Liquidity and the Threat of Fraudulent Assets*

Yiting Li Guillaume Rocheteau Pierre-Olivier Weill

October 1, 2011

Abstract

We study an over-the-counter (OTC) market with bilateral meetings and bargaining where the usefulness of assets, as means of payment or collateral, is limited by the threat of fraudulent practices. We assume that agents can produce fraudulent assets at a positive cost, which generates endogenous upper bounds on the quantity of each asset that can be sold, or posted as collateral in the OTC market. Each endogenous, asset-specific, resalability constraint depends on the vulnerability of the asset to fraud, on the frequency of trade, and on the current and future prices of the asset. In equilibrium, the set of assets can be partitioned into three liquidity tiers, which differ in their resalability, their prices, their sensitivity to shocks, and their responses to policy interventions. The dependence of an asset's resalability on its price creates a pecuniary externality, which leads to the result that some policies commonly thought to improve liquidity can be welfare reducing.

J.E.L. Classification: D82, D83, E40, E50, G10

Keywords: search, payments, collateral, fraud, liquidity premia.

*A previous version of this paper was circulated under the title "Liquidity Constraints." We thank two anonymous referees and Rob Shimer for comments that greatly improved the paper. We also thank Veronica Guerrieri for an insightful discussion, Andy Atkeson, Bill Branch, Dean Corbae, Manolis Galenianos, Hugo Hopenhayn, Thomas Philippon, Shouyong Shi, and seminar participants at the Bank of Canada, the Federal Reserve Bank of Chicago, the Federal Reserve Bank of Cleveland, the National University of Seoul, the National Taiwan University, the Singapore Management University, the University of California at Riverside and Santa Cruz, the University of Wisconsin, and the Wharton school at the University of Pennsylvania for useful comments. We thank Monica Crabtree-Reusser for editorial assistance. Yiting Li: National Taiwan University, yitingli@ntu.edu.tw; Guillaume Rocheteau: University of California, Irvine and Federal Reserve Bank of Cleveland, grochete@uci.edu; Pierre-Olivier Weill: University of California, Los Angeles, NBER, and CEPR, poweill@econ.ucla.edu.

1 Introduction

Liquidity premia, or convenience yields, are key determinants of asset prices. This point is uncontroversial for fiat money, which derives its value solely from its liquidity services. According to Krishnamurthy and Vissing-Jorgensen (2010), the same is true for government securities, high-grade corporate bonds, and agency bonds. In this paper we present a theory of asset liquidity and convenience yields, based on the following premise: an asset's liquidity—the extent to which it can facilitate exchange, as means of payment or as collateral—depends on its vulnerability to fraud. We address a class of questions related to the cross-sectional dispersion and time-variation of liquidity premia, such as what fundamental characteristics make some assets have higher turnover and lower yields than others? What shocks prompt investors to suddenly shift their portfolios towards the most liquid assets, which leads to widening yield spreads? Are liquid assets more susceptible of exhibiting excess volatility? And, what types of open-market operations and financial regulations are effective to mitigate aggregate liquidity shortages?

The threat of fraud has been a pervasive friction throughout history. Classical examples include the clipping of coins in ancient Rome and medieval Europe, and the counterfeiting of banknotes during the first half of the 19$^{\text{th}}$ century in the United States (Sargent and Velde, 2002; Mihm, 2007). Modern financial assets are no less susceptible to fraud. Intangible means of payment suffer from identity thefts (Schreft, 2007), and mortgage-backed securities are subject to moral hazard problems and lax incentives that plague the process of securitization (see, among others, Keys, Mukherjee, Seru, and Vig, 2010).[1] Similarly, the fact that some investors can spend resources to cherry-pick the collateral used to secure risk-sharing arrangements is a concern for participants in OTC derivative markets.[2]

We introduce the threat of fraud into a search-theoretic model of asset markets, building on

[1] The 2010 "Performance and Activity Report" of the SEC details many cases of financial fraud related to mortgage-based securities. Frauds and moral hazard problems in the mortgage market are not new. Snowden (2010) describes the US mortgage crisis of the late 20s and 30s and the earlier forms of securitization in the 20s. Real estate bond houses were overappraising properties, they violated underwriting standards, and they substituted bad loans for performing mortgages in their mortgage pools.

[2] The International Swap and Derivatives Association (2010) reported that over 78% of OTC derivatives trades are collateralized. Importantly, market participants consider some asset classes (e.g., cash or government securities) to be of higher collateral quality than others. Collateral quality depends on various factors such as volatility, credit risk, and pricing ease.

recent work in monetary and financial economics (e.g., Lagos and Wright, 2005; Duffie, Gârleanu, and Pedersen, 2005). In the first period, agents trade an arbitrary number of assets in a competitive market. In the second period, they trade goods and services in an over-the-counter (OTC) market, with bilateral meetings and bargaining. Because of the frictions caused by a lack of commitment and limited enforcement, agents use assets as means of payment, or as collateral, in the OTC market. However, the extent to which an asset can play such a role is limited by the threat of fraud: after incurring an asset-specific fixed cost, an agent can produce fraudulent assets, which are worthless and indistinguishable from their genuine counterparts. In order to solve the resulting OTC bargaining problem under asymmetric information, we assume that the asset holder makes a take-it-or-leave-it offer, and we use the recent methodology of In and Wright (2011) for signaling games with hidden choices to select an equilibrium.

A key insight of our analysis is that the threat of fraud generates asset-specific, endogenous resalability constraints. While there are no exogenous restrictions on the transfer of assets in bilateral matches in the OTC market, if the quantity of an asset offered is above some threshold, then the trade is rejected with positive probability because of the rational fear that the asset might be fraudulent. In equilibrium, agents never find it optimal to offer more of the asset than what can be accepted with certainty, which prevents fraud from taking place. The resulting endogenous resalability constraint has three determinants: the asset's vulnerability to fraud, the difference between the asset's price and the discounted value of its cash flows, and the frequency of trades in OTC markets. We emphasize three main implications of these endogenous resalability constraints below.

First, because an asset's resalability depends on its own vulnerability to fraud, prices and measures of liquidity vary across assets with identical cash flows. We obtain an endogenous three-tier categorization of assets: illiquid, partially liquid, and liquid assets, which differ in their resalability, their price, as well as their sensitivity to shocks and policy interventions. While the price of an illiquid asset is equal to the present value of its cash flows, the price of a partially liquid or liquid asset is strictly larger than the present value of its cash flows; i.e., this asset enjoys a liquidity premium. This premium increases with the asset's recognizability but decreases with its supply, which is consistent with the downward-sloping aggregate demand for Treasury debt documented in Krishnamurthy and Vissing-Jorgensen (2010). Finally, while the prices of illiquid and partially liquid assets are constant in the absence of shocks to fundamentals, the prices of liquid assets can exhibit self-fulfilling fluctuations.

Second, in a similar spirit as Guerrieri and Shimer (2011), our model identifies shocks that generate phenomena akin to flights to liquidity, whereby investors shift their asset demands from less liquid to more liquid assets, widening the liquidity spread between the two types of assets (see Longstaff, 2004, and Dick-Nielsen, Feldhutter, and Lando, 2010). For instance, we consider an increase in the frequency of liquidity needs in the OTC market that results in higher demand for collateral. Such a shock increases the value of holding assets, as they are more likely to be used as means of payment or as collateral, but it also has the countervailing effect of increasing fraud incentives. We show that the first effect dominates for liquid assets and raises their prices, while the second effect dominates for partially liquid assets and lowers their prices. Moreover, the set of liquid assets endogenously shrinks, meaning that agents shift their demand to the most recognizable assets, in accordance with a flight to liquidity. The same phenomenon can be generated in our model by a shock that raises the threat of fraud for some partially liquid or liquid assets, thereby reducing their resalability.

The third main implication of our results concerns policies aimed at managing the aggregate supply of liquidity through open-market operations or financial regulations. In our model, an open-market operation has a positive welfare effect if and only if it increases a simple measure of aggregate liquidity—a weighted sum of asset supplies. Therefore, a substitution of liquid assets with other liquid assets is irrelevant. An open-market purchase of illiquid assets with liquid ones, on the other hand, raises aggregate liquidity and output. However, under a balanced budget requirement, a purchase of partially liquid assets with liquid ones *reduces* aggregate liquidity, the yield of liquid assets, and output. This paradoxical result arises because of a "pecuniary externality," according to which an increase in the price of an asset reduces its resalability, which in turn can lower its liquidity premium *below* the true marginal social value of its liquidity services. Due to this externality, a balanced budget open-market purchase syphons out more liquidity than it is injecting in. This result can shed some light on quantitative easing, which consists of injecting reserves in exchange for less liquid assets (Krishnamurthy and Vissing-Jorgensen, 2011). According to our model, for such policies to successfully increase aggregate liquidity, they must target the most illiquid assets. In a similar vein, we study retention requirements that were introduced by the Dodd-Frank Act to mitigate moral hazard problems in the securitization process. In the context of our model, such requirements are welfare improving only if applied to illiquid assets.

1.1 Literature review

Kiyotaki and Moore (2001, 2005) study limited resalability by assuming that each period, agents cannot sell more than an exogenous proportion of their asset holdings. While such exogenous resalability constraints can be chosen to replicate our distribution of asset prices, they generate markedly different comparative statics and policy recommendations (see Supplementary Appendix E). For instance, with proportional resalability constraints, an increase in the frequency of trading needs weakly increases the prices of all assets, while in our model it has asymmetric effects: it increases the prices of liquid assets, and decreases the prices of partially liquid assets, consistent with evidence on flight to liquidity. As another example, with proportional liquidity constraints, an open-market purchase of partially liquid assets with liquid ones increases liquidity, asset yields, and welfare. In our model, because of a new pecuniary externality, we obtain the opposite effects, consistent with evidence on quantitative easing.

In Holmstrom and Tirole's (2011, and references therein) corporate finance model, a moral hazard problem generates endogenous borrowing constraints, i.e., resalability constraints in the primary asset market. In the secondary market, corporate claims with identical cash flows enjoy the same liquidity premium. In our model, by contrast, we focus on moral hazard in secondary markets. We highlight the fact that agents' incentives to take hidden actions depend on contemporaneous secondary market prices and on OTC market frictions, and we generate cross-sectional differences in liquidity premia between assets with identical cash flows.

The search-theoretic literature on the liquidity structure of asset returns includes, e.g., Wallace (1998, 2000), Weill (2008), and Lagos (2010), and related work on the rate-of-return-dominance puzzle. Our approach goes beyond this earlier search literature by showing how cross-sectional differences in liquidity arise from fraud-based endogenous resalability constraints.[3] Lester, Postle-

[3] Wallace (1998, 2000) emphasizes assets' indivisibilities, Weill (2008) assumes increasing returns in the matching

waite, and Wright (2011) consider a private information problem where agents can recognize the quality of an asset at some cost, but to determine the terms of trade under asymmetric information they make the simplifying assumption that unrecognized assets are not accepted in a bilateral match.[4] They address this issue in an extension that follows our methodology closely.

