Private and Public Liquidity Provision in Over-the-Counter Markets*

David M. Arseneau
Federal Reserve Board

David E. Rappoport W.
Federal Reserve Board

Alexandros P. Vardoulakis
Federal Reserve Board

January 31, 2020

Abstract

We show that trade frictions in OTC markets result in inefficient private liquidity provision. We develop a dynamic model of market-based financial intermediation with a two-way interaction between primary credit markets and secondary OTC markets. Private allocations are generically inefficient due to a congestion externality operating through market liquidity in the OTC market. This inefficiency can lead to liquidity that is suboptimally low or high compared to the second best, providing a rationale for the regulation and public provision of liquidity. Moreover, our model characterizes a transmission channel of quantitative easing or tightening operating through liquidity premia.

Keywords: Liquidity provision, market liquidity, over-the-counter markets, OTC, quantitative easing, quantitative tightening, monetary policy normalization.

JEL classification: E44, G18, G30.

*We thank the editor, Florian Scheuer, and three anonymous referees for excellent comments and suggestions that greatly improved the paper. In addition, we thank Garth Baughman, Zachary Bethune, Regis Breton, Max Bruche, Francesca Carapella, Giovanni Favara, Zhiguo He, Nobu Kiyotaki, Michal Kowalik, Ricardo Lagos, Benjamin Lester, Konstantin Milbradt, Cecilia Parlatore, Lasse H. Pedersen, Skander van den Heuvel, Chris Waller, Randy Wright, and seminar participants at Econometric Society World Economic Congress, SED, Cowles General Equilibrium Conference, EEA, EFA, Federal Reserve Day Ahead Conference, LACEA, IMF, Society for Advancement of Economic Theory, DC Search and Matching Conference, Summer Workshop on Money, Banking, Payments, and Finance, Federal Reserve Board, Federal Reserve System Conference on Financial Structure and Regulation, Federal Reserve System Conference on Macroeconomics, St Louis Fed, Banque de France, Athens University of Economics and Business, University of Piraeus, Bank of Greece, and University of Chile for comments. All errors herein are ours. A previous version of this paper circulated under the title "Secondary Market Liquidity and the Optimal Capital Structure". The views expressed in this paper are those of the authors and do not necessarily represent those of Federal Reserve Board of Governors or anyone in the Federal Reserve System.
Emails: david.m.arseneau@frb.gov, david.e.rappoport@frb.gov, alexandros.vardoulakis@frb.gov.

1 Introduction

Public liquidity provision is warranted when the private sector is unable to produce enough liquid assets to diversify aggregate liquidity risk (Holmström and Tirole, 1998). Alternatively, it is warranted when liquidity shortages arise as a result of, for example, fire sales (Allen and Gale, 1994, 2004, Lorenzoni, 2008, Schleifer and Vishny, 2011, He and Kondor, 2016, and others). But, is there a role for public liquidity provision when liquid assets are abundant and the prospects of fire sales unlikely? This question is particularly important following the global financial crisis, where unconventional monetary policies such as quantitative easing (QE)—implemented *well after* the onset of the crisis—are thought to have implications for market liquidity (Krishnamurthy and Vissing-Jorgensen, 2011).

We develop a dynamic model of market-based financial intermediation to address this question. In our model, credit is provided in the primary market, where firms issue long-term bonds which are then retraded in an imperfectly liquid secondary market. Thus, secondary market liquidity affects investors' supply of credit to firms.[1] At the same time, the demand for credit, i.e., the issuance of illiquid bonds, affects secondary market liquidity through the composition of investors' portfolios as they must allocate limited financial resources between liquid and illiquid assets. This trade-off between credit provision in the primary market and liquidity provision in secondary OTC markets is a key contribution of our analysis.

The primary financial friction is the presence of search frictions in the secondary OTC market. The empirical evidence suggests that search frictions are the main driver of illiquidity in OTC markets for bonds (Edwards et al., 2007, and Bao et al., 2011) and, accordingly, have been the focus of a large theoretical literature (Duffie et al., 2005, Lagos and Rocheteau, 2009, He and Milbradt, 2014, Atkeson et al., 2015, and others). In addition to search frictions in the secondary market, agency frictions in the primary credit market are modeled using the costly state verification (CSV) framework (Townsend, 1979, Gale and Hellwig, 1985, Bernanke and Gertler, 1989), such that debt emerges as the optimal contract for firms' financing. The specific nature of the agency friction in the CSV framework is not crucial for our results. What is key is that there is a downward sloping demand for credit in the primary market and the CSV framework delivers this in a way that is tractable and well understood in the literature. The novelty of our framework is that the search and agency frictions interact to determine equilibrium credit supply and market liquidity.

The model has three periods and two types of agents: firms and investors. In the first period, firms need financing for productive projects that pay off in the last period. Financing is obtained

[1] We focus on credit provision through capital markets, which has become increasingly important for non-financial firms and households in the U.S. Data from the Financial Accounts of the United States reveal that the fraction of credit provided by the market has increased in the last 25 years and stands at over 60 percent for non-financial firms and at about 50 percent for households.

from investors who only value future consumption but are subject to preference shocks: a fraction becomes impatient and would like to consume in the interim period, while the rest remain patient and are willing to buy the assets of impatient investors. However, asset exchange between patient and impatient investors takes place in an OTC market characterized by search frictions, so that counterparties are only found with some probability. These probabilities are determined endogenously as a function of the ratio of liquid assets relative to illiquid assets available for trade. The trading frictions in the OTC market introduce a liquidity premium in firms' external financing and, thus, affect credit supply through a *liquidity premium channel* in the cost of credit.

At the same time, the quantity of bonds issued in primary markets reduces liquid resources available for trade in the secondary market. This is reflected in the portfolios of investors who allocate limited financial resources between illiquid bonds and liquid assets. Other things equal, market liquidity is lower as the composition of investors' portfolios shifts toward illiquid bonds. Thus, bond issuance affects secondary market liquidity through a *liquidity provision channel*.

We provide a full analytic characterization of the interaction between the liquidity premium and liquidity provision channels. Moreover, we establish in an existence proof that the two channels interact to determine the unique equilibrium in the primary and secondary markets. Intuitively, firms issue more debt exactly when market liquidity is high and, hence, the liquidity premium is low. But, as investors hold more of this debt, their portfolios shift away from liquid assets, thereby reducing liquidity provision in the OTC market. The liquidity premium channel dominates, while the liquidity provision channel acts as an automatic stabilizer such that an improvement or a deterioration in market liquidity cannot perpetually increase or decrease bond issuance. This mechanism, which is fundamentally different from models which feature an amplification between funding and market liquidity stemming from binding collateral constraints and limits to arbitrage (e.g., Brunnermeier and Pedersen, 2009, Gromb and Vayanos, 2002), is consistent with empirical evidence presented in Bao et al. (2011), who find that more issuance by firms is correlated with better secondary market liquidity.

Our normative analysis reveals that the private equilibrium is generically constrained inefficient owing to a congestion externality in the secondary market. The congestion externality operates through the trading probabilities as investors' fail to internalize how their portfolio choices affect the ease with which they can trade in the secondary market. For example, a patient investor that holds an extra unit of storage fails to internalize that she makes it easier for impatient investors to trade their bonds, and vice versa. Market liquidity is (generically) either suboptimally low or high. When liquidity is lower than the social optimum, firms are over-leveraged and write excessively risky bond contracts. This over-abundance of long-term bonds leads to an under-provision of liquidity in the secondary market. The opposite is true when liquidity is suboptimally high.

This type of congestion externality is well known in the search and matching literature. Indeed, Lester et al. (2015) highlight that efficiency in OTC markets can be restored when the trade surplus is split according to the elasticity of the matching function, satisfying the Hosios (1990) condition. Moreover, the authors show that this efficient split of the surplus arises endogenously in a competitive search equilibrium with posted prices. Accordingly, they argue in favor of a market microstructure that promotes price transparency.

In contrast, in our framework, even though price posting delivers a surplus split that satisfies the classic Hosios (1990) condition, this is not the efficient outcome. The reason is that there is a wedge between the private and social valuation of the trade surplus. The private surpluses are determined by the value at which investors are willing to trade a bond in the OTC market. Price posting maximizes the total *private* valuation of the trade surpluses. In contrast, the social planner in our model wants to maximize the total *social* surplus, measured in terms of discounted lifetime utility. In our model, when the interest rate is strictly positive, the private and social surpluses differ in so far as patient agents gain additional utility from a trade because they consume in the future. In light of this, the planner wants to reduce the share of the surplus allocated to patient investors so that impatient investors receive a share that is at least as large as the elasticity of the matching function and is increasing in the risk free rate. We refer to the resulting socially efficient surplus split as the *intertemporal Hosios condition*.

We examine the ability of a social planner to regulate the private provision of liquidity to implement the constrained efficient equilibrium. When private liquidity is inefficiently low, optimal regulation calls for a tax on leverage to restrict illiquid bond issuance by firms coupled with a subsidy on storage to provide an incentive for investors to hold a more liquid portfolio. Because of the congestion externality, a change in liquidity can generate ex ante welfare gains to investors. These welfare gains allow the planner to reduce firms' financing costs and increase their profits despite having to operate on a smaller scale. In contrast, when private liquidity is inefficiently high the opposite it true: a leverage subsidy combined with a storage tax are able to align the private and social incentives.

In addition to regulation, we also examine how the optimal provision of public liquidity, as implemented by quantitative policies like QE, can alleviate trading frictions and improve economic efficiency. We show that QE in our model works through a portfolio rebalancing mechanism: any public policy that alters the aggregate liquidity of private portfolios will affect the compensation investors require to purchase illiquid assets, i.e., the liquidity premium. In other words, QE transfers liquidity risk from private investors to the public sector which, in turn, influences savings and investment decisions in the real economy (see Stein, 2014, for a general discussion). While our analysis focuses on QE policies implemented by a central bank that uses liquid reserves to purchase less liquid assets from investors, quantitative tightening (QT) is expected to operate through the same mechanism.

A key insight of our model is that public liquidity management (QE or QT) is inherently different from liquidity regulation. Both policies affect the level of market liquidity, but whereas regulation trades off liquidity and credit provision, public liquidity management implies that public liquidity and credit provision move in tandem. This is because liquidity management enhances the intermediation technology of the economy. However, the feasibility of QE requires that the liquidity premium be high enough to offset the social costs created by this policy, which in our model derive from the additional monitoring cost incurred by the central bank relative to private agents. High liquidity premia are likely to arise during financial crisis rationalizing QE interventions in these states of the world rather that in normal times. The difference between liquidity regulation and management opens the door for these two policies to coexist. Indeed, our analysis shows that quantitative policies should be supplemented with optimal regulation to generate even larger welfare gains. In this sense, liquidity regulation and liquidity management should be viewed as complements, not substitutes, in the policy toolkit.

Related Literature. This paper is closely related to Holmström and Tirole (1998) in the sense that the role of regulation and provision of public liquidity is a central part of our analysis. Our paper is also related to other studies of the public role for liquidity provision (see for example, Allen and Gale, 1994, 2004, Lorenzoni, 2008, Farhi, Golosov, and Tsyvinski, 2009, Schleifer and Vishny 2011, Hart and Zingales, 2015, He and Kondor, 2016). These studies suggest that private liquidity provision is suboptimally low, with the exception of Hart and Zingales (2015) who finds that it is suboptimally high. We contribute to this literature by analyzing optimal liquidity provision with trade frictions in OTC market and show that under the *same* financial frictions private liquidity provision can be *either* suboptimally high or suboptimally low.

Our paper is also related to the literature studying frictional OTC trade in financial markets (Duffie et al. 2005, Lagos and Rocheteau 2009, Geromichalos and Herrenbrueck 2012, He and Milbradt 2014, and Atkeson et al. 2015, among others). This literature has primarily focused on how search frictions affect bid-ask spreads in OTC markets. Some studies have explored the role of endogenous market liquidity by considering the cost of secondary market participation (Shi, 2015, Bruche and Segura, 2017, and Cui and Radde, 2019) or endogenous asset issuance (Geromichalos and Herrenbrueck 2016, and Bethune et al. 2018). Other studies consider the effect of monetary policy on trade frictions by changing agents' money holdings used in decentralized exchange (Lagos and Wright, 2005, Lagos and Zhang, forthcoming, among others). Our focus is different. We consider the trade-off between credit and liquidity provision in primary credit markets as the main determinant of secondary market liquidity.

Focusing on this trade-off relates our work to the literature on the interaction between default and liquidity for corporate bonds traded in OTC markets with search frictions. He and Milbradt (2014) and Chen et al. (forthcoming) show that the interplay between default risk and market liquidity creates an amplifying feedback loop whereby shocks that increase default risk make secondary markets more illiquid, impeding the ability of firms to rollover existing debt and bringing them closer to default. In contrast, we show that there is an equilibrating relationship between debt issuance and secondary market liquidity, abstracting from debt rollover risk.

Finally, our paper contributes to the literature studying the channels through which quantitative policies transmit to the real economy. These channels include scarcity of safe assets, financial distress, or credit risk transfers between the private and public sector (Gertler and Karadi 2011, Williamson 2012, Farhi and Caballero 2013, Moreira and Savov 2017).[2] Another channel is the portfolio rebalancing mechanism, whereby changes in the net supply of assets affects the relative compensation that agents require to hold them (Tobin 1969). Papers in this branch of the literature have modeled the portfolio rebalancing mechanism assuming a reduced form portfolio adjustment cost (Andres, Lopez-Salido, and Nelson 2004, and Chen, Curdia, and Ferrero 2012). We contribute to this literature by presenting a model where portfolio adjustment costs are endogenous, and more importantly, they are affected by the QE policies. This is important as it characterizes the limits of these quantitative policies.

The rest of the paper proceeds as follows. Section 2 presents the model and establishes the existence and uniqueness of the equilibrium. Section 3 describes the effect of OTC trade on bond premia and primary credit markets. Section 4 presents the social planner's problem, describes the externalities operating through secondary market liquidity. Section 5 derives the set of policy instruments that can implement the constrained efficient outcome. Section 6 describes the transmission of quantitative policies and its optimal design. Finally, section 7 concludes. All proofs are relegated to the Appendix.

2 A Model of Market-Based Intermediation

2.1 Physical Environment

There are three time periods $t = 0, 1, 2$, a single consumption good, and two types of agents: entrepreneurs and investors. Entrepreneurs have long-term investment projects and may fund these projects with internal funds or with external funds received from investors. Ex ante identical investors provide funds to entrepreneurs, but once lending has taken place and while production is underway, investors are subject to a preference (liquidity) shock which reveals whether they are impatient, and hence prefer to consume earlier rather than later, or patient. Investors can trade their assets in an OTC market with search frictions to meet their liquidity needs (see Figure 1).

