium as the spreads between AAA and BAA bonds, assuming that these types of bonds are equally (il)liquid. But if certain assets carry different liquidity premia *because* they have different safety characteristics (as indicated by the conventional wisdom and confirmed by our theory), bonds of "equal liquidity" may be tricky to identify. Finally, policy makers and financial regulators are often concerned about liquidity in certain assets markets. If safety implies liquidity, could we just improve safety and let liquidity follow?

[1] The examples are numerous, so for the sake of brevity we highlight just two. From the IMF's 2012 Global Financial Stability Report: "Safe assets are a desirable part of a portfolio from an investor's perspective, as they [...] are highly liquid, permitting investors to liquidate positions easily." And at the 2017 American Economic Association meeting, one session was titled: "How safe and liquid assets impact monetary and financial policy".

[2] Although there are economists who adopt slightly different definitions for both of these terms. For instance, Gorton and Ordonez (2022) emphasize that an important aspect of *safe* assets is that they are "information insensitive". Also, a large number of papers in the New Monetarist literature, assume that an asset's *liquidity* refers to the ease with which that asset can be used to purchase consumption, e.g., by serving as a means of payment; see Lagos, Rocheteau, and Wright (2017). For a careful comparison of the various approaches, see the Literature Review (Section 1.1).

These questions reveal that it is essential to carefully study the relationship between asset safety and asset liquidity, rather than just assume that one implies the other. To do so, we build a multi-asset model in which an asset's safety and liquidity are well-defined and *distinct* from one another. Treating safety as a primitive, we examine the relationship between an asset's safety and its liquidity. We show that the commonly held belief that "safer assets will be more liquid" is generally justified, but with important exceptions. We then describe the conditions under which a riskier asset can be more liquid than its safe(r) counterparts, and use our model to rationalize a prominent safety-liquidity reversal observed in the data. Finally, we highlight a surprising implication of our model about the effect of an increase in the supply of safe assets on welfare.

To answer the research question at hand we build a dynamic general-equilibrium model with two assets, A and B. The concept of asset safety is straightforward in our framework: asset A is "safe" in the sense that it always pays the promised cash flow, whereas asset B may default with a certain probability, known to everyone. The concept of liquidity is more involved; specifically, we define an asset's liquidity as the ease with which an agent can sell it for cash (if needed). To capture this idea, we employ the monetary model of Lagos and Wright (2005), extended to incorporate asset trade in over-the-counter (OTC) secondary asset markets subject to search and bargaining frictions, á la Duffie, Gârleanu, and Pedersen (2005). Another important ingredient we introduce is an entry decision made 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. Thus, in our model, an asset's liquidity depends on the *endogenous* choice of agents to visit the secondary market where that asset is traded, not the exogenous characteristics of that market.

After agents make their portfolio decisions, two shocks are realized. The first is an idiosyncratic shock that determines whether an agent will have a consumption opportunity in that period, and the second is an aggregate shock that determines whether asset B will default in that period. Since purchasing the consumption good necessitates the use of a medium of exchange (i.e., money) and carrying money is costly, in equilibrium, agents who receive a consumption opportunity will visit a secondary market to sell assets and boost their cash holdings. Hence, assets have *indirect liquidity* properties (they can be sold for cash, although they do not serve directly as means of payment), and their equilibrium price in the primary market will typically contain a *liquidity premium*, i.e., it will exceed the fundamental value of holding the asset to maturity.

Our first result is that, other things equal, the safer asset carries a higher liquidity premium, and that premium is increasing in the default probability of asset B. The intuition is as follows. 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 can only sell assets they own and are therefore committed to the market corresponding to their chosen asset, asset buyers are free to visit any market they wish, since their money is good to buy any asset. As a result, if asset B defaults, all asset buyers will rush into the market for asset A. A second, more subtle force is that agents who specialize in the (riskier) asset B endogenously choose to carry more money for precautionary reasons. Consequently, asset B sellers need a smaller liquidity boost, which means that the surplus generated from trades in the market for asset B is lower. This force attracts more buyers to the market for asset A *even if asset B does not default*. Taken together, these forces imply that ex post (and regardless of the default shock realization) more buyers are attracted to the market for asset A, which in turn encourages agents to specialize in asset A ex ante, precisely because this asset will be easier to sell down the road. Through this channel a small default probability for asset B can be magnified into a big endogenous liquidity advantage for asset A, even with constant returns to scale (CRS) in the OTC matching technology.

So far we have assumed that all parameters other than asset safety are kept equal. Allowing for varying asset supplies delivers the second important result of the paper.[3]

[3] There are two more parameters held equal in the background: the efficiency of matching in each OTC market and the bargaining power of buyers versus sellers in each OTC market. Since our goal is to develop

Even with slight increasing returns to scale (IRS) in OTC matching, demand curves can be *upward sloping*, because an asset in large supply is now more likely to be liquid. Consequently, asset B can be more liquid than asset A, despite being less safe, as long as the supply of the former is large enough compared to the latter.

The mechanism is as follows. When the supply of (say) asset B increases, so does the trading volume in its associated OTC market, which in turn makes this market more attractive to enter for both buyers and sellers. With CRS in the matching function, the larger numbers of buyers and sellers balance out so that the matching probabilities of either side are unchanged. And as more agents choose to specialize in asset B, each one of them will hold a smaller amount, whereas the reverse is true for holders of asset A, so that in equilibrium the surplus in any single trade is also balanced out across the two OTC markets. (We dub this the *dilution effect* since the asset in larger supply is 'diluted' among more agents holding it.) Now, with IRS in the matching function, the matching probabilities will *not* remain constant when more buyers and sellers enter market B, but increase for both sides. As a result, following an increase in the supply of asset B, entry into this market will remain attractive until the surplus of a single trade is substantially smaller than it would be in the market for asset A. In other words, with IRS, increased entry into a market tends to make this market even more attractive on the extensive margin, so entry will continue until the intensive margin becomes sufficiently unattractive via the dilution effect – i.e., agents trading in the thicker market must be trading small amounts, while agents in the smaller market trade larger amounts and get closer to satisfying their liquidity needs. But both of these forces imply that the asset in larger supply (and therefore with the thicker OTC market) will command a higher liquidity premium: first, because of its higher matching probability, and second, because a lower quantity of asset sold increases the marginal value of selling more, which is to say, the marginal value

a theory that links asset safety and asset liquidity in an unbiased way, we assume that these parameters are always equal in both OTC markets. This guarantees that any difference in liquidity between the two assets is driven exclusively by differences in safety and not by exogenous market characteristics.

of carrying more of the asset in the first place.

Our model can shed some light on the puzzling empirical observation that, in the U.S., the virtually default-free AAA bonds are less liquid than (the less safe) AA corporate bonds.[4] In recent years, regulations introduced to improve the stability and transparency of the financial system (e.g., the Dodd-Frank Act) have made it especially hard for corporations to attain the AAA score. As a result, the supply of such bonds has fallen dramatically. During the same time, the yield on AA corporate bonds has been *lower* than that on AAA bonds, even without controlling for the risk premium associated with the riskier AA bonds. While it is plausible to attribute this differential to a higher liquidity premium enjoyed by AA corporate bonds – and this is precisely what practitioners have claimed – existing models of asset liquidity do not capture this stylized fact (for details, see Section 1.1). Our approach of combining the 'indirect liquidity' mechanism with endogenous market entry can capture this stylized fact, and not only qualitatively: in Section 5, we calibrate our model and show it can reproduce the AA-AAA yield reversal quantitatively as well.

The model also delivers a surprising result regarding welfare. A large body of recent literature highlights that the supply of safe assets has been scarce, and that increasing this supply would be beneficial for welfare (see for example Caballero, Farhi, and Gourinchas, 2017). In our model this result is not necessarily true: there exists a region of parameter values for which welfare is decreasing in the supply of the safe asset. The intuition is as follows. In our model agents have the opportunity to acquire additional cash by selling assets in the secondary market. When the safe asset becomes more plentiful, agents expect that it will be easier to acquire extra cash *ex-post* and, thus, choose to hold less money *ex-ante*. This channel depresses money demand, which, in turn, decreases the values of

[4] For more details and empirical evidence, see Section 5.1. Moreover, there are other cases where the common belief "safety and liquidity go together" is violated. Christensen and Mirkov (2019) document yet another class of bonds – Swiss Confederation Bonds – that are extremely safe, but not particularly liquid. And, vice versa, Beber, Brandt, and Kavajecz (2008) report that Italian government bonds are among the most liquid, but also the most risky of Euro-area sovereign bonds.

money and of the trade that the existing money supply can support.

1.1 Literature review

Our paper is related to the "New Monetarist" literature (see Lagos et al., 2017) that highlights the importance of liquidity for the determination of asset prices; see Geromichalos, Licari, and Suárez-Lledó (2007), Lagos (2011), Nosal and Rocheteau (2013), Andolfatto, Berentsen, and Waller (2014), and Hu and Rocheteau (2015). Unlike these papers, where assets serve *directly* as means of payment, here liquidity is *indirect*, since agents sell assets for money in secondary asset markets. This approach is empirically relevant and integrates the concepts of liquidity adopted by monetary economics and finance (Geromichalos and Herrenbrueck, 2016). The indirect liquidity approach is employed in several recent papers, including Berentsen, Huber, and Marchesiani (2014, 2016), Han (2015), Mattesini and Nosal (2016), Herrenbrueck and Geromichalos (2017), Herrenbrueck (2019a), and Madison (2019).

Our work is also related to the literature that studies how frictions in OTC markets affect asset prices and trade, such as Weill (2007, 2008), Lagos and Rocheteau (2009), Chang and Zhang (2015), and Üslü (2019). Vayanos and Weill (2008) is conceptually related to our paper, since the authors use a model with trade in OTC markets to rationalize a puzzling empirical observation known as the "on-the-run" phenomenon. A key ingredient in their mechanism is the existence of short-sellers. Since short-sellers initially sell an asset and eventually buy that asset back, their concentration in one asset (in conjunction with increasing returns to scale in matching) increase the trading volume of that asset and therefore its liquidity.[5]

[5] Vayanos and Weill (2008) show that if assets differ enough in their supplies, there exists a unique equilibrium where short-sellers concentrate in the largest-supply asset, thus making it more liquid. However, in their model agents can only hold 0 or 1 unit of an asset, thus, an increase in an asset's supply automatically means that more agents must hold it. In our model, the mechanism is different, since asset holdings are perfectly divisible and agents are free to hold any asset they wish regardless of the supply. As we have already explained, an asset in larger supply implies a thicker OTC market for that asset, and a higher liquidity premium, both because of a higher matching probability and because of a higher per trade surplus generated in the OTC market. Also, notice that in Lester, Postlewaite, and Wright (2012) an asset in larger

Our paper is also related to the growing literature that studies the role of safe assets in the macroeconomy. Examples include Gorton, Lewellen, and Metrick (2012), Piazzesi and Schneider (2016), Caballero et al. (2017), Gorton (2017), Caballero and Farhi (2018), and He, Krishnamurthy, and Milbradt (2019). None of these papers study the relationship between asset safety and asset liquidity. Two recent exceptions are Infante and Ordoñez (2020) and Ozdenoren, Yuan, and Zhang (2021). The former paper shows that increased volatility and illiquidity in the economy will raise the price of safe assets due to their ability to hedge idiosyncratic risks. The latter shows that safe assets earn a liquidity premium due to their ability to mitigate the adverse selection problem in asset markets.

Andolfatto and Martin (2013) consider a model where an asset, whose expected return is subject to a news shock, serves as medium of exchange. They show that the non-disclosure of news can enhance the asset's property as an exchange medium. As already mentioned, the concept of (indirect) liquidity adopted here is different, and so is the concept of safety. (The idea that some assets are more "information sensitive" is close to the definition of safety adopted by Gorton and Ordonez (2022).) Here, an asset's safety is simply the (ex-ante) probability with which the assets will pay the promised cash flow, which is public knowledge and can be thought of (or approximated by) a credit rating agency's score. Rocheteau (2011) studies a model where bonds serve as media of exchange alongside with money. The author shows that if the bond holders (and goods buyers) have private information about the bond's return, then money will endogenously arise as more liquid asset (i.e., a better medium of exchange). In our paper the various assets' safety characteristics are public knowledge.

Our paper is also related to Lagos (2010), who considers a model where bonds, whose return is deterministic, and stocks, whose return is stochastic, compete as media of ex-

supply is bound to be more liquid (even though the authors never discuss such a result). In that model, a real asset competes with money as a medium of exchange, and producers can invest in a technology that allows them to recognize and consequently accept the asset. Thus, a larger supply makes an asset more likely to be accepted as a medium of exchange (thus more liquid), simply because it is more likely to make it worthwhile for producers to invest in that technology. Again, our mechanism is completely different. Also, neither of these papers study the relationship between asset safety and liquidity.

change. The author quantitatively demonstrates that the equity premium puzzle can be explained through a liquidity differential between the safe and the risky asset. Jacquet (2021) employs a similar model, but includes a larger variety of asset classes and ex-ante heterogeneous agents. The author shows that the equilibrium displays a "class structure" in the sense that agents with different liquidity needs will only be willing to hold assets of a certain risk structure. Our paper differs from the aforementioned papers, not only because it employs a different model of liquidity, but also because it predicts that an asset in large(r) supply may carry a higher liquidity premium. This result is not present in Lagos (2010) and Jacquet (2021), as in these papers the asset demand curve is decreasing. Thus, our model of indirect liquidity and endogenous market entry can rationalize why assets in limited supply can be highly illiquid, even when they enjoy a high credit rating (as in the case of AAA corporate bonds in the U.S.), i.e., it can rationalize the reversal between asset safety and liquidity.

He and Milbradt (2014) study a one-asset model where defaultable corporate bonds are traded in an OTC secondary market, and show that the inverse bid-ask spread, which is their proxy for bond liquidity, is positively related with credit ratings. However, in their model the probability of trade between agents is exogenous. We define liquidity as the ease with which an investor can sell her assets, if needed. We build a two-asset (easily extended to an N-asset) model, where the probability of selling an asset depends on the *endogenous* decision of agents to visit the various asset markets, which, in turn, is a function of each asset's safety characteristics. Also, He and Milbradt (2014) employ the model of Duffie et al. (2005) where assets are indivisible, i.e., agents can hold either 0 or 1 units of the asset. Our model also incorporates OTC secondary asset trade *à la* Duffie et al. (2005), but does so within the monetary model of Lagos and Wright (2005), which allows us to study perfectly divisible asset holdings, and opens up a number of new insights. Such insights include the possibility of upward-sloping demand curves, the possibility that a riskier asset can be more liquid in general equilibrium, and the fact that welfare can be decreasing in the supply of safe assets.