There is a literature that emphasizes adverse selection problems in asset markets with search frictions (e.g., Hopenhayn and Werner, 1996). The most closely related papers are Rocheteau (2009) who introduces an adverse selection problem in a monetary model to explain the illiquidity of risky assets, and Guerrieri, Shimer, and Wright (2010), who consider a competitive search environment to illustrate how trading delays emerge endogenously to screen high- and low-quality assets. Guerrieri and Shimer (2011) extend the previous paper to a general equilibrium framework and, among other results, provide an explanation for flights to liquidity based on a dynamic adverse selection problem. While the distinction between adverse selection and moral hazard in asset markets is often subtle, the methodologies for capturing the two frictions differ profoundly. We take the view that informational asymmetries in asset markets often result from strategic behavior, which allows us to focus the model more squarely on the effects of the threat of fraud on asset liquidity. At a more theoretical level, an important distinction between adverse selection and moral hazard is that the type distribution is exogenous with the former, but is endogenous with the latter. With an exogenous type distribution, under some conditions, agents can mitigate the asymmetric information friction by holding broadly diversified asset portfolios. As our model demonstrates, when the type distribution is endogenous, the asymmetric information friction remains relevant.

The next section presents the model. Section 3 solves the bargaining game under the threat of fraud. Section 4 solves for asset prices, and Section 5 presents three main implications. The appendix contains omitted proofs, and the supplementary appendix presents additional results and extensions.

technology, and Lagos (2010) introduces exogenous restrictions on the use of some assets as means of payment. Similarly, Shi (2008) studies the pricing of bonds in a search economy where exogenous legal restrictions prevent bonds from being used in payments in a fraction of trades.

[4] There is also a related literature on counterfeiting, e.g., Green and Weber (1996), Williamson and Wright (1994), and Nosal and Wallace (2007). In those studies, there is a single asset, asset holdings are restricted to $\{0, 1\}$, and assets are indivisible, while those restrictions are all relaxed in our model.

2 The model

The economy lasts for two periods, $t \in \{0, 1\}$, and is populated by a continuum of agents who trade sequentially in two markets: in a centralized market (CM) at $t = 0$, and in a decentralized over-the-counter market (DM) at $t = 1$. There are two perfectly divisible and perishable goods. The first good, which we take to be the *numéraire*, is produced and consumed at $t = 0$ and at the end of $t = 1$. The second good, labeled the DM good, is produced and consumed in bilateral meetings in the DM. There is a finite set of assets indexed by $s \in S$. Each asset pays off at the end of $t = 1$ a dividend normalized to one unit of the numéraire.

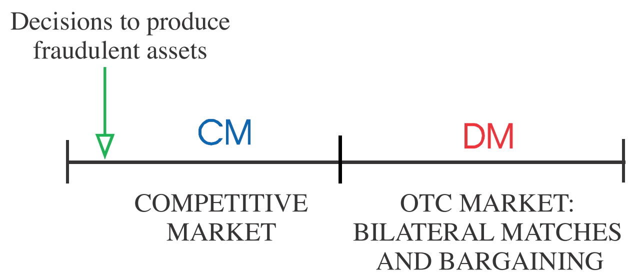

Figure 1: Timing of the game

Agents are divided evenly into two types, called *buyers* and *sellers*. Buyers wish to consume in the DM but cannot produce, while sellers have the technology to produce goods in the DM but do not want to consume. Together with frictions described below, this preference structure creates a need for liquidity: buyers will acquire assets in the CM in order to finance the purchase of goods produced by sellers in the DM. The utility of a buyer is:

$$x_0 + \beta \left[u(q_1) + x_1 \right], \tag{1}$$

where $x_t \in \mathbb{R}$ is the consumption of the numéraire good at time t, with $x_t < 0$ being interpreted as production, $q_1 \in \mathbb{R}_+$ is the consumption of the DM good, and $\beta \equiv (1 + r)^{-1} \in (0, 1)$ is a discount factor. The utility function, $u(q)$, over the DM good is twice continuously differentiable, with $u(0) = 0$, $u'(q) > 0$, $u'(0) = \infty$, $u'(\infty) = 0$, and $u''(q) < 0$. The utility of a seller is:

$$x_0 + \beta \left(-q_1 + x_1 \right), \tag{2}$$

where q_1 is the seller's production in a pairwise meeting in the DM. Let $q^* = \arg\max_q [u(q) - q] > 0$ denote the output level that maximizes the match surplus, so $u'(q^*) = 1$.

The CM is a perfectly competitive market, where agents trade the numéraire good and assets. The DM, on the other hand, is an over-the-counter market, where a fraction $\sigma \in (0, 1]$ of buyers are matched bilaterally and at random with an equal fraction of sellers. Because of a lack of commitment and limited enforcement, buyers purchase DM goods with assets or, equivalently, with loans collateralized by assets (see footnote 11).

Terms of trade in pairwise meetings in the DM are determined according to a simple bargaining game, in which the buyer makes a take-or-leave-it offer.[5] The buyer, whose asset holdings are private information, asks for a given amount of the DM good in exchange for some specified portfolio of assets.[6] The seller accepts or rejects the offer. If the seller accepts the offer, then the trade is implemented, provided that the asset transfer is feasible given the buyer's asset holdings. Matched agents split apart before assets pay off.

We introduce the possibility of asset fraud as follows. In the CM at $t = 0$, a buyer can pay a fixed cost $k(s) > 0$ to produce any quantity of fraudulent asset of type s. Fraudulent assets have zero terminal value and, in the DM, cannot be distinguished by sellers from genuine assets.

2.1 Interpretations

Counterfeiting of a means of payment. A literal interpretation of the model concerns assets used as means of payment, such as coins or banknotes, for which the fraud consists of producing counterfeits.[7] During the first half of the 19th century, the fixed cost to produce fake banknotes included the cost to acquire plates and dies. See, e.g., Mihm (2007). Nowadays, this cost corresponds

[5] In her discussion of our paper, Veronica Guerrieri investigated a version of the model with competitive search and showed that this alternative pricing mechanism generates the same liquidity constraint as the one obtained under our simple bargaining game.

[6] By assuming that asset holdings are unobservable, we reduce the set of signals sent by a buyer, which simplifies the analysis. As shown in the earlier version of our working paper, results are robust to alternative assumptions regarding the observability of asset holdings. Also, we do not allow buyers to offer lotteries over allocations. In our context we conjecture that it is with no loss in generality, but such lotteries could be useful in the presence of alternative cost structures of producing fraudulent assets—see Supplementary Appendix C on variable costs.

[7] Our model can accommodate fraud on unsecured credit in bilateral matches. In this case, an agent has the option to produce a fake identity in the CM at a fixed cost (e.g., the cost incurred by a computer hacker to steal the identity of someone else) and he can issue an IOU in the DM if matched. The repayment of genuine IOUs can be enforced in the following CM. In contrast, IOUs based on fake identities are not repaid.

to the price of photo-editing software and copy machines.

Collateral fraud. An alternative interpretation is that buyers use assets as collateral to secure loans to be repaid at the end of $t = 1$. If the asset is a house, the transaction in the DM is an equity extraction loan to finance consumption. An example of mortgage fraud that closely resembles our model is the property flipping scheme, whereby a buyer obtains a high-loan-to-value mortgage based on a fake property appraisal, and the bank is left with worthless collateral.[8] In this example, the cost of producing fraudulent assets represents the cost of creating false documentation about the borrower and the property. The DM can also be interpreted as an OTC market for credit derivatives, such as the market for credit default swaps or interest rate swaps. In that context, the goods traded in the DM are risk-sharing services, and collateral is used to mitigate counterparty risk.[9] The cost of producing fraudulent assets is the informational cost incurred by the buyer to identify bad collateral. This cost is related to the complexity of the asset, its issuer, and the quality and quantity of information released about the asset's cash flows.

Securitization fraud. In this context buyers represent mortgage securitizers who originate and package loans in the CM. Sellers represent final asset holders who acquire securitized assets in the DM. There are gains from trading assets in the DM because it allows mortgage securitizers to spread the risk of the underlying loans to final asset holders.[10] In this example, the cost of producing fraudulent assets is the cost of generating false documentation about the underlying security, bribing an agency for a good rating, or engaging in accounting frauds.

3 Bargaining under the threat of fraud

In this section we solve for the equilibrium of the game between a buyer and a seller matched at random. The game starts in the CM at $t = 0$ and ends in the DM at $t = 1$. For now we take as

[8] See http://www.fbi.gov/about-us/investigate/white_collar/mortgage-fraud/mortgage_fraud.

[9] In Supplementary Appendix G, we provide an explicit model of risk-sharing arrangements, where the DM good can be interpreted as risk-sharing services.

[10] In Supplementary Appendix H, we provide such a model of securitization, where agents have Constant Absolute Risk Aversion (CARA) utilities. This model confirms, albeit with different functional forms, that $u(q)$ can be interpreted as the utility of reducing the securitizer's risk position, and $q = c(q)$ can be interpreted as the cost of increasing the final asset holder's risk position.

given asset prices in the CM, $\phi(s)$, $s \in S$, and we anticipate that, in equilibrium, they will satisfy $\phi(s) \geq \beta$; i.e., the rate of return of asset s is no greater than the discount rate, which would be the "fundamental price" of the asset in a frictionless economy.

The sequence of moves is as follows: (i) In the CM at $t = 0$, the buyer chooses a portfolio of $\{a(s)\}$ genuine and $\{\tilde{a}(s)\}$ fraudulent assets, subject to $a(s) \geq 0$ and $\tilde{a}(s) \geq 0$; (ii) In the DM at $t = 1$, the buyer is matched with a seller with probability σ, in which case he makes an offer $(q, \{d(s)\})$, where q represents the output produced by the seller and $d(s)$ is the transfer of assets of type s (genuine or fraudulent) from the buyer to the seller; (iii) The seller decides whether to accept the offer; (iv) If the offer is accepted, the seller delivers q units of goods to the buyer, and the buyer delivers $\tau(s) \in [0, a(s)]$ genuine and $\tilde{\tau}(s) \in [0, \tilde{a}(s)]$ fraudulent units of asset s to the seller, with $\tau(s) + \tilde{\tau}(s) = d(s)$.[11] The extensive form of the game, for the $\sigma = 1$ case, is illustrated in the left panel of Figure 2. Arcs indicate that the action set at a given node is infinite, while a dotted line represents an information set.