[2] Empirical evidence on the efficacy of the QE program implemented by the Federal Reserve includes Gagnon et al. (2011) and Krishnamurthy and Vissing-Jorgensen (2011) while Joyce et al. (2011) and Christensen and Rudebusch

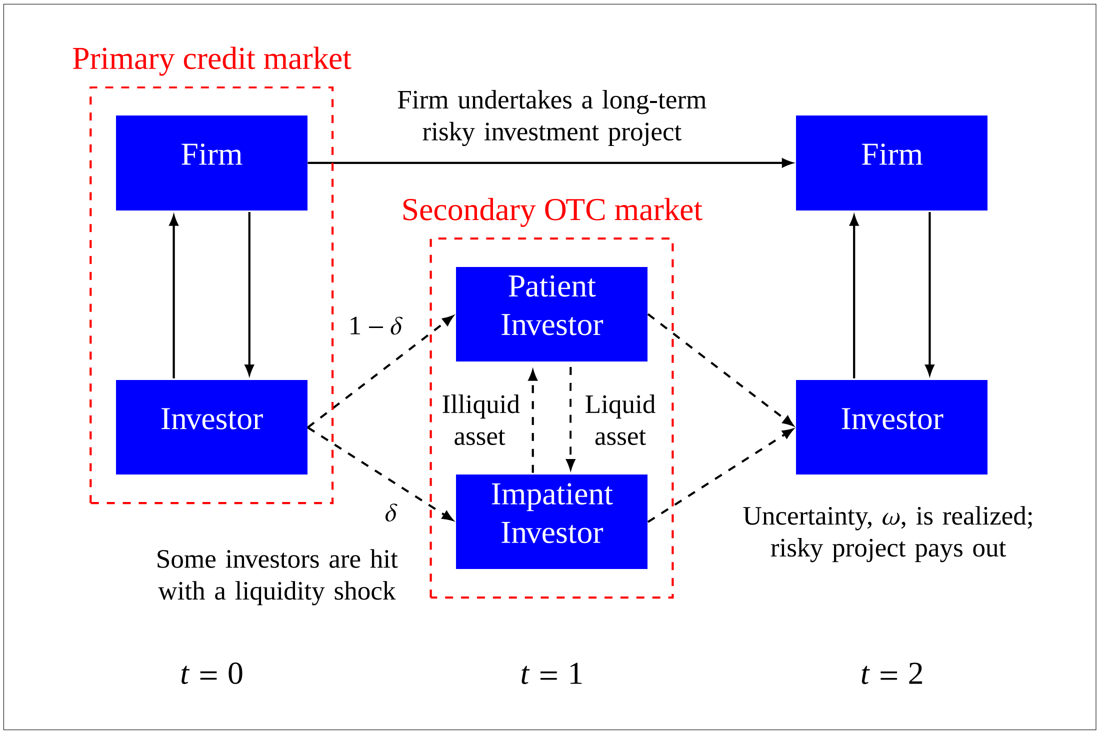

Figure 1: Timeline.

There is a mass one of ex ante identical entrepreneurs, who are endowed with n_0 units of consumption at $t = 0$. Entrepreneurs invest to maximize expected consumption at $t = 2$, which amounts to maximizing the return on their investment. They operate a linear technology, which delivers $R^k\omega$ at $t = 2$ per unit of consumption invested at $t = 0$. The random variable ω is an idiosyncratic productivity shock that hits after the project starts, and is distributed according to the cumulative distribution function F, with unit mean. The productivity shock is privately observed by the entrepreneur at $t = 2$, but investors can learn about it if they pay a monitoring cost μ as a fraction of assets. The (expected) gross return R^k is assumed to be known at $t = 0$, as there is no aggregate uncertainty. In order to produce, the firm must finance investment, denoted k_0, either through its own resources or by issuing two-period financial contracts to investors.[3] So,

(2012) present evidence on the program implemented by the Bank of England.

[3]We restrict firms to issue only long-term contracts because we are interested in analyzing the role of search frictions in the OTC market, where these long-term contracts are retraded. Alternatively, the firm could issue short-term contracts or borrow long-term from a bank who, in turn, funds itself short-term. In the absence of roll-over risk and bank runs, these two alternatives would yield the efficient sharing of liquidity risk described in Diamond and Dybvig (1983). But as Jacklin (1987) argued, this efficient outcome could also be achieved under frictionless secondary markets for long-term contracts. Thus, our paper could be seen as studying the limits to Jacklin's argument when there are search frictions in secondary markets. But we abstract from a holistic analysis of firms' capital structure under both illiquid secondary markets for long-term contracts and fragile short-term funding.

profits equal total revenue in period 2, $R^k\omega k_0$, minus payment obligations from financial contracts. Entrepreneurs represent the corporate sector in our model, so we refer to entrepreneurs and firms interchangeably.

There is a mass one of ex ante identical investors, who are endowed with e_0 units of consumption at $t = 0$. Investors have unknown preferences at $t = 0$. At $t = 1$ investors learn their preferences as a fraction δ turn out to be impatient and a fraction $1 - \delta$ patient. Following Diamond and Dybvig (1983), the preference shocks are private information and are not contractible ex ante. Patient consumers have preferences only for consumption in $t = 2$, $u^P(c_1, c_2) = c_2$, whereas impatient consumers have preferences for both consumption in $t = 1$ and 2, but discount period 2 consumption at rate $\beta \leq 1$, $u^I(c_1, c_2) = c_1 + \beta c_2$.

Investors in both period 0 and 1 have access to a storage technology with yield $r > 0$, i.e., every unit of consumption stored yields $1 + r$ units of consumption in the next period. The amount stored in period t is denoted s_t. In addition, at $t = 0$, investors can purchase financial contracts issued by entrepreneurs; and, at $t = 1$, they can exchange consumption for financial contracts in an OTC market subject to search frictions (see Figure 1). The terms of trade in the OTC market are influenced by the bargaining power of impatient investors, denoted by ψ.

Finally, note that the expected return on financial contracts will be known in period 0 and 1, since there is no aggregate uncertainty or new information arriving after investors and firms have agreed on the terms of these contracts. This means that asymmetric information considerations will not play a role in the OTC market.[4]

Throughout the remaining of the paper we make the following assumptions.

Assumption 1 (Relative Returns) *The discounted expected return on firms' projects is larger than the return on storage and the return on storage plus the return on secondary markets, i.e.,* $1 + r \leq \beta R^k$ *and* $(1 + r)[\psi/(1 + r) + (1 - \psi)\beta]^{-1} \leq R^k$. *In addition, monitoring costs are such that* $R^k(1 - \mu) < (1 + r)^2$.

Assumption 2 (Productivity Distribution) *Let* $h(\omega) = dF(\omega)/(1 - F(\omega))$ *denote the hazard rate of the productivity distribution. It is assumed that* $\omega h(\omega)$ *is strictly increasing.*

Assumption 3 (Impatience) *The fraction of impatient investors is positive,* $\delta > 0$*; impatient investors discount future consumption at a higher rate than the return on storage,* $1/\beta - 1 > r$ *or* $\beta < 1/(1 + r)$*; and impatient investors do not have total bargaining power, so that* $0 \leq \psi < 1$.

Assumption 4 (Investors Deep Pockets) *It is assumed that investors' (total) endowment* e_0 *is significantly higher than entrepreneurs' (total) endowment* n_0*, i.e.,* $e_0 >> n_0$.

Assumption 1 is necessary for the issuance of financial contracts in equilibrium. On the one

[4] See Gorton and Pennacchi (1990), Eisfeldt (2004), and Bigio (2015) for models of adverse selection and illiquid secondary markets.

hand, $1 + r \leq \beta R^k$, allows firms to offer a return that is higher than the cumulative two-period return on storage.[5] On the other hand, $(1 + r)[\psi/(1 + r) + (1 - \psi)\beta]^{-1} \leq R^k$, allows firms to offer a higher return than that which can be earned on storage, even when the prospective return on the OTC market is taken into account. Furthermore, this assumption rules out equilibria where entrepreneurs are always monitored, $(1 + r)^2 > R^k(1 - \mu)$. Assumption 2 rules out credit rationing in equilibrium. As a result, in equilibrium, there is a positive trade-off between the riskiness of the financial contract and the firm's leverage. Assumption 3 makes impatient investors have a (strict) preference for current versus future consumption when the interest rate is r, $\beta < 1/(1 + r)$, such that trade frictions in the OTC market are not irrelevant.[6] Assumption 4 ensures that investors can meet the credit demand of entrepreneurs. Together these assumptions ensure the existence and uniqueness of equilibrium, as we discuss below.

2.2 The Financial Contract and the Demand for Credit

The demand for credit is modeled using the off-the-shelf CSV framework of Townsend (1979), Gale and Hellwig (1985), and Bernanke and Gertler (1989). Hence, the exposition of this part of the model is kept brief.

Entrepreneurs finance their investment k_0 using either internal resources, n_0, or by selling long-term financial contracts to investors. These contracts specify an amount, b_0^d, borrowed from investors at $t = 0$ and a *promised* gross interest rate, Z, made upon completion of the project at $t = 2$. In our notation, b_0^d corresponds to the demand for credit in the primary market. If entrepreneurs cannot make the promised interest payments, investors can take all firm's proceeds, paying a monitoring cost equal to a fraction μ of the value of assets. Then, the $t = 0$ budget constraint for the entrepreneur is given by $k_0 \leq n_0 + b_0^d$. For what follows, it will be useful to define the entrepreneur's leverage, l_0, as the ratio of assets to (internal) equity k_0/n_0.

An entrepreneur is protected by limited liability, so her profits are always non-negative. Thus, the entrepreneur's expected profit in period $t = 2$ is given by $\mathbb{E}_0 \max\left\{0, R^k\omega k_0 - Zb_0^d\right\}$. Limited liability implies that the entrepreneur will default on the contract if the realization of ω is sufficiently low such that the payoff of the long-term project is smaller than the promised payout: $R^k\omega k_0 < Zb_0^d$. This condition defines a threshold productivity level, $\bar{\omega}$, such that the entrepreneur defaults when $\omega < \bar{\omega} = (Z/R^k)(l_0 - 1)/l_0$. The productivity threshold measures the credit risk of the financial contract, as it increases the firm's probability of default. Note that the productivity threshold is

[5]To see this, note that $1 + r \leq \beta R^k$, together with Assumption 3 imply that $(1 + r)^2 < R^k$. If these conditions are not met, then the firm return is dominated by the return of the storage technology from the perspective of investors that hold the financial contract to maturity.

[6]Note that $\beta < 1/(1 + r)$ is necessary but not sufficient for the relevance of OTC trade frictions in the model. In addition, it is required that $\psi < 1$, so the financial contract is not a perfect hedge against liquidity risk for impatient investors, when secondary market liquidity is sufficiently abundant.

increasing in the promised return Z and leverage l_0, and it is decreasing in the expected return R^k. Intuitively, as the entrepreneur makes a larger promise, or, alternatively, increases reliance on debt financing, it will need a higher realization of the idiosyncratic productivity shock to break even. Conversely, as the expected return increases the entrepreneur will be able to honor his promised debt payments even with smaller realizations of the productivity shock.

For notational convenience, we define $G(\bar{\omega}) \equiv \int_0^{\bar{\omega}} \omega dF(\omega)$ and $\Gamma(\bar{\omega}) \equiv \bar{\omega}(1 - F(\bar{\omega})) + G(\bar{\omega})$, which allows us to express the value of debt and equity as integrals over the regions where there is, and there is no, default. The function $G(\bar{\omega})$ equals the truncated expectation of entrepreneurs' productivity given default. The function $\Gamma(\bar{\omega})$ equals the expected value of a random variable equal to ω if there is default ($\omega < \bar{\omega}$) and equal to $\bar{\omega}$ when there is not ($\omega \geq \bar{\omega}$). It follows that $R^k k_0 \Gamma(\bar{\omega})$ corresponds to the expected transfers from entrepreneurs to investors. Then, the firms' objective, to maximize expected consumption, can be expressed as $1/n_0\, \mathbb{E}_0 \max\{0, R^k \omega k_0 - Z b_0^d\} = [1 - \Gamma(\bar{\omega})]\, R^k l_0$, i.e., the firm maximizes expected profits per unit of endowment.

Similarly, the expected gross *hold-to-maturity* return, R^b, can be expressed as a function of leverage and risk as

$$R^b(l_0, \bar{\omega}) = \frac{1}{b_0}\left[\int_{\bar{\omega}}^{\infty} Z b_0^d dF(\omega) + (1-\mu)\int_0^{\bar{\omega}} R^k \omega k_0 dF(\omega)\right] = \frac{l_0}{l_0 - 1} R^k \left[\Gamma(\bar{\omega}) - \mu G(\bar{\omega})\right] \ . \tag{1}$$

Clearly $R^b(l_0, \bar{\omega})$ is decreasing in l_0 as leverage dilutes lenders claim on the firm's assets. Moreover, in equilibrium it will be increasing in risk, $\bar{\omega}$, as detailed below.

We continue the presentation of the model, deriving the demand (and supply) of credit in the primary market. Our motivation to proceed in this manner is that our existence proof is going to use a fixed point argument for the mapping that concatenates the demand and inverse supply of credit. We will show later that the maximum attainable utility is proportional to the hold-to-maturity return, with a constant of proportionality that is endogenous and depends on the level of secondary market liquidity. This allows us to simplify the investors' indifference condition into a required hold-to-maturity return (equation (6)). Using this notation we can write down the firm's problem of choosing the optimal contract, which determines the demand for credit, as[7]

$$\max_{l_0, \bar{\omega}} \, [1 - \Gamma(\bar{\omega})]\, R^k l_0 \qquad \text{s.t.} \qquad R^b(l_0, \bar{\omega}) = R^b \ , \tag{2}$$

where, in a slight abuse of notation, R^b denotes the hold-to-maturity return. This return in equilibrium will also reflect the compensation that investors require to hold a bond that is not perfectly

[7]Our CSV problem may appear non-standard, but it is not. To see this, consider that substituting the break-even condition (6) into the firms' problem (2) recovers the standard representation. We have chosen to present the problem this way for two reasons: (1) it facilitates proving existence and uniqueness (Theorem 1); and (2) allows us to represent the equilibrium graphically (Figure 2).

liquid in the presence of trade frictions. By contrast, $R^b(l_0, \bar{\omega})$ denotes the hold-to-maturity return offered by the firm, for contract terms l_0 and $\bar{\omega}$, as described by equation (1). To avoid confusion we will make explicit the dependence on the contract characteristics for the latter throughout the paper.

Note that the firm's problem also defines the demand for credit in the primary market $b_0^d(R^b) = (l_0(R^b) - 1)n_0$, which is, as we show below, a strictly decreasing function of the expected hold-to-maturity return R^b. As is well established in the CSV literature, the optimal financial contract will take the form of a debt contract. Therefore, we refer to these contracts as bonds.

2.3 The OTC Market and the Supply of Credit

At $t = 0$, ex ante identical investors are endowed with e_0 units of the consumption good and allocate their wealth across two assets: a storage technology, s_0, and financial contracts, b_0^s, which corresponds to the supply of credit in the primary market. Thus, their budget constraint is given by $s_0 + b_0^s = e_0$, where $s_0, b_0^s \geq 0$, i.e., borrowing at the storage rate or short-selling bonds are not allowed.

The storage technology pays a fixed rate of return $1 + r$ at $t = 1$ in units of consumption. The proceeds of this investment, if not consumed, can be reinvested to earn an additional return of $1 + r$ between period 1 and 2. In this sense, storage is a liquid asset, as it can be costlessly transformed into consumption at any point in time. In contrast, bonds pay a hold-to-maturity return R^b and are illiquid due to trade frictions in the OTC market.

2.3.1 The OTC Market

The ex post heterogeneity introduced by the preference shock generates potential gains from trade in the secondary market. Impatient investors want to exchange illiquid long-term bonds for consumption (Assumption 3). Patient investors are willing to take the other side of the trade if the return from buying bonds in the OTC market is greater than the return on the storage technology.

To model the exchange in the OTC market we consider that each investor represents a large family of small *traders*, in the spirit of Shi (1997), Atkeson et al. (2015), or Bianchi and Bigio (2014). In particular, each investor is comprised by a continuum of infinitesimal traders, where each trader is restricted to trade only one bond. That is, traders' portfolios are restricted to one bond for sellers and q_1 units of consumption for buyers, where q_1 denotes the price of the bond determined by bargaining between patient and impatient traders. Traders are paired up according to a matching technology. Impatient investors send a mass of b_0^s traders to sell their bonds in the OTC market. Patient investors send a mass of $(1 + r)s_0/q_1$ traders to buy bonds. This is akin to a situation where impatient investors submit b_0^s sell orders and patient investors submit $(1 + r)s_0/q_1$

buy orders, so for ease of exposition we will refer to this trading process as submitting orders.