In related empirical work, Krishnamurthy and Vissing-Jorgensen (2012) distinguish between asset safety and liquidity, and extract safety and liquidity premia from the data to explain why Treasury yields have been decreasing. Their model identifies the safety premium through the spreads between AAA and BAA bonds, assuming that both of these types of bonds are equally illiquid. The present paper demonstrates that certain assets may carry different liquidity premia precisely because they are characterized by different default risks. This, in turn, highlights that we need more theory that studies the relationship between asset safety and liquidity, and we view the present paper as a part of this important agenda.

The rest of this paper is organized as follows. In Section 2, we describe the baseline model, and we solve it in Section 3. We present our main results in Section 4, and calibrate the model in Section 5. The Web Appendix contains further details including microfoundations for key assumptions of our model, and proofs.

2 The Model

Our model is a hybrid of Lagos and Wright (2005) (henceforth, LW) and Duffie et al. (2005). Time is discrete and continues forever. Each period is divided into three subperiods, characterized by different types of trade (for an illustration, see Figure 1 below). In the first subperiod, agents trade in OTC secondary asset markets. In the second subperiod, they trade in a decentralized goods market (DM). Finally, in the third subperiod, agents trade in a centralized market (CM). The CM is the typical settlement market of LW, where agents settle their old portfolios and choose new ones. The DM is a decentralized market characterized by anonymity and imperfect commitment, where agents meet bilaterally and trade a special good. These frictions make a medium of exchange necessary, and we assume that only money can serve this role. The OTC markets allow agents with different liquidity needs to rebalance their portfolio by selling assets for money.

Agents live forever and discount future between periods, but not subperiods, at rate $\beta \in (0, 1)$. There are two types of agents, consumers and producers, distinguished by their roles in the DM. The measure of each type is normalized to the unit. Consumers consume in the DM and the CM and supply labor in the CM; producers produce in the DM and consume and supply labor in the CM. All agents have access to a technology that transforms one unit of labor in the CM into one unit of the CM good, which is also the numeraire. The preferences of consumers and producers within a period are given by $\mathcal{U}(X, H, q) = X - H + u(q)$ and $\mathcal{V}(X, H, q) = X - H - q$, respectively, where X denotes consumption of CM goods, H is labor supply in the CM, and q stands for DM goods produced and consumed. We assume that u is twice continuously differentiable, with $u' > 0$, $u'(0) = \infty$, $u'(\infty) = 0$, and $u'' < 0$. The term q^* denotes the first-best level of trade in the DM, i.e., it satisfies $u'(q^*) = 1$. All goods are perishable between periods.

Notice that the agents dubbed "producers" will never choose to hold any assets, as long as these assets are priced at a premium for their liquidity. (Why pay this premium when they know they will never have a liquidity need in the DM?) As a result, all the interesting portfolio choices are made by the "consumers". Thus, henceforth, we will refer to the "consumers" simply as "agents". When we use the terms "buyer" and "seller", it will be exclusively to characterize the role of these agents in the secondary asset market.

There is a perfectly divisible object called fiat money that can be purchased in the CM at the price φ in terms of CM goods. The supply of money is controlled by a monetary authority, and follows the rule $M_{t+1} = (1+\mu)M_t$, with $\mu > \beta - 1$. New money is introduced if $\mu > 0$, or withdrawn if $\mu < 0$, via lump-sum transfers 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. Using the Fisher equation, we summarize the money growth rate by $i = (1 + \mu - \beta)/\beta$; this rate will be a useful benchmark as the yield on a completely illiquid asset. (Thus, i should not be thought of as representing the yield on T-bills; see Geromichalos and Herrenbrueck, 2022.)

There are also two types of assets, asset A and asset B. These are one-period, nominal bonds with a face value of one dollar; their supply is exogenous and denoted by S_A, S_B. Asset $j = \{A, B\}$ can be purchased at price p_j in the CM, which we think of as the primary market. After leaving the CM agents receive an idiosyncratic consumption shock (discussed below) and may trade these assets before maturity in a secondary OTC market. Each asset j trades in a distinct secondary market, which we dub OTC$_j$, $j = \{A, B\}$. To make things tractable, we assume that agents can only hold asset A or asset B, and can visit only one OTC market per period. (In Section 2.1 we discuss this assumption in detail, and we relax it in Web Appendix C.) Thus, we say they "specialize" in asset A or B. However, agents are free to choose any quantity of money and the asset of their choice.

The economy is characterized by two shocks, both of which are revealed after the CM closes and before the OTC round of trade opens. The first is an aggregate shock that determines whether asset B will default or not in that period.[6] More precisely, we define

[6] The timing adopted here, i.e., assuming that the aggregate shock is revealed before agents go to the OTC, captures the idea that when an asset defaults, trade in its secondary market will also be impeded, which will affect its liquidity. Exploring this link between default and asset liquidity is one of the main goals of our paper. Thus, assuming that agents find themselves in the secondary round of trade after

the aggregate state $s = \{n, d\}$ (n for "normal" and d for "default"); with probability π, $s = n$ and each unit of asset B pays the promised dollar, but with probability $1 - \pi$, $s = d$ where asset B defaults and pays nothing. Throughout the paper we assume that asset A is a perfectly safe and default-free asset.[7] This aggregate default shock is *iid* across time.

The second shock is an idiosyncratic consumption shock determining whether an agent will have an opportunity to consume in the DM. Only a fraction $\ell < 1$ of agents obtains this opportunity. Thus, a measure ℓ of agents will be of type C ("Consuming"), and a measure $1 - \ell$ of agents will be of type N ("Not consuming"). This shock is *iid* across agents and time. Since the various types are realized after agents have made their portfolio choices in the CM, N-types will typically hold some cash that they do not need in the current period, and C-types may find themselves short of cash, since carrying money is costly. Thus, in the OTC market, C-types will play the role of assets *sellers* and N-types will be the asset *buyers*. The goal of this market is to reallocate money into the hands of the agents who need it most, i.e., the C-types.[8]

As we have discussed, agents can only trade in one OTC market per period, and they will choose to trade in the market where they expect to find the best terms. Suppose that a measure C_j of C-types and a measure N_j of N-types have chosen to trade in the market for asset $j = \{A, B\}$. Then, the matching technology:

$$f(C_j, N_j) = \left(\frac{C_j N_j}{C_j + N_j}\right)^{1-\eta} (C_j N_j)^{\eta}, \quad \eta \in [0, 1], \tag{1}$$

determines the measure of successful matches in OTC$_j$. The suggested matching function satisfies $f(C, N) \leq \min\{C, N\}$, and is useful because it admits both constant and increas-

learning that asset B has defaulted better corresponds to our research question.

[7] Our results are robust to different model specifications. For instance, modeling asset A as a default-free asset is done for simplicity and because many real-world assets characterized as AAA are virtually default free. However, all one needs is that asset A defaults with a lower probability than asset B. To simplify the presentation in the main text, we assume for now that when asset B defaults, it defaults completely; however, we derive the model with partial default (fraction $\rho < 1$ of asset B's face value can be recovered) in Web Appendix A.8, and we account for this in the calibration.

[8] The first paper to incorporate this idea into the LW framework is Berentsen, Camera, and Waller (2007), but there the reallocation of money takes place through a competitive banking system.

ing returns to scale (CRS and IRS, respectively) as subcases: when $\eta = 0$, the matching technology features CRS, while $\eta > 0$ implies IRS. Within each successful match the buyer and seller split the available surplus based on proportional bargaining (Kalai, 1977), with $\theta \in (0, 1)$ denoting the seller's (C-type's) bargaining power.[9] Notice that the matching technology and the bargaining protocol are identical in both OTC markets. This guarantees that any differences in liquidity between assets A and B will be driven by differences in safety, and not by exogenous market characteristics (see footnote 3).

Since all the action of the model takes place in the CM and, more importantly, the OTC markets, 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 they make a take-it-or-leave-it offer.

Figure 1 summarizes the timing of events and the important economic actions of the model. A few details are worth emphasizing. First notice that agents who turn out to be C-types are *committed* to visit the OTC market of the asset they chose to specialize in. (One cannot sell asset B in OTC$_A$.) However, this is not true for N-types: an agent who turns out to be an N-type can visit either OTC market, because *her money is good* to buy any type of assets. This has an important consequence. In the default state (see Figure 1 (b)), OTC$_B$ will shut down so *all* N-types will rush into OTC$_A$. And what about the agents who specialized in asset B and turned out to be C-types? Unfortunately, they must proceed to the DM only with the money that they carried from the CM. But it is important to remember that agents are aware of this possibility and may choose to hold asset B anyway. Part of what makes this choice optimal is that they may pay a low(er) price for asset B and choose to carry more money as a precaution.

[9] The proportional bargaining solution is more tractable than the Nash solution. Moreover, in recent work, Rocheteau, Hu, Lebeau, and In (2021) solve a sophisticated model of bargaining with strategic foundations, and find that, under fairly general conditions, their solution converges to the proportional one.

2.1 Discussion of modeling assumptions and microfoundations

A key assumption in our analysis that makes our model tractable and enables us to get analytical results is that secondary markets are segmented and agents must specialize in one asset in the CM. The first part of this assumption is certainly realistic: different classes of assets trade in distinct secondary markets, and most fixed income dealers do not intermediate multiple kinds of securities, since there is a cost to becoming an expert in a specific security. (For instance, in investment banks different "desks" are in charge of trading different types of securities, so they are committed ex-ante to a specific class of assets.) The second part of the assumption, according to which agents must specialize in one asset in the CM, is just a stark way to capture the idea that even if investors can hold multiple assets and visit multiple secondary markets, they will visit more frequently the market where they expect to find better trading conditions.[10]

There is an extensive empirical finance literature reporting that asset markets are segmented from one another (Allen and Gale, 1994; Edmond and Weill, 2012). Many of these papers study specific examples of assets that have (almost) identical characteristics, yet are traded in distinct secondary markets with different degrees of liquidity, and thus sell at different prices (e.g., Warga (1992) for "on-the-run" vs "off-the-run" Treasury bonds that have identical payouts; and Andreasen, Christensen, and Riddell (2021) for Treasury Inflation Protected Securities, which sell at a significant discount compared to common Treasuries after controlling for expected inflation).

Thus, adopting a model with some market segmentation is not only theoretically interesting, but also empirically relevant. However, one may still wonder whether we need this strong version of market segmentation, whereby each agent can only hold one asset at a time, and we investigate this question in several steps. First, in Section 4.3, we discard the segmentation assumption altogether and assume that agents can trade both

[10] This assumption is analogous to a discrete choice model; such models are popular in economics (especially in Industrial Organization) because they offer a tractable way of modeling individual consumer choice over various goods while permitting a realistic description of aggregate market shares.

assets in a consolidated secondary market. We show that market segmentation is not necessary for our first result (safety implies liquidity), but it is essential for our second result (a possible "disconnect" between asset safety and liquidity). Then, in the Web Appendix, we consider versions of the model with weaker segmentation assumptions.

In Web Appendix C.1, we relax the assumption that each agent must hold only one asset, and replace it with the assumption that they can hold both assets but visit only one OTC market per period. This allows agents to "specialize" in asset B (and visit OTC$_B$ in the normal state), while carrying some asset A for precautionary motives, i.e., with an intention to sell it in OTC$_A$ in the case that asset B defaults. We find that despite giving agents the option to carry asset A as a precaution, they choose to not exercise it. The main reason is that asset A does not provide cheap insurance; it carries a hefty liquidity premium. Moreover, agents in our baseline model already had a way of insuring themselves against the default shock: they could bring money.

In Web Appendix C.2, we develop a model with two distinct *marketplaces*. Agents can pay an ex-ante cost, κ_1, and visit the "segmented marketplace", which consists of two submarkets, OTC$_A$ and OTC$_B$. Once in this marketplace, agents must choose to visit either OTC$_A$ or OTC$_B$, exactly as in our baseline model. Alternatively, agents can pay an ex-ante real cost, κ_2, and visit a "consolidated marketplace", where they can trade both assets A and B in a single OTC market. Since the consolidated marketplace offers agents the option to trade both assets, one may expect that all agents would be attracted there. Our results show this is not the case because our model also contains a force that makes the segmented marketplace desirable. The intuition is as follows. As an asset seller the agent may (or may not) prefer the consolidated marketplace because it offers a form of insurance in case asset B defaults. However, as an asset buyer the agent always prefers the segmented marketplace because it offers her the *ex-post* flexibility of visiting either submarket, and, crucially, avoiding OTC$_B$ in the default state. In contrast, in the consolidated marketplace, the whole market, including the N-types, is subject to the default of asset B, and there is no way to avoid it.

In sum, we show that the strong market segmentation adopted in the baseline model as an *assumption* arises as an *endogenous equilibrium outcome* in a more general model where agents are given the option to avoid such segmentation, and is consistent with empirically plausible parameters.

3 Analysis of the Model

3.1 Value functions, bargaining, and matching probabilities

In order to streamline the analysis, we relegate the derivations of the value functions and the solutions of the various bargaining problems to Web Appendix A.1-A.2. Here is a summary of this analysis. As is standard in models that build on LW, all agents have linear value functions in the CM, a result that follows from the (quasi) linear preferences. This makes the bargaining solution in the DM easy to characterize. Consider a DM meeting between a producer and a C-type agent who carries m units of money, and define $m^* \equiv q^*/\varphi$ as the amount of money that (given the price φ) allows the agent to purchase the first-best quantity, q^*. Then, either $m \geq m^*$, and the buyer can purchase q^*, or $m < m^*$, and she spends all of her money to purchase the amount $q = \varphi m < q^*$.