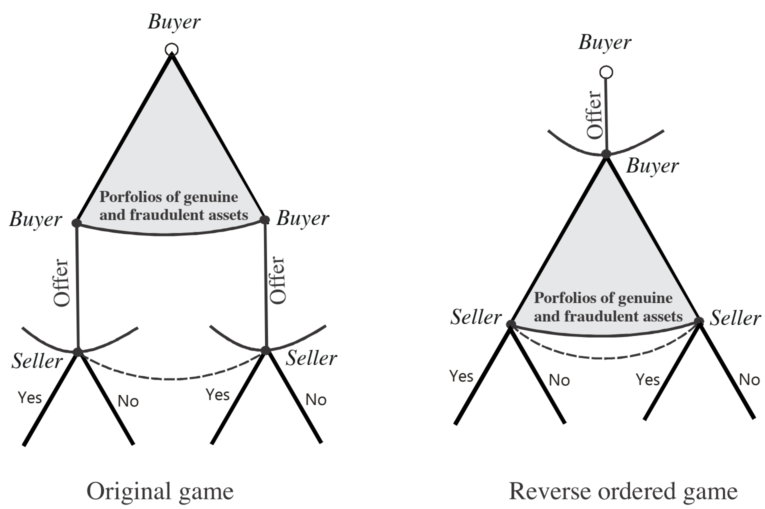

Figure 2: Game trees

[11] We can reinterpret the payment, $(q, \{d(s)\})$, as a fully collateralized loan, where the buyer promises to repay $\sum_{s \in S} d(s)$ units of the CM output at the end of period 1. In order to secure the repayment of the loan, the buyer posts $d(s)$ units of asset s as collateral with a third party. If one asset is fraudulent, then the buyer will choose to default on his obligation, in which case the seller seizes the assets that serve as collateral. If all assets are genuine, then the buyer is indifferent between repaying his debt or defaulting.

Payoffs. The Bernoulli payoff of the buyer is:

$$-\sum_{s\in S}\Big\{k(s)\mathbb{I}_{\{\tilde{a}(s)>0\}}+\phi(s)a(s)\Big\}+\beta\mu\Big\{u(q)+\sum_{s\in S}\left[a(s)-\tau(s)\right]\Big\}+\beta(1-\mu)\sum_{s\in S}a(s),$$

where $\mathbb{I}_{\{\tilde{a}(s)>0\}}=1$ if the buyer produces fraudulent assets of type s, $\tilde{a}(s)>0$, and zero otherwise. In the above, $\mu=1$ if the buyer meets a seller who accepts his offer, and $\mu=0$ otherwise. The first term is the payoff of the buyer at $t=0$. In order to accumulate $\tilde{a}(s)>0$ fraudulent units of asset s, the buyer must incur the fixed cost $k(s)$. In order to accumulate $a(s)$ units of genuine asset s, he must produce $\phi(s)a(s)$ units of the numéraire good in the CM. The second term is the discounted payoff at $t=1$ if $\mu=1$; i.e., if the buyer meets a seller in the DM and his offer is accepted. He then enjoys the utility of DM good consumption, $u(q)$, as well as the payoff from his net holding of genuine assets, $a(s)-\tau(s)$, the initial amount purchased net of the asset transfer to the seller, keeping in mind that each unit of genuine asset pays off one unit of the numéraire good at the end of $t=1$. The last term is, similarly, the discounted payoff of the buyer at $t=1$ if $\mu=0$. Collecting terms, we can rewrite the payoff as

$$-\sum_{s\in S}\left\{k(s)\mathbb{I}_{\{\tilde{a}(s)>0\}}+\Big[\phi(s)-\beta\Big]a(s)\right\}+\beta\mu\left\{u(q)-\sum_{s\in S}\tau(s)\right\}. \tag{3}$$

Similarly, the Bernoulli payoff of the seller is

$$\beta\mu\left\{-q+\sum_{s\in S}\tau(s)\right\}, \tag{4}$$

where we anticipate that, in equilibrium, sellers will not find it optimal to accumulate assets in the CM.[12] If the seller accepts the offer ($\mu=1$), he suffers the disutility of producing, q, and receives $\tau(s)$ genuine units of asset s.

Equilibrium concept. Our equilibrium concept is Perfect Bayesian Equilibrium: actions are sequentially rational following every history, and beliefs accord with Bayes's rule whenever it is possible. The notion of Perfect Bayesian Equilibrium imposes little discipline on the seller's belief in the DM regarding the decision of the buyer in the initial stage of the game to produce fraudulent assets, conditional on an off-equilibrium offer being made. Our approach to circumvent this

[12]Sellers have no strict incentives to accumulate assets if $\phi(s)\geq\beta$, because their asset holdings are not observable and hence do not affect the terms of trade offered by the buyer.

difficulty consists of adopting a notion of strategic stability, according to which any equilibrium of the original game should also be an equilibrium of the reverse-ordered game, whose timing is shown in Figure 2: (i) The buyer determines his DM offer, $(q, \{d(s)\})$, before making any decision in the CM (e.g., one interpretation is that he posts an offer at the beginning of the CM for the next DM); (ii) He chooses his portfolio composed of genuine and fraudulent assets; (iii) He is matched with a seller who chooses whether to accept or reject the offer.[13] This reordered game captures the idea that upon seeing the buyer's offer, the seller will infer that the buyer's unobservable actions (portfolio and production of fraudulent assets) were chosen optimally with the offer in mind. The refinement is intuitive in that it selects an equilibrium of the original game that yields the highest payoff to the player making the offer, in our case the buyer. Moreover, it improves tractability as subgame perfection becomes sufficient to solve the game.

Solving for equilibrium. The analysis of the game can be simplified by making two observations. First, because of the fixed cost, the buyer will either produce the quantity of fraudulent assets that is necessary to execute the offer in a match or he will produce no fraudulent asset at all. Consequently, $\tilde{\tau}(s) = [1 - \chi(s)]\, d(s)$ and $\tau(s) = \chi(s)d(s)$, where $\chi(s) = 0$ if the buyer produces fraudulent assets, and $\chi(s) = 1$ otherwise. Moreover, the buyer must be able to cover his intended transfer of genuine assets; i.e., $a(s) \geq \chi(s)d(s)$.

Second, we can solve for the buyer's optimal asset demand before solving for equilibrium offers. Indeed, if $\phi(s) = \beta$, it follows from the buyer's payoff, (3), that any demand satisfying the constraint $a(s) \geq \chi(s)d(s)$ is optimal. If $\phi(s) > \beta$, it is costly to hold assets, and so it is optimal to demand $a(s) = \chi(s)d(s)$. In both cases, substituting the optimal asset demands into the objective amounts

[13] The re-ordering methodology, called the *reordering invariance refinement*, was developed by In and Wright (2011) for signaling games with unobservable choices. This refinement is based on the invariance condition of strategic stability from Kohlberg and Mertens (1986), which requires that the solution of a game should also be the solution of any game with the same reduced normal form. (The intuitive criterion does not apply to our game because in contrast to standard signaling games types are endogenous.) Beside being powerful in selecting equilibria and tractable (because subgame perfection becomes sufficient to solve the game), this equilibrium notion has a strong decision-theoretic justification and nice normative properties. Specifically, in our model the reordered game captures the idea that upon seeing the buyer's offer, the seller will infer that the buyer's unobservable actions (portfolio and production of fraudulent assets) were chosen optimally with the offer in mind. (This forward induction logic is reminiscent to the one of most refinements in the signaling literature.) From a normative viewpoint, this refinement has the appealing property of selecting an equilibrium of the original game that yields the highest payoff to the buyer, the agent making the offer. A more detailed description of the merits of this approach is provided in In and Wright (2011).

to replacing $a(s)$ with $\chi(s)d(s)$.

With these observations in mind, a buyer's strategy specifies the following two objects: the offer, $(q, \{d(s)\})$, and conditional on any offer, a probability distribution over $\{\chi(s)\} \in \{0,1\}^S$, denoted by η. The seller's strategy specifies, conditional on any offer $(q, \{d(s)\})$, the probability of accepting, denoted by π.

The game is solved by backward induction. Following an offer, $(q, \{d(s)\})$, the seller's decision to accept a trade must be optimal given the buyer's decision to produce fraudulent assets; i.e.,

$$\pi \in \arg\max_{\hat{\pi}\in[0,1]} \hat{\pi}\left\{-q + \sum_{s\in S} \eta(s)d(s)\right\}, \tag{5}$$

where $\eta(s)$ denotes the marginal probability of bringing genuine assets of type s.[14] The seller's value of accepting the offer depends on the disutility of producing q units of goods and on the expected quality of the asset transfer, determined by η.

Similarly, following an offer $(q, \{d(s)\})$, the buyer's decision to bring genuine or fraudulent assets is optimal given the seller's probability of accepting; i.e.,

$$\{\eta(s)\} \in \arg\max_{\{\hat{\eta}(s)\}} -\sum_{s\in S}\left\{k(s)\left[1-\hat{\eta}(s)\right] + \left[\phi(s)-\beta\right]\hat{\eta}(s)d(s) + \beta\sigma\pi\hat{\eta}(s)d(s)\right\}, \tag{6}$$

where the expression that is maximized consists of the terms in the buyer's payoff that depend on η. It shows that there are two gains from producing fraudulent assets: the savings in the holding cost, $\phi(s)-\beta$, and the savings in the expected cost of transferring genuine assets to a seller.

Finally, given equilibrium decision rules $\{\eta(s)\}$ and π, the optimal offer, $(q, \{d(s)\})$, maximizes the following objective

$$-\sum_{s\in S}\left\{k(s)\left[1-\eta(s)\right] + \left[\phi(s)-\beta\right]\eta(s)d(s)\right\} + \beta\sigma\pi\left\{u(q) - \sum_{s\in S}\eta(s)d(s)\right\}. \tag{7}$$

A perfect Bayesian equilibrium that satisfies the reordering invariance refinement is a pair of buyer's and seller's strategies satisfying (5), (6), and (7). The next proposition provides a simple joint characterization of the asset demands and the offers made in any equilibrium.

[14]Note that, after replacing $a(s)$ and $\tau(s)$ with $\chi(s)d(s)$ in (3) and (4), the payoffs of buyers and sellers become linear functions of the binary actions $\{\chi(s)\}$. Therefore, taking expectations with respect to η amounts to replacing $\chi(s)$ with the marginal probability $\eta(s)$.

Proposition 1 *The asset demands, $\{a(s)\}$, and the equilibrium offers, $(q, \{d(s)\})$, solve:*

$$\max_{q,\{a(s),d(s)\}} \left\{ -\sum_{s\in S} [\phi(s)-\beta]\, a(s) + \beta\sigma\, [u(q)-q] \right\} \tag{8}$$

$$s.t. \quad \sum_{s\in S} d(s) - q = 0 \tag{9}$$

$$d(s) \leq \frac{k(s)}{\phi(s)-\beta+\beta\sigma}, \quad \text{for all } s\in S \tag{10}$$

$$d(s) \in [0, a(s)]\,, \qquad \text{for all } s\in S. \tag{11}$$

Moreover, following any equilibrium offer, the buyer transfers genuine assets with probability one, $\eta(s) = 1$ for all s, and the seller accepts the offer with probability one, $\pi = 1$.