Casting the model in terms of a family of small traders allows us to model the effect of bond issuance on trading probabilities, while keeping the model tractable. In fact, if we were to consider that the matching takes place at the investor level, then trading probabilities will be determined by the likelihood of preference shocks and will be independent of bond issuance in the primary market. However, this assumption would eliminate the liquidity provision channel that we explore in this paper. A shortcoming of our modeling approach using a family of small traders is that it does not provide a complete microfoundation for the pricing mechanism, and for the associated ex ante problem of choosing how much assets and goods to bring to the OTC market. Instead, we assume that buyers and sellers send quantities of assets and goods to the market that are consistent with their expectations about the terms of trade.[8]

Suppose, in aggregate, there are A sell (or ask) orders and B buy orders. The matching function is assumed to be constant returns to scale, as long as the number of matches does not exceed the number of orders on each side of the market. It is given by $m(A, B) = \min\left\{A, B, \nu A^{\alpha} B^{1-\alpha}\right\}$ with $0 < \nu$, a scaling constant, and $0 < \alpha < 1$, the elasticity of matches with respect to sell orders.

We define a concept of *market liquidity* through the ratio of buy orders to sell orders, or $\theta = B/A$. This notion of liquidity—defined by the ease with which trade takes place in the OTC market—has different implications for traders on opposing sides of the market. For example, when θ is large, a bond in the secondary market is relatively liquid; that is, buy orders are abundant relative to sell orders so that it is relatively easy for sellers to trade. But, at the same time, it is relatively hard to execute a buy order. Note that our notion of liquidity is related to, but distinct from, the easiness to trade for *all* market participants, which is captured in our framework by the efficiency of the matching technology ν. Increasing (decreasing) ν makes it easier (harder) for participants on both sides of the market to trade in a symmetric fashion.

Using the matching function, the probability that a sell order is executed is expressed as $f(\theta) = m(1, \theta)$, and the probability that a buy order is executed is expressed as $p(\theta) = m(\theta^{-1}, 1)$.

The fact that matches are bounded by the minimum number of orders defines two liquidity thresholds, $\underline{\theta} = \min\{\nu^{1/\alpha}, 1\}$ and $\overline{\theta} = \max\{\nu^{-1/(1-\alpha)}, 1\}$. When market liquidity is such that $\theta \leq \underline{\theta}$ all buy orders are executed; buyers trade with probability $p(\theta) = 1$, whereas sellers trade with probability $f(\theta) = \theta$. Alternatively, when $\theta \geq \overline{\theta}$ all sell orders are executed; thus the probabilities are $f(\theta) = 1$ and $p(\theta) = \theta^{-1}$. When liquidity is in the interval $[\underline{\theta}, \overline{\theta}]$ matches are given by the constant returns to scale matching function and the probabilities are $f(\theta) = \nu\theta^{1-\alpha}$ and $p(\theta) = \nu\theta^{-\alpha}$.

From the seller's perspective, a trading match yields additional consumption from the sale of the bond at price q_1. If the seller walks away from the match she will hold the bond to maturity and

[8] See Lagos and Wright (2005) and, more recently, Lebeau (2019) for the problem of how many goods to bring to trade in decentralized markets.

receive an expected payoff βR^b in $t = 2$. Thus, the impatient investor is willing to sell the bond to the patient investor in exchange for storage for any price above βR^b. The surplus that accrues to an impatient investor is given by $S^I(q_1) = q_1 - \beta R^b$. Similarly, the value of a trading match to a buyer is the present value of the (discounted) payout of the bond, net of the price that needs to be paid for each bond in the secondary market, $S^P(q_1) = R^b/(1 + r) - q_1$. The patient investor is willing to pay up to $R^b/(1 + r)$ for providing liquidity to the impatient investor.

We follow Duffie et al. (2005) and the subsequent literature on OTC markets in assuming the price of the debt contract on the secondary market is determined by Nash bargaining. The Nash price maximizes the product of the respective surpluses, $\max_{q_1} \left(S^I(q_1)\right)^{\psi} \left(S^P(q_1)\right)^{1-\psi}$, where $\psi \in [0, 1)$ is the bargaining power of impatient investors.

The solution yields the following bond price in the secondary market

$$q_1(R^b) = \left(\frac{\psi}{1 + r} + (1 - \psi)\beta\right) R^b \ . \tag{3}$$

From equation (3) it follows that the return patient investors make in the secondary market, per executed buy order, is given by $\Delta = R^b/q_1(R^b) = (\psi/(1 + r) + (1 - \psi)\beta)^{-1} > 1 + r$.

The price q_1 depends on the equilibrium hold-to-maturity return, R^b, which in equilibrium is a function of contract terms, $(l_0, \bar{\omega})$. Yet, the expected return in the secondary market, Δ, depends in equilibrium only on exogenous parameters that govern bargaining, ψ, and those that pin down traders' outside options, r and β. This is a consequence of the outside options being linear in the expected bond return. It adds tractability, allowing us to derive closed form analytical results.[9]

2.3.2 Investors' Portfolio Choice and the Supply of Credit

Utility maximization on the part of investors can be summarized by an optimal portfolio allocation in the initial period. To describe this portfolio choice problem, it is useful to first consider the optimal behavior of impatient and patient investors when they arrive in period $t = 1$ with a generic portfolio of storage and bonds (s_0, b_0^s).

Impatient Investors. By Assumption 3 impatient investors want to consume immediately in period $t = 1$. They can consume the proceeds from the resources they put into storage, $s_0(1 + r)$, plus the additional proceeds from placing b_0^s sell orders in the OTC market. These orders are executed with probability $f(\theta)$ and each executed order yields q_1 units of consumption. Thus, the consumption of impatient investors in period 1 is given by $c_1^I = s_0(1 + r) + f(\theta) q_1 b_0^s$.

[9] See Bianchi and Bigio (2014) for a model where multiple rounds of trade reintroduces the dependence on market conditions of the price in the secondary market. See Mattesini and Nosal (2016) for a model where renegotiation between investors and brokers introduces a dependence of market conditions on the price in the secondary market.

On the other hand, with probability $1 - f(\theta)$ orders are not matched and impatient investors have to carry bonds into period 2. Therefore, consumption in the final period is given by $c_2^I = (1 - f(\theta))R^b b_0^s$, with the utility derived from c_2^I discounted by β.

Patient Investors. Patient investors only value consumption in the final period and are willing to place buy orders in the OTC market if there is a surplus to be made. The price determination in the OTC market guarantees this is always the case ($1 + r < \Delta$), thus patient investor would ideally like to exchange all of their consumption for bonds.

But, the buy orders of patient investors will be executed only with probability $p(\theta)$. That is, they place $s_0(1 + r)/q_1$ buy orders, of which a fraction $p(\theta)$ are executed. So, patient investors expect to increase their bond holding by $p(\theta)s_0(1 + r)/q_1$ units. It follows that expected storage held at the end of $t = 1$, s_1^P, is equal to a fraction $1 - p(\theta)$ of the available liquid funds $s_0(1 + r)$, i.e., $s_1^P = (1 - p(\theta))s_0(1 + r)$. Then, consumption in the final period is given by $c_2^P = (1 - p(\theta))s_0(1 + r)^2 + [b_0^s + p(\theta)s_0(1 + r)/q_1]R^b$. That is, the payout from consumption that was stored and not traded away in the OTC market plus the expected payout from bond holdings.

Investors' Portfolio Choice. In the initial period investors solve a portfolio allocation problem, taking the liquidity in the OTC market θ as given. They choose between storage and bonds to maximize their expected lifetime utility $\mathbb{U} = \delta(c_1^I + \beta c_2^I) + (1 - \delta)c_2^P$, subject to the period 0 budget constraint and the expected consumption of impatient and patient investors presented above.

Using the expressions for optimal expected consumption, we can rewrite the expected lifetime utility as $\mathbb{U} = U_s s_0 + U_b b_0^s$, where U_s and U_b denote the expected utility from investing in storage and bonds in period 0, respectively, and are given by

$$U_s(\theta) = \delta(1 + r) + (1 - \delta)\left[(1 - p(\theta))(1 + r)^2 + p(\theta)(1 + r)\Delta\right] , \tag{4}$$

$$\text{and} \qquad U_b(R^b, \theta) = u_b(\theta)R^b , \tag{5}$$

where $u_b(\theta) \equiv \delta\left[f(\theta)\Delta^{-1} + (1 - f(\theta))\beta\right] + (1 - \delta)$ corresponds to the expected loss a bond investor expects to incur relative to the hold-to-maturity bond return, with $u_b(\theta) \geq \beta$. That is, the expected utility of holding bonds in period 0 can be decomposed as the product of the expected hold-to-maturity return on the bond, R^b, and the expected loss owing to holding an illiquid bond, $u_b(\theta)$. By contrast, the utility of holding storage in the initial period only depends on market liquidity, through the probability that a buy order is executed, $p(\theta)$.

Using these definitions, we can express the asset demand correspondence that maximizes the investors portfolio problem as $b_0^s = e_0$ if $U_b > U_s$, $b_0^s \in [0, e_0]$ if $U_b = U_s$, and $b_0^s = 0$ if $U_b < U_s$, where $s_0 = e_0 - b_0^s$ in all cases. That is, if the expected utility of holding bonds in period 0 is greater than the utility of holding storage in period 0, then investors will demand only bonds in the initial period. On the contrary, if the expected utility of holding bonds is smaller than that of holding storage, then investors will chose an initial portfolio consisting only of the storage asset. Finally, if the expected benefits are equal, investors will be indifferent between investing in storage and bonds initially, and their demands will be an element of the set of feasible portfolio allocations: $s_0, b_0^s \in [0, e_0]$, such that the total value of assets equal the initial endowment.

Given our assumptions, the expected utility from investing in storage, U_s, is higher than the expected utility in financial autarky, given by $\delta(1+r)+(1-\delta)(1+r)^2$, owing to the return of buying a bond in the secondary market, $\Delta > 1 + r$. Hence, the equilibrium portfolio allocation will be interior, i.e., $U_s = U_b$ with $s_0, b_0^s > 0$. For future reference we label this condition the *investors' break-even condition:*

$$U_s(\theta) = U_b(R^b, \theta) = u_b(\theta)R^b \ . \tag{6}$$

The upshot of writing the investors' break-even condition in this way is that, from the perspective of a firm that takes the liquidity in the OTC markets as given, the break-even condition amounts to ensuring that investors receive a hold-to-maturity return which satisfies $R^b(l_0, \bar{\omega}) = U_s(\theta)/u_b(\theta)$. Moreover, this condition describes the *aggregate* credit supply, $b_0^s(R^b)$. In fact, it is implicitly defined by $U_s(\theta(b_0^s, R^b)) = u_b(\theta(b_0^s, R^b))R^b$, where $\theta(b_0^s, R^b) = (1-\delta)(e_0 - b_0^s)(1+r)\Delta/(\delta b_0^s R^b)$. To be clear, our concept of aggregate credit supply is not just the sum of the individual investors' supply of credit, but is one where the consistency of market liquidity is taken into account, i.e., θ is a function of (b_0^s, R^b). As shown below, aggregate credit supply is strictly increasing in the expected hold-to-maturity bond return.

2.4 Equilibrium

The equilibrium of the model is defined as follows.

Definition 1 (Private Equilibrium) *We say that $(b_0^d, \bar{\omega}, b_0^s, s_0, R^b, \theta, q_1)$ is a private equilibrium iff:*

1. *Given the outcome in the secondary market (θ, q_1), the financial contract is described by $(l_0 = b_0^d/n_0 + 1, \bar{\omega})$ that maximizes entrepreneurs' return on investment subject to investors' break-even condition (6), with R^b satisfying equation (1).*

2. *Given the outcome in the primary and secondary market (R^b, θ, q_1), investors' portfolios (b_0^s, s_0) are optimal.*

3. *The credit market clears, i.e., $b_0^d = b_0^s \equiv b_0$.*

4. *Market liquidity corresponds to $\theta = (1-\delta)(1+r)s_0/(q_1\delta b_0^s)$.*

5. *q_1 is determined via the surplus sharing rule (3).*

6. *All agents have rational expectations about q_1 and θ.*

Our strategy to find the equilibrium is as follows. First, note that in the unique interior equilibrium the investors' break-even condition (6), together with equation (1), implicitly defines a function $\bar{\omega}^{pe} \equiv \bar{\omega}^{pe}(l_0)$, with

$$\frac{d\bar{\omega}^{pe}}{dl_0} = -\frac{\partial R^b(l_0, \bar{\omega})/\partial l_0}{\partial R^b(l_0, \bar{\omega})/\partial \bar{\omega}}. \tag{7}$$

This implicit function is useful for expositional purposes because it helps us to more easily characterize the inefficiencies in the model. For this reason, we introduce the superscript *pe* to denote this implicit function in the *private equilibrium*.

Substituting $\bar{\omega}^{pe}$ into the Lagrangian of the firm's problem, we get, $\mathcal{L}^{pe}(l_0) = [1 - \Gamma(\bar{\omega}^{pe})] R^k l_0$. Then, at the optimum

$$[1 - \Gamma(\bar{\omega}^{pe})] = \Gamma'(\bar{\omega}^{pe})\, l_0 \frac{d\bar{\omega}^{pe}}{dl_0}. \tag{8}$$

That is, the entrepreneur's privately optimal choice of leverage trades off the marginal increase in profits that comes from increasing leverage against the cost of having to compensate investors for the additional risk that comes with providing that leverage.

Next, equation (8), the investor break even condition (which defined the implicit function $\bar{\omega}^{pe}$), as well as the other equilibrium conditions, jointly characterize the optimal financial contract $(l_0, \bar{\omega})$. In particular, using these equilibrium conditions we can write market liquidity as

$$\theta = \frac{(1-\delta)s_0(1+r)}{\delta b_0 q_1} = \frac{(1-\delta)(1+r)\Delta\,(e_0 - n_0(l_0 - 1))}{\delta n_0(l_0 - 1)R^b(l_0, \bar{\omega})}, \tag{9}$$

where we have substituted the equilibrium price, q_1, and the credit market clearing condition.

The following theorem establishes the existence and uniqueness of equilibrium in our model.

Theorem 1 (Existence and Uniqueness of Private Equilibrium) *Under the maintained assumptions there exists a unique private equilibrium of the model. Furthermore, in the unique equilibrium credit is not rationed.*[10]

The proof uses a fixed-point argument on a mapping that concatenates the demand for credit with the inverse supply of credit, imposing the other equilibrium conditions. Thus, a fixed point of this mapping is an equilibrium of the model. The proof proceeds in three steps. The first step shows that the optimal financial contract defines a credit demand function, i.e., each offered return yields a unique demand for credit or level of bond issuance by firms. This step derives results that are similar to results found in the CSV literature. The second step shows that the aforementioned mapping is continuous and maps the interval of expected returns $[(1+r)^2, R^k]$ on itself, thus having a fixed point and establishing the existence of equilibrium. These derivations generalize previous

[10] In a credit rationed equilibrium the firm leverage is pinned down by the maximum amount of credit investors are willing to provide, i.e., the relationship between leverage, risk, and hold-to-maturity return breaks down. With a fixed bond issuance there will be no scope for the portfolio composition of investors to affect secondary market liquidity.

results to the case when financial contracts are retraded in OTC markets, and they show that the aggregate credit demand is strictly decreasing in bonds' expected returns. Finally, the third step establishes that multiple equilibria do not arise due to the retrading in the OTC market. We establish uniqueness by showing that when the matching function exhibits constant returns to scale, the aforementioned mapping is strictly decreasing. That is, when the expected return offered by firms declines, the borrowing by firms increases, which lowers market liquidity and increases the expected hold-to-maturity return required by investors. The last result suggests that aggregate credit supply is upward sloping. In fact, the proof of Theorem 1 allows us to establish the following results.

Corollary 1 (The Optimal Financial Contract and The Demand for Credit) *The optimal leverage l_0 and risk $\bar{\omega}$ of the financial contract, are strictly decreasing in the expected hold-to-maturity return R^b. That is, the demand for credit $b_0^d(R^b)$ is strictly decreasing.*

In addition, we can establish that the aggregate supply of credit is strictly increasing in the expected hold-to-maturity return, in the relevant part of the parameter space.