Next, consider a meeting in OTC$_j$, $j = \{A, B\}$, 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. Since money is costly to carry, in equilibrium we will have $m < m^*$, and the C-type will want to acquire the amount of money that she is missing in order to reach m^*, namely, $m^* - m$. Whether she will be able to acquire that amount of money depends on her asset holdings. If her asset holdings are enough (of course, how much is "enough" depends on the bargaining power θ), then she will acquire exactly $m^* - m$ units of money. If not, then she will give up all her assets to obtain an amount of money $\xi(m, d_j) < m^* - m$, which is increasing in d_j (the more assets she has, the more money she can acquire) and decreasing in m (the more money she carries, the less she needs to acquire through OTC trade). This last case, where assets are scarce, is especially interesting, because it is precisely then that having a few more assets would have allowed the agent to alleviate the binding cash constraint, which is why an asset price will carry a liquidity premium.

This discussion assumes that $m + \tilde{m} \geq m^*$, i.e., that the money holdings of the C-type and the N-type pooled together are enough to allow the C-type to purchase the first best quantity q^*. We restrict attention to this case for two reasons. First, our calibration in Section 5 reveals that this case is the relevant one (see footnote 25). Second, the equilibrium associated with the complementary case, where $m + \tilde{m} < m^*$, would imply that none of the assets carry a positive liquidity premium, a result which is clearly unrealistic. The reason is simple: when $m + \tilde{m} < m^*$, the money of the two agents pooled together is so scarce that asset holdings become relatively plentiful. Given these two points, we choose to restrict attention to $m + \tilde{m} \geq m^*$. This simplifies the analysis since it implies that we do not need to keep track of N-types' money holdings.

One of the innovations of our paper is that the OTC matching probabilities are a function of agents' entry choices. Let $e_C \in [0, 1]$ be the fraction of C-type agents who specialize in asset A and are thus committed to trading in OTC$_A$, no matter the eventual aggregate state. And let $e_N^s \in [0, 1]$ be the fraction of N-type agents who enter OTC$_A$ in aggregate state $s = \{n, d\}$. (These terms will be carefully described in Section 3.4.) Then, in state s, $e_C\ell$ is the measure of C-types and $e_N^s(1-\ell)$ is the measure of N-types who enter OTC$_A$. Similarly, $(1-e_C)\ell$ is the measure of C-types and $(1-e_N^s)(1-\ell)$ is the measure of N-types who enter OTC$_B$. As a result, we need to track eight different matching probabilities α_{ij}^s, of an i-type who enters OTC$_j$ in state s (though $\alpha_{CB}^d = \alpha_{NB}^d = 0$ will be trivially true in the full-default version of our model).

The matching probability of a C-type who enters OTC$_A$ in a non-default state provides a representative example:

$$\alpha_{CA}^n \quad \equiv \quad \frac{f[e_C\ell, e_N^n(1-\ell)]}{e_C\ell} \quad = \quad \left[\frac{e_N^n(1-\ell)}{e_C\ell + e_N^n(1-\ell)}\right] \cdot [e_C\ell + e_N^n(1-\ell)]^\eta.$$

It is the product of two terms: a standard *market tightness* term whereby an asset seller benefits from asset buyers entering on the other side of the market but is congested by asset sellers entering on the same side, and a *market size* term whereby any market participant benefits from any other market participant's presence. This market size term is governed by the elasticity η, and in particular it is shut down when $\eta = 0$.

3.2 Optimal portfolio choice

As is standard in models that build on LW, all agents choose their optimal portfolio in the CM independently of their trading histories in previous markets. In our model, 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 the shocks have been realized. The agent's choice can be analyzed with an objective function, $J(\hat{m}, \hat{d}_A, \hat{d}_B)$, which we derive in Web Appendix A.3 and reproduce here for convenience:

$$\begin{aligned} J(\hat{m}, \hat{d}_A, \hat{d}_B) \equiv & -\varphi(\hat{m} + p_A \hat{d}_A + p_B \hat{d}_B) + \beta\hat{\varphi}(\hat{m} + \hat{d}_A + \pi \hat{d}_B) \\ & + \beta\ell\Big(u(\hat{\varphi}\hat{m}) - \hat{\varphi}\hat{m} + \pi \max\{\alpha^n_{CA}\mathcal{S}_{CA}, \alpha^n_{CB}\mathcal{S}_{CB}\} + (1-\pi)\alpha^d_{CA}\mathcal{S}_{CA}\Big), \end{aligned}$$

where $\mathcal{S}_{Cj}$ is the surplus of an agent who turns out to be a C-type and trades in OTC$_j$:

$$\mathcal{S}_{Cj} = u(\hat{\varphi}(\hat{m} + \xi_j(\hat{m}, \hat{d}_j))) - u(\hat{\varphi}\hat{m}) - \hat{\varphi}\chi_j(\hat{m}, \hat{d}_j).$$

In the above expression, ξ_j stands for the amount of money that the agent can acquire by selling assets, and χ_j stands for the amount of assets sold in OTC$_j$, $j = \{A, B\}$.

The interpretation of J is straightforward. The first term is the cost of choosing the portfolio $(\hat{m}, \hat{d}_A, \hat{d}_B)$. This portfolio yields the expected payout $\hat{\varphi}(\hat{m} + \hat{d}_A + \pi\hat{d}_B)$ in next period's CM (the second term of J). The portfolio also offers certain liquidity benefits, but these will only be relevant if the agent turns out to be a C-type; thus, the term in the second line of J is multiplied by ℓ. The C-type can enjoy at least $u(\hat{\varphi}\hat{m}) - \hat{\varphi}\hat{m}$ just with the money she brought from the CM. Furthermore, she can enjoy an additional benefit by selling assets for cash in the secondary market. How large this benefit is depends on the market choice of the agent (the term inside the $\max$ operator) and on the realization of the aggregate shock: if asset B defaults, an event that happens with probability $1 - \pi$, a C-type who specialized in that asset has no benefit. A default of asset B is not the only reason why the C-type may not trade in the OTC markets; it may just be that she did not match with a trading partner. This is why the various surplus terms $\mathcal{S}_{Cj}$ are multiplied by the α-terms, i.e., the matching probabilities discussed in Section 3.1.[11]

3.3 Equilibrium

We focus on steady-state equilibria. Before we move on to characterizing possible equilibria, we first need to understand their structure. We have twelve endogenous variables to be determined in equilibrium (not including the terms of trade in the OTCs):

- prices: $\varphi,\, p_A,\, p_B$
- real balances: $z_A,\, z_B$
- entry choices: $e_C(\equiv e_C^n = e_C^d),\, e_N^n,\, e_N^d$
- DM production: $q_{0A}(\equiv q_{0A}^n = q_{0A}^d),\, q_{1A}(\equiv q_{1A}^n = q_{1A}^d),\, q_{0B}(\equiv q_{0B}^n = q_{0B}^d),\, q_{1B}(\equiv q_{1B}^n)$

In this list of equilibrium variables, the asset prices are obvious, and z_j, $j = \{A, B\}$, is simply the real balances held by an agent who chooses to specialize in asset j. The remaining terms deserve some discussion. First, notice that the fraction of C-types who enter OTC$_A$, e_C, does not depend on the aggregate state $s = \{n, d\}$. This is because C-types are committed to visiting the OTC market of the asset they chose to specialize in (and this choice is effectively made before the realization of the shock).

Regarding the DM production quantities q_{kj}, $k = \{0, 1\}$ indicates whether the C-type did ($k = 1$) or did not ($k = 0$) trade in the preceding OTC market, and $j = \{A, B\}$ indicates the asset in which she specializes. For example, q_{0A} is the amount of DM good purchased by an agent who specialized in asset A and did not match in OTC$_A$, and so

[11] There are two reasons why the objective function does not contain any term that represents the event in which the agent is an N-type. First, and most obviously, N-types are defined as the agents who do not get to consume in the DM. Second, the OTC terms of trade, χ and ξ, depend only on the portfolio of the C-type. An intuitive explanation was presented in Section 3.1; for the details, see Web Appendix A.2.

on. These quantities do not depend on the aggregate state $s = \{n, d\}$. To see why, notice that q_{0A} depends only on the amount of real balances that the agent carried from the CM (this agent did not trade in the OTC), and that choice was made before s was realized. The same reasoning applies to q_{0B}. How about the term q_{1A}? This term depends on the real balances that the agent carried from the CM (which, we just argued, is independent of the shock realization), and on the amount of assets this agent carries from the CM (see Section 3.1). How many assets does this agent carry? The answer is S_A/e_C: the exogenous asset supply, S_A, divided by the measure of agents who specialize in asset A. Since S_A is a parameter, and e_C is independent of the state s, the same will be true for the term q_{1A}.

To simplify the exposition of the equilibrium analysis, we focus on the block of variables $\{q_{0A}, q_{1A}, q_{0B}, q_{1B}, e_C, e_N^n\}$, which we refer to as the "core" variables. In Web Appendix A.6, we establish that the remaining variables, $\{z_A, z_B, \varphi, p_A, p_B\}$, follow immediately once the core variables have been determined. Of course, e_N^d is always equal to 1. To determine our six core variables, we proceed as follows.

First, we have two money demand equations for agents who specialize in asset A and B: Equations (A.14)-(A.15) in Web Appendix A.6, derived in Web Appendix A.5. What is important to remember here is that agents who choose to specialize in different assets will typically carry different amounts of money; not surprisingly, agents who choose to carry the less safe asset B self-insure against the probability of default (and the shutting down of OTC_B) by carrying more money.

Next, we use asset market clearing to solve for the outcome of the OTC bargaining protocol in a form that involves only core equilibrium variables: Equations (A.16)-(A.17) in Web Appendix A.6. Intuitively, they state that if the agent's asset holdings are large, then $q_{1j} = q^*$, because the agent will acquire (through selling assets) the money necessary to purchase the first-best quantity, and no more. In contrast, if the agent's asset holdings are scarce, she will give up all her assets and purchase an amount of DM good equal to q_{0j} (the amount she could have purchased without any OTC trade) plus the additional amount she can now afford by selling assets for extra cash.

Our last two equilibrium conditions come from the optimal OTC market entry decisions of agents. An important remark is that the OTC surplus of N-types does not depend on their portfolios (see Section 3.1 or A.2), whereas the OTC surplus of C-types does depend on their portfolios. Hence, in making their entry decisions, C-types consider not only the expected surplus of entering in either market, as is the case for N-types, but also the cost associated with each entry decision. Another way of stating this is to say that e_C is determined ex ante and represents the decision to specialize in asset A, while e_N^n is determined ex post and represents the fraction of N-types who enter OTC$_A$ in the normal state. Therefore, the optimal entry of C-types is characterized by:

$$e_C = \begin{cases} 1, & \tilde{\mathcal{S}}_{CA} > \tilde{\mathcal{S}}_{CB} \\ 0, & \tilde{\mathcal{S}}_{CA} < \tilde{\mathcal{S}}_{CB} \\ \in [0,1], & \tilde{\mathcal{S}}_{CA} = \tilde{\mathcal{S}}_{CB} \ , \end{cases} \tag{2}$$

where the *ex-ante* surplus terms $\tilde{\mathcal{S}}_{CA}$ and $\tilde{\mathcal{S}}_{CB}$ are defined in Equations (A.18)-(A.19). The optimal entry of N-types is characterized by:

$$e_N^n = \begin{cases} 1, & \alpha_{NA}^n \mathcal{S}_{NA} > \alpha_{NB}^n \mathcal{S}_{NB} \\ 0, & \alpha_{NA}^n \mathcal{S}_{NA} < \alpha_{NB}^n \mathcal{S}_{NB} \\ \in [0,1], & \alpha_{NA}^n \mathcal{S}_{NA} = \alpha_{NB}^n \mathcal{S}_{NB} \ , \end{cases} \tag{3}$$

in the normal state, where the *ex-post* surplus terms $\mathcal{S}_{NA}$ and $\mathcal{S}_{NB}$ are defined in Equations (A.20)-(A.21), and by:

$$e_N^d = \begin{cases} 1, & e_C > 0 \\ \in [0,1], & e_C = 0 \ , \end{cases} \tag{4}$$

in the default state.

We can now define the steady-state equilibrium of the model:

Definition 1. For given asset supplies $\{S_A, S_B\}$, the steady-state equilibrium for the core variables of the model consists of the equilibrium quantities and entry choices, $\{q_{0A}, q_{1A}, q_{0B}, q_{1B}, e_C, e_N^n\}$, such that (A.14), (A.15), (A.16), (A.17), (2) and (3) hold. The equilibrium real balances, $\{z_A, z_B\}$, satisfy (A.10), the equilibrium price of money, φ, solves (A.11), and the equilibrium asset prices, $\{p_A, p_B\}$, solve (A.12) and (A.13).

3.4 Equilibrium market entry

In this section we analyze the optimal entry decision of agents, which is at the heart of our model. To begin with, we fix a value of e_C (the fraction of agents who specialize in asset A and, if they should turn out to be C-types, are thus committed to sell in OTC$_A$), compute the rest of the core equilibrium variables optimally (including the entry decision of N-types in the normal state, $e_N^n(e_C)$), and then evaluate the best response of a representative agent to our initial guess of e_C. This task becomes easier by recognizing that there are three opposing forces at work. We dub them the congestion effect, the coordination effect, and the dilution effect; they are described in detail in Web Appendix A.7, so we focus on an intuitive summary here.

First, a high initial e_C will *discourage* the representative agent from holding asset A because it implies that there will be many sellers in OTC$_A$; holding the measure of buyers constant, it becomes harder to match for each of the sellers. This is the "congestion effect". However, a high e_C also attracts buyers to market A, i.e., it implies a high e_N^n, which in turn improves sellers' matching probability and *encourages* the representative agent to hold asset A; this is the "coordination effect". With CRS the two effects tend to balance out, while with IRS the coordination effect becomes stronger. A more subtle force is the "dilution effect". When e_C is high, many agents specialize in asset A, and each one of them carries a small fraction of the (fixed) asset supply. As a result, the surplus generated per meeting in OTC$_A$ will be small, another force that *discourages* the representative agent from holding asset A when many others already do so.

Moving to the formal analysis, we construct equilibria as fixed points of e_C. To be specific: first, we fix a level of e_C; then we solve for the optimal portfolio choices through Equations (A.14)-(A.17) and (3); and finally, we define the C-types' *best response function*:

$$G(e_C) \equiv \tilde{\mathcal{S}}_{CA} - \tilde{\mathcal{S}}_{CB},$$

where the surplus terms have the optimal choices substituted. This function measures the relative benefit to an *individual* C-type from specializing in asset A over asset B, assuming a proportion e_C of all *other* C-type agents specialize in A, and all other decisions (portfolios and entry of N-types) are conditionally optimal. We say that a value of e_C is part of an "interior" equilibrium if $e_C \in (0,1)$ and $G(e_C) = 0$, or a "corner" equilibrium if $e_C = 0$ and $G(0) \leq 0$ (B-corner) or $e_C = 1$ and $G(1) \geq 0$ (A-corner).