Proposition 1 shows that equilibrium asset demands and offers maximize the buyer's expected utility subject to three constraints. First is the individual rationality constraint, (9), which states that the seller must be indifferent between accepting and rejecting the offer, given that the buyer's assets are genuine. The seller's expected payoff is zero since the bargaining protocol specifies that the buyer makes a take-it-or-leave-it offer. Second is the incentive compatibility constraint, (10), which states that the buyer must find it optimal to accumulate genuine assets with probability one, given that the seller accepts with probability one. Third is the feasibility constraint, (11), which states that the buyer must hold enough genuine assets to cover his transfer to the seller.

To understand why the buyer finds it optimal to bring genuine assets with probability one, consider a candidate equilibrium in which he brings genuine assets of type s_0 with a probability $\eta(s_0) \in (0,1)$.[15] In this candidate equilibrium, the buyer's payment capacity is slack. To see this, notice that the buyer could deviate and demand higher consumption in the DM, $q' > q$, keep the same $\{d(s)\}$, and compensate the seller by bringing genuine assets of type s_0 with higher probability, $\eta'(s_0) > \eta(s_0)$. This deviation would not change the buyer's expected cost of transferring assets, since he is indifferent between genuine or fraudulent assets of type s_0. Moreover, by (6), indifference implies:

$$k(s_0) = [\phi(s)-\beta+\beta\sigma\pi]\, d(s_0) \quad \Longrightarrow \quad \pi = \frac{k(s_0) - [\phi(s_0)-\beta]\, d(s_0)}{\beta\sigma d(s_0)};$$

i.e., the seller's probability of acceptance, π, is pinned down by the transfer $d(s_0)$, and is unaffected by the increase in q. Taken together, these observations mean that the buyer could increase his

[15] Looking at $\eta(s_0) > 0$ is without loss. See the proof of Proposition 1 for details.

payoff by raising his offer q without changing his expected cost of transferring the asset, and without changing the seller's acceptance probability.

Lastly, the proposition shows that, in equilibrium, the buyer always finds it optimal to make an offer that is accepted with probability one. This result is not obvious because offering more assets than the threshold of equation (10) has two effects going in opposite directions. The positive effect is that the buyer can demand a higher q in exchange for a higher d. The negative effect is that a larger offer increases fraud incentives, and hence it has a positive probability of being rejected. Our proof shows that, with the fixed cost of producing fraudulent assets, the negative effect always dominates.[16]

Endogenous resalability constraints. Perhaps the most important result of Proposition 1 is that the incentive-compatibility constraints, (10), take the form of resalability constraints, specifying upper bounds on the transfer of assets from buyers to sellers.[17] The resalability constraints depend on the cost of producing fraudulent assets, $k(s)$, the holding cost of an asset, $\phi(s) - \beta$, and the frequency of trades in the DM, σ.

From (10), an asset which is more susceptible to fraud is subject to a more stringent resalability constraint. To illustrate this point, suppose that there are no search frictions, $\sigma = 1$. Then, the resalability constraint of asset s is $\phi(s)d(s) \leq k(s)$. The real value of the asset that can be transferred in a bilateral match is simply the cost of producing fraudulent assets. In accordance with the Wallace (1998) dictum, the liquidity of an asset depends on its intrinsic properties, which here are captured by the ease of producing fraudulent assets.

The resalability constraints also depend on the frequency of trade in the DM. Increasing the frequency of trade exacerbates the threat of fraud because the trade surplus of a con artist, $u(q)$, is greater than the match surplus of an honest buyer, $u(q) - q$. Therefore, the upper bound must

[16] In Supplementary Appendix C, we show that the negative effect also dominates if we add proportional costs of producing fraudulent assets provided that those costs are not too large. If the proportional costs are large relative to the fixed costs, then there can be situations where fraud generates rationing both at the intensive margin (the quantity of assets that can be transferred in a match) and at the extensive margin (the number of matches in which trade occurs).

[17] If the asset is interpreted as an IOU (see Footnote 7), $s = \ell$, then one can set $\phi(\ell) = \beta$ since an IOU is issued in the DM and there is no cost of holding it. In this case the incentive-compatibility constraint, (10), takes the form of a borrowing constraint, $d(\ell) \leq \frac{k(\ell)}{\beta\sigma}$.

be lowered to keep incentives in line. To give a concrete example, if the process of securitization implies that an asset can be retraded more frequently, then an increase in securitization raises the threat of fraud and makes resalability constraints more likely to bind.[18]

Finally, the holding cost of the asset, $\phi(s) - \beta$, enters the resalability constraint, because lack of commitment forces agents to accumulate assets before liquidity needs occur. An increase in the asset price raises the holding cost, which raises the buyer's incentives to produce fraudulent versions of the asset for a given size of the trade.

4 The liquidity structure of asset returns

In this section we study the implications of our model for cross-sectional liquidity premia. We endogenize asset prices in the CM and show that the endogenous resalability constraints derived in Proposition 1 create liquidity and price differences across assets, even if they have the same cash flows. Our results help explain differences in asset prices that cannot be fully accounted for by risk, and shed light on a variety of evidence on the positive relationship between liquidity and asset prices.[19]

4.1 The liquidity-return trade-off

Assume that each asset $s \in S$ comes in fixed supply, denoted by $A(s)$. We define a symmetric equilibrium to be a collection of prices, $\{\phi(s)\}$, asset demands, $\{a(s)\}$, and a DM offer, $(q, \{d(s)\})$, such that the asset demands and the offer solve the buyer's problem (8)-(11) given prices, and the asset market clears; i.e., $a(s) = A(s)$ for all $s \in S$.[20]

Guessing that $a(s) \geq 0$ and $d(s) \geq 0$ do not bind, the first-order conditions of the buyer's

[18] Keys, Mukherjee, Seru, and Vig (2010) establish evidence that the securitization of subprime loans led to lax screening. Purnanandam (2009) finds that banks involved highly in the originate-to-distribute market, where the originator of loans sells them to third parties, originated excessively poor-quality mortgages.

[19] Since Amihud and Mendelson (1986), liquidity (level and risk) has been shown to explain risk-adjusted asset return differentials. For recent studies, see, e.g., Chordia, Huh, and Subrahmanyam (2009).

[20] The symmetry restriction that all buyers have the same asset demands serves to pin down portfolios when some assets are priced at their fundamental values, $\phi(s) = \beta$.

problem are:

$$\xi = \beta\sigma\left[u'(q) - 1\right] = \lambda(s) + \nu(s) \quad (12)$$

$$\phi(s) = \beta + \nu(s), \quad (13)$$

for all $s \in S$, where $\xi \geq 0$ is the Lagrange multiplier of the seller's participation constraint, (9), $\lambda(s) \geq 0$ is the multiplier of the resalability constraint, (10), and $\nu(s) \geq 0$ is the multiplier of the feasibility constraint, (11). The multiplier, ξ, measures the net utility of spending an additional unit of asset in the DM, if matched with a seller with probability σ. The increased consumption yields marginal utility $u'(q)$ to the buyer, and the asset transfer has an opportunity cost equal to one.

Taken together, (12) and (13) imply the following bounds on asset prices:

$$\beta \leq \phi(s) \leq \beta + \xi. \quad (14)$$

The upper bound is the present value of the asset's cash flow, β, which we refer to as the "fundamental value" of the asset, augmented by the net utility of spending an additional unit of the asset in the DM, ξ. The lower bound is the "fundamental value" of the asset, β, since a buyer can always hold onto any unit of the asset and consume its cash flow at the end of $t = 1$. Assuming for now that $q < q^*$, so that $\xi > 0$, these first-order conditions imply that there are three categories of assets.

Liquid assets. For this type of asset, the feasibility constraint is binding, $\nu(s) > 0$, but the resalability constraint is slack, $\lambda(s) = 0$. Therefore, the asset price is equal to the upper bound, $\beta + \xi$. The asset is said to be perfectly liquid in the following sense: if the buyer holds an additional unit of the asset, he would spend it in the DM. Substituting the market clearing condition, $a(s) = A(s)$, and the price, $\phi(s) = \beta + \xi$, into the binding feasibility constraint and the slack resalability constraint, we obtain $d(s) = A(s) \leq \frac{k(s)}{\xi+\beta\sigma}$. This last inequality can be equivalently written as $\kappa(s) \geq \beta\sigma + \xi$, where $\kappa(s) \equiv k(s)/A(s)$ is the cost of fraud per unit of the asset.

partially liquid assets. For this type of asset, both the resalability and feasibility constraints bind, $\lambda(s) > 0$ and $\nu(s) > 0$. In equilibrium, a buyer spends all his holdings of the asset. However, if he were to acquire an additional unit, he would choose not to spend it in the DM, for otherwise there would be a positive probability of the trade being rejected. The asset is thus said to be partially liquid and its price must be lower than the upper bound. From (10), $d(s) = A(s) = \frac{k(s)}{\phi(s)-\beta+\beta\sigma}$, which leads to $\phi(s) = \beta + \kappa(s) - \beta\sigma$, keeping in mind that $\kappa(s) = k(s)/A(s)$. The conditions $\lambda(s) = \xi + \beta - \phi(s) > 0$ and $\nu(s) = \phi(s) - \beta > 0$ can be written as $\beta\sigma < \kappa(s) < \beta\sigma + \xi$.

Illiquid assets. Lastly, there are assets for which the resalability constraint binds, $\lambda(s) > 0$, but the feasibility constraint is slack, $\nu(s) = 0$. In equilibrium the buyer does not spend a fraction of his asset holdings even though he is liquidity constrained. Therefore, the asset is said to be illiquid, and its price is equal to the lower bound, $\phi(s) = \beta$. The binding resalability constraint implies that $d(s) = \frac{k(s)}{\beta\sigma}$. Substituting this expression into the slack feasibility constraint, we obtain that $\kappa(s) \leq \beta\sigma$.