Proposition 1 (Aggregate Credit Supply Elasticity) *The aggregate credit supply, $b^s(R^b)$, has a finite and strictly positive interest rate elasticity.*

Proposition 1 characterizes the role of secondary market liquidity on the aggregate supply of credit in the primary market for bonds. As the expected hold-to-maturity bond return increases, for investors to be indifferent between illiquid bonds and liquid storage, market liquidity needs to drop so the return on storage increases and the expected loss from holding illiquid bonds increases. Market liquidity drops only if investors' portfolios become more illiquid, which is the case when investors' bond holdings increase. That is to say that the supply of credit increases. Note that in our model, where individual investors—who are risk neutral—supply credit totally elastically, the interaction of trade frictions and investors' limited liquid resources generates an increasing aggregate supply of credit.

The previous results are useful to analyze the model through the aggregate demand and supply for credit, i.e., the demand of credit by firms from investors and investors' supply of credit to firms in the primary bond market, depicted in Figure 2. This representation can be used to contrast our model with previous work. In the CSV literature it is typically the case that aggregate credit supply is perfectly elastic at the expected rate of return of risky debt, e.g., Bernanke et al. (1999). In the context of our model, when search frictions in the OTC market are irrelevant, the expected return on debt will equal $(1 + r)^2$, depicted by the dashed line in Figure 2. Recall that the expected return in the classic CSV literature, as in our paper, is different from the promised return, Z. In practice, bond promised returns, or yields, are comprised of default and liquidity premia. Below, we provide

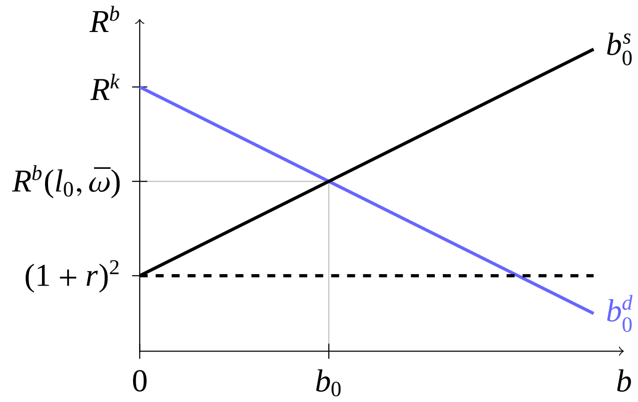

Figure 2: Aggregate Demand and Supply of Credit in the Primary Market.

an analytical characterization of these premia in our model. In terms of this characterization, the classic CSV literature corresponds to the case where there is no liquidity premium. In other models of OTC trade with search frictions, such as Duffie et al. (2005), where the trading probabilities and market liquidity are functions of exogenous parameters, the aggregate credit supply will be totally elastic at some rate $R^b > (1+r)^2$ and $R^b < R^k$.

3 Characterization of Liquidity and Default Premia

It is useful to define a benchmark interest rate that is the return on a two-period bond that could be traded in a perfectly liquid secondary market. In the absence of arbitrage, such a contract needs to deliver the same return in expectation as a strategy of investing only in storage both in the initial and interim periods, i.e., $\delta R^\ell/(1+r) + (1-\delta)R^\ell = \delta(1+r) + (1-\delta)(1+r)^2$. This gives rise to the following definition.

Definition 2 (Liquid Two-period Rate) *The liquid two-period rate is defined as the gross interest rate on a perfectly liquid two-period bond* $R^\ell \equiv (1+r)^2$.

The benchmark rate allows us to decompose the total gross return on the financial contract written by the firm into a default and a liquidity premium. In order to do this, express the total bond premium as the gross return of the firm's contract relative to the liquid two-period rate, $\Phi^t \equiv Z/R^\ell$. Then, this total premium is decomposed into a component owing to default risk, Z/R^b, and a component owing to liquidity risk, R^b/R^ℓ. With this decomposition, we have the following definitions for the default and liquidity premia, respectively.

Definition 3 (Default and Liquidity Premia) *The bond default premium* Φ^d *and the bond liquidity premium* Φ^ℓ *are given by* $\Phi^d \equiv Z/R^b$ *and* $\Phi^\ell \equiv R^b/R^\ell$*.*

Consequently, the total bond premium is $\Phi^t = \Phi^d\, \Phi^\ell$. These definitions provide sharp characterizations of both the default and liquidity premia, which are convenient to help trace out the underlying economic mechanisms in our model.

From the definition of the default premium we have that

$$\Phi^d(\bar{\omega}) = \frac{\bar{\omega}}{\Gamma(\bar{\omega}) - \mu G(\bar{\omega})}\ . \tag{10}$$

It follows that in our model, as in the classic CSV model, the default premium is an increasing function of credit risk $\bar{\omega}$, as formalized in the next proposition.

Proposition 2 (Credit Risk and the Default Premium) *Under the maintained assumptions, the default premium* $\Phi^d(\bar{\omega})$ *is a strictly increasing function of credit risk* $\bar{\omega}$*.*

Intuitively, investors demand a higher default premium for financial contracts that are more likely to default (i.e., contracts that are more risky, or specify a higher productivity threshold $\bar{\omega}$ for paying out the full promised value). The more subtle part of the argument is that leverage does not directly affect the default premium, as is the case in the classic CSV model, though leverage and risk are jointly determined in equilibrium. This is due to the fact that, for a fixed threshold productivity, $\bar{\omega}$, leverage affects both the face value of the contract, Z, and the hold-to-maturity return for investors, R^b, in the same way (equation (10)).

Moreover, from the definition of the liquidity premium and the investors' break-even condition (6), we have that

$$\Phi^\ell(\theta) = \frac{U_s(\theta)}{(1+r)^2 u_b(\theta)}\ . \tag{11}$$

The liquidity premium is the spread between the bond hold-to-maturity return, which is given by the ratio between the expected utility from liquidity provision $U_s(\theta)$ and the product of the expected utility loss due to bond illiquidity $u_b(\theta)$, and the liquid two-period return $(1+r)^2$. Equation (11) provides an analytical characterization of the relationship between the liquidity premium and secondary market liquidity θ. This relationship is characterized in the following Lemma.

Lemma 1 (Secondary Market Liquidity and the Liquidity Premium) *The liquidity premium,* Φ^ℓ*, or equivalently, the hold-to-maturity return,* R^b*, is a decreasing function of secondary market liquidity,* θ*. Moreover, the elasticity of the liquidity premium,* Φ^ℓ*, with respect to secondary market liquidity,* θ*, is lower than* 1 *in absolute value.*

Lemma 1 formalizes the intuition that the price of liquidity risk is inversely proportional to the amount of liquidity in secondary OTC markets. It also plays a key role in our analysis because it establishes an important link between liquidity in the secondary market and debt issuance in the primary market.

To understand this link, consider that when secondary market liquidity is low, sell orders are more difficult to execute ($f(\theta)$ is low), implying that impatient investors expect a higher utility loss from holding illiquid bonds ($u_b(\theta)$ is low). By the same token, buy orders are easier to execute ($p(\theta)$ is high), implying that patient investors expect a higher utility benefit from holding storage to provide liquidity in the secondary market ($U_s(\theta)$ is high). Both imply that investors require a higher premium to hold illiquid bonds over liquid storage, $\Phi^{\ell}(\theta)$.

This relationship is at the heart of the *liquidity premium channel* because it describes how secondary market liquidity, θ, affects liquidity premia, which, in turn, feeds back into the equilibrium in the primary credit market (b_0, R^b). In particular, our model shows in a very tractable way how the liquidity premium shapes credit terms in the primary market as summarized by R^b, and it affects the demand for credit b_0^d.

At the same time, the equilibrium in the credit market (b_0, R^b) also influences secondary market liquidity, θ, through the composition of investors portfolios. If the firm's demand for credit declines, in equilibrium investors will shift their portfolio composition away from illiquid bonds and into liquid storage. In this sense, supplying fewer bonds in equilibrium crowds in secondary market liquidity. We call this the *liquidity provision channel* because it describes how the equilibrium in the primary credit market (b_0, R^b) feeds into secondary market liquidity θ. Indeed, equation (9) characterizes how θ is a function of the volume and expected return of bonds in the credit market (b_0, R^b). The liquidity provision channel is novel to the literature analyzing financial markets with trade frictions.

The liquidity premium channel and the liquidity provision channel work to offset one another to determine the joint equilibrium of the primary and secondary credit markets. In the uniqueness proof we establish that the liquidity premium channel dominates, while the liquidity provision channel acts as an automatic stabilizer such that an improvement or a deterioration in market liquidity cannot perpetually increase or decrease bond issuance.

The analytical tractability of our model has additional advantages. One is that it allows us to describe the outcomes when trade frictions become irrelevant. When liquidity considerations are absent (i.e., the liquidity premium is such that $\Phi^{\ell} = 1$) the hold-to-maturity return and liquid two-period rate are equal. In this special case, which we have ruled out with our assumptions, our model collapses to the classic CSV model in which the supply of credit is perfectly elastic at the expected return of debt, as represented by the dashed line in Figure 2. In addition, in this case the relationship between the primary and secondary markets breaks down. Another advantage of the analytic tractability is that it facilitates the analysis of comparative statics. We take advantage of this later in the paper in section 6 to describe the effects of quantitative easing policies.

4 Efficient Liquidity Provision in OTC Markets

We consider a planner that maximizes the utility of entrepreneurs and investors and is constrained by the search frictions and the structure of trade in the OTC market. For expositional convenience, and without loss of generality, we consider the planner's Pareto problem, as opposed to a linear social welfare function with weights for each type of agent.[11] This formulation of the planner's problem delivers constrained efficient allocations and more easily compares to the private equilibrium problem studied in section 3. In particular, the planner maximizes the utility of the firm under the consideration that investors are not worse-off compared to the private equilibrium. In other words, the social planner chooses the optimal contract to maximize expected consumption (profits) of the firm while internalizing how the contract choice impacts secondary market liquidity.

To formalize the planner's problem let $(l_0, \bar{\omega}, \theta, q_1)$ denote the socially efficient outcomes and let $(l_0^*, \bar{\omega}^*, \theta^*, q_1^*)$ denote the private equilibrium described in section 3. After substituting in equation (3) for the price in the OTC market where appropriate, the planner's problem can be written as $\max_{\bar{\omega}, l_0, \theta} [1 - \Gamma(\bar{\omega})] R^k l_0$, subject to equation (9), and

$$\mathbb{U}(l_0, \bar{\omega}, \theta) \geq \mathbb{U}(l_0^*, \bar{\omega}^*, \theta^*) \; . \tag{12}$$

Condition (12) says that the planner cannot choose equilibrium allocations that result in lower expected utility for investors compared to the private equilibrium, i.e., it forces the planner to optimize along the investors' indifference curves. Moreover, we have imposed the pricing equation (3) as a restriction in the planner's problem to focus on interventions in the initial period similar to the normative analysis of financial markets as in Allen and Gale (2004) or Lorenzoni (2008).

The social planning problem differs from the private equilibrium in two respects. First, the planner need not respect the investors' break-even condition (6), but cannot make investors worse off, i.e., needs to satisfy (12). Second, whereas both investors and firms take liquidity as given in solving for the private equilibrium, the planner internalizes how allocations in the initial period affect liquidity in the secondary market by explicitly considering (9) as a constraint. We will restrict attention to the more interesting case where $\theta \in (\underline{\theta}, \bar{\theta})$, so trading probabilities depend on the matching function.

Our strategy for finding the constrained efficient equilibrium follows closely the approach outlined in section 2.4 for solving the private equilibrium. The Pareto improvement condition (12)

[11] Alternatively, one could consider a utilitarian social planner who maximizes a linear social welfare function under some (exogenously chosen) weights assigned to firm's and investors' welfare. The solution to these utilitarian problems can be easily characterized conditional on the social welfare weights. Moreover, the welfare weights could be bounded to ensure Pareto optimality. Our formulation yields the solution of the utilitarian problem when the relative weight between investors and the firm is equal to the Lagrange multiplier on the constraint that investors' welfare is at least as high as in the private equilibrium (constraint (12)).

implicitly defines a function $\bar{\omega}^{sp} \equiv \bar{\omega}^{sp}(l_0)$, with

$$\frac{d\bar{\omega}^{sp}}{dl_0} = -\frac{b_0 u_b \left(\partial R^b / \partial l_0\right) + (\partial \mathbb{U}/\partial \theta)\,(\partial \theta / \partial l_0) + n_0 \left(u_b R^b - U_s\right)}{b_0 u_b\,(\partial R^b / \partial \bar{\omega}) + (\partial \mathbb{U}/\partial \theta)\,(\partial \theta / \partial \bar{\omega})} \,. \tag{13}$$

We introduce superscript *sp* to denote this implicit function in the *social planner's* problem.

Substituting $\bar{\omega}^{sp}$ along with (9) into the Lagrangian of the planner's problem, we get, $\mathcal{L}^{sp}(l_0) = [1 - \Gamma(\bar{\omega}^{sp})]\, R^k l_0$. Then, at the optimum

$$[1 - \Gamma(\bar{\omega}^{sp})] = \Gamma'(\bar{\omega}^{sp})\, l_0 \frac{d\bar{\omega}^{sp}}{dl_0} \,. \tag{14}$$

Equation (14), together with the constraint on investors total expected utility (12), describes the socially optimal debt contract. Note that equation (14) takes a similar form as its counterpart in the private equilibrium, given by equation (8). In other words, the social optimum also trades off the marginal gain from increasing the firm's leverage against the marginal cost required to compensate investors for providing that leverage.

We are ready to establish the generic inefficiency of the private provision of liquidity.[12]

Proposition 3 (Generic Constrained Inefficiency of Liquidity Provision) *Consider a planner that designs an optimal financial contract, as described by (12), (14), (3) and (9). Given the parameters (α, ψ, r) belonging to a generic set $\mathcal{P}$, the planner will set a level of secondary market liquidity that is different from the private equilibrium. That is, the private equilibrium is generically constrained inefficient.*

The proof is relegated to the appendix, but it is instructive to outline the main steps to gain intuition about the source of the inefficiency. Comparing equations (8) and (14), it follows that the private equilibrium is constrained efficient if $\partial\bar{\omega}^{pe}/\partial l_0 = \partial\bar{\omega}^{sp}/\partial l_0$. That is, the investors' break-even condition and the Pareto improvement constraint need to impose the same trade-off between leverage and risk. Comparing equations (7) and (13), it follows that efficiency obtains under two conditions. First, when investors are indifferent to providing liquidity or credit in the initial period, i.e., $u_b R^b = U_s$, as it was the case in the interior equilibrium studied in section 3. Second, when changes to market liquidity do not change investors' expected utility, i.e., $\partial\mathbb{U}/\partial\theta = 0$. Hence, $\partial\mathbb{U}/\partial\theta = 0$ is the key condition for constrained efficiency, and as we characterize in the proof, this is equivalent to

$$\frac{R^b - q_1(1+r)}{q_1 - \beta R^b} = \frac{1-\alpha}{\alpha} \qquad \Leftrightarrow \qquad \psi = \frac{\alpha(1+r)}{1+\alpha r} \,. \tag{15}$$

The generic inefficiency follows from the observation that the set of parameters that satisfies (15)

[12] See also Geanakoplos and Polemarchakis (1986) for a general characterization of constrained inefficiency.

is of measure zero.

The source of the inefficiency in our model stems from a congestion externality in the OTC market. To see this note that equation (15) specifies that the ratio of gains for patient and impatient investors of an additional executed trade equals the ratio of the elasticities of the matching function with respect to a sell and a buy order. In fact, the latter is given by $(1-\alpha)/\alpha = |f'(\theta)\theta/f(\theta)|/|p'(\theta)\theta/p(\theta)|$. Note that $[R^b - q_1(1+r)]/[q_1 - \beta R^b] = (1+r)S^P(q_1)/S^I(q_1)$ is the ratio of the respective social surpluses for patient and impatient investors from executing a trade. It is important to make two observations. First, note that the social gain for patient investors is the private surplus from an executed trade, $S^P(q_1)$, times the discounted utility of investing in storage $1+r$.[13] This reflects that patient investors consume in the final period after accruing at least the return on storage, $1+r$. Second, when $r = 0$ equation (15) reduces to $\psi = \alpha$, the Hosios (1990) condition. In light of these, we call equation (15) the *intertemporal Hosios condition*.