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

(a) There exists a corner equilibrium where $e_C = 0$, $e_N^n = e_N^d = 0$.

(b) There exists a corner equilibrium where $e_C = 1$, $e_N^n = e_N^d = 1$.

(c) Assume $\eta = 0$ *(CRS) and asset supplies are low enough so that assets are scarce in OTC trade. Then,* $\lim_{e_C \to 0+} G(e_C) > 0 > G(0)$*; the equilibrium at the* B*-corner is not robust to small trembles.*

(d) Assume $\eta = 0$ *(CRS) and asset supplies are low enough so that assets are scarce in OTC trade. Then,* $\lim_{e_C \to 1} G(e_C) < G(1)$*. If* $\pi \to 1$*, then the limit is negative, and the equilibrium at the* A*-corner is not robust, either.*

(e) Assume $\eta = 0$ *(CRS),* $\pi \to 1$*, and asset supplies are low enough so that assets are scarce in OTC trade. Then, there exists at least one interior equilibrium which is robust.*

(f) Given $\eta > 0$ *(IRS),* $\lim_{e_C \to 0+} G(e_C) \neq G(0)$.

(g) Given $\eta > 0$ *(IRS),* $\lim_{e_C \to 1-} G(e_C) = G(1) > 0$*; the equilibrium at the* A*-corner is robust.*

Proof. See Web Appendix B.1. □

Figures 2 and 3 illustrate these results for the CRS and the IRS case, respectively. The left panel of each figure depicts the individual C-type's best response function, $G(e_C)$. Since this function depends not only on the behavior of fellow C-types, but also on that of N-types, on the right panel of each figure we show the optimal entry choice of N-types, e_N^n, as a function of e_C. The figures also illustrate how equilibrium entry is affected by changes in the supply of asset A, keeping the supply of asset B constant.

As indicated in the right panel of each figure, we have $e_N^n(0) = 0$ and $e_N^n(1) = 1$: when all C-types are concentrated in one market, the N-types will follow. Generally, the higher e_C is, the more N-types would like to go to market A: this is just the coordination effect and it tends to make $e_N^n(e_C)$ increasing. Whether it will be strictly increasing or not, ultimately depends on the strength of the dilution effect relative to the coordination effect. This is why in both figures, $e_N^n(e_C)$ is increasing when A is large: it is a large asset supply that weakens the dilution effect.[12]

Next, we have $G(0) < 0$ and $G(1) > 0$, while $e_N^n(0) = 0$ and $e_N^n(1) = 1$. This illustrates parts (a) and (b) of Proposition 1; the corners are always equilibria (marked with circles on the left panel of the figures). However, with CRS these equilibria are not robust because G is discontinuous at the corners; this illustrates parts (c) and (d) of the proposition.[13] Also, with CRS the congestion effect is so dominant that the G-function is globally decreasing in the interior, as shown in part (e) of the proposition and illustrated in Figure 2. Therefore, there exists a robust interior equilibrium where the representative C-type is indifferent between entering market A or B; i.e., $G(e_C) = 0$. As the supply of asset A increases, so does the equilibrium value of e_C, because a larger asset supply weakens the dilution effect

[12] There is also a difference between the two figures. In Figure 2 (CRS case), $e_N^n(e_C)$ is strictly increasing in its entire domain. However, in Figure 3 (IRS case), and for the case of large S_A, $e_N^n(e_C)$ reaches 1 for a rather small value of e_C and becomes flat afterwards. This is because with IRS, the desire of N-types to go to the market with many C-types, i.e., the coordination effect, is supercharged.

[13] More precisely, they are not "trembling hand perfect" Nash equilibria. Consider for example the equilibrium with $e_N^n = e_C = 1$ (a similar argument applies to the one with $e_N^n = e_C = 0$). Since all N-types visit market A, the representative C-type also wishes to visit that market. (Why try to trade in a ghost town, which OTC_B is in this case?) However, if an arbitrarily small measure ε of N-types visited market B by error, the representative C-type would have an incentive to deviate to market B, where her chance of matching is now extremely high (since $e_C = 1$, she would be the only C-type in that market).

and increases the incentives of agents to concentrate on market A.

Moving on to the IRS case, the two corner solutions are still equilibria, and since IRS strengthen the coordination effect, the equilibrium where all agents go to OTC$_A$ ($e_N^n = e_C = 1$) is now robust (part (g) of the proposition). This may or may not be true for the B-corner ($e_N^n = e_C = 0$), depending on the values of π and η.[14] Figure 3 demonstrates the case of non-robustness; as shown in part (f) of the proposition, the best response function is discontinuous at the B-corner, though it is continuous at the A-corner. With the coordination effect amplified, multiple interior equilibria are typical (as in the case of "small S_A" and "medium S_A"). However, the only robust interior equilibrium is the one where G has a negative slope. A rise in S_A will lead to an increase in the (interior and robust) equilibrium value of e_C. But with IRS, another interesting possibility arises: if S_A is large enough, the desire of agents to coordinate on OTC$_A$ is so strong that interior equilibria cease to exist. This is depicted in the "large S_A" case, where one can see that the A-corner (with $e_N^n = e_C = 1$) is the unique robust equilibrium entry outcome.

3.5 Liquidity premia

Most of our main results will be about the *liquidity premia* assets A and B may carry. In equilibrium, asset prices consist of a "fundamental value" multiplied by a premium that reflects the possibility of selling the asset in the OTC market. We define the fundamental value of an asset as the equilibrium price that would emerge if this possibility was eliminated. In that case, agents would value the assets only for their payouts at maturity, and the equilibrium prices would be given by $1/(1+i)$, for asset A, and $\pi/(1+i)$, for asset B.

[14] Consider first the equilibrium with $e_N^n = e_C = 1$. With IRS the desire to go to OTC$_A$ (where all agents are concentrated) is so strong that, even if some N-types visit OTC$_B$ by error, the representative C-type no longer has an incentive to deviate to that market (unlike the CRS case; see footnote 13). But the channel described so far is relevant for both corners. So why is the equilibrium where all agents go to OTC$_B$ not always robust as well? Because OTC$_B$ is the market of the asset that may default. When that happens (ex-post), all N-types will rush to market A, i.e., $e_N^d = 1$, and this creates an additional incentive for the representative C-type to deviate to market A (a decision made ex-ante). This additional incentive will be relatively large, when π is low (high default probability) and η is low (weak coordination effect). Therefore, the equilibrium with $e_N^n = e_C = 0$ is likely to be non-robust for relatively low values of π and η.

The liquidity premium of asset j, denoted by L_j, is therefore defined as the percentage difference between an asset's price and its fundamental value:

$$p_A = \frac{1}{1+i}(1+L_A), \qquad\qquad p_B = \frac{\pi}{1+i}(1+L_B), \tag{5}$$

where p_A and p_B are described by the asset pricing equations (A.12) and (A.13), and:

$$\begin{aligned} L_A &= \ell \cdot \left(\pi\alpha^n_{CA} + (1-\pi)\alpha^d_{CA}\right) \cdot \frac{\theta}{\omega_\theta(q_{1A})} \cdot (u'(q_{1A}) - 1), \\ L_B &= \ell \cdot \alpha^n_{CB} \cdot \frac{\theta}{\omega_\theta(q_{1B})} \cdot (u'(q_{1B}) - 1). \end{aligned} \tag{6}$$

Each liquidity premium is the product of four terms. First, the probability that an agent turns out to be a C-type and thus needs liquidity at all (ℓ). Second, given that the agent is a C-type, the expected probability of matching in the respective OTC market, conditional on entering that market. Third, the share of the marginal surplus captured by the C-type (θ/ω_θ), which is endogenous but constrained to the interval $(0, \theta]$. And fourth, the marginal surplus of the match: the utility gained by a consumer who brings one more unit of real balances into the DM, net of the production cost ($u'(q_{1j}) - 1$).

Thus, there are two ways a liquidity premium can be zero: either the relevant OTC market is closed ($\alpha_{Cj} = 0$), or assets are so plentiful that selling an extra asset in the OTC does not create additional surplus in the DM ($q_{1j} = q^*$, thus $u'(q_{1j}) = 1$). In the latter case, the asset is still "liquid", but its liquidity is *inframarginal* so it does not affect the price.

4 Main Results

4.1 Result 1: Safe and liquid

The first result of the paper is that, other things equal, the safer asset (A) tends to be more liquid. We demonstrate this result adopting the liquidity premium as the measure of liquidity. However, in Web Appendix B.4, we confirm that this result remains valid if one adopts trade volume in the OTC market as the measure of asset liquidity (this approach is common in the finance literature). Throughout Section 4.1, we assume that the supplies of the two assets are equal ($S_A = S_B$), in order to focus on liquidity differences purely due to the assets' safety differential. Because of the complexity of our model, we break our analysis into two stages.[15] First, we take a local approximation of our model around $\pi \to 1$, assuming CRS ($\eta = 0$). The perturbation of this specification with small changes in π can be solved in closed form (see Web Appendix B.2). Second, in order to obtain global results away from $\pi \to 1$, and to conduct comparative statics with respect to η, we solve the model numerically. The following proposition summarizes our analytical results:

Proposition 2. *Assume that asset supplies S_A and S_B are equal and are low enough so that assets are scarce in OTC trade. Then:*

(a) At $\pi = 1$, there exists a symmetric equilibrium where $e_C = e_N = 0.5$, $q_{0A} = q_{0B}$, $q_{1A} = q_{1B}$, and $L_A = L_B$.

(b) Assume $\eta = 0$ (CRS) and $(1 - \ell)\theta$ is sufficiently large. Then, locally, $\pi < 1$ implies $L_A > L_B$: the safer asset is more liquid.

Proof. See Web Appendix B.2. □

[15] Our model has six 'core' equilibrium variables, most of which show up in multiple equations; these equations are non-linear and include kinks, due to the various branches of the bargaining solutions and the agents' market entry decisions. Simply put, every time a parameter value changes, all six endogenous variables are affected by simultaneous and, typically, opposing forces. For more detail, one can inspect matrix equation (B.8) in the Web Appendix, which describes the effect of changes in π on the core variables in general equilibrium, keeping in mind that this matrix is evaluated at the limit as $\pi \to 1$.

Naturally, when $\pi = 1$, the two assets are perfect substitutes and their equilibrium prices (and liquidity premia) will be equal. However, as π falls below 1, the liquidity premium of asset A generally exceeds that of asset B. Near the symmetric equilibrium, the derivative of the difference between the liquidity premia with respect to π is:

$$\left.\frac{d(L_A - L_B)}{d\pi}\right|_{\pi\to 1} = \ell\theta\frac{u'(q_1)-1}{w_\theta(q_1)}(\alpha^n_{CA} - \alpha^d_{CA}) \ \ldots$$
$$+ \ \ell\theta\frac{u''(q_1)}{w_\theta(q_1)^2} \times \frac{d(q_{1A} - q_{1B})}{d\pi} \ + \ \ell\theta\frac{u'(q_1)-1}{w_\theta(q_1)} \times \frac{d(\alpha^n_{CA} - \alpha^n_{CB})}{d\pi}$$

The first term on the right-hand side represents a negative *direct effect*: the probability of meeting a buyer for asset A is always lower in the normal state than in the state where B defaults ($\alpha^n_{CA} < \alpha^d_{CA}$), therefore the liquidity advantage of asset A increases as B becomes less safe ($\pi \downarrow$, i.e., $d\pi < 0$). But this liquidity advantage is magnified by the endogenous responses of agents to *perceived* default risk, which affect what happens even in the normal state. Consider the second term in the equation. An agent who specializes in asset B despite the default risk will self-insure by carrying more money, which translates (after OTC trade) to a higher q_{1B}, resulting in a lower marginal utility of selling asset B (indicated by multiplication with $u''(q_1) < 0$) and thus a lower liquidity premium for asset B; thus, this *intensive margin effect* is also negative and always reinforces the direct effect.

Finally, there is the the third term in the equation, representing an *extensive margin effect*: generally, when $\pi < 1$, N-types respond more strongly to the lower trading surplus in the B-market, thus the matching probability for C-types is higher in OTC$_A$. If so, then all three effects point in the same direction and thus the overall sign of the equation is negative, as per part (b) of Proposition 2. Analytically, we can show that this is indeed the case when $(1-\ell)\theta$ is sufficiently large; numerically, we can find counterexamples, but the overall negative sign is still the predominant result.[16]

[16] To be precise, we checked the sign for all combinations of θ and ℓ in $\{0.1, 0.5, 0.9\}$, and asset supplies of $S_A = S_B \in \{.02, .05, .10, .15\}$, with $\eta = 0$, $i = .1$, and $M = 1$ maintained. Out of these 36 parameter combinations, in four of them the assets are so plentiful that both liquidity premia are zero for any π; in

Figure 4 illustrates our result for a range of π, and for both CRS and an intermediate degree of IRS. In each of these cases, the difference between L_A and L_B is positive and strictly decreasing in π. It is important to remind the reader that this differential is purely due to liquidity; it is not a risk premium. Indeed, decreasing π makes agents less willing to hold asset B because that asset is now at higher risk of default, but that effect is already included in the fundamental value of the assets (see Equations 5). The new result here is that as asset B becomes less safe it also enjoys a smaller liquidity premium on top of the smaller fundamental value.

The intuition behind Result 1 is as follows. Unlike C-types, who are committed to visit the market of the asset in which they chose to specialize, N-types are free to visit any market they wish, since their money is good to buy any asset. Consequently, *if asset B defaults*, all the N-types (even those who had chosen to specialize in asset B) will rush into OTC$_A$. But there is also a more subtle force at work: precisely because asset B may default, agents who specialize in that asset B endogenously choose to carry more money as insurance. Consequently, asset B sellers need a smaller liquidity boost, which means that the surplus generated from asset trades in OTC$_B$ is lower. This is a force that attracts more N-types to OTC$_A$ *even if asset B does not default*. Taken together, these two forces imply that *ex-post* (and regardless of the default shock realization) more buyers are attracted to OTC$_A$. This, in turn, incentivizes agents to specialize in asset A *ex-ante*, as they realize that in this market they will have a high trade probability, if they turn out to be sellers.