The next step is to determine ξ and verify that $q < q^*$. From the above, we have:

$$d(s) = \min\left[A(s), \frac{k(s)}{\beta\sigma}\right] = \theta(s)A(s), \quad \text{where} \quad \theta(s) = \min\left[1, \frac{\kappa(s)}{\beta\sigma}\right].$$

That is, the buyer either transfers all his holdings of asset s, or the maximum holding consistent with the resalability constraint and the no-arbitrage restriction that $\phi(s) \geq \beta$. Substituting the expression for $d(s)$ into the seller's binding participation constraint, (9), we obtain

$$q = L \equiv \sum_{s \in S} \theta(s)A(s). \tag{15}$$

The aggregate liquidity, L, is a weighted average of asset supplies, with endogenous weights depending on trading frictions and assets' recognizability characteristics.[21] Given q, the convenience yield of liquid assets, ξ, is determined by (12). One can easily verify that, if $L < q^*$, the above asset prices, offer, and asset demands constitute a symmetric equilibrium. The condition $L < q^*$ means that the aggregate liquidity is not large enough to satiate buyers' liquidity needs, represented by q^*. If $L \geq q^*$, then the equilibrium has $q = q^*$ and $\phi(s) = \beta$ for all $s \in S$. Summarizing:

[21] This approach is consistent with a definition of the quantity of money suggested by Friedman and Schwartz (1970) as "the weighted sum of the aggregate value of all assets, the weights varying with the degree of moneyness." Our definition of aggregate liquidity is also related to the Divisia monetary aggregates (e.g., Barnett, Fisher, and Serletis, 1992). A key difference is that in our approach the weight assigned to an asset in order to calculate liquidity changes is not equal to its holding cost, which has normative implications that we discuss in Section 5.

Proposition 2 ***(The liquidity-return relationship)*** *There exists a unique symmetric equilibrium. If* $L \geq q^*$*, then* $q = q^*$ *and* $\phi(s) = \beta$ *for all* $s \in S$*. If* $L < q^*$*, then* $q < q^*$*,* $\xi \equiv \beta\sigma\,[u'(q) - 1] > 0$*. Letting* $\underline{\kappa} \equiv \beta\sigma$*, and* $\overline{\kappa} \equiv \beta\sigma + \xi$*, there are three categories of assets:*

1. *Liquid assets: for any* $s \in S$*, such that* $\kappa(s) \geq \overline{\kappa}$*,*

$$\phi(s) = \beta + \xi \tag{16}$$

$$\theta(s) = 1. \tag{17}$$

2. *Partially liquid assets: for any* $s \in S$*, such that* $\kappa(s) \in (\underline{\kappa}, \overline{\kappa})$*,*

$$\phi(s) = \beta + [\kappa(s) - \beta\sigma] \tag{18}$$

$$\theta(s) = 1. \tag{19}$$

3. *Illiquid assets: for any* $s \in S$*, such that* $\kappa(s) \leq \underline{\kappa}$*,*

$$\phi(s) = \beta \tag{20}$$

$$\theta(s) = \frac{\kappa(s)}{\beta\sigma} < 1. \tag{21}$$

The central implication of Proposition 2 is that, whenever there is a liquidity shortage, $L < q^*$, assets with identical cash flows can have different prices. See Figure 3 for a graphical representation of these price differences. This departure from the no-arbitrage principle is another formulation of the rate-of-return dominance puzzle, according to which monetary assets coexist with other assets with similar risk characteristics that generate a higher yield. In our model price differentials across assets are attributed to differences in the cost of fraud. An asset which is more recognizable—in the sense of not being sensitive to fraudulent activities—as captured by a high cost of fraud, is used more intensively to finance random spending opportunities. Relative to assets that are less recognizable, this asset generates some non-pecuniary liquidity services, $\nu(s) = \phi(s) - \beta$, also referred to as a convenience yield, and is sold at a higher price.[22]

[22] To see why the price differentials do not represent arbitrage opportunities, relax the short-selling constraint and assume that, in order to sell an asset he does not own, an agent has to borrow it from someone else in exchange for a fee, to be determined in equilibrium. The agent who borrows the asset can use it in the DM, but the agent who lends it cannot. The equilibrium remains unchanged, and the fee clearing the market for borrowing asset $s \in S$ is equal to its convenience yield, $\phi(s) - \beta$. Indeed, an agent who borrows a liquid or partially liquid asset must compensate the lender for his forgone liquidity services in the DM.

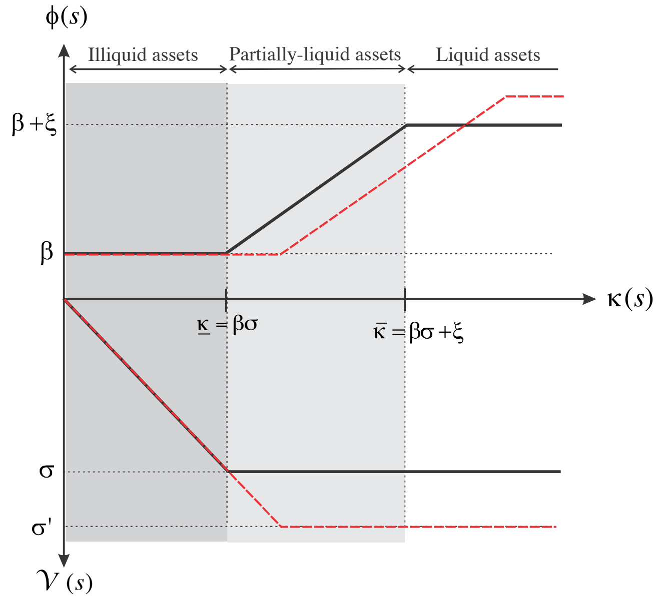

Figure 3: Liquidity structure of asset returns ($L < q^*$)

Krishnamurthy and Vissing-Jorgensen (2010, 2011) document the existence of convenience yields for Treasury securities and, to a lesser extent, highly-rated bonds. They argue that a safety-premium (which they view as distinct from a standard risk premium) is an important component of asset prices. Through the lens of our model, we can interpret this safety premium as the premium offered by assets that are highly recognizable and that are less sensitive to informational asymmetries and moral hazard considerations. Similarly, Vickery and Wright (2010) argue about the existence of a liquidity premium for agency mortgage-backed securities, which are better protected against the informational asymmetries that plague the process of securitization.

Proposition 2 also has insights for cross-sectional differences in transaction velocity, a standard measure of liquidity in monetary economies. In our model, transaction velocity in the DM is $\mathcal{V}(s) \equiv \frac{\sigma d(s)}{A(s)} = \sigma\theta(s)$. Proposition 2 predicts a positive relationship between the price of an asset and its velocity. The most liquid assets (i.e., any asset s such that $\kappa(s) \geq \bar{\kappa}$) trade at the highest price, and their velocity is maximum and equal to the frequency of spending opportunities in the DM, σ. Illiquid assets (i.e., any asset s such that $\kappa(s) <\underline{\kappa}$), however, have the highest rate of return, equal to the rate of time preference, and the lowest velocities, less than σ. This result is consistent with the view that bonds that are used more intensely as collateral in OTC markets tend to have higher prices (Duffie, 1996).

In reality, a myriad of assets are not used as means of payment or collateral. This observation is consistent with our results if there is a mass of assets that do not circulate in the DM, $\theta(s) = 0$. From (21) such assets must be characterized by $\kappa(s) = 0$: these are assets for which agents have so little knowledge about their mere existence or attributes, that even simple, costless frauds can be deceptive.[23]

5 Applications and Extensions

In this last section we apply our model of the liquidity structure of asset returns to analyze flight-to-liquidity phenomena and to assess the effectiveness of aggregate liquidity management policies. Moreover, we extend the model to an infinite time horizon in order to study time variations in liquidity premia.

5.1 Flights to liquidity

A flight to liquidity occurs when market participants seek to reallocate their portfolios toward highly liquid assets, which leads to a widening yield spread between liquid and less liquid assets.[24] In what follows, we apply our analysis on the liquidity structure of asset returns to identify the shocks that can generate a simultaneous increase in the prices of the most-liquid assets and a reduction in the prices of less-liquid ones—a phenomenon resembling a flight to liquidity.

According to our model, a flight to liquidity can be explained by an exogenous reduction in $k(s)$ for some initially liquid or partially liquid assets that make them become illiquid. For instance, agents might realize that some assets (e.g., MBS) can be subject to a broader set of fraudulent

[23] That assets, or claims on those assets, can be counterfeited at no cost has been the standard explanation in monetary theory for why capital goods are illiquid, since Freeman (1985), and more recently, Lester, Postlewaite, and Wright (2011).

[24] During the 1998 Russian-default crisis, many investors shifted their funds into the more liquid U.S. Treasury market, widening the yield spread between Treasury bonds and less-liquid debt instruments (Longstaff, 2004). Evidence also shows that, during the subprime crisis, the flight-to-quality was confined to AAA-rated bonds, and the illiquidity component of the rate of return of bonds with lower grades rose sharply (Longstaff, 2010; Dick-Nielsen, Feldhutter, and Lando, 2010).

practices than previously thought.[25] The resalability and velocity of these assets decrease, which causes aggregate liquidity and output to fall, and the liquidity premium on liquid assets, ξ, to increase.[26] The prices of partially liquid assets do not change, except for the ones that are characterized by a lower cost of fraud. In addition, an increase in the threat of fraud can shrink the set of liquid assets, while it expands the set of illiquid and partially liquid ones. Indeed, the threshold $\bar{\kappa} = \beta\sigma + \xi$ and the interval $\bar{\kappa} - \underline{\kappa} = \xi$ are increasing functions of the size of the liquidity premium on liquid assets. Therefore, during a flight to liquidity, market demand for assets is concentrated on a smaller set of highly recognizable assets.

An alternative explanation for a flight to liquidity is an increase in σ that formalizes an aggregate liquidity demand shock, e.g., an increase in counterparty risk, leading to an increase in the demand for collateral for OTC transactions.[27] From (16) and (18) when σ increases the prices of liquid assets rise, whereas the prices of partially liquid assets fall. The increase in the prices of liquid assets occurs due to two effects going in the same direction. There is a direct effect according to which liquid assets are used more often as collateral or means of payment, which raises their liquidity value. The indirect effect is to reduce aggregate liquidity: from (15), an increase in σ lowers the weights of illiquid assets in L, which reduces the output in bilateral matches and makes liquid assets even more useful; i.e., the term $\beta\left[u'(q) - 1\right]$ in (12) goes up. For partially liquid assets the increase in σ has the additional markedly different effect of exacerbating fraud incentives. As a result, their prices have to fall so that their resalability constraints hold, re-establishing buyers' incentives to bring genuine assets. As shown in Figure 3, the set of illiquid and partially liquid assets expands (because $\underline{\kappa}$ increases with σ and $\bar{\kappa} - \underline{\kappa}$ increases with ξ) while the set of liquid assets shrinks (because $\bar{\kappa}$ increases in ξ).

[25] For some prominent economists this type of shock is a central explanation for the financial crisis of 2008. In an interview to the Wall Street journal (09/24/2011), Robert Lucas argued that "the shock came because complex mortgage-related securities minted by Wall Street and certified as safe by rating agencies had become part of the effective liquidity supply of the system. All of a sudden, a whole bunch of this stuff turns out to be crap".

[26] Some recent studies (e.g., Ajello, 2010; Shi, 2011) formulate the hypothesis that recessions are driven by liquidity shocks formalized by a reduction in the exogenous resalability of some assets. In contrast to our approach, these models have the counterfactual implication that the prices of the assets that become more difficult to resell increase.