While the inefficiency stems from a congestion externality, the fact that firms' borrowing behavior affects their cost of debt via the liquidity provision and liquidity premium channels implies that our model also exhibits a pecuniary externality. To see this consider the problem of a firm that internalizes the effect of its contract terms on the liquidity premium $\Phi^\ell(\theta)$. This firm will choose lower leverage, which will implement a higher market liquidity through the liquidity provision channel.[14] A higher level of market liquidity will only make investors better off when $\partial\mathbb{U}/\partial\theta > 0$. So, in general, the fact that firms internalize their effect on secondary market liquidity will not necessarily deliver an efficient outcome.

As it is well understood in models with search and matching frictions, congestion externalities can typically be corrected through price posting and directed search (see, for example, Moen, 1997, or Wright et al. 2017). Moreover, Lester, Rocheteau, and Weill (2015) have argued this type of price transparency implements efficient outcomes in OTC markets. Thus, next we examine whether price transparency, as implemented by price posting with directed search, can address the core inefficiency in our model.

Consider the problem of a *trader selling* a bond. Following Moen (1997), she wants to maximize her trade surplus, while providing an incentive for buyers to participate. Suppose, *traders buying* can, in expectation, obtain a payoff $\overline{U}_{s,1}$, then we can write down the directed search prob-

[13]Recall that we normalize the discount factor of patient investors to 1, so $(1+r)$ should be interpreted as the rate of return on storage net of the intertemporal discount of future consumption by patient investors.

[14]Consider a firm that is constrained by the investors' break-even condition, but internalizes how her choices affect secondary market liquidity, i.e., $\max_{l_0,\bar{\omega}}[1-\Gamma(\bar{\omega})]R^k l_0$ subject to $R^b(l_0,\bar{\omega}) = \Phi(\theta(l_0,\bar{\omega}))(1+r)^2$. Note that the constraint to this problem implicitly defines a function $\bar{\omega}^F \equiv \bar{\omega}^F(l_0)$ with $d\bar{\omega}^F/dl_0 = -(\partial R^b/\partial l_0 - \Phi'(\theta)(\partial\theta/\partial l_0)(1+r)^2)/(\partial R^b/\partial\bar{\omega} - \Phi'(\theta)(\partial\theta/\partial\bar{\omega})(1+r)^2)$. Using this, it is possible to establish that $d\bar{\omega}^F/dl_0 > d\bar{\omega}^{pe}/dl_0$, implying that this firm will choose a lower leverage compared to the private equilibrium. Furthermore, $d\bar{\omega}^F/dl_0 \neq d\bar{\omega}^{sp}/dl_0$ given by (13), i.e., a firm that internalizes the liquidity premium will not choose the constrained efficient allocation.

lem as

$$\max_{q_1,\theta} \; f(\theta)\left(q_1 - \beta R^b\right) \qquad \text{s.t.} \qquad \overline{U}_{s,1} \leq p(\theta)\left(\frac{R^b}{1+r} - q_1\right) \tag{16}$$

From the constraint in problem (16), we can solve for the price as a function of market liquidity $q_1(\theta) = \left[R^b/(1+r) - \overline{U}_{s,1}/p(\theta)\right]$. We can then substitute this expression for $q_1(\theta)$ in the objective of a trader selling a bond and obtain the following optimality condition,

$$\frac{R^b/(1+r) - q_1}{q_1 - \beta R^b} = \frac{1-\alpha}{\alpha} \; . \tag{17}$$

We can solve equation (17) for q_1 and compare the resulting expression to the price obtained under Nash bargaining, given by equation (3). Doing so reveals that the two are equal as long as the classic Hosios condition, $\psi = \alpha$, is satisfied for the bargaining problem. Accordingly, because $r > 0$, the equilibrium under price posting is inefficient. Hence, additional interventions are needed to implement the planner's solution (see section 5).

The intuition is that $r > 0$ drives a wedge between the private and social valuation of the trade surplus. The solution to the price posting problem—or, equivalently, the bargaining problem under the classic Hosios condition—is the one that maximizes the total size of the *private* surplus, determined by the investors' willingness to pay in OTC exchange. In contrast, the *social* surpluses are determined by investors' discounted lifetime utility. This distinction opens a wedge between the private and social valuation of the trade surplus because the return on storage net of the rate of preference of patient investors is positive, $r > 0$. In this case, patient agents stand to gain additional utility by waiting to consume in the final period. In fact, the social surplus is given by the product of the discounted return on storage $(1 + r)$ and the private surplus, S^P. From the perspective of the optimal posted price , the additional utility of patient investors is immaterial as it amounts to a linear transformation of utility.[15] By contrast, from the perspective of the planner, this additional utility matters because patient investors will stand to gain a surplus that is greater by a factor of $(1+r)$. The planner understands this and, relative to the privately optimal allocation, implements a higher price in order to redistribute some surplus towards impatient investors. In fact, the efficient value of ψ in our intertemporal Hosios condition is larger than the value that obtains in the classic Hosios condition because $\alpha(1 + r)/(1 + \alpha r) > \alpha$ for $r > 0$. This distinction between private and social trade surplus arises naturally in the study of liquidity provision.[16]

[15] In fact, note that changing the definition of the surplus of patient traders from $\overline{U}_{s,1} \leq p(\theta)(R^b/(1 + r) - q_1)$ to $\tilde{U}_{s,1} \leq p(\theta)(R^b - q_1(1 + r))$ in the directed search problem (16) does not change the conclusion that equation (17) characterizes the solution.

[16] Another approach to shed light on the distinction between the efficient provision of liquidity and directed search in the OTC market, is to consider the problem of a date 1 planner, who chooses market liquidity, takes the decisions in the primary market as given, and is constrained by the search frictions and the structure of trade in the secondary market. Let $U_{s,1} = p(\theta)R^b + (1 - p(\theta))(1 + r)q_1$ be the expected utility per buy order for a patient investor, and

Before we continue, note that our results do not rule out that price posting can lead to a Pareto improvement relative to the private equilibrium with bargaining. Instead, our results establish that even under price posting a Pareto improvement can be achieved when $r > 0$.

The intertemporal Hosios condition allows us to provide a complete characterization of the under- or over-provision of liquidity in our model, as stated in the following proposition.

Proposition 4 (Under- and Over-Provision of Liquidity) *If $\psi > \alpha(1+r)/(1+\alpha r)$, liquidity is under provided and the planner chooses $\theta > \theta^*$, $l_0 < l_0^*$, and $\bar{\omega} < \bar{\omega}^*$. The opposite is true, if $\psi < \alpha(1+r)/(1+\alpha r)$. Finally, if $\psi = \alpha(1+r)/(1+\alpha r)$ liquidity is efficiently provided and equilibrium is constrained efficient, i.e., $(l_0, \bar{\omega}, \theta) = (l_0^*, \bar{\omega}^*, \theta^*)$.*

The planner internalizes the congestion externality and exploits it to increase the profitability of firms without harming investors. Proposition 4 tells us that the ability to do so depends on the parameters (α, r, ψ). Specifically, it depends on the value of the bargaining power of impatient investors, ψ, relative to what is prescribed by the intertemporal Hosios condition (15).

For example, consider the case when the bargaining power of impatient investors is sufficiently large such that $\psi > \alpha(1+r)/(1+\alpha r)$. Increasing market liquidity makes investors ex ante better off relative to the private equilibrium, i.e., $\partial\mathbb{U}/\partial\theta > 0$. Higher liquidity generates ex ante welfare gains for impatient investors because it is easier to sell unwanted bonds in the secondary market. At the same time, patient investors suffer as it becomes more difficult to earn a higher return by purchasing bonds at a discounted price. Provided the bargaining power of impatient investors ψ is sufficiently large, it will be the case that the gains for impatient investors outweigh the losses for patient investors, so $\partial\mathbb{U}/\partial\theta > 0$.

A social planner that understands this would like to rewrite the terms of the debt contract offered by the firm in the private equilibrium in a way that reduces both primary bond issuance, $l_0 < l_0^*$, and funding costs, $\bar{\omega} < \bar{\omega}^*$. Restricting bond issuance alters the composition of investors' portfolios in a way that generates higher liquidity in the secondary market and, hence, higher ex ante welfare for investors. The planner redistributes these welfare gains back to firms in the form of higher profits. Despite the fact that the firm is forced to operate at a smaller scale (owing to lower bond issuance), it is able to retain a larger share of profits in expectation through a reduction in financing costs. Financing costs are lower owing to both reduced risk of the financial contract, hence the default premium is lower, as well as higher liquidity in the secondary market, hence the liquidity premium is lower. In other words, higher secondary *market* liquidity, increases *funding* liquidity in the primary market via a reduction in the liquidity premium and, thus, in the total bond premium.

---

let $U_{b,1} = f(\theta)q_1 + (1 - f(\theta))\beta R^b$ be the expected utility per bond for an impatient investor. It is straightforward to show the solution to $\max_\theta BU_{s,1} + AU_{b,1}$ yields our intertemporal Hosios condition. In contrast, the solution to $\max_\theta BU_{s,1}/(1+r) + AU_{b,1}$ yields the classic Hosios condition.

The opposite intuition applies when the bargaining power of impatient investors is sufficiently small such that $\psi < \alpha(1 + r)/(1 + \alpha r)$. In this case, liquidity is over-provided in the private equilibrium and the planner would like to correct this by increasing primary bond issuance. The resulting welfare gain to patient investors (who find it easier to trade in the secondary market) outweighs the losses to impatient investors (who find trading more difficult) such that investors in the initial period are better off, i.e., $\partial\mathbb{U}/\partial\theta < 0$. As above, the planner redistributes these welfare gains back to the firm through the contract terms, as bond issuance and financing costs are both higher relative to the private equilibrium, $l_0 > l_0^*$ and $\bar{\omega} > \bar{\omega}^*$. In this case, firm profits rise because the increased scope of the firm more than offsets the higher financing costs, as both the default and the liquidity premium increase.

Regardless of the direction of the inefficiency, a general intuition is that in both cases the planner exploits the congestion externality to create additional surplus for investors which is then redistributed back to firms. In the knife-edge case, where the bargaining power of impatient investors coincides with our intertemporal Hosios condition, $\psi = \alpha(1+r)/(1+\alpha r)$, the planner cannot exploit the congestion externality and the private equilibrium allocation is constrained efficient.

Finally, these efficiency results were derived under the assumption of a constant returns to scale matching function. While this assumption was motivated by analytic tractability, we note that the congestion externality is likely to generalize to equilibria under an increasing returns matching function. (See Section 8.3 of the survey by Shimer et al. 2004 in the context of search and matching frictions in the labor market).

5 Optimal Liquidity Regulation

We analyze the optimal liquidity regulation under the assumption that the planner has access to a complete set of tax instruments. Specifically, we introduce a proportional tax on storage and leverage in period 0, denoted τ^s and τ^l, and lump sum transfers to firms and investors, T^l and T^s. Negative taxes correspond to subsidies.

With these instruments, the objective of firms and investors becomes, respectively, $[1-\Gamma(\bar{\omega})]\,R^k l_0 - \tau^l l_0 + T^l$ and $\mathbb{U} = b_0 U_b + s_0 U_s(1 - \tau^s) + T^s$. Lump-sum transfers are introduced to offset the income effect from distortionary taxation on each group of agents. Thus, in equilibrium $T^l = \tau^l l_0$ and $T^s = \tau^s s_0 U_s$. In this case, as in the private equilibrium, the relevant constraint faced by firms corresponds to the investors' break-even condition. However, when private liquidity is regulated, this constraint is distorted by the tax on storage, $U_s(1 - \tau^s) = u_b R^b$.

Proposition 5 characterizes the optimal regulation of liquidity provision.

Proposition 5 (Optimal Liquidity Regulation) *The planner's solution can be implemented by* *levying distortionary taxes on the portfolio allocation decision of investors and the financing decision of firms, with the optimal taxes on storage, τ^s, and leverage, τ^l, given by:*

$$\tau^s = \frac{e_0}{b_0}\left(1 - \frac{U_s(\theta^*)}{U_s(\theta)}\right) \qquad and \qquad \tau^l = \Gamma'(\bar{\omega})R^k l_0 \left[\frac{d\bar{\omega}^{sp}(l_0)}{dl_0} - \frac{d\bar{\omega}^{pe}(l_0)}{dl_0}\right] \tag{18}$$

When liquidity is under supplied, i.e., $\psi > \alpha(1+r)/(1+\alpha r)$, is optimal to tax leverage ($\tau^l > 0$) and subsidize storage ($\tau^s < 0$). The opposite is true when liquidity is over supplied. And when liquidity is efficient, i.e., $\psi = \alpha(1+r)/(1+\alpha r)$, the optimal $\tau^l = \tau^s = 0$.

The role of the tax on storage is to make investors internalize the effect of their portfolio decisions on market liquidity. Doing so, requires introducing a wedge in the investors' break-even condition (6), so investors' are no longer indifferent to providing credit and liquidity, i.e., $U_b \neq U_s$. The lump sum transfer to investors ensures they are not made better or worse off. The role of the tax on leverage is to make the firm internalize the effect of its borrowing on its financing cost.

Using the insights from Proposition 4 above, the statement in Proposition 5 is intuitive. For example, when $\psi > \alpha(1+r)/(1+\alpha r)$, the planner wants to implement higher liquidity relative to the private equilibrium, $\theta > \theta^*$. Accordingly, the optimal regulation needs to be designed in a way that results in investors holding a more liquid portfolio and firms issuing fewer bonds in the primary market. This can be achieved through a tax on leverage, $\tau^l > 0$, and a storage subsidy, $\tau^s < 0$.

6 Optimal Public Liquidity Management

We now examine how the optimal management of public liquidity can alleviate trading frictions and improve economic efficiency beyond what can be achieved by liquidity regulation, as studied in the previous section. Our model features a portfolio rebalancing mechanism where the OTC search frictions endogenize the costs of portfolio adjustment. Any public policy that alters the aggregate liquidity of private portfolios will affect the relative compensation investors require to purchase illiquid assets, i.e., the liquidity premium. Changes in the liquidity premium, in turn, influence savings and investment decisions in the real economy. In practice, important policies that have altered the aggregate liquidity of private portfolios are quantitative easing or large scale asset purchases, as the ones implemented during the Great Recession in the U.S., Europe, and Japan.

To model quantitative policies we make two assumptions. First, there is a central bank (CB) that purchases firms' financial contracts but is not subject to liquidity-preference shocks. This creates a technological advantage for the planner relative to private agents, so it should not be surprising that a CB might be able to improve upon the constrained efficient outcome. However, our analysis does provide new insights about the optimal design of quantitative policies. Second, assets purchased by the CB are less liquid relative to reserves, i.e., the liabilities issued to finance these assets. This could be viewed as a strong assumption for the Federal Reserve QE program, which was limited to U.S. Treasuries and Agency Mortgage Backed Securities. Nonetheless, the evidence suggests that these assets become less liquid as they are *retraded* in secondary markets.[17] This assumption is less of a concern when considering QE programs in other jurisdictions, like Europe or Japan, where central banks have purchased non-government guaranteed assets, perceived as less liquid than central bank reserves.

6.1 The Central Bank and Quantitative Policies

We model quantitative policies through direct purchases and sales by the central bank (CB) of long-term illiquid assets (the financial contracts issued by firms and which are retraded by investors in OTC markets). These purchases are financed by the issuance of short-term liquid liabilities, referred to as reserves. Alternatively, asset sales are used to redeem existing reserves and reduce the size of the CB's balance sheet. This seems a reasonable approximation for the policies implemented by the Federal Reserve during the Great Recession, where lending facilities and asset purchases were financed primarily with redeemable liabilities in the form of reserves (see Carpenter et al. 2013).