The discussion following Equations (6) reveals why this is important for liquidity: an agent who buys an asset today (in the primary market) is willing to pay a higher price if she expects that it will be easy to sell that asset 'down the road', and, importantly, it is the C-types who sell assets down the road. Through this channel, any positive default probability for asset B translates into a matching advantage for C-types in OTC$_A$. This,

three of them, all with maximal ℓ and minimal asset supplies, the sign is reversed so that $L_A < L_B$ when $\pi < 1$, i.e., the safer asset is less liquid; in the remaining 29 cases, we have the 'normal' result where the safer asset is more liquid. For more details, see Web Appendix B.2.

in turn, translates into a higher liquidity premium for asset A, because that premium depends on the (anticipated) ease with which the agent can sell the asset if she turns out to be a C-type. Naturally, this channel, and the liquidity differential between the two assets, will be magnified if matching is characterized by IRS.

4.2 Result 2: Safer yet less liquid

The previous section established that a safer asset will also be more liquid – other things being equal. Of course, other things are not always equal, and of those we are particularly interested in asset supplies. Allowing for differences in asset supplies delivers the second important result of the paper: even with slight IRS in OTC matching, the coordination channel becomes so strong that asset demand curves can be upward sloping. Consequently, asset B can carry a higher liquidity premium than the safe asset A, as long as the supply of the former is sufficiently larger than that of the latter.

Figure 5 depicts the liquidity premia for assets A and B as functions of the supply S_B, keeping S_A fixed, and for various degrees of IRS in matching. First, notice that the liquidity premium on asset A is always decreasing in S_B. With CRS (top-left panel), this is also true for the liquidity premium on asset B, as is standard in existing models of asset liquidity. However, with even a small degree of IRS, the asset demand curves have upward-sloping segments. And if S_B exceeds S_A by a large enough amount, we observe $L_B > L_A$, i.e., the less safe asset emerges as more liquid.

The mechanism of this result is as follows. As we have seen, our model has a channel whereby a safer asset also enjoys an endogenous liquidity advantage. However, whether this advantage will materialize depends on the relative strength of other channels. Consider a change in an asset's supply (say, $S_B\uparrow$): in equilibrium, this will be accommodated by a shift in the proportion of agents holding this asset ($e_C \downarrow$) so that each individual agent, whether holding A or B, can obtain a similar surplus when selling that asset in the corresponding OTC. This is the dilution effect described in the previous section, and as long as we have constant returns in matching the other two effects (congestion and coordination) offset, so this is the end of the story.

With IRS, however, the coordination effect (whereby entry of sellers in an OTC market begets the entry of buyers, and vice versa) is strengthened and dominates the congestion effect (see Web Appendix A.7). What does this mean for the effects of an increase in S_B? Following the logic described above, e_C decreases (more agents hold asset B and thus enter OTC$_B$ as sellers), and following the logic described in Section 3.4, e_N^n decreases too (assuming asset B does not default, more buyers enter OTC$_B$). But with a strong coordination effect, this means that the matching probability increases for *both* sides in OTC$_B$, and decreases for both sides in OTC$_A$, which begets a further shift in entry rates. We illustrate this channel in Figure 6.[17] (In principle, the coordination effect could be so strong as to dominate both the congestion and dilution effects combined, in which case the asset in larger supply would be the only liquid one, irrespective of its safety; however, this only happens for values of η far higher than our empirical study in Section 5 allows.) Thus, with IRS in matching, there is a "thick market" channel which is so strong that the premium an agent is willing to pay for an asset can be increasing in that asset's supply.

Figure 7 summarizes Results 1 and 2. It depicts the liquidity premia of assets A and B as functions of S_B, keeping S_A fixed, with a slight degree of IRS, $\eta = 0.2$. When the supplies of the two assets are equal ($S_B = S_A$), asset A carries a higher liquidity premium (Result 1). However, as S_B increases further, we enter the region where the demand for asset B becomes upward sloping, until eventually L_B surpasses L_A (Result 2).

[17] Close inspection of the figure reveals another interesting effect: the sell-probability always responds more strongly to any impulse than the buy-probability. This reflects the fact that N-types always respond more elastically to any change in market attractiveness, which happens because their market entry choice is made *ex post* and any cost of holding money or assets is already sunk; the C-types, on the other hand, must take the cost of holding assets into account. This fact serves to further *amplify* our result of upward-sloping demand curves, but the result would still hold if we forced $e_C = e_N^n$ in the background.

4.3 Consolidated secondary asset market

In this section, we abandon the segmentation assumption and study a model where agents can trade both assets in a single *consolidated* OTC market. The purpose of this section is to help the reader understand exactly which results the segmentation assumption is responsible for. The main difference between this section and the baseline model is that now there exists a unique consolidated secondary market where agents can liquidate assets. With a unique OTC market, agents have a trivial entry decision, i.e., they all visit the consolidated market. Within that market, the matching technology and the bargaining protocol remain unaltered. The details of the model, including characterization of the equilibrium, are presented in Web Appendix B.3. Here we discuss the main result.

Proposition 3. *We restrict attention to equilibria with positive liquidity premia. In the consolidated market model, regardless of the individual asset supplies, S_A, S_B, it is always true that $L_A > L_B$. Thus, Result 1 holds in the model with a unique consolidated market, but Result 2 fails.*

Proof. See Web Appendix B.3.3. □

Proposition 3 implies that market segmentation is not a necessary condition for Result 1 (other things equal, safety implies liquidity), but it is a necessary condition for Result 2 (a less safe asset could be more liquid). The reason why Result 1 holds here is different, and simpler, than the justification provided in Section 4.1: a liquidity premium reflects downward sloping demand for liquidity, and, in the default state, when there is less liquidity, agents value the remaining liquidity (the one provided by the safe asset A) more. However, since $L_A > L_B$ for *any* asset supplies, the riskier asset B can never be more liquid than asset A. We conclude that some form of segmentation is necessary for the safer asset to be less liquid.

4.4 Result 3: Safe asset supply and welfare

In our final result, we highlight an important implication of our model about the effect of an increase in the supply of safe assets on welfare. A large body of recent literature highlights that the supply of safe assets has been scarce, and that increasing this supply would be beneficial for welfare (see, for example, Caballero et al., 2017). In our model this result is not necessarily true. In particular, welfare may not be monotonic in S_A.

First, let us define the welfare function of this economy, which is the C-type agent's surplus in the DM, averaged between agents who had the opportunity to rebalance their portfolios in the OTC round of trade, and those who did not.[18] Clearly, one also needs to remember that here we have agents who chose to specialize in different assets, and two possible aggregate states (default and no-default). In the normal state, welfare is:

$$\begin{aligned}\mathcal{W}^n = {} & e_C\ell \cdot \left[(1-\alpha^n_{CA})\big(u(q_{0A})-q_{0A}\big)+\alpha^n_{CA}\big(u(q_{1A})-q_{1A}\big)\right] \\ & + (1-e_C)\ell \cdot \left[(1-\alpha^n_{CB})\big(u(q_{0B})-q_{0B}\big)+\alpha^n_{CB}\big(u(q_{1B})-q_{1B}\big)\right],\end{aligned}$$

and in the default state, it is:

$$\mathcal{W}^d = e_C\ell \cdot \left[(1-\alpha^d_{CA})\big(u(q_{0A})-q_{0A}\big)+\alpha^d_{CA}\big(u(q_{1A})-q_{1A}\big)\right] + (1-e_C)\ell \cdot \big(u(q_{0B})-q_{0B}\big).$$

We define aggregate welfare as:

$$\mathcal{W} = \pi\mathcal{W}^n + (1-\pi)\mathcal{W}^d.$$

Figure 8 plots equilibrium welfare as a function of the supply of the safe asset, and highlights the case in which welfare is non-monotonic in S_A. This result may seem surprising at first. A higher supply of asset A enhances the liquidity role of that asset (or,

[18] In models that build on LW, steady-state welfare depends only on the volume of DM trade. Hence, a sufficient statistic for welfare is how close the average DM production is to the first-best quantity, q^*.

equivalently, allows for more secondary market asset trade), which, in turn, should allow agents to purchase more goods in the DM. While not wrong, this intuition is incomplete. What is missing is that when the safe asset becomes more plentiful, agents expect that it will be easier to acquire extra cash ex-post and, thus, they choose to hold less of it ex-ante. In other words, our model is characterized by an externality: agents prefer to carry assets rather than money, and they wish to acquire money in the secondary market(s) only after they have learned that they really need it (i.e., only if they have turned out to be a C-type). But someone has to bring the money, and that someone will not be adequately compensated. This channel depresses the demand for money, which, in turn, decreases the value of money and the volume of trade that the existing money supply can support.

An interesting detail seen in Figure 8 is that welfare always decreases when S_A is large enough. This feature of equilibrium can be explained a follows. As S_A increases, the amount of DM goods purchased by an agent who traded in OTC$_A$, q_{1A}, also increases, because that agent was able to sell more assets and boost her money holdings. On the other hand, as S_A increases, the amount of DM goods purchased by an agent who did not trade in OTC$_A$, q_{0A}, decreases, because the higher asset supply induced that agent to carry fewer money balances *ex-ante* (see previous paragraph). Hence, an increase in S_A generates two opposing effects on welfare: the surplus term $u(q_{1A}) - q_{1A}$ (involving agents who traded in OTC$_A$) increases, but the surplus term $u(q_{0A}) - q_{0A}$ (involving agents who did not trade in OTC$_A$) decreases.[19] While it is hard to know which effect prevails for any value of S_A, what is certain is that if S_A keeps rising, there will come a point where the *marginal* liquidity benefit of more A-assets will be zero (because $q_{1A} \to q^*$ implies $u'(q_{1A}) \to 1$). Near that point, an increase in S_A still hurts welfare by depressing $u(q_{0A}) - q_{0A}$ (because $u'(q_{0A}) \gg 1$), but now it generates no countervailing benefit.

[19] Of course, this is a general equilibrium model where any change in S_A affects not only the terms q_{0A}, q_{1A}, but also the terms q_{0B}, q_{1B}. However, the latter is a secondary effect which turns out to be quantitatively not too important.

5 Quantitative Analysis

5.1 Motivation: The illiquidity of AAA corporate bonds

An interesting fact that has recently drawn the attention of practitioners (but not so much that of academic researchers yet) is that, in the U.S., the virtually default-free AAA bonds are less liquid than (the riskier) AA corporate bonds. Figure 9 plots the time-series yields of AAA versus AA corporate bonds (as well as their difference) on the top panel, and, as a reference point, it does the same for the AAA versus AA municipal bond yields in the bottom panel. The bottom panel is consistent with what one would expect to see: the riskier AA municipal bonds command a higher yield than the one on AAA municipal bonds, because investors who choose to hold the former want to be compensated for their higher default probability.

Interestingly, this logical pattern is reversed in the case of corporate bonds. On the top panel of the figure, we see that in the last 5 years of our sample, the yield on AA corporate bonds has been consistently lower than that on AAA bonds. Why do investors command a higher yield to hold (the virtually default-free) AAA corporate bonds? Many practitioners have claimed that this is so because the secondary market for AAA corporate bonds is extremely illiquid.[20] Our model could shed some light on this empirical observation, if it was the case that AAA corporate bonds have a scarce supply relative to AA corporate bonds. This turns out to be overwhelmingly true. In the years following the financial crisis, regulations introduced to improve the stability and transparency of the financial system (such as the Dodd-Frank Act) have made it especially hard for corporations to attain the AAA score. This resulted in a large decrease in the outstanding supply of this class of bonds. As a benchmark of comparison, in June 2018, the outstanding sup-

[20] This narrative is consistent with the observations depicted in Figure 9, and it is supported by further evidence. For instance, He and Milbradt (2014) document that the bid-ask spread in the market for AAA corporate bonds is higher than the one in the market for AA corporate bonds. Additionally, in recent years Bloomberg has ceased constructing its price index for AAA-rated corporate bonds, due to the dearth of outstanding bonds and the lack of secondary market trading. Of course, a high bid-ask spread and a low trade volume are both strong indicators of an illiquid market.

ply of AAA over AA corporate bonds was 1/10, while the same statistic for municipal bonds was 1/3.

While it is plausible to attribute the irregularity observed on the top panel of Figure 9 to ‘some liquidity story’, existing models of liquidity cannot help us understand this puzzling observation (see a review of the literature in Section 1.1). In these papers, the asset demand curves are decreasing, hence, an asset in large (small) supply will tend to have a low (high) liquidity premium. Our model formalizes the idea that an asset in very scarce supply will be illiquid, even if it maintains an excellent credit rating. And our ‘indirect liquidity’ approach, coupled with endogenous market entry, is key for delivering this empirically relevant result.[21]

5.2 Quantitative exercise

Motivated by the dramatic decrease in the number of AAA-rated corporations in the early 2010s, discussed in Section 5.1, and the resulting fall in the supply of AAA bonds, we want to test whether the model can quantitatively match the observed changes in the spread between AAA and AA yields. To do so, we calibrate the model to data on the U.S. economy and bond markets, split into a “before” (2004-2011) and “after” (2012-2018) period. We feed the model data on bond supplies before and after the split, and show that it is capable of matching the spreads observed in those periods.[22]

For the utility function, we define $u(q) = q^{1-\sigma}/(1-\sigma)$. Thus, we need to calibrate ten parameters: the supplies of bonds and money (S_A, S_B, and M), the elasticity of marginal

[21] The case of AAA versus AA US corporate bonds is not the only one where the commonly held belief that “safety and liquidity go together” is violated. Christensen and Mirkov (2019) highlight yet another class of bonds – Swiss Confederation Bonds – that are considered extremely safe, yet not particularly liquid. Furthermore, Beber et al. (2008) report that Italian government bonds are among the most liquid, despite also being among the riskiest Euro-area sovereign bonds. The authors justify this observation by pointing to the large supply of Italian debt, which is consistent with our model’s prediction.