[27] Suppose, for instance, that a fraction σ_u of the trades in the DM can be financed with unsecured debt (e.g., because commitment/enforcement is available in those meetings) while a fraction σ_s of the trades require collateral to be posted because of counterparty risk (e.g., sellers in those meetings cannot commit or cannot be forced to repay their debt.) An increase in counterparty risk can be formalized as an increase in σ_s such that $\sigma_s + \sigma_u$ is unchanged.

5.2 Liquidity management

In this section we use our model to study the effectiveness of policies aimed at managing the supply of liquidity in the economy. These policies can take the form of open-market operations by the central bank, which are intended to substitute liquid assets for less-liquid ones, or regulatory measures that reduce the threat of frauds and relax resalability constraints for some assets.

Measuring the social value of assets' liquidity services. Much of the analysis that follows is based on the following theoretical observation. In competitive models with reduced-form demand for liquidity (e.g., cash-in-advance or money-in-the-utility function), the convenience yield of an asset not only measures the *marginal private value* of its liquidity services, but also its *marginal social value*.[28] In our model this property holds true for illiquid and liquid assets, but fails to hold for partially liquid assets.

The marginal social value of the liquidity services provided by a unit of asset s is $\frac{\partial L}{\partial A(s)}\xi$, which is equal to ξ for liquid *and* partially liquid assets, and 0 for illiquid assets. Therefore, the convenience yield of partially liquid assets, $\phi(s) - \beta < \xi$, underestimates the true marginal social value of their liquidity services. The reason for this discrepancy is that an increase in the price of an asset reduces its demand in two ways: by raising the holding cost, $\phi(s) - \beta$, and by tightening the resalability constraint. The latter effect creates a negative "pecuniary externality," which can depress asset prices below the marginal social value of the asset's liquidity services.[29] As we show below, this observation implies that liquidity management policies targeting partially liquid assets can be welfare reducing, because they underestimate these assets' true contribution to aggregate liquidity. By contrast, when targeting illiquid assets, the same policies are welfare improving.

Open-market purchases. Central banks routinely engage in aggregate liquidity management, by issuing (or withdrawing) reserves, the most liquid assets, in exchange for Treasuries and, in recent

[28] This logic is underlying the calculation for the welfare cost of inflation in Lucas (2000), the measure of the liquidity services provided by Treasuries in Krishnamurthy and Vissing-Jorgensen (2010), and Barnett, Fisher, and Serletis's (1992) definition of Divisia monetary aggregates.

[29] By contrast, with the exogenous proportional resalability constraint, there is no such pecuniary externality, and asset convenience yields coincide with the marginal social value of the asset's liquidity services. See Supplementary Appendix E.

years, a wider range of less liquid assets, including agency bonds and mortgage-backed securities. Consider a policy-maker in the CM, who sells a quantity, $\Delta A(s)$, of some liquid asset s from his portfolio, and simultaneously purchases a quantity, $\Delta A(s')$, of some other asset s'. A small open-market operation has a small effect on prices, so that the budget constraint of the policy-maker is, to a first-order approximation, $\phi(s)\Delta A(s) + \phi(s')\Delta A(s') = 0$. The welfare effect of such a policy is

$$\Delta L \times \xi = \left[\frac{\partial L}{\partial A(s)}\Delta A(s) + \frac{\partial L}{\partial A(s')}\Delta A(s')\right]\xi = \left[1 - \frac{\partial L}{\partial A(s')}\frac{\phi(s)}{\phi(s')}\right]\Delta A(s) \times \xi.$$

Suppose first that $\kappa(s') > \bar{\kappa}$, so both s and s' are liquid assets. Then, $\phi(s) = \phi(s')$, $\frac{\partial L}{\partial A(s')} = 1$, and $\Delta L = 0$. Such an open-market operation is irrelevant: it does not change aggregate liquidity and welfare, and hence it has no effect on output and asset prices. So liquidity management has real effects only if it involves assets with different degrees of liquidity.

Suppose next that $\kappa(s') <\underline{\kappa}$, asset s' is illiquid. In this case aggregate liquidity does increase because the purchase of illiquid assets has no consequence on aggregate liquidity; i.e., $\Delta L \times \xi = \Delta A(s) \times \xi > 0$. Thus, welfare increases, the price of liquid assets decreases, and the price of illiquid assets is unaffected.

Finally, suppose that $\kappa(s') \in (\underline{\kappa}, \bar{\kappa})$; i.e., asset s' is partially liquid. Then, $\phi(s') < \phi(s)$ and $\Delta L \times \xi = \left[1 - \frac{\phi(s)}{\phi(s')}\right]\Delta A(s) \times \xi < 0$, implying that such a policy *reduces* aggregate liquidity and welfare. The intuition is in line with our earlier observation: while partially liquid and liquid assets have different prices, they contribute *equally* to aggregate liquidity. At the same time, because it has a higher price, one share of a liquid asset buys more than one share of a partially liquid one. Thus a balanced-budget open-market operation ends up syphoning out more liquidity than it is injecting in; i.e., aggregate liquidity is reduced. The welfare effect of this open-market operation is of the opposite sign of the yield difference between the asset that is withdrawn and the asset that is injected, and the prices of both assets s and s' increase.

The results above can help interpret some of the findings in Krishnamurthy and Vissing-Jorgensen (2011) regarding the effect of quantitative easing. They find that the purchases of Treasuries, agency bonds, and highly-rated corporate bonds in exchange for reserves led to a drop in interest rates but it did not affect the yields on relatively illiquid assets (Baa corporate bonds). This finding is consistent with our results if we interpret Baa corporate bonds as illiquid assets, Treasuries and highly rated bonds as partially liquid, and reserves as fully liquid. Furthermore, according to our findings, the drop in interest rates indicates that quantitative easing reduced liquidity and welfare.

Regulatory measures. Some of the leading regulatory measures of the Dodd-Frank Act aim to curb fraud incentives in the securitization industry.[30] One of these measures is a requirement for securitizers to retain at least 5 percent of the credit risk they originate. Importantly, some asset-backed securities, deemed of higher quality, are exempted from this requirement. In this section, we study the optimality and welfare impact of retention requirements. We show that the regulator faces a trade-off between the role these requirements play as a discipline mechanism and the distortion they introduce by increasing the costs of holding assets. We demonstrate that the first effect dominates for illiquid assets, while the second effect dominates for partially liquid and liquid assets. Hence, our model suggests that retention requirements should be confined to the least liquid assets, i.e., the ones more susceptible to fraud.

Under a retention requirement policy, a buyer who wishes to transfer $d(s)$ units of asset s in the DM must hold $1+\rho(s)$ units of the asset; i.e., $d(s) \leq \frac{a(s)}{1+\rho(s)}$, where $\rho(s)$ is the retention rate associated with asset s. The policy imposes that the asset kept in retention is the exact same asset as the one transferred in a match, i.e., if the asset transferred is fraudulent, so is the asset in retention.[31] The cost of producing $d(s)$ units of fraudulent asset is of the form $k_f(s) + k_v(s)d(s)$, where the variable cost component, $k_v(s)d(s)$, was introduced to provide a channel through which the regulatory measure can reduce agents' incentive to commit fraud. We let $k_f(s) > 0$ and take $k_v(s)$ to be small enough so that, as before, equilibrium offers are accepted with probability one (see Supplementary Appendix B). The resalability constraint of asset s becomes:

$$k_f(s) + k_v(s)\left[1+\rho(s)\right]d(s) \geq \left[\phi(s) - \beta\right]\left[1+\rho(s)\right]d(s) + \beta\sigma d(s). \tag{22}$$

The left side of (22) is the cost of fraud on $d(s)$ units of asset s. If $k_v(s) > 0$, then policy increases

[30] The Dodd-Frank Act, enacted in July 2010 in response to the 2007-08 financial crisis, institutes a wide array of new regulations for the financial services industry.

[31] In the context of securitization (see Supplementary Appendix H), a retention requirement means that the securitizer (represented in the model by the buyer) needs to retain assets from the same issue of asset-based securities he is offering to the general public (represented in the model by the seller).

the cost of fraud and, therefore, reduces fraud incentives. The right side of (22) is the cost of holding $[1+\rho(s)]\,d(s)$ genuine units of asset s. Thus, if the asset is liquid or partially liquid, $\phi(s)-\beta>0$, the retention requirement generates a distortion by increasing the effective holding cost of the asset.

In Supplementary Appendix B, we solve for equilibrium following the same steps as before. We show that a retention requirement has asymmetric effects on the resalability of an asset, depending on its liquidity status. For illiquid assets, equilibrium resalability becomes:

$$\theta(s)=\frac{k_f(s)/A(s)}{\beta\sigma-k_v(s)\,[1+\rho(s)]}. \tag{23}$$

It is an *increasing* function of $\rho(s)$ because when $\phi(s)=\beta$ retention rates raise the cost of committing fraud but do not affect assets' effective holding costs. For liquid and partially liquid assets resalability becomes:

$$\theta(s)=\frac{1}{1+\rho(s)}, \tag{24}$$

which is a *decreasing* function of $\rho(s)$. Thus, for liquid or partially liquid assets, the distortionary effect of retention rates dominates the incentive effect, reducing velocity and welfare. In the case of liquid assets, this result is straightforward since the threat of fraud is not a binding constraint. In the case of partially liquid assets, retention requirements have the partial equilibrium effect of relaxing resalability constraints. But this effect simultaneously increases the demand for partially liquid assets. Therefore, for asset markets to clear, the prices of partially liquid assets must increase, tightening back the resalability constraints and eliminating the positive incentive effect of the retention policy. Taken together, the above results suggest that retention requirements should target illiquid assets. Other, more recognizable assets, should be exempted, in line with the prescriptions of the Dodd Frank Act.

5.3 Dynamics of liquidity premia

This section provides a dynamic extension of our static model. We show that the prices of all liquid assets covary with a common liquidity premium.[32] This common liquidity premium can be

[32] References on the empirical literature on the co-movements of liquidity across assets are included in Acharya and Pedersen (2005) who develop an asset pricing model in which a per-share cost of selling securities can vary over time.

the subject of self-fulfilling fluctuations, creating excess volatility in the price of liquid assets. The prices of illiquid and partially liquid assets are, however, immune to such fluctuations.