At the beginning of the initial period, the CB credibly commits to purchase a quantity $\bar{b}_0$ of bonds from investors and hold them to maturity.[18] These bond purchases are financed through the issuance of $\bar{s}_0$ units of reserves that pay interest $\bar{r}$. In the interest of space, we focus on the case of QE. But, the model lends itself to the analysis of quantitative tightening (QT), endowing the CB with some initial holdings of illiquid bonds and investors with reserves balances in period 0. In our model, reserves offer the same liquidity as storage from the point of view of investors, thus in an equilibrium where investors hold reserves, $\bar{r} \geq r$. The CB waits for the primary debt market to clear and then meets with investors to exchange reserves for bonds. The QE operation is conducted in a frictionless market that meets after bonds have been issued but before the OTC market opens. That is, we have abstracted from trade frictions between the CB and investors, as in practice the Federal Reserve announces in advance its intention to buy bonds and has readily available trading counterparties. Investors can freely trade reserves for consumption with the CB at any point.[19] We

[17]Vayanos and Weill (2008) argue that the off-the-run phenomenon can be explained by trade frictions in U.S. Treasury markets. Vickery and Wright (2013) describe the TBA market and the the market for "specified pool" agency MBS as OTC markets.

[18]Note that the CB buys bonds at face value. This implies that the effect of QE does not rely on the purchase of bonds at distressed values. In addition, if the CB purchases bonds at a discount it will increase the expected return on asset purchases and relax its no loss constraint.

[19]This is isomorphic to a model in which trade in the Fed Funds market is frictionless. Another literature studies

also assume that the CB allocates reserves evenly across investors who demand them.

We make two additional assumptions, which as discussed below, will prevent the CB from disintermediating the OTC market. First, we preclude the CB from injecting real resources into the economy. For that, we assume that the CB can only finance its operations with the proceeds from issuing reserves and the payout from its illiquid investments. Thus, in the initial period, the CB budget constraint is simply $\bar{b}_0 = \bar{s}_0$. In order to keep its bond holdings, the CB needs to roll over its outstanding reserves and pay interest on them in period 1. The CB will have to borrow an amount equal to $(1+\bar{r})\bar{s}_0$. Finally, in period 2 the CB receives the debt payout from the financial contract and expends $(1+\bar{r})^2\bar{s}_0$ in interest and principal on outstanding reserves. The period 2 budget constraint imposes a no loss constraint on the CB, i.e., the total payment on reserve balances cannot exceed the hold-to-maturity return to the CB.

Second, we assume that the CB is at a disadvantage relative to the private sector in monitoring investment projects. It thus needs to pay a higher monitoring cost relative to investors, denoted by $\bar{\mu} > \mu$. This implies the CB expects to receive $\bar{R}^b\bar{b}_0$ for its bond holdings, with $\bar{R}^b(l_0, \bar{\omega}) = l_0/(l_0-1)R^k\left[\Gamma(\bar{\omega}) - \bar{\mu}G(\bar{\omega})\right]$. Let $L(l_0, \bar{\omega}) = l_0/(l_0-1)R^k G(\bar{\omega})(\bar{\mu}-\mu)$ denote the additional expected monitoring costs paid by the CB relative to private investors. Using these we can express the CB *no loss constraint* as

$$(1+\bar{r})^2 + L(l_0, \bar{\omega}) \leq R^b(l_0, \bar{\omega}) \qquad \Leftrightarrow \qquad 1 + L(l_0, \bar{\omega})/(1+r)^2 \leq \Phi^{\ell}(\theta)\,. \tag{19}$$

That is, the interest paid on reserves and the additional expected monitoring costs cannot exceed the private hold-to-maturity return. In other words, the liquidity premium needs to be higher than the anticipated additional monitoring costs in present value. Because we have abstracted from aggregate risk, a pool of firms' contracts will have no credit risk, and constraint (19) guarantees that the central bank never suffers ex post losses.

6.2 Quantitative Policies and Portfolio Rebalancing

This section presents a positive analysis of quantitative policies. As we will show, these policies operate through a portfolio rebalancing mechanism as in Tobin (1969). As the CB purchases illiquid assets, the risk premium on these assets declines, which in turn influences real investment decisions. Contrary to much of the literature, adjustment costs in our model are determined endogenously through the search frictions in the OTC market.

The Optimal Financial Contract. As in the previous sections, we consider that firms choose the terms of the optimal contract in the private equilibrium. We follow the preceding analysis by

frictional trade in the Fed Funds market, see, for example, Afonso and Lagos (2015) or Bianchi and Bigio (2014).

writing the relevant constraint imposed by investors on firms as a required hold-to-maturity return, $R^b(l_0, \bar{\omega})$.

The OTC Market. In period 0 investors allocate their wealth across two assets: the storage technology and bonds. So, the budget constraint at $t = 0$ is given by $s_0 + b_0^s = e_0$, with $s_0, b_0^s \geq 0$. Subsequently, investors exchange $\bar{b}_0$ bonds for $\bar{s}_0$ reserves with the central bank. Following the approach of Section 2, we consider the optimal behavior of impatient and patient investors in $t = 1$ when they arrive with a generic portfolio of storage, reserves, and bonds $(s_0, \bar{s}_0, b_0^s - \bar{b}_0)$.

Impatient Investors. By Assumption 3 impatient investors want to consume all their wealth at $t = 1$. They can consume the payout of their liquid assets: $(1 + r)s_0 + (1 + \bar{r})\bar{s}_0$; in addition, they can consume the proceeds from their sell orders in the OTC market: q_1 units of consumption for each order executed. Thus, the expected consumption of impatient investors in periods 1 and 2, respectively, is given by $c_1^I = (1 + r)s_0 + (1 + \bar{r})\bar{s}_0 + f(\theta)q_1(b_0^s - \bar{b}_0)$ and $c_2^I = (1 - f(\theta))R^b(b_0^s - \bar{b}_0)$.

Patient Investors. Patient investors only value consumption in the final period and, as a result, are willing to place buy orders in the OTC market because the return from doing so, Δ, is strictly greater than the return on storage, $1 + r$. Moreover, it is also the case that the return on reserves, $1 + \bar{r}$, is at least as large as that on storage, so patient investors are willing to allocate liquid wealth to reserves. Accordingly, liquidity provision in the secondary market will depend on the return on OTC trade, Δ, relative to the return on reserves, $1 + \bar{r}$. Specifically, if $1 + \bar{r} < \Delta$ patient investors will pledge *all* their liquid wealth to place buy orders in the OTC market. On the other hand, if $1 + \bar{r} > \Delta$ patient investors will use their liquid wealth to buy higher yielding reserves first, and then allocate the remainder of their liquid wealth to placing buy orders in the OTC market. For expositional purposes, we assume throughout the remainder of the paper that $1 + \bar{r} < \Delta$ (although for the main results of this section—stated below in Propositions 6 and 7—we trace out the proofs over the entire parameter space of the model, where appropriate).

When the anticipated return to OTC trade exceeds the return on reserves, patient investors use $(1+r)s_0+(1+\bar{r})\bar{s}_0$ units of consumption to place buy orders. A fraction $p(\theta)$ are matched allowing patient investors to exchange consumption for bonds, while the $1 - p(\theta)$ unmatched portion needs to be reinvested in liquid assets in period $t = 1$. Given that the central bank needs to finance itself in the interim period, it will reallocate reserves to patient investors, hence total reserve holdings for patient investors are given by $\bar{s}_1^P = (1+\bar{r})\bar{s}_0/(1-\delta)$. All remaining units of consumption are placed into the lower yielding storage technology, so $s_1^P = (1-p(\theta))\,[(1 + r)s_0 + (1 + \bar{r})\bar{s}_0]-(1+\bar{r})\bar{s}_0/(1-\delta)$, which is strictly positive from Assumption 4. It follows that expected consumption of patient investors equals $c_2^P = s_1^P(1 + r) + (1 + \bar{r})^2\bar{s}_0/(1 - \delta) + \left\{b_0^s - \bar{b}_0 + p(\theta)((1 + r)s_0 + (1 + \bar{r})\bar{s}_0)/q_1\right\} R^b$.

Investors' Portfolio Choice. Using the optimal behavior of investors in period 1, derived above, we can rewrite the expected lifetime utility as the sum of the expected utilities from each of the three assets available in the initial period: $\mathbb{U} = U_s s_0 + U_{\bar{s}} \bar{s}_0 + U_b(b_0^s - \bar{b}_0)$. As before, the expected utility of investing in storage and bonds, U_s and U_b, are given by equations (4) and (5), respectively. However, given the additional option to investors to purchase reserves, U_s now represents the marginal return on investing in storage (before marginal and average returns were equal). This representation is useful as it allows us to draw on previous results to characterize the equilibrium with quantitative policies. Finally, the expected total utility from reserves is given by

$$U_{\bar{s}} = \delta(1+\bar{r}) + (1-\delta)(1+\bar{r})\left[(1-p(\theta))(1+r) + \frac{\bar{r}-r}{1-\delta} + p(\theta)\Delta\right] . \tag{20}$$

Reserves yield $1+\bar{r}$ for impatient investors. For patient investors, there is additional compensation that comes from the expected return from buy orders in the secondary market, plus the spread between reserves and storage, $\bar{r} - r \geq 0$, for the inframarginal reserves bought in period 1. It follows, that in equilibrium, investors will be indifferent between a marginal unit invested in storage or bonds, as described by equation (6).

Equilibrium. A given QE policy $(\bar{b}_0, \bar{r})$ is said to be feasible if it satisfies the CB constraints. For a feasible QE policy, $(\bar{b}_0, \bar{r})$, equilibrium is defined as in Section 2.4. When QE policy is given by $(\bar{b}_0, \bar{r})$ and the equilibrium in the credit market is given by (b_0, R^b), market liquidity corresponds to

$$\theta = \frac{(1-\delta)[(1+r)s_0 + (1+\bar{r})\bar{s}_0]}{\delta\left(b_0 - \bar{b}_0\right)q_1} = \frac{(1-\delta)\Delta\left[(1+r)\left(e_0 - n_0(l_0-1)\right) + (1+\bar{r})\bar{b}_0\right]}{\delta R^b\left(n_0(l_0-1) - \bar{b}_0\right)} . \tag{21}$$

This expression establishes a link between QE and secondary market liquidity which we summarize in the following proposition.

Proposition 6 (Quantitative Policies and Portfolio Rebalancing) *QE, i.e., increasing CB bond purchases, $\bar{b}_0$, increases secondary market liquidity θ, reduces the liquidity premium, Φ^ℓ, and increases firms' investment.*

The intuition is straightforward. Each bond bought by the CB will be held to maturity and, therefore, reduces the number of sell orders in the secondary market. At the same time, these bonds need to be financed with reserves, which patient investors can use to submit additional buy orders in the secondary market.[20] So, a bond buying program has a *direct effect* on secondary market liquidity because it alters the composition of investor's portfolios away from illiquid bonds toward publicly provided liquid assets. The resulting reduction in the liquidity premium demanded by investors pushes down the cost of financing for firms, who respond by taking on higher leverage and

[20] Christensen and Krogstrup (2016) present empirical evidence for the effect of QE through the creation of reserves.

risk in equilibrium. This later *indirect effect* attenuates the effect of the QE program as increased bond issuance by firms crowds out public and private liquidity.

It should also be noted that Proposition 6 is presented from the perspective of a CB that wants to increase liquidity by expanding reserves in order to purchase illiquid bonds. But, this result is more general. A central bank that starts with an initial endowment of bonds could remove liquidity by becoming a net seller to investors of illiquid bonds in exchange for reserves. A quantitative tightening (QT) program such as this would effectively withdraw public liquidity and reduce secondary market liquidity.

6.3 Optimal Public Liquidity Management via Quantitative Policies

To describe the optimal design of quantitative policies, we consider a planner that maximizes the utility of entrepreneurs and investors and is constrained by search frictions, the structure of trade in the OTC market, and the feasibility of QE policies. In addition, as in Section 4, we consider the planner's Pareto problem, where investors cannot be made worse off. To formalize this problem, let $\mathbb{U}(l_0, \bar{\omega}, \theta, \bar{b}_0, \bar{r})$ be the expected utility of investors when the equilibrium is described by $(l_0, \bar{\omega}, \theta)$, with the secondary market price given by (3), and the QE program described by $(\bar{b}_0, \bar{r})$. Similarly, let $\mathbb{U}(l_0^*, \bar{\omega}^*, \theta^*)$ be the expected utility of investors in the private equilibrium, when the secondary market price is given by (3). We refer to this planner that have access to QE policies as the CB. Then, the CB's problem can be written as $\max_{l_0,\bar{\omega},\theta,\bar{b}_0,\bar{r}}\ [1 - \Gamma(\bar{\omega})] R^k l_0$, subject to the feasibility of QE, equation (21), and

$$\mathbb{U}(l_0, \bar{\omega}, \theta, \bar{b}_0, \bar{r}) \geq \mathbb{U}(l_0^*, \bar{\omega}^*, \theta^*). \tag{22}$$

We show when quantitative policies can improve upon the constrained efficient allocation described in section 4, which we denote by $(l_0^{**}, \bar{\omega}^{**}, \theta^{**})$. The following proposition characterizes the optimal design of QE policies.

Proposition 7 (Optimal QE Policy) *If $\psi > \alpha(1+r)/(1+\alpha r)$ and $\Phi^\ell(\theta^{**}) > 1 + L(l_0^{**}, \bar{\omega}^{**})/(1+r)^2$, then QE improves upon the constrained efficient allocation and it is characterized by $\bar{b}_0 > 0$ and $\bar{r} > r$. Alternatively, if $\psi \leq \alpha(1+r)/(1+\alpha r)$ or $\Phi^\ell(\theta^{**}) \leq 1 + L(l_0^{**}, \bar{\omega}^{**})/(1+r)^2$, then QE does not improve upon the constrained efficient allocation and optimally $\bar{b}_0 = 0$.*

The intuition for the optimal QE policy follows from the previously established results. From Proposition 6 we know that QE, or CB bond purchases $\bar{b}_0$, increases secondary market liquidity θ. In addition, in section 4 we established that when $\psi > \alpha(1+r)/(1+\alpha r)$ higher market liquidity increases investors expected utility $\partial\mathbb{U}/\partial\theta > 0$. Thus, when $\psi > \alpha(1+r)/(1+\alpha r)$ a CB will implement a QE policy if it is feasible, as it increases investors' expected utility and the resulting gains can be redistributed to firms. Intuitively, the QE program transfers illiquid bonds from investors, who value liquidity, to the CB, who does not value liquidity because it is a long-term investor and is not subject to runs.

Moreover, Proposition 7 shows that QE will only be used when the liquidity premium is larger than the additional monitoring costs incurred by the CB relative to the private sector. This condition implies that QE will be optimal only in times when the liquidity premium is relatively high. High liquidity premia are likely to arise during financial crises, rationalizing QE interventions in these states of the world rather that in normal times. Although we do not study aggregate fluctuations, our model can be extended in that direction. Moreover, this condition draws a link between optimal quantitative policies and financial spreads, connecting our analysis to the literature that studies the effect of quantitative policies in macro models (see Carlstrom, Fuerst, and Paustian 2017, and Gertler and Karadi 2011).

There are a few additional points worth mentioning. First, the public provision of liquidity is inherently different from liquidity regulation. Both policies affect the level of market liquidity, but regulation trades off liquidity and credit provision, whereas public liquidity management implies that public liquidity provision and credit provision move in tandem. This is due to the fact that public liquidity provision enhances the intermediation technology of the economy, as the transfer of liquidity risk between the public and private sector can only be achieved in the model through quantitative policies.

While we have shown that the optimal management of public liquidity can lead to a Pareto improvement, these quantitative policies do not explicitly address the externalities identified in Section 4. Indeed, a QE program is optimal when liquidity is inefficiently low, or equivalently, when firm's leverage and the riskiness of the contracts it offers to investors are inefficiently high. While QE is effective at boosting liquidity, it does so at the expense of encouraging firms to take on even more leverage and write even riskier contracts. This opens the door for optimal liquidity management (through quantitative policies) to coexist with optimal liquidity regulation, echoing a similar result found in Holmström and Tirole (1998).