[22] The number of AAA-rated corporations in the U.S., never high, decreased to four – Automatic Data Processing, Exxon Mobil, Johnson & Johnson, and Microsoft – in 2011. (Automatic Data Processing got downgraded in 2014, and Exxon Mobil in 2016. Today, there are only two AAA-rated companies.) We also repeated our analysis with the sample split a year earlier or later, and obtained similar results.

utility (σ), the nominal interest rate on an illiquid bond ($i = (1 + \mu - \beta)/\beta$, which subsumes time preference and expected inflation), 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 (η), and the default probability ($1 - \pi$). For the calibration we use the version of our model that allows for partial default (presented in Web Appendix A.8), thus, we also need to calibrate the recovery rate in case of default (ρ). Let us point out that in this section, to ease the presentation, we focus on the version of our model where agents must hold only one asset and visit only one OTC market (full segmentation). However, in Web Appendix C.2.1 we repeat our calibration exercise for the general model where agents can trade both assets and avoid market segmentation, and we discuss the parameter values under which the unique equilibrium of the general model coincides with the specification considered here.

Some of the model parameters have straightforward empirical targets, while others do not. For the latter, we start by showing that for some (arguably) reasonable choices, our model delivers a safer-yet-less-liquid yield reversal similar to the one we observe in the data. Additionally, we perform the inverse experiment: what combinations of parameters are required in order to match the yield spreads *exactly*, before and after the change in relative supply?

First, for the bond supplies S_A and S_B, we use the market valuation of the ICE BofA AAA US Corporate Index (C0A1) and ICE BofA AA US Corporate Index (C0A2), and for M, we use the MZM monetary aggregate (from FRED). We divide the bond supplies by the money stock for each year, and then average over 2004-2011 and 2012-2018 to obtain our "before" and "after" values of AAA (0.0109 vs 0.0049) and AA (0.0477 vs 0.0445) bond supplies, 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 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 recovery rate ρ, we use 0.4 (Moody's Investors Service, 2003); for the probability that asset B pays out, we set $\pi = 0.996$.[23]

This leaves us with the bargaining power θ and the scale elasticity η, which have no direct counterparts in the data. For θ, since it is the bargaining power of asset sellers in the OTC markets, it scales liquidity premia on both bonds approximately proportionally (see Equation 6), without favoring one or the other. Since it is precisely these liquidity premia that make our model work, we want θ to be on the high side (while still less than 1 so that asset buyers have a meaningful market entry decision); thus, we set $\theta = 0.8$. For η, we already know that some amount of IRS is necessary for our Result 2, thus it must be positive. We also know from our numerical examples (e.g., Figure 6) that η does not have to be particularly high, and indeed if η is too high then one of the assets will attract all secondary market trade, and the other will be completely illiquid. Thus, we set $\eta = 0.1$.[24] A summary of all the calibrated parameter values can be found in Table 1.

Table 2 summarizes our findings. As we can see, the supply of AA corporate bonds did not change much across the two periods, but the supply of AAA bonds decreased by more than half. The spread between AA and AAA yields was positive in the earlier period, as every textbook would have predicted, then became negative in the later period. The final row of the table shows the outcome of our calibrated model, which captures

[23] The 1982-2003 issuer-weighted mean recovery rate of all bonds was 40.2%; this is the most recent data we could find. For AA bonds specifically, the rate is 41.1% if default happens five or more years later – at which point the bond might not be rated AA anymore – but higher if default happens closer in time to the latest AA rating. Furthermore, eventual recovery rates do not necessarily cover all costs (legal, etc.) associated with recovering assets after a default. Historical default probabilities are lower than our value for π implies, but similarly hard to define for the reason discussed above: defaults are usually preceded by a downgrade. Of course, from the point of view of an investor, it does not matter if a bond defaults outright or if it is downgraded before eventually defaulting.

[24] There is also the question of how we deal with the multiplicity of equilibrium (see Proposition 1). Numerically, we use $e_C^0 \equiv S_A/(S_A + S_B)$ as a starting point, then iterate the function $G(e_C)$ in the direction of its sign, until $G(e_C) \to 0$ or until reaching a corner. If the corners are not robust, this procedure will always find a robust interior equilibrium.

both the positive spread in the earlier period when AAA supplies were relatively high and the negative spread in the later period when AAA supplies had shrunk. More than just getting the sign right, the table shows that our calibrated model can also capture the magnitude of the spreads, and their reversal.[25]

Having shown that our model can quantitatively capture the change in AA-AAA spreads following the big reduction in the supply of AAA corporate bonds, for reasonable choices of parameters that do not have clear empirical targets, we now perform the inverse exercise: what would those parameters have to be in order to match the spreads exactly? We have three parameters (θ, π, and η) and two targets (the AA-AAA spread in the earlier and later periods). Since we have an extra degree of freedom, we fix a value of the bargaining power θ and let our model (and the data) tell us what the implied default probability and scale elasticity should be.

The results of this exercise are shown in Table 3. For $\theta < 0.6$, no exact match is possible, reflecting the fact that low values of θ result in low liquidity premia (and thus low power for our model). For $\theta > 0.6$, our model can match the data perfectly, up to $\theta \to 1$ in the limit. The levels of π and η adjust as we vary θ: π in order to match the level of the spreads (since lower π, everything else equal, increases the spread via the default premium), and η in order to match the size of the change (since η governs how sensitive the spread is to the difference in bond supplies).

The resulting parameter values are all plausible; the most novel and interesting one is the scale elasticity of the matching function (which to our knowledge has never been estimated for financial markets). Our calibrated value of η ranges from 0.069 to 0.212, which corresponds to an elasticity of the matching function with respect to market size of $[1.069, 1.212]$. This can be interpreted as saying that if the measure of traders in an

[25] Our calibration implies that $q_{0A}/q^* = 0.8886$ and $q_{0B}/q^* = 0.8922$ for the 2004-2011 period and $q_{0A}/q^* = 0.8839$ and $q_{0B}/q^* = 0.8869$ for the 2012-2018 period. These numbers are significantly higher than 0.5, which is the threshold below which the combined money holdings of the C-type and the N-type are not sufficient to purchase q^*. (Given the take-it-or-leave-it assumption in the DM, q_{0j} coincides with the real balances of an agent specializing in asset j; see Equation (A.10).) This verifies our claim, made on page 19, that the range for which $m + \tilde{m} < m^*$ is not relevant for the analysis.

OTC market (buyers *and* sellers) increases by 1%, overall matches increase by 1.069% to 1.212% – or, equivalently, the matching probabilities for individual traders increase by 0.069% to 0.212%. Quantitatively, the estimates in this range are closer to CRS than to the congestion-free matching functions employed in most of the finance search literature (which our formulation nests for $\eta = 1$; e.g., Duffie et al., 2005, Vayanos and Wang, 2007, and Vayanos and Weill, 2008). But they are still robustly different from CRS, enough so that an asset class in dramatically smaller supply than a competitor – for example, AAA bonds versus AA bonds, or Swiss bonds compared to German or Italian ones – can end up being substantially less liquid in equilibrium.

6 Conclusion

We argue that understanding the link between an asset's safety and its liquidity is crucial. To this end, we present a general equilibrium model where asset safety and asset liquidity are well-defined and *distinct* from one another. Treating safety as a primitive, we examine the relationship between an asset's safety and liquidity. We show that the commonly held belief that "safety implies liquidity" is generally justified, but there may be exceptions. In particular, we highlight that a safe asset in scarce supply may be less liquid than a less-safe asset in large supply. Next, we calibrate our model to rationalize the puzzling observation that AAA corporate bonds in the U.S. are less liquid than (the riskier) AA corporate bonds, and show that our model can explain this reversal as resulting from a recent decrease in the relative supply of AAA bonds. Finally, and contrary to a recent literature on the role of safe assets, we show that in our model increasing the supply of the safe asset is not always beneficial for welfare.

Data Availability

Data and codes replicating the tables and figures in this article can be found in Geromichalos, Herrenbrueck, and Lee (2022) in the Harvard Dataverse, https://doi.org/10.7910/DVN/ZS00JR.

References

Allen, F. and D. Gale (1994). Limited Market Participation and Volatility of Asset Prices. *The American Economic Review 84*(4), 933–955.

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.

Andreasen, M. M., J. H. Christensen, and S. Riddell (2021). The TIPS Liquidity Premium. *Review of Finance 25*(6), 1639–1675.

Beber, A., M. W. Brandt, and K. A. Kavajecz (2008). Flight-to-Quality or Flight-to-Liquidity? Evidence from the Euro-Area Bond Market. *Review of Financial Studies 22*(3), 925–957.

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.

Caballero, R. J. and E. Farhi (2018). The Safety Trap. *Review of Economic Studies 85*, 223–274.

Caballero, R. J., E. Farhi, and P.-O. Gourinchas (2017). The Safe Assets Shortage Conundrum. *Journal of Economic Perspectives 31*(3), 29–46.

Chang, B. and S. Zhang (2015). Endogenous Market Making and Network Formation. *Available at SSRN 2600242*.

Christensen, J. H. and N. Mirkov (2019). The Safety Premium of Safe Assets. Federal Reserve Bank of San Francisco.

Duffie, D., N. Gârleanu, and L. H. Pedersen (2005). Over-the-Counter Markets. *Economet-*

rica 73(6), 1815–1847.

Edmond, C. and P.-O. Weill (2012). Aggregate Implications of Micro Asset Market Segmentation. *Journal of Monetary Economics 59*(4), 319–335.

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: A New Monetarist DSGE Approach. *Review of Economic Dynamics 45*, 134–167.

Geromichalos, A., L. Herrenbrueck, and S. Lee (2022). Replication Data for: 'Asset Safety versus Asset Liquidity'. Harvard Dataverse, https://doi.org/10.7910/DVN/ZS00JR.

Geromichalos, A., J. M. Licari, and J. Suárez-Lledó (2007). Monetary Policy and Asset Prices. *Review of Economic Dynamics 10*(4), 761–779.

Gorton, G. (2017). The History and Economics of Safe Assets. *Annual Review of Economics 9*, 547–586.

Gorton, G., S. Lewellen, and A. Metrick (2012). The Safe-Asset Share. *American Economic Review 102*(3), 101–06.

Gorton, G. B. and G. Ordonez (2022). The Supply and Demand for Safe Assets. *Journal of Monetary Economics 125*, 132–147.

Han, H. (2015). Over-the-Counter Markets, Intermediation, and Monetary Policy. Working Paper.

He, Z., A. Krishnamurthy, and K. Milbradt (2019). A Model of Safe Asset Determination. *American Economic Review 109*(4), 1230–1262.

He, Z. and K. Milbradt (2014). Endogenous Liquidity and Defaultable Bonds. *Econometrica 82*(4), 1443–1508.

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.

Infante, S. and G. Ordoñez (2020). The Collateral Link between Volatility and Risk Sharing. NBER Working Paper 28119.

Jacquet, N. L. (2021). Asset Classes. *Journal of Political Economy 129*(4), 1100–1156.

Kalai, E. (1977). Proportional Solutions to Bargaining Situations: Interpersonal Utility Comparisons. *Econometrica 45*(7), 1623–1630.

Krishnamurthy, A. and A. Vissing-Jorgensen (2012). The Aggregate Demand for Treasury Debt. *Journal of Political Economy 120*(2), 233–267.

Lagos, R. (2010). Asset Prices and Liquidity in an Exchange Economy. *Journal of Monetary Economics 57*(8), 913–930.

Lagos, R. (2011). Asset Prices, Liquidity, and Monetary Policy in an Exchange Economy. *Journal of Money, Credit and Banking 43*, 521–552.

Lagos, R. and G. Rocheteau (2009). Liquidity in Asset Markets with Search Frictions. *Econometrica 77*(2), 403–426.

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). A Unified Framework for Monetary Theory and Policy Analysis. *Journal of Political Economy 113*(3), 463–484.

Lester, B., A. Postlewaite, and R. Wright (2012). Information, Liquidity, Asset Prices, and Monetary Policy. *Review of Economic Studies 79*(3), 1209–1238.

Madison, F. (2019). Frictional Asset Reallocation under Adverse Selection. *Journal of Economic Dynamics and Control 100*, 115–130.

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.

Moody's Investors Service (2003). Recovery Rates on Defaulted Corporate Bonds and Preferred Stocks, 1982–2003.

Nosal, E. and G. Rocheteau (2013). Pairwise Trade, Asset Prices, and Monetary Policy. *Journal of Economic Dynamics and Control 37*(1), 1–17.

Ozdenoren, E., K. Yuan, and S. Zhang (2021). Safe Assets as Balance Sheet Multipliers. Working Paper.

Piazzesi, M. and M. Schneider (2016). Payments, Credit and Asset Prices. Working Paper, Stanford University.

Rocheteau, G. (2011). Payments and Liquidity under Adverse Selection. *Journal of Monetary Economics 58*(3), 191–205.

Rocheteau, G., T.-W. Hu, L. Lebeau, and Y. In (2021). Gradual Bargaining in Decentralized Asset Markets. *Review of Economic Dynamics 42*, 72–109.

Rocheteau, G., R. Wright, and C. Zhang (2018). Corporate Finance and Monetary Policy. *American Economic Review 108*(4–5), 1147–1186.

Üslü, S. (2019). Pricing and Liquidity in Decentralized Asset Markets. *Econometrica 87*(6), 2079–2140.

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. *Journal of Finance 63*(3), 1361–1398.

Warga, A. (1992). Bond Returns, Liquidity, and Missing Data. *Journal of Financial and Quantitative Analysis*, 605–617.

Weill, P.-O. (2007). Leaning Against the Wind. *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.