Given agents' quasilinear preferences, it is straightforward to introduce an infinite time horizon using the setup of Lagos and Wright (2005). Time is indexed by $t \in \mathbb{N}$. Each period is divided into two subperiods, a DM followed by a CM. Each unit of the asset s pays off a dividend normalized to one unit of the numéraire at the beginning of each CM. The technology to produce fraudulent assets in period $t-1$ becomes obsolete in period t, and all fraudulent assets produced in period t are confiscated by the government before agents enter the CM of period t.[33]

As shown in supplementary Appendix D, these assumptions allow us to apply the analysis of the static model, where the terminal value of the asset is equal to the cum-dividend value, $1+\phi_t(s)$, of reselling the asset in the CM in period t. Focusing on equilibrium with $q_t < q^*$, Proposition 2 generalizes as follows.[34] There are three classes of assets, of which the prices solve:

$$\phi_{t-1}(s) = [1+\phi_t(s)] \times \begin{cases} \beta + \xi_t & \kappa_t(s) \geq \bar{\kappa}_t \\ \beta + [\kappa_t(s) - \beta\sigma] \quad \text{if} & \kappa_t(s) \in (\underline{\kappa}, \bar{\kappa}_t) \\ \beta & \kappa_t(s) \leq \underline{\kappa} \end{cases}, \tag{25}$$

where $\kappa_t(s) \equiv \frac{k(s)}{[1+\phi_t(s)]A(s)}$, $\underline{\kappa} \equiv \beta\sigma$, $\bar{\kappa}_t \equiv \xi_t + \beta\sigma$, and

$$\xi_t = \beta\sigma \left[u'(q_t) - 1\right] \tag{26}$$

$$q_t = L = \sum_{s \in S} \theta_t(s) [1+\phi_t(s)] A(s), \quad \text{where} \quad \theta_t(s) = \min\left[1, \frac{\kappa_t(s)}{\beta\sigma}\right]. \tag{27}$$

The equilibrium equations are the same as in the static model, but with an endogenous terminal value of $1+\phi_t(s)$. This difference is substantial because expectations of future liquidity premia, capitalized in $\phi_t(s)$, feed back into the current liquidity premium, $\phi_{t-1}(s) - \beta[1+\phi_t(s)]$.[35]

We now characterize equilibria in a neighborhood of the unique steady state, $\langle\{\bar{\phi}(s)\}, \bar{q}, \bar{\xi}\rangle$. In such a neighborhood, the sets of liquid, partially liquid, and illiquid assets do not change. Moreover,

[33] This assumption, borrowed from Nosal and Wallace (2007), is made for tractability to prevent fraudulent assets from circulating across periods.

[34] If aggregate liquidity is abundant, there exists a unique equilibrium in which the resalability constraint and the feasibility constraint do not bind at any date, $q_t = q^*$, and each asset is priced at its fundamental value; i.e., $\phi_t(s) = 1/r$.

[35] This effect, as is well known, can lead to an equilibrium in which an asset has positive value even if it pays no dividend; i.e., a positive liquidity premium can be a self-fulfilling phenomenon. In Supplementary Appendix D we consider such an economy with fiat money.

from (25), one can verify that $\frac{d\phi_{t-1}(s)}{d\phi_t(s)} \in [0,1)$ for all $\kappa_t(s) < \bar{\kappa}_t$, so that the prices of illiquid and partially liquid assets are equal to their steady-state values in any dynamic equilibrium. This need not be the case for liquid assets. To see this point, let us linearize the equilibrium equations near the steady state. We obtain, from (25), that the price of liquid assets solves:

$$\hat{\phi}_{t-1}(s) = \left(\beta + \bar{\xi}\right)\hat{\phi}_t(s) + \left[\bar{\phi}(s) + 1\right]\hat{\xi}_t, \tag{28}$$

where $\hat{\phi}_t(s) \equiv \phi_t(s) - \bar{\phi}(s)$ and $\hat{\xi}_t \equiv \xi_t - \bar{\xi}$. The first term on the right side of (28) is the discounted value of the future price of the asset, with the discount rate augmented by a liquidity premium; the second term captures the change in the liquidity premium. Linearizing (26) and (27) in the neighborhood of the steady state:

$$\hat{\xi}_t = \beta\sigma u''(\bar{q})\hat{q}_t \quad \text{where} \quad \hat{q}_t = \sum_{s:\kappa(s)\geq\bar{\kappa}} \hat{\phi}_t(s)A(s), \tag{29}$$

and $\hat{q}_t \equiv q_t - \bar{q}$. From (29), the size of the liquidity premium, relative to its steady-state value, depends negatively on changes in the market capitalization of liquid assets.

Multiplying both sides of (28) by $A(s)$ and taking the sum over all liquid assets, we obtain

$$\hat{\xi}_{t-1} = \gamma\hat{\xi}_t, \quad \text{where} \quad \gamma = \beta + \bar{\xi} + \beta\sigma u''(\bar{q}) \sum_{s:\kappa(s)\geq\bar{\kappa}} \left[\bar{\phi}(s) + 1\right] A(s). \tag{30}$$

The nature of the dynamics depends on γ. If $\gamma > -1$, then $\hat{\xi}_t = \bar{\xi}$ for all t, and the liquidity premium is constant over time. If $\gamma < -1$, in contrast, there exists a continuum of equilibria indexed by the initial value of $\hat{\xi}$ in the neighborhood of zero that converges to the steady state. Along these equilibria, $\hat{\xi}_t$ alternates between positive and negative values. The price of liquid assets covaries and exhibits excessive volatility relative to fundamentals, whereas the prices of partially liquid assets and illiquid ones remain constant. As can be seen from (29), the fluctuating liquidity premium is a self-fulfilling phenomenon. If agents anticipate that the liquidity premium will be high in the future so that $\hat{\phi}_t > 0$, then aggregate liquidity and output are high, $\hat{q}_t > 0$. But, at the margin, agents do not value much the asset's liquidity services, and so the current liquidity premium is low, $\hat{\xi}_t < 0$. If the constant multiplying $\hat{\xi}_t$ in (28) is large enough, then $\hat{\phi}_{t-1} < 0$. The same reasoning implies that the liquidity premium one period before will be high, $\hat{\xi}_{t-1} > 0$, and the fluctuations will continue.

6 Conclusion

In this paper we have proposed a theory of the cross-sectional distribution and time-variation of liquidity premia by taking seriously the possibility of asset fraud in an economy with limited commitment and enforcement. We have shown the emergence of asset-specific resalability constraints that take the form of upper bounds on the transfer of assets in OTC market trades. These bounds are not invariant to policy shifts (e.g., the composition of asset supplies and regulation on assets' retention requirements), and they depend on some characteristics of the assets such as their vulnerability to fraud, as well as the frequency of trading opportunities. Our model generates a liquidity structure of asset returns based on a three-tier classification of assets. This classification is relevant for the comparative statics of asset prices, the dynamics of liquidity premia, explanations of flight-to-liquidity phenomena, and the analysis of open-market operations and regulations of the OTC market.

References

[1] Ajello, Andrea (2010). "Financial Intermediation, Investment Dynamics and Business Cycle," manuscript.

[2] Acharya, Viral V. and Lasse Heje Pedersen (2005). "Asset Pricing with Liquidity Risk," Journal of Financial Economics, Elsevier 77, 375-410.

[3] Amihud, Yakov and Haim Mendelson (1986). "Asset Pricing and the Bid Ask Spread," Journal of Financial Economics 17, 223-249.

[4] Barnett, William A., Douglas Fisher, and Apostolos Serletis (1992). "Consumer Theory and the Demand for Money," Journal of Economic Literature 30, 2086-2119.

[5] Chordia, Tarun, Huh Sahn-Wook and Avanidhar Subrahmanyam (2009). "Theory-Based Illiquidity and Asset Pricing," Review of Financial Studies 22, 3629-3668.

[6] Dick-Nielsen, Jens, Peter Feldhutter, and David Lando (2010). "Corporate Bond Liquidity Before and After the Onset of the Subprime Crisis," manuscript.

[7] Duffie, Darell (1996). "Special Repo Rates" Journal of Finance 51, 493-526.

[8] Duffie, Darrell, Nicolae Garleanu and Lasse Heje Pedersen (2005). "Over-the-counter Markets," Econometrica 73, 1815-1847.

[9] Freeman, Scott (1985). "Transactions Costs and the Optimal Quantity of Money," Journal of Political Economy 93, 146-157.

[10] Green, Edward and Warren Weber (1996). "Will the New $100 Bill Decrease Counterfeiting?" Federal Reserve Bank of Minnepolis Quarterly Review 20, 3-10.

[11] Guerrieri, Veronica and Robert Shimer (2011). "Dynamic Adverse Selection: A Theory of Illiquidity, Fire Sales, and Flight to Quality," mimeo.

[12] Guerrieri, Veronica, Robert Shimer, and Randall Wright (2010). "Adverse Selection in Competitive Search Equilibrium," Econometrica 78 , 1823-1862.

[13] Holmstrom, Bengt and Jean Tirole (2011). Inside and Outside Liquidity, Princeton, Princeton University Press.

[14] Hopenhayn, Hugo A. and Ingrid M. Werner (1996). "Information, Liquidity, and Asset Trading in a Random Matching Game," Journal of Economic Theory 68, 349-379.

[15] In, Younghwan and Julian Wright (2011). "Signaling Private Choices," manuscript.

[16] Keys, Benjamin J., Tanmoy K. Mukherjee, Amit Seru, and Vikrant Vig (2010). "Did securitization lead to lax screening? Evidence from subprime loans," The Quarterly Journal of Economics 125, 307-62.

[17] Kiyotaki, Nobuhiro and John Moore (2001). "Liquidity, Business Cycles and Monetary Policy," manuscript.

[18] Kiyotaki, Nobuhiro and John Moore (2005). "Liquidity and asset prices," International Economic Review 46, 317-349.

[19] Kohlberg, Elon and Jean-Francois Mertens (1986). "On the strategic stability of equilibria," Econometrica 54, 1003-1037.

[20] Krishnamurthy, Arvind and Annette Vissing-Jorgensen (2010). "The aggregate demand for Treasury debt," manuscript.

[21] Krishnamurthy, Arvind and Annette Vissing-Jorgensen (2011). "The Effects of Quantitative Easing on Interest Rates," manuscript.

[22] Lagos, Ricardo (2010). "Asset prices and liquidity in an exchange economy," Journal of Monetary Economics 57, 913-930.

[23] Lagos, Ricardo and Randall Wright (2005). "A unified framework for monetary theory and policy analysis," Journal of Political Economy 113, 463–484.

[24] Lester, Benjamin, Andrew Postlewaite and Randall Wright (2011). "Information, Liquidity, Asset Prices, and Monetary Policy," Review of Economic Studies, forthcoming.

[25] Longstaff, Francis A. (2004). "The Flight-to-Liquidity Premium in U.S. Treasury Bond Prices," Journal of Business 77, 511-526.

[26] Longstaff, Francis A. (2010). "The subprime credit crisis and contagion in financial markets," Journal of Financial Economics 97, 436-450.

[27] Lucas, Robert E. (2000). "Inflation and Welfare," Econometrica 68, 247-274.

[28] Friedman, Milton and Anna J. Schwartz (1970). Monetary statistics of the United States: estimates, sources, methods, NBER Studies in business cycles, no. 20, New York: National Bureau of Economic Research.