Second, note that the proposition suggests QE is effective when the interest rate on storage is sufficiently low, $r < (\psi - \alpha)/(\alpha - \alpha\psi)$. Although it is beyond the scope of this model, these conditions indicate that QE may be an effective policy response in a protracted low interest rate environment. By the same token, the proposition also suggests that QT can be optimal when the interest rate is sufficiently high, so that $r > (\psi - \alpha)/(\alpha - \alpha\psi)$. In the context of the current policy debate, our framework offers support for a strategy of raising interest rates prior to unwinding the size of the balance sheet.

Finally, its useful to note that when QE is effective, the absence of constraints that limit the size of the program could lead to an extreme outcome in which the CB disintermediates the bond market. That is, the optimal policy is for the CB to buy all the bonds offered by the firm and offer the corresponding amount of reserves to investors. Doing so would allow the CB to replicate the frictionless benchmark. However, we prevent this by limiting the size of the QE program through the constraints imposed on the CB.

7 Conclusion

We show that search frictions in OTC markets provide a rationale for the regulation and public management of market liquidity. In our model, investors face a trade-off between liquidity and credit provision linking the outcomes in primary and secondary credit markets. Because of a congestion externality the private equilibrium is generically inefficient. A novel aspect of our analysis is that private liquidity can be either inefficiently high or inefficiently low, depending on the incentives faced by investors. We provide an analytic characterization of the distortions, show that price transparency as implemented through price posting does not resolve the inefficiency, and establish how the socially efficient equilibrium can be decentralized with tax instruments. Finally, we derive the conditions under which the quantitative policies, such as QE and QT, that manage public liquidity can enhance welfare.

There are a number of directions for future work. First, our paper opens up new avenues for research on optimal liquidity provision when financial intermediation is conducted in markets with OTC characteristics. For example, our framework could be potentially used to examine the behavior of banks trading in other OTC markets, such as Fed funds or repo markets, and to study the implications for bank capital and liquidity regulations. Second, it would be interesting to explore the quantitative relevance of the mechanism described in this paper. To this end, we have deliberately stayed very close to the quantitative model of Bernanke et al. (1999). Finally, another interesting extension would be to jointly consider bank- and bond-financing and study the interaction of these two sources of financing for the real economy, as well as spillovers from bank (liquidity) regulation on market liquidity (bank credit provision).

Appendix

Proof of Theorem 1: We need to show that there is a unique equilibrium and that, in this equilibrium, credit is not rationed. We proceed in three parts. In Part 1, we establish that the privately optimal contract is an interior solution to the firm's optimization problem. In Part 2, we establish existence of equilibria. In Part 3, we establish uniqueness.

Part 1. The privately optimal contract is interior. First of all, note that from the definition of $\Gamma(\omega)$ and $G(\omega)$ it follows that for any $\bar{\omega} > 0$: $\Gamma(\bar{\omega}) > 0$; $1 - \Gamma(\bar{\omega}) = \mathbb{P}(\omega \geq \bar{\omega})\mathbb{E}[\omega - \bar{\omega}|\omega \geq \bar{\omega}] > 0$;

$1 > \Gamma'(\bar{\omega}) = 1 - F(\bar{\omega}) > 0$; $\Gamma''(\bar{\omega}) = -dF(\bar{\omega}) < 0$; $0 < G(\bar{\omega}) < 1$; $\mu G(\bar{\omega}) < G(\bar{\omega}) < \Gamma(\bar{\omega})$; $G'(\bar{\omega}) = \bar{\omega}dF(\bar{\omega}) > 0$; $G''(\bar{\omega}) = dF(\bar{\omega}) + \bar{\omega}d(dF(\bar{\omega}))/d\bar{\omega}$; $\lim_{\bar{\omega}\to 0}\Gamma(\bar{\omega}) = 0$; $\lim_{\bar{\omega}\to\infty}\Gamma(\bar{\omega}) = \bar{\omega}\mathbb{P}(\omega \geq \bar{\omega}) + \mathbb{P}(\omega < \bar{\omega})\mathbb{E}[\omega|\omega < \bar{\omega}] = 1$; $\lim_{\bar{\omega}\to 0} G(\bar{\omega}) = 0$; and $\lim_{\bar{\omega}\to\infty} G(\bar{\omega}) = 1$.

In addition, from Assumption 2, $\bar{\omega}dF(\bar{\omega})/(1-F(\bar{\omega}))$, is increasing so, $1-\mu\bar{\omega}dF(\bar{\omega})/(1-F(\bar{\omega}))$, has only one root, which is strictly positive and is denoted by $\bar{\bar{\omega}} > 0$. Then, $\Gamma'(\bar{\omega}) - \mu G'(\bar{\omega}) = (1-F(\bar{\omega}))\,(1 - \mu\bar{\omega}dF(\bar{\omega})/(1 - F(\bar{\omega})))$ is greater than zero if $\bar{\omega} < \bar{\bar{\omega}}$, is zero if $\bar{\omega} = \bar{\bar{\omega}}$, and is smaller than zero if $\bar{\omega} > \bar{\bar{\omega}}$.

The value of the firm, $[1 - \Gamma(\bar{\omega})]\,R^k l_0$, is increasing in leverage, l_0, and decreasing in risk, $\bar{\omega}$. If investors' required expected (hold-to-maturity) return is $R^b \in [(1 + r)^2, R^k]$, then, for a given level of risk, $\bar{\omega}$, the leverage yielding this return, $l_0^{pe}(\bar{\omega})$, is given by $R^b = R^b(l_0^{pe}(\bar{\omega}), \bar{\omega})$, or

$$l_0^{pe}(\bar{\omega}) = \frac{R^b}{R^b - R^k[\Gamma(\bar{\omega}) - \mu G(\bar{\omega})]}\ . \tag{A.1}$$

Since $\Gamma(\bar{\omega}) - \mu G(\bar{\omega})$ attains a maximum at $\bar{\bar{\omega}}$, we get that $l_0^{pe}(\bar{\omega}) \leq R^b/(R^b - R^k[\Gamma(\bar{\bar{\omega}}) - \mu G(\bar{\bar{\omega}})]) \leq (1+r)^2/((1+r)^2 - R^k[\Gamma(\bar{\bar{\omega}}) - \mu G(\bar{\bar{\omega}})]) \equiv \bar{\bar{l}}_0$. It follows that the firm will never choose risk above $\bar{\bar{\omega}}$, as additional risk, which reduces firm's value, does not allow the firm to increase leverage. Therefore, the firm chooses a level of risk $0 \leq \bar{\omega} \leq \bar{\bar{\omega}}$ and value of leverage $l_0^{pe}(0) = 1 \leq l_0 \leq \bar{\bar{l}}_0 = l_0^{pe}(\bar{\bar{\omega}})$.

Given an R^b, the optimal contract is given by choosing $\bar{\omega}$ to maximize the value of the firm, $[1 - \Gamma(\bar{\omega})]\,R^k l_0^{pe}(\bar{\omega})$, subject to $0 \leq \bar{\omega} \leq \bar{\bar{\omega}}$. Note that since the firm's objective is continuous the maximum is achieved in the closed set defined by the constraint. Now we want to establish that the maximum is interior, i.e., $0 < \bar{\omega} < \bar{\bar{\omega}}$. We write the Lagrangian for this problem as $\mathcal{L} = [1 - \Gamma(\bar{\omega})]\,R^k l_0^{pe}(\bar{\omega}) + \check{\eta}\bar{\omega} - \hat{\eta}[\bar{\omega} - \bar{\bar{\omega}}]$. Then, the first-order condition with respect to $\bar{\omega}$ is, $-\Gamma'(\bar{\omega})R^k l_0^{pe}(\bar{\omega}) + [1 - \Gamma(\bar{\omega})]\,R^k(dl_0^{pe}/d\bar{\omega}) + \check{\eta} - \hat{\eta} = 0$, where $dl_0^{pe}/d\bar{\omega} = l_0^{pe}R^k[\Gamma'(\bar{\omega} - \mu G'(\bar{\omega}))]/\{R^b - R^k[\Gamma(\bar{\omega}) - \mu G(\bar{\omega})]\} > 0$.

Suppose the optimal financial contract is not interior. Then, it must be that $\bar{\omega} = 0$ and $\check{\eta} > 0$ or $\bar{\omega} = \bar{\bar{\omega}}$ and $\hat{\eta} > 0$. Suppose first that $\bar{\omega} = 0$ and $\check{\eta} > 0$. Note that $l_0^{pe}(0) = 1$, $\Gamma'(0) = 1 - F(0) = 1$, and $\Gamma(0) = G(0) = 0$. Then, from the first-order condition $\check{\eta} = R^k(R^b - R^k)/R^b \leq 0$, which is a contradiction. So we conclude that $\bar{\omega} > 0$. Similarly, if $\bar{\omega} = \bar{\bar{\omega}}$ and $\hat{\eta} > 0$. In this case, $l_0^{pe}(\bar{\bar{\omega}}) = \bar{\bar{l}}_0$ and $dl_0^{pe}(\bar{\bar{\omega}})/d\bar{\omega} = 0$. Then from the first-order condition $\hat{\eta} = -\Gamma'(\bar{\bar{\omega}})R^k\bar{\bar{l}}_0 < 0$, which is a contradiction. So we conclude that $\bar{\omega} < \bar{\bar{\omega}}$ and, thus, the privately optimal contract is an interior solution to the firm's optimization problem, with the optimal leverage $1 < l_0 < \bar{\bar{l}}_0$. Since $\bar{\omega} < \bar{\bar{\omega}}$ and $l_0 < \bar{\bar{l}}_0$ we conclude that credit will not be rationed and that $R^b(l_0, \bar{\omega})$ given in equation (1) is increasing in $\bar{\omega}$.

Part 2. Existence of equilibria. Let $J : C \to \mathbb{R}$, with $C = [(1 + r)^2, R^k]$. For $R^b \in C$, $J(R^b)$ is defined as follows. Given R^b define $(l_0(R^b), \bar{\omega}(R^b))$ as the privately optimal contract, which is interior, as shown in Part 1 above, and feasible, because $\bar{\bar{l}}_0 < e_0/n_0 + 1$ from Assumption 4. Use $(l_0(R^b), \bar{\omega}(R^b))$ to calculate $b_0(R^b) = n_0(l_0(R^b) - 1)$ and $s_0(R^b) = e_0 - b_0(R^b)$ and $\theta(R^b)$ as $\theta(R^b) = (1-\delta)s_0(R^b)(1+r)\Delta/(\delta b_0(R^b)R^b)$. Then, $J(R^b) = U_s(\theta(R^b))/u_b(\theta(R^b))$. Intuitively, for any hold-to-maturity two-period return R^b, the function $J(R^b)$ gives the hold-to-maturity return that makes investors indifferent between liquid storage and illiquid two-period bonds, given that: (i) firms optimally choose the contract, and the demand for credit, given R^b; and (ii) that the level of secondary market liquidity is consistent with the investors portfolios that support the optimal firms' bond issuance $b_0(R^b)$, i.e., the level of market liquidity is consistent with market clearing in the primary debt market. It follows that a fix point of J constitutes a private equilibrium. Thus, we want to show that J is a continuous single valued function and $J(\mathcal{C}) \subset \mathcal{C}$, so J has a fixed point $R^b = J(R^b)$, which constitutes a non-rationing equilibrium from Part 1 above.

First, we show that J is a single valued function. For that it suffices to show that the optimal contract as a function of R^b is a single valued function. The objective of the firms' problem is concave as $\Gamma'(\bar{\omega}) > 0$. But the feasible set defined by the constraints to the firm's optimization problem, given R^b, is not convex, so we need to rule out that the firm's indifference curves and the investors' expected return condition, $R^b = R^b(l_0, \bar{\omega})$, intersect more than once. Note that the firm's indifference curves are described by $l_0^{ic}(\bar{\omega}) = L/\left(R^k[1-\Gamma(\bar{\omega})]\right)$, where L is a constant that describes the level of profits at the indifference curve. Then, $dl_0^{ic}/d\bar{\omega} = L\,\Gamma'(\bar{\omega})/\left(R^k[1-\Gamma(\bar{\omega})]^2\right)$. At the optimal contract, these two curves intersect and have the same derivative. We use the condition that the two curves intersect to express L in terms R^b. In fact, $l_0^{ic} = l_0^{pe}$ imply $L = R^b R^k[1-\Gamma(\bar{\omega})]/\left(R^b - R^k[\Gamma(\bar{\omega}) - \mu G(\bar{\omega})]\right)$. Moreover, these two curves have the same slope, i.e., $dl_0^{ic}/d\bar{\omega} = dl_0^{pe}/d\bar{\omega}$, if and only if $\mathcal{H}(\bar{\omega}) = (R^k - R^b)/\left(\mu R^k\right)$, where $\mathcal{H}(\bar{\omega}) \equiv G(\bar{\omega}) + [1 - \Gamma(\bar{\omega})]G'(\bar{\omega})/\Gamma'(\bar{\omega})$. The function $\mathcal{H}(\bar{\omega})$ is a strictly increasing function of $\bar{\omega}$. In fact, $\mathcal{H}'(\bar{\omega}) = [1-\Gamma(\bar{\omega})]\,[\Gamma'(\bar{\omega})G''(\bar{\omega}) - G'(\bar{\omega})\Gamma''(\bar{\omega})]\,/\Gamma'(\bar{\omega})^2 = d(\bar{\omega}h(\bar{\omega}))/d\bar{\omega}\cdot[1-\Gamma(\bar{\omega})][1-F(\bar{\omega})]^2/\Gamma'(\bar{\omega})^2 > 0$. Then we conclude that there is only one solution to the firms' maximization problem. Therefore, $(l_0(R^b), \bar{\omega}(R^b))$ are single valued functions and so are $b_0(R^b)$, $s_0(R^b)$, $\theta(R^b)$, and $J(R^b)$.

It follows from above that $J(R^b)$ is also continuous. In fact, $\mathcal{H}(\bar{\omega})$ is a continuous function with $\mathcal{H}'(\bar{\omega}) > 0$, so by the Implicit Function Theorem $\bar{\omega}(R^b)$ is a continuous strictly decreasing function in $\mathcal{C}$, i.e., $\bar{\omega}'(R^b) < 0$. That is, the risk of the optimal contract is decreasing in the expected hold-to-maturity return offered to investors. Then, $l_0(R^b) \equiv l_0^{pe}(\bar{\omega}(R^b))$ is a continuous function in $\mathcal{C}$, given that $R^b = R^b(l_0, \bar{\omega}) > R^k[\Gamma(\bar{\omega}) - \mu G(\bar{\omega})]$. It follows, that $b_0(R^b)$ and $s_0(R^b)$ are continuous in $\mathcal{C}$, and thus, that $\theta(R^b)$ and $J(R^b)$ are continuous in $\mathcal{C}$.

Now we show that $J(R^b) \geq (1+r)^2$ and $J(R^b) \leq R^k$. From $(1+r) \leq \Delta \leq \beta^{-1}$, we have that $\delta(1+r)+(1-\delta)(1+r)^2 \leq U_s(\theta(R^b)) \leq \delta(1+r)+(1-\delta)(1+r)\Delta$ and $\delta\beta+1-\delta \leq u_b(\theta(R^b)) \leq \delta\Delta^{-1}+1-\delta$. On the one hand, from Assumptions 1 and 3, we have that $\delta[1+r-\beta R^k] \leq 0 \leq (1-\delta)[R^k-(1+r)\Delta]$. Rearranging, $J(R^b) \leq (\delta(1+r)+(1-\delta)(1+r)\Delta)\,/\,(\delta\beta+1-\delta) \leq R^k$. On the other hand, since

$\Delta \geq 1+r$ we have that $J(R^b) \geq \left(\delta(1+r)+(1-\delta)(1+r)^2\right)\Big/\left(\delta\Delta^{-1}+1-\delta\right) \geq (1+r)^2$.

Hence, $J(R^b)$ has a fixed point in the feasible domain and a non-rationing equilibrium exists.