	Description	Value
S_A	supply of AAA corporate bonds in the "before" period	0.0109
	supply of AAA corporate bonds in the "after" period	0.0049
S_B	supply of AA corporate bonds in the "before" period	0.0477
	supply of AA corporate bonds in the "after" period	0.0445
M	money supply	1
i	nominal interest rate on an illiquid bond (quarterly)	1.75%
ℓ	fraction of C-type agents	0.5
σ	elasticity of marginal utility	0.34
ρ	recovery rate in case of default	0.4
π	$1 -$ default probability	0.996
θ	relative bargaining power of C-types	0.8
η	elasticity of the OTC matching function	0.1

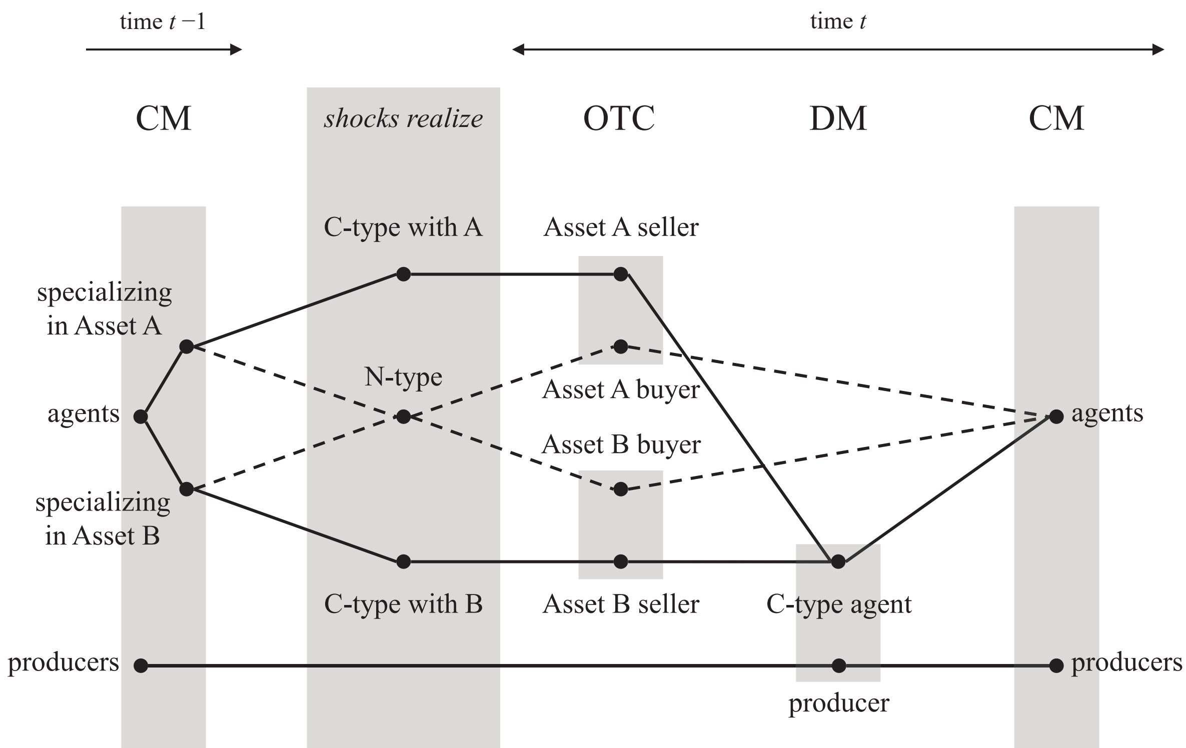
Table 1: Key parameter values

	2004-2011	2012-2018	Change
AAA supply / M, observed	0.0109	0.0049	-55%
AA supply / M, observed	0.0477	0.0445	-7%
Spread AA-AAA (%), observed	0.2245	-0.0409	-0.2654
Spread AA-AAA (%), predicted	0.1961	-0.0648	-0.2609

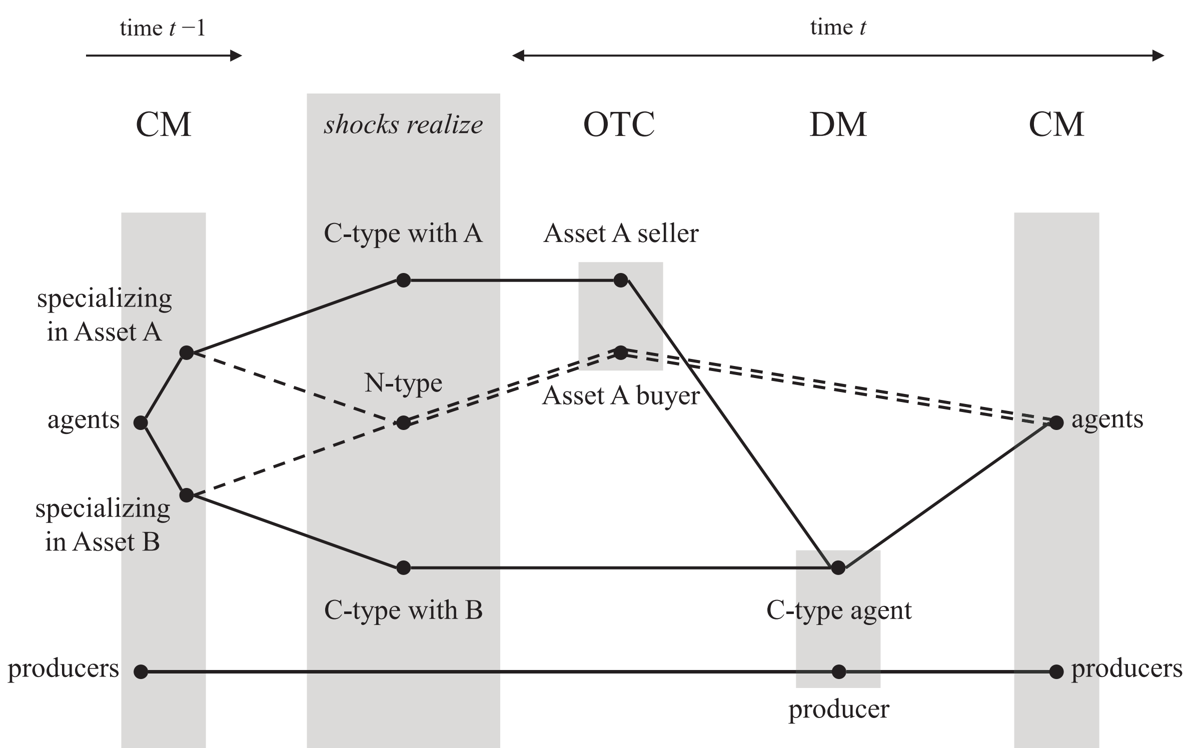
Table 2: Prediction of the model with Table 1 parameters

θ	π	η
0.6	0.9949	0.212
0.7	0.9957	0.135
0.8	0.9958	0.103
0.9	0.9958	0.083
0.99	0.9956	0.069

Table 3: Combinations of θ, π, and η that result in an exact match of our model to the observed AA-AAA spreads in the "before" and "after" periods



(a) When both assets pay out



(b) When asset B defaults

Figure 1: Timeline

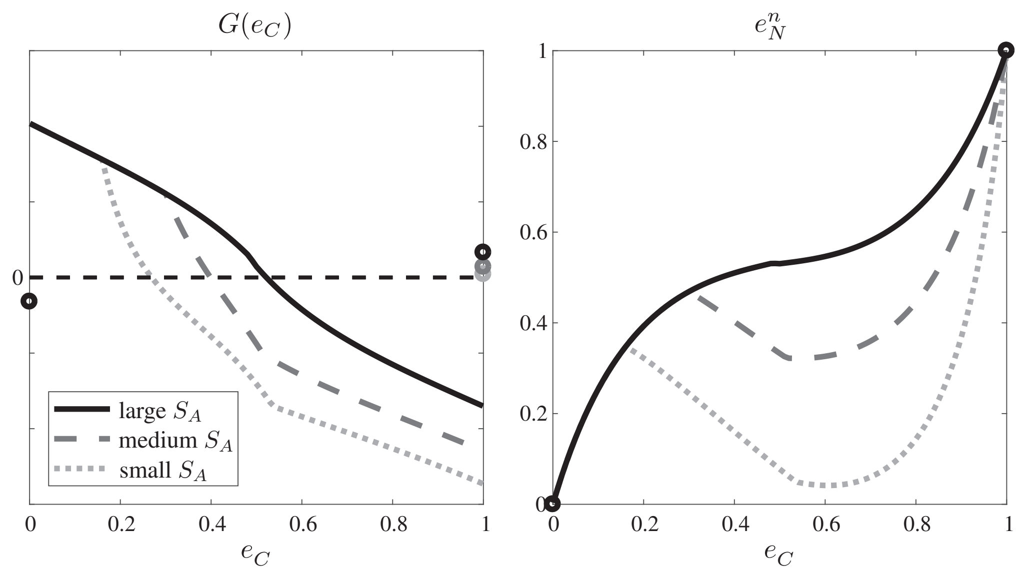


Figure 2: C-types' incentive to deviate and N-types' optimal entry choice, given e_C, for the case of CRS

Notes: The figure depicts the function $G(e_C) \equiv \tilde{\mathcal{S}}_{CA} - \tilde{\mathcal{S}}_{CB}$ (left panel) and the optimal response of N-types, e_N^n (right panel), as functions of aggregate e_C, assuming CRS in matching ($\eta = 0$). Equilibrium entry is illustrated for three levels of asset supply S_A, keeping the supply of asset B constant. The parameters used in the figure are: $M = 1$, $i = 0.1$, $\ell = 0.5$, $\theta = 0.5$, $\pi = 0.95$, $S_A = \{0.05, 0.09, 0.14\}$, $S_B = 0.14$, with log utility. We refer to the corner with $e_C = 0$ (all C-types are specializing in asset B) as the B-corner, and the corner with $e_C = 1$ (all C-types are specializing in asset A) as the A-corner.

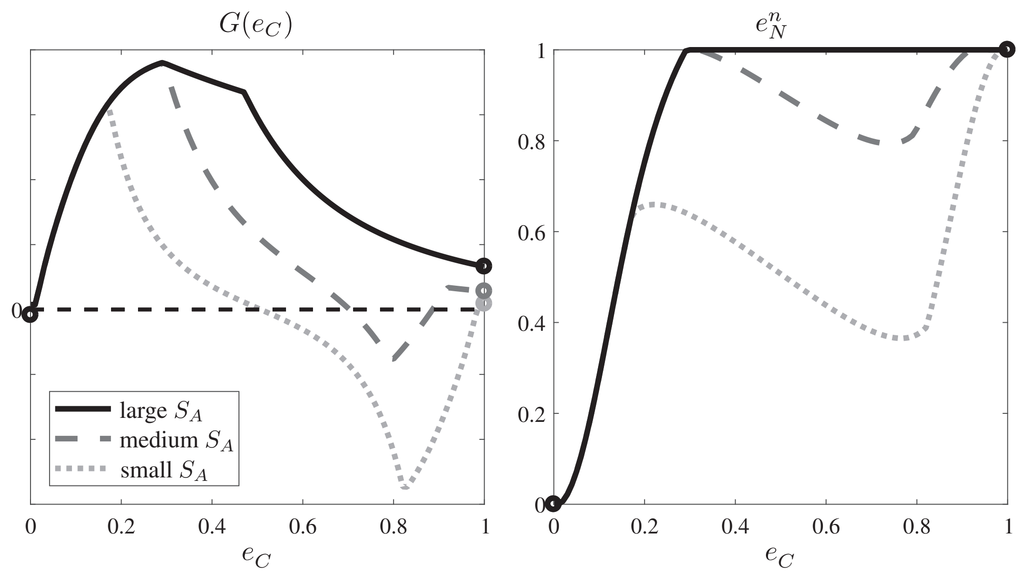


Figure 3: C-types' incentive to deviate and N-types' optimal entry choice, given e_C, for the case of IRS

Notes: The figure depicts the function $G(e_C) \equiv \tilde{\mathcal{S}}_{CA} - \tilde{\mathcal{S}}_{CB}$ (left panel) and the optimal response of N-types, e_N^n (right panel), as functions of aggregate e_C, assuming IRS in matching ($\eta = 0.3$). Equilibrium entry is illustrated for three levels of asset supply S_A, keeping the supply of asset B constant. The parameters used in the figure are: $M = 1$, $i = 0.1$, $\ell = 0.5$, $\theta = 0.5$, $\pi = 0.95$, $S_A = \{0.05, 0.09, 0.14\}$, $S_B = 0.05$, with log utility. We refer to the corner with $e_C = 0$ (all C-types are specializing in asset B) as the B-corner, and the corner with $e_C = 1$ (all C-types are specializing in asset A) as the A-corner.

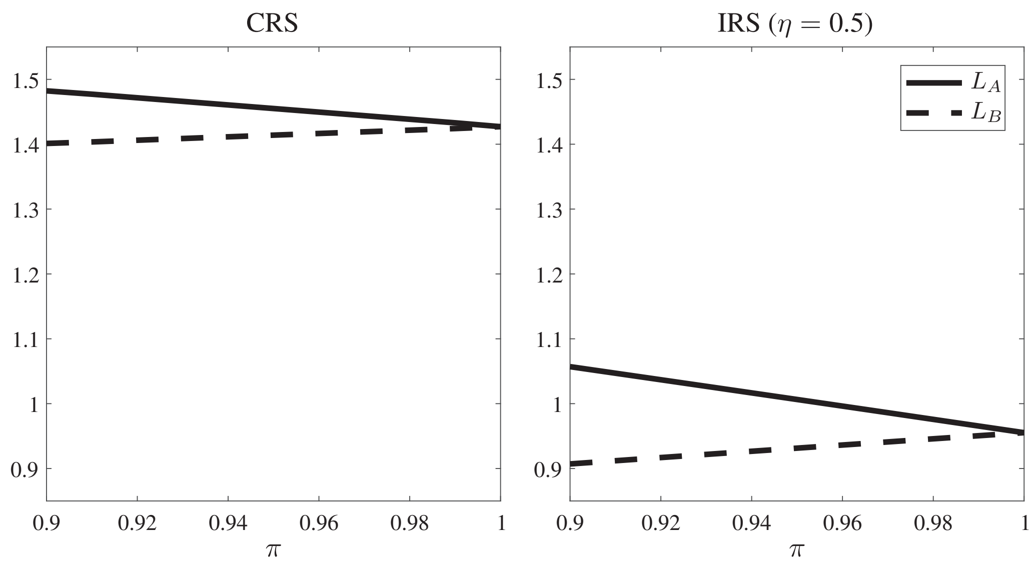


Figure 4: Liquidity premia as functions of π

Notes: The figure depicts the liquidity premia of assets A, B as functions of π, assuming symmetric asset supplies. The left panel illustrates the case of a CRS matching technology, and the right panel represents the the case of IRS ($\eta = 0.5$). The parameters used in the figure are: $M = 1$, $i = 0.1$, $\ell = 0.5$, $\theta = 0.5$, $S_A = S_B = 0.05$, with log utility.