[29] Mihm, Stephen (2007). A nation of counterfeiters: Capitalists, con men, and the making of the United States, Cambridge, Harvard University Press.

[30] Nosal, Ed and Neil Wallace (2007). "A model of (the threat of) counterfeiting," Journal of Monetary Economics 54, 229-246.

[31] Purnanandam, Amiyatosh K. (2009). "Originate-to-Distribute Model and the Subprime Mortgage Crisis," manuscript.

[32] Rocheteau, Guillaume (2009). "A Monetary Approach to Asset Liquidity," Working Papers 09-01, Federal Reserve Bank of Cleveland.

[33] Sargent, Thomas J. and Francois R. Velde (2002). The Big Problem of Small Change, Princeton: Princeton University Press.

[34] Schreft, Stacey L. (2007). "Risks of identity theft: Can the market protect the payment system?" Federal Reserve Bank of Richmond, Economic Review, Fourth Quarter, 5-40.

[35] Shi, Shouyong (2008). "Efficiency improvement from restricting the liquidity of nominal bonds," Journal of Monetary Economics 55, 1025– 1037.

[36] Shi, Shouyong (2011). "Liquidity, Assets and Business Cycles," manuscript.

[37] Snowden, Kenneth A. (2010) “Covered Farm Mortgage Bonds in the Late Nineteenth Century U.S,” NBER Working Papers 16242

[38] Vickery, James and Joshua Wright (2010). “TBA Trading and Liquidity in the Agency MBS Market,” manuscript.

[39] Wallace, Neil (1998). “A dictum for monetary theory,” FRB Minneapolis Quarterly Review 22: 20-26.

[40] Wallace, Neil (2000). “A model of the liquidity structure based on asset indivisibility,” Journal of Monetary Economics 45, 55-68.

[41] Weill, Pierre-Olivier (2008). “Liquidity premia in dynamic bargaining markets,” Journal of Economic Theory 140, 66-96.

[42] Williamson, Stephen and Randall Wright (1994). “Barter and Monetary Exchange under Private Information,” American Economic Review 84, 104-123.

A Proof of Proposition 1

We define an outcome of the game as an offer $(q, \{d(s)\})$ made by the buyer, probabilities $\{\eta(s)\}$ of bringing genuine assets, and a probability $\pi \in [0,1]$ that the seller accepts the offer. Let us consider the auxiliary problem of choosing an outcome in order to maximize the expected utility of a buyer,

$$-\sum_{s\in S}\Big\{k(s)\left[1-\eta(s)\right]+\left[\phi(s)-\beta\right]\eta(s)d(s)\Big\}+\sigma\beta\pi\left[u(q)-\sum_{s\in S}\eta(s)d(s)\right], \tag{31}$$

subject to the constraint that the probabilities π and $\{\eta(s)\}$ are the basis of an equilibrium in the sub-game following offer $(q, \{d(s)\})$; that is:

$$\pi \in \arg\max_{\hat{\pi}\in[0,1]} \hat{\pi}\Big\{-q+\sum_{s\in S}\eta(s)d(s)\Big\} \tag{32}$$

$$\eta(s) \in \arg\max_{\hat{\eta}\in[0,1]} \hat{\eta}\Big\{k(s)-\left[\phi(s)-\beta+\beta\sigma\pi\right]d(s)\Big\}, \text{ for all } s\in S. \tag{33}$$

We start by showing that:

Claim 1 *Any solution of the auxiliary problem has the property that* $\eta(s)=1$ *and* $q=\sum_{s\in S} d(s)$.

Proof. Consider first any feasible outcome $\langle q, d, \eta, \pi\rangle$ such that $\eta(s_0) < 1$ for some s_0. If $\eta(s_0) = 0$, then consider the alternative outcome, $\langle q', d', \eta', \pi'\rangle$, such that: (i) $q' = q$, $d'(s) = d(s)$ for all $s \neq s_0$, $d'(s_0) = 0$; (ii) $\eta'(s) = \eta(s)$ for all $s \neq s_0$ and $\eta'(s_0) = 1$; (iii) $\pi' = \pi$. The incentive constraint of the seller, (32), is satisfied since it only depends on the product $\eta(s)d(s)$. The incentive constraint of the buyer, (33), is obviously satisfied for $s \neq s_0$. For $s = s_0$ we have $k(s_0) > \left[\phi(s) - \beta + \beta\sigma\right] d'(s_0) = 0$ and so $\eta'(s_0) = 1$ is optimal for the buyer. One can then verify that, with this alternative outcome, the expected utility of the buyer increases by $k(s_0) > 0$.

Next, consider any feasible outcome such that $\eta(s_0) \in (0,1)$: the buyer is indifferent between counterfeiting asset s_0 or not. We then increase $\eta(s_0)$ by $\varepsilon \in (0,1]$ and q by $\varepsilon d(s_0)$, which is positive since the incentive constraint of the buyer, (33), binds. The incentive constraint of the seller, (32) is satisfied because his payoff conditional on accepting the offer does not change. Because the buyer is indifferent between counterfeiting asset s_0 or not, an increase in $\eta(s_0)$ affects neither his payoff, (31), nor his incentive constraint, (33). The corresponding increase in q, however, increases his payoff strictly.

Next, consider any feasible outcome $\langle q, d, \eta, \pi \rangle$ such that $\eta(s) = 1$ for all s, but $q < \sum_{s \in S} d(s)$. Then the alternative outcome with $q' = \sum_{s \in S} > q$, $\eta'(s) = 1$, and $\pi' = \pi$, increases the expected payoff to the buyer by $\sigma \beta \pi \left[u(q') - u(q) \right] > 0$ and satisfies all the constraints. ■

This claim implies that we can rewrite the auxiliary problem as

$$\max_{q,d,\pi} - \sum_{s \in S} \left[\phi(s) - \beta \right] d(s) + \sigma \beta \pi \left[u(q) - q \right] \tag{34}$$

$$\text{s.t.} \quad \sum_{s \in S} d(s) - q = 0 \tag{35}$$

$$k(s) \geq \left[\phi(s) - \beta + \beta \sigma \pi) \right] d(s), \text{ for all } s \in S. \tag{36}$$

The second condition is the first-order necessary and sufficient condition for (33) evaluated at $\eta(s) = 1$. Next, we show:

Claim 2 *Any solution of the auxiliary problem, (31)-(33), has the property that* $u'(q) \geq 1$ *and* $\pi = 1$.

Proof. The first claim holds because otherwise one could reduce the quantity produced, increase the expected utility of the buyer, and satisfy all the constraints. To prove the second claim suppose, towards a contradiction, that $\pi < 1$. Note first that the value of the auxiliary problem must be positive: a small offer $q' = d'(s_0) > 0$, $d'(s) = 0$ for $s \neq s_0$, and $\pi' = 1$ yields a positive payoff. This implies that both $q > 0$ and $\pi > 0$. Moreover, at least one of the incentive constraints, (36), must be binding since otherwise one could increase π without violating any of the incentive constraints, and improve the objective. Let $S_B \subseteq S$ be the set of binding IC constraints. Since $\left[\phi(s) - \beta + \beta \sigma \pi \right] d(s) = k(s)$ for all $s \in S_B$, it follows that $d(s) > 0$. Now consider the following variational experiment: increase π by some small ε and decrease the payments $d(s)$, for all $s \in S_B$, so that all the incentive constraints continue to bind. Up to second-order terms, the decrease in $d(s)$ is equal to $m(s)\varepsilon$, where

$$m(s) \equiv \frac{\beta \sigma d(s)}{\phi(s) - \beta + \beta \sigma \pi},$$

is the marginal rate of substitution between π and $d(s)$ in the IC constraint for asset $s \in S_B$. Lastly, to satisfy the participation constraint, the decrease in q must be, up to second order terms,

$\sum_{s\in S_B} m(s)\varepsilon$. The change in the buyer's expected utility is, up to second-order terms, equal to $\Delta U \times \varepsilon$, where

$$\begin{aligned}\Delta U &= \sum_{s\in S_B} [\phi(s)-\beta]\, m(s) - \beta\sigma\pi \left[u'(q)-1\right] \sum_{s\in S_B} m(s) + \beta\sigma \left[u(q)-q\right] \\ &> \sum_{s\in S_B} [\phi(s)-\beta]\, m(s) + \beta\sigma \left[u'(q)-1\right] \left[q - \pi \sum_{s\in S_B} m(s)\right] \\ &\geq \sum_{s\in S_B} [\phi(s)-\beta]\, m(s) + \beta\sigma \left[u'(q)-1\right] \sum_{s\in S_B} [d(s)-\pi m(s)] = \sum_{s\in S_B} [\phi(s)-\beta]\, m(s) u'(q) \geq 0,\end{aligned}$$

where we move from the first line to the second line using $u(q)-q > q\,[u'(q)-1] \geq 0$ (the equality is strict because of two facts: first, $u(q)$ is strictly concave and second, $q>0$, since the value of the auxiliary problem is positive); from the second line to the third line using $q \geq \sum_{s\in S_B} d(s)$; and from the third to the fourth line by noting that $d(s) - \pi m(s) = [\phi(s)-\beta]\, m(s)/(\beta\sigma)$. ■

From Claims 1-2 and the result according to which $a(s) \geq \chi(s)d(s)$ if $\phi(s) = \beta$, and $a(s) = \chi(s)d(s)$ if $\phi(s) > \beta$, it follows that the auxiliary problem, (31)-(33), reduces to the maximization problem of Proposition 1, (8)-(11). Now we note that the solution to the auxiliary problem is an upper bound on the value of the buyer in any equilibrium of the game. Let $(\tilde{q}, \{\tilde{d}(s)\})$ be one solution of the auxiliary problem. Because, as argued above, the value of the auxiliary problem is positive, it must satisfy $\tilde{q} > 0$ and $\tilde{d}(s) > 0$ for some $s \in S$. Consider, for any $\varepsilon > 0$ small enough, the offer $d^{\varepsilon}(s) = \max\{\tilde{d}(s)-\varepsilon, 0\}$ and $q^{\varepsilon} = \tilde{q} - (S+1)\varepsilon$. By construction, this offer is such that $[\phi(s)-\beta+\beta\sigma]\, d^{\varepsilon}(s) < k(s)$, and $q^{\varepsilon} < \sum_{s\in S} d^{\varepsilon}(s)$. Thus, $\pi = 1$ and $\eta(s) = 1$ is the unique equilibrium in the subgame following $(q^{\varepsilon}, \{d^{\varepsilon}(s)\})$. By letting ε go to zero and making the above offer, the buyer can achieve a value arbitrarily close to that of the auxiliary problem. Therefore, in any equilibrium, the buyer's value must be equal to that of the auxiliary problem. Moreover, any equilibrium outcome satisfies (32) and (33). Therefore, any equilibrium outcome must solve the auxiliary problem.