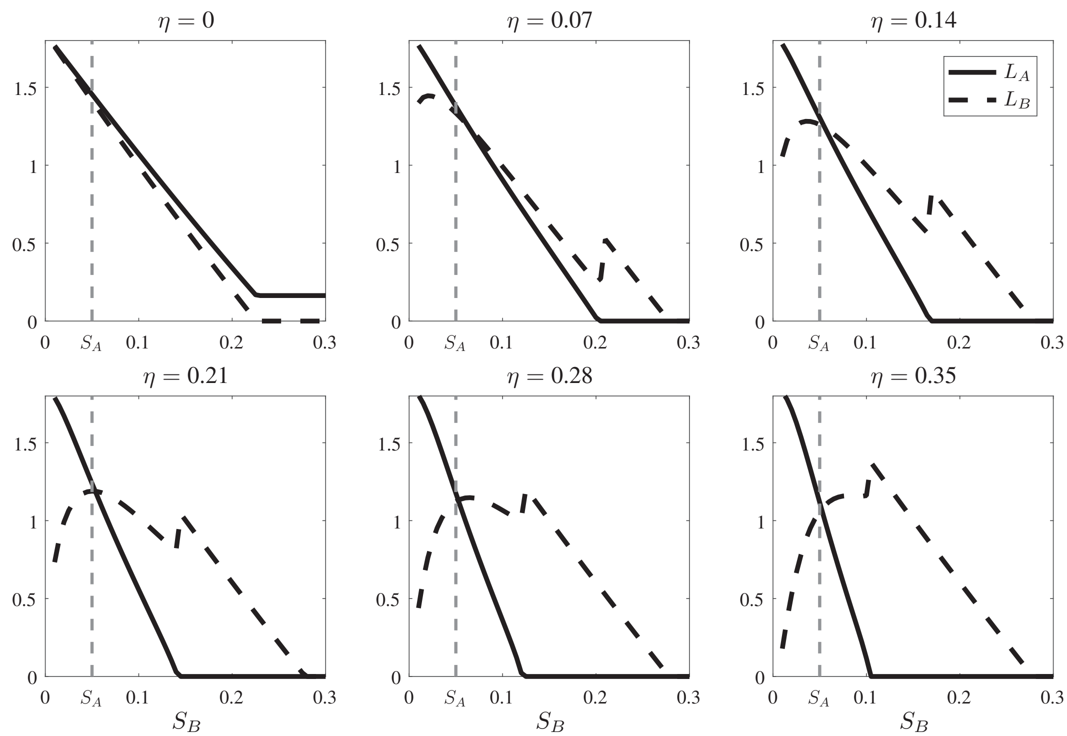


Figure 5: Liquidity premia with varying degrees of IRS

Notes: The figure depicts the liquidity premia of assets A and B as functions of S_B, for a constant S_A, and for varying degrees of IRS. The dashed vertical line indicates the (fixed) supply of asset A. The parameters used in the figure are: $M = 1$, $i = 0.1$, $\ell = 0.5$, $\theta = 0.5$, $\pi = 0.95$, $S_A = 0.05$, with log utility.

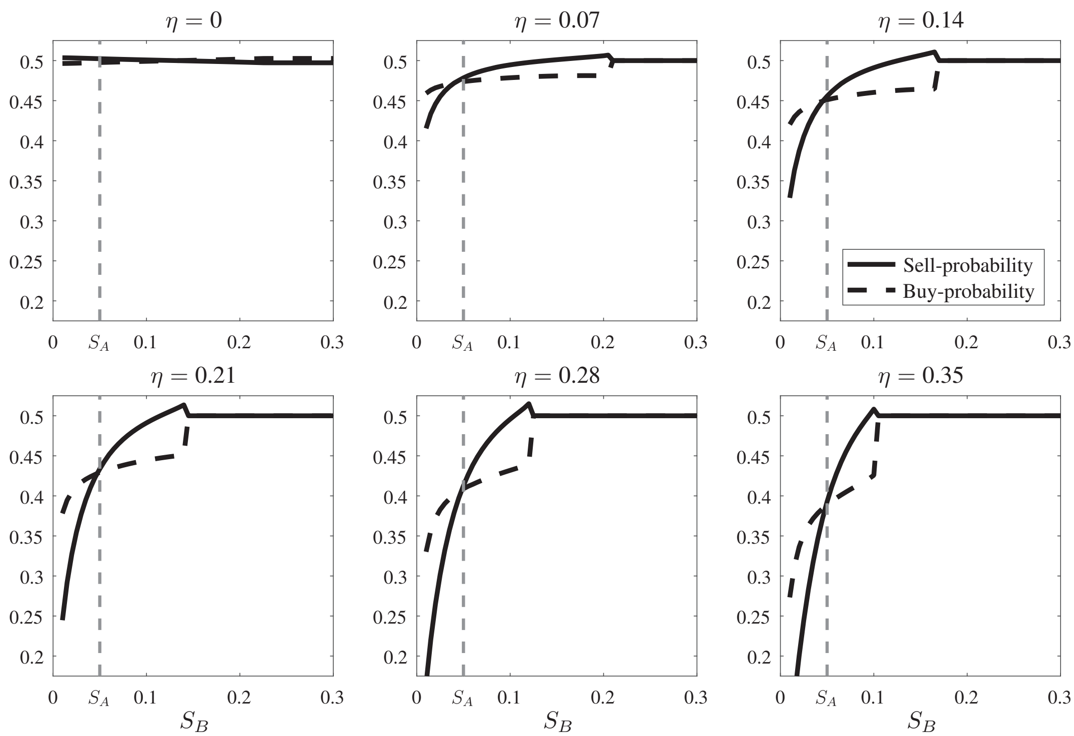


Figure 6: Sell- and buy-probabilities in OTC$_B$

Notes: The figure depicts the sell-probability, α^n_{CB}, and the buy-probability, α^n_{NB}, in the secondary market <u>for asset B</u>, in the normal state, as a function of S_B (and for varying degrees of IRS). The dashed vertical line indicates the (fixed) supply of asset A. The parameters used in the figure are: $M = 1$, $i = 0.1$, $\ell = 0.5$, $\theta = 0.5$, $\pi = 0.95$, $S_A = 0.05$, with log utility.

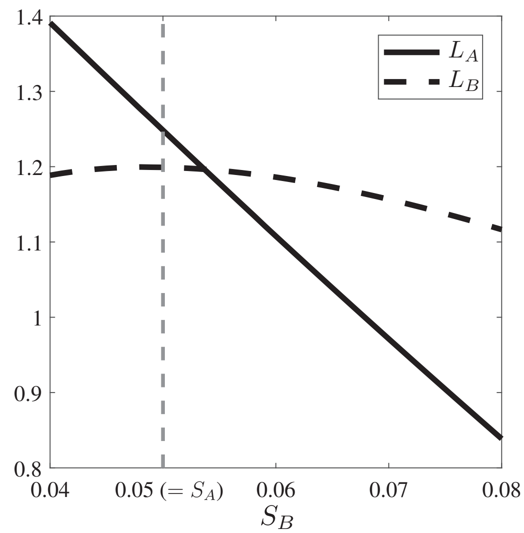


Figure 7: Liquidity premia

Notes: The figure depicts the liquidity premia of assets A and B as functions of S_B, for a constant S_A, and for $\eta = 0.2$. The dashed vertical line indicates the (fixed) supply of asset A. The parameters used in the figure are: $M = 1$, $i = 0.1$, $\ell = 0.5$, $\theta = 0.5$, $\pi = 0.95$, $S_A = 0.05$, with log utility.

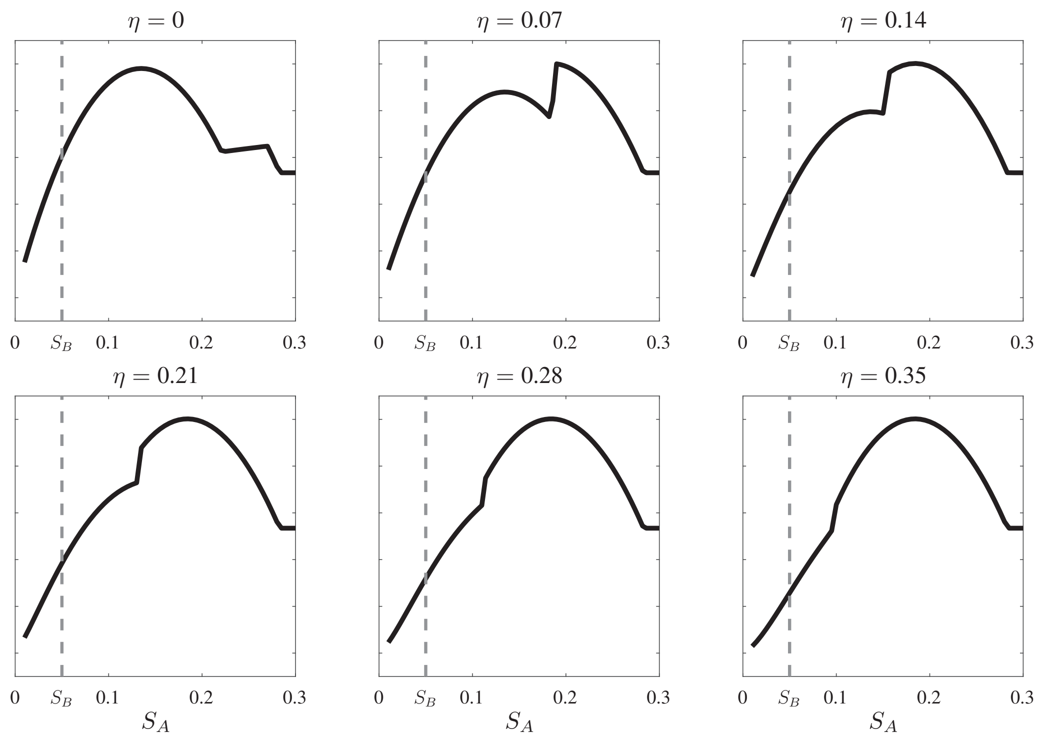


Figure 8: Safe asset supply and welfare

Notes: The figure depicts equilibrium welfare as a function of S_A, for various values of η, including $\eta = 0$ (CRS). The dashed vertical line indicates the (fixed) supply of asset B. The parameters used in the figure are: $M = 1$, $i = 0.1$, $\ell = 0.5$, $\theta = 0.5$, $\pi = 0.95$, $S_B = 0.05$, with log utility.

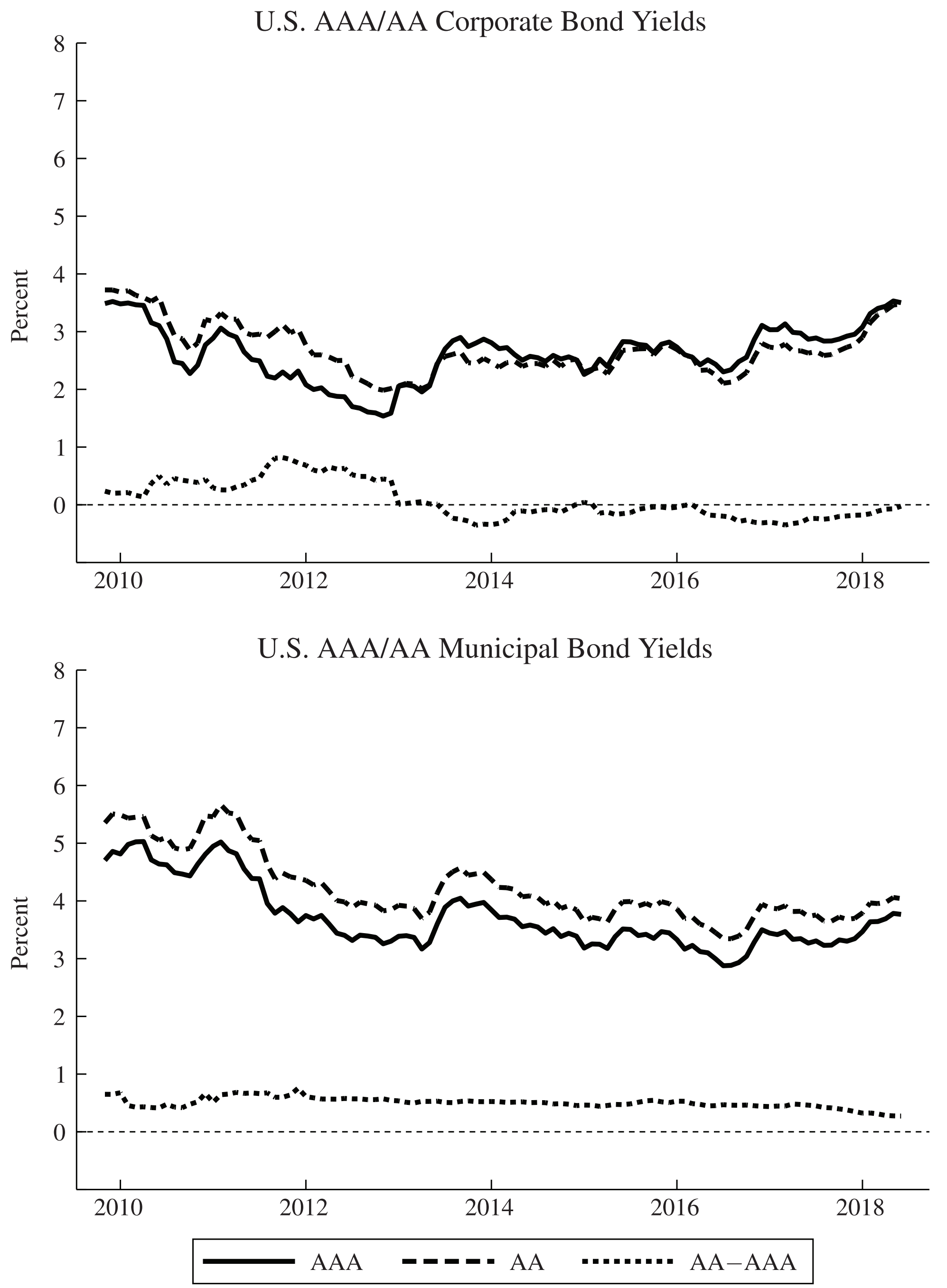


Figure 9: Historical yields of AAA/AA corporate/municipal bonds

Notes: The data on municipal bonds comes from Standard & Poor's, and the data on corporate bonds comes from Federal Reserve Economic Data (FRED). The original data is on a daily base, but, to make the graphs more legible, it is converted to a monthly base. The graphs show the historical yields for the past 10 years. *Sources:* FRED; Standard & Poor's.