Part 3. Uniqueness: Show that $J(R^b)$ is decreasing in C. Differentiating we obtain

$$\frac{dJ(R^b)}{dR^b} = J(R^b)\left[\frac{1}{U_s(\theta(R^b))}\frac{dU_s(\theta(R^b))}{d\theta} - \frac{1}{u_b(\theta(R^b))}\frac{du_b(\theta(R^b))}{d\theta}\right]\frac{d\theta(R^b)}{dR^b}\,. \tag{A.2}$$

To sign this derivative note that $dU_s/d\theta = (1-\delta)(1+r)p'(\theta)\,[\Delta-(1+r)] \leq 0$ and $du_b/d\theta = \delta f'(\theta)\left[\Delta^{-1}-\beta\right] \geq 0$, where the inequalities follow from $p'(\theta) \leq 0$, $f'(\theta) \geq 0$, and $(1+r) < \Delta \leq \beta^{-1}$. Thus, the term in square brackets in (A.2) is negative. We are left to show that $d\theta/dR^b > 0$. For that note that $d\theta/dR^b = (\theta/s_0)\,ds_0/dR^b - (\theta/b_0)\,db_0/dR^b - \theta/R^b$. In addition, from above we had that $\bar{\omega}'(R^b) < 0$ and $dl_0/dR^b = R^bR^k[\Gamma'(\bar{\omega}(R^b)) - \mu G'(\bar{\omega}(R^b))]\bar{\omega}'(R^b)/\left(R^b - R^k[\Gamma(\bar{\omega}(R^b)) - \mu G(\bar{\omega}(R^b))]\right)^2 < 0$. Note that the previous inequality imply that the demand for credit by firms is downward slopping, and it shows that the leverage of the optimal contract is decreasing in the expected hold-to-maturity return offered to investors. Using that $ds_0/dR^b = -db_0/dR^b$, $db_0/dR^b = n_0 dl_0/dR^b$, and $dR^b/dR^b = 1 = \left(\partial R^b/\partial l_0\right)dl_0/dR^b + \left(\partial R^b/\partial\bar{\omega}\right)d\bar{\omega}/dR^b$, we get that

$$\frac{R^b}{\theta}\frac{d\theta}{dR^b} = -\frac{dl_0}{dR^b}\left[\frac{\left(R^b - R^k[\Gamma(\bar{\omega}(R^b)) - \mu G(\bar{\omega}(R^b))]\right)^2}{R^k[\Gamma(\bar{\omega}(R^b)) - \mu G(\bar{\omega}(R^b))]} + \frac{R^b(e_0+n_0)}{l_0(e_0 - n_0(l_0-1))}\right] > 0\,, \tag{A.3}$$

where we used that in an interior contract $l_0 < e_0/n_0 + 1$. Hence, the non-rationing equilibrium is unique. ■

Proof of Proposition 1: The aggregate credit supply is determined by the investors' break-even condition (6). Taking the total derivative with respect to R^b, we get that the interest rate elasticity of the aggregate credit supply is

$$\varepsilon_{b_0^s,R^b} \equiv \frac{R^b}{b_0^s}\frac{db_0^s}{dR^b} = \frac{\left[\dfrac{\theta}{u_b}\dfrac{du_b}{d\theta} - \dfrac{\theta}{U_s}\dfrac{dU_s}{d\theta}\right]\dfrac{R^b}{\theta}\dfrac{\partial\theta}{\partial R^b} + 1}{\left[\dfrac{\theta}{U_s}\dfrac{dU_s}{d\theta} - \dfrac{\theta}{u_b}\dfrac{du_b}{d\theta}\right]\dfrac{b_0^s}{\theta}\dfrac{\partial\theta}{\partial b_0^s}}\,. \tag{A.4}$$

From the definition of market liquidity (9), we get that $\partial\theta/\partial b_0^s = -\left(\theta/b_0^s\right)e_0/(e_0 - b_0^s) < 0$ and $\partial\theta/\partial R^b = -\theta/R^b$, where the inequality follows from Assumption 4. Moreover, $dU_s/d\theta = (1-\delta)(1+r)p'(\theta)\,[\Delta-(1+r)] \leq 0$ and $du_b/d\theta = \delta f'(\theta)\left[\Delta^{-1}-\beta\right] \geq 0$. Thus, the denominator in (A.4) is non-negative; still we need to rule out that the denominator is zero, which is the case when $dU_s(\theta)/d\theta = du_b(\theta)/d\theta = 0$. This is the case when either $\theta > \bar{\theta}$ and $\psi = 1$, or $\theta < \underline{\theta}$

and $\psi = 0$. The first case violates the assumption that $\psi < 1$. The second case corresponds to the case where the liquidity premium is fixed at $\beta^{-1}/(1+r) > 1$. In this case $\theta < \underline{\theta}$ implies that $(1-\delta)(1+r)(e_0 - b_0)\beta^{-1}/\left(\delta b_0 R^k\right) \leq (1-\delta)(1+r)(e_0 - b_0)\beta^{-1}/\left(\delta b_0 R^b\right) < \min\{1, \nu^{1/\alpha}\}$. But, rearranging and using that $b_0 \leq n_0(\bar{l}_0 - 1)$ we get $e_0 \leq \left[\delta\beta R^k \min\{1, \nu^{1/\alpha}\}/\left((1-\delta)(1+r)\right) + 1\right] b_0 \leq \left[\delta\beta R^k \min\{1, \nu^{1/\alpha}\}/\left((1-\delta)(1+r)\right) + 1\right] n_0(\bar{l}_0 - 1)$, which contradicts Assumption 4, $e_0 >> n_0$. That is, the deep pocket assumption prevents liquidity from having a finite upper bound and we conclude that liquidity cannot be smaller than $\underline{\theta}$. Thus, we are left to show that the numerator in (A.4) is positive, i.e., $(\theta/u_b)\, du_b/d\theta - (\theta/U_s)\, dU_s/d\theta < 1$. This follows from Lemma 1 below where we prove that $\left|\varepsilon_{\Phi^\ell,\theta}\right| < 1$. ■

Proof of Proposition 2: Taking the derivative of $\Phi^d(\bar{\omega})$ in equation (10) with respect to $\bar{\omega}$ yields $d\Phi^d(\bar{\omega})/d\bar{\omega} = [\Gamma(\bar{\omega}) - \mu G(\bar{\omega}) - \bar{\omega}(\Gamma'(\bar{\omega}) - \mu G'(\bar{\omega}))] / [\Gamma(\bar{\omega}) - \mu G(\bar{\omega})]^2 = [(1-\mu)G(\bar{\omega}) + \bar{\omega}\mu G'(\bar{\omega})] / [\Gamma(\bar{\omega}) - \mu G(\bar{\omega})]^2 > 0$, given that $\Gamma(\bar{\omega}) = \bar{\omega}\Gamma'(\bar{\omega}) + G(\bar{\omega})$, $1 - \mu > 0$, $G(\bar{\omega}) \geq 0$ and $G'(\bar{\omega}) = \bar{\omega} dF(\bar{\omega}) > 0$, for any $\bar{\omega} > 0$. ■

Proof of Lemma 1: We want to show that the derivative of the liquidity premium with respect to secondary market liquidity is negative. Note that $U_s(\theta), u_b(\theta) > 0$, since the trading probabilities and returns are non-negative. In addition, note that $dU_s(\theta)/d\theta = (1-\delta)(1+r)\,[\Delta - (1+r)]\, dp(\theta)/d\theta \leq 0$ and $du_b(\theta)/d\theta = \delta\left[\Delta^{-1} - \beta\right] df(\theta)/d\theta \geq 0$, where the inequalities follow from $\beta \leq 1/(1+r)$, the definition of $f(\theta)$ and $p(\theta)$, and the fact that the matching function $m(A, B)$ is increasing in both arguments. From equation (11) we have that $d\Phi^\ell(\theta)/d\theta = \Phi^\ell(\theta)\left[U_s(\theta)^{-1} dU_s(\theta)/d\theta - u_b(\theta)^{-1} du_b(\theta)/d\theta\right] \leq 0$ using the previously established inequalities.

Turning to the second part of the Lemma, the elasticity of the liquidity premium, Φ^ℓ, with respect to the secondary market liquidity, θ, is written as $\varepsilon_{\Phi^\ell,\theta} = (\theta/\Phi^\ell) d\Phi^\ell/d\theta = [(\theta/U_s(\theta)) dU_s(\theta)/d\theta - (\theta/u_b(\theta)) du_b(\theta)/d\theta]$. Since the elasticity is negative, $\left|\varepsilon_{\Phi^\ell,\theta}\right| < 1$ requires that $\mathbb{X} \equiv U_s(\theta)u_b(\theta) + \theta u_b(\theta)(dU_s(\theta)/d\theta) - \theta U_s(\theta)(du_b/d\theta) > 0$. First, lets consider the case where $\theta \in \left(\underline{\theta}, \overline{\theta}\right)$. In this case, $f(\theta) = \nu\theta^{1-\alpha}$ and $p(\theta) = \nu\theta^{-\alpha}$ and, thus, $\theta df(\theta)/d\theta = (1-\alpha)f(\theta)$ and $\theta dp(\theta)/d\theta = -\alpha p(\theta)$. It follows that $\theta dU_s(\theta)/d\theta = -\alpha U_s(\theta) + \alpha(1+r)[\delta + (1-\delta)(1+r)] \leq 0$ and $\theta du_b(\theta)/d\theta = (1-\alpha)u_b(\theta) - (1-\alpha)[\beta\delta + (1-\delta)] \geq 0$. The latter yield $\mathbb{X} = \alpha u_b(\theta)(1+r)[\delta + (1-\delta)(1+r)] + (1-\alpha)U_s(\theta)[\beta\delta + (1-\delta)] > 0$. Second, consider the case where $\theta < \underline{\theta}$. In this case, $p(\theta) = 1$ and $f(\theta) = \theta$, so $df(\theta)/d\theta = 1$ and $dp(\theta)/d\theta = dU_s(\theta)/d\theta = 0$. Thus, we want to show that $\mathbb{X} = u_b(\theta) - \theta(du_b(\theta)/d\theta) > 0$. From above $du_b(\theta)/d\theta = \delta[\Delta^{-1} - \beta]$, which yields $\mathbb{X} = \delta\beta + (1-\delta) > 0$. Finally, consider the case where $\theta > \overline{\theta}$. In this case, $df(\theta)/d\theta = du_b(\theta)/d\theta = 0$ and $p(\theta) = \theta^{-1}$. Thus, we want to show that $\mathbb{X} = U_s(\theta) + \theta(dU_s(\theta)/d\theta) > 0$. From above, $\theta dU_s(\theta)/d\theta = -\theta^{-1}(1-\delta)(1+r)[\Delta - (1+r)]$, which yields $\mathbb{X} = \delta(1+r) + (1-\delta)(1+r)^2 > 0$. ■

Proof of Proposition 3: We want to show that if the private equilibrium is constrained efficient, then $(\alpha, \psi, r) \in \emptyset$, a set of measure zero. Suppose that the private equilibrium $(l_0^*, \bar{\omega}^*, \theta^*, q_1^*)$

is constrained efficient, then from equations (8) and (14), $d\bar{\omega}^{pe}(l_0^*)/dl_0 = d\bar{\omega}^{sp}(l_0^*)/dl_0$. Since $(l_0^*, \bar{\omega}^*, \theta^*, q_1^*)$ is a private equilibrium the investor break-even condition (6) holds, i.e., $U_s = U_b$, and we get that $(\partial R^b/\partial l_0)/(\partial R^b/\partial\bar{\omega}) = [b_0^{pe}u_b(\partial R^b/\partial l_0) + (\partial\mathbb{U}/\partial\theta)(\partial\theta/\partial l_0)]/[b_0^{pe}u_b(\partial R^b/\partial\bar{\omega}) + (\partial\mathbb{U}/\partial\theta)(\partial\theta/\partial\bar{\omega})]$, which is the case *iff* $(\partial\mathbb{U}/\partial\theta)[(\partial\theta/\partial l_0)(\partial R^b/\partial\bar{\omega}) - (\partial\theta/\partial\bar{\omega})(\partial R^b/\partial l_0)] = 0$, where all expressions are evaluated at the private equilibrium allocation.

Note that $(\partial\theta/\partial l_0)(\partial R^b/\partial\bar{\omega}) - (\partial\theta/\partial\bar{\omega})(\partial R^b/\partial l_0) < 0$: from equation (9) $\partial\theta/\partial l_0, \partial\theta/\partial\bar{\omega} < 0$; $\partial R^b/\partial l_0 = -R^b/\,(l_0(l_0-1)) < 0$; and following Part 1 of Theorem 1 $\partial R^b/\partial\bar{\omega} = R^b[\Gamma' - \mu G']/\,[\Gamma - \mu G] > 0$. Then, if the private equilibrium is constrained efficient we have that $\partial\mathbb{U}/\partial\theta = 0$, or, after some algebra and substituting the definition of θ^* from equation (9), $\alpha[R^b - q_1(1+r)] = (1-\alpha)[q_1 - \beta R^b]$, or equivalently, $\alpha[\Delta - (1+r)] = (1-\alpha)[1-\beta\Delta]$. Using that $\Delta^{-1} = \psi/(1+r) + (1-\psi)\beta$, we get that when the private equilibrium is constrained efficient then $\psi(1+\alpha r) = \alpha(1+r)$. The set of (α, ψ, r) satisfying this condition is, thus, of measure zero. ■

Proof of Proposition 4: Our proof proceeds in four parts.

Part 1: The sign of the marginal social gains to increasing leverage l_0 from the private equilibrium l_0^* depends on the sign of $d\bar{\omega}^{sp}(l_0^*)/dl_0 - d\bar{\omega}^{pe}(l_0^*)/dl_0$. Recall that $\mathcal{L}^{sp}(l_0)$ denotes the Lagrangian of the planner's problem. Differentiating, evaluating at l_0^*, and using (8), we get $d\mathcal{L}^{sp}(l_0^*)/dl_0 = -\Gamma'(\bar{\omega}^*)l_0^*R^k[d\bar{\omega}^{sp}(l_0^*)/dl_0 - d\bar{\omega}^{pe}(l_0^*)/dl_0]$, with $\Gamma' > 0$. So the planner wants to decrease leverage at the private equilibrium allocation, when $d\bar{\omega}^{sp}(l_0^*)/dl_0 - d\bar{\omega}^{pe}(l_0^*)/dl_0 > 0$.

Part 2: Show that $d\bar{\omega}^{sp}(l_0^*)/dl_0 - d\bar{\omega}^{pe}(l_0^*)/dl_0$ is proportional to $\partial\mathbb{U}(l_0^*, \bar{\omega}^*, \theta^*, q_1^*)/\partial\theta$. Note that the private equilibrium lies in the Pareto improvement constraint, with the price and the liquidity in the secondary market liquidity given by (3) and (9), respectively. So $\bar{\omega}^{sp}(l_0^*) = \bar{\omega}^{pe}(l_0^*)$. Then, when evaluating at l_0^* we can cancel the term $n_0(u_bR^b - U_s)$ in (13) and, after some algebra, we obtain

$$\frac{d\bar{\omega}^{sp}(l_0^*)}{dl_0} - \frac{d\bar{\omega}^{pe}(l_0^*)}{dl_0} = \frac{\frac{\partial\mathbb{U}}{\partial\theta}\left[\frac{\partial\theta}{\partial\bar{\omega}}\frac{\partial R^b}{\partial l_0} - \frac{\partial\theta}{\partial l_0}\frac{\partial R^b}{\partial\bar{\omega}}\right]}{\frac{\partial R^b}{\partial\bar{\omega}}\left[b_0^*u_b\frac{\partial R^b}{\partial\bar{\omega}} + \frac{\partial\mathbb{U}}{\partial\theta}\frac{\partial\theta}{\partial\bar{\omega}}\right]}, \tag{A.5}$$

where all the derivatives on the RHS are evaluated at $(l_0^*, \bar{\omega}^*, \theta^*, q_1^*)$. In the proof of Proposition 3 we noted that $(\partial\theta/\partial\bar{\omega})(\partial R^b/\partial l_0) - (\partial\theta/\partial l_0)(\partial R^b/\partial\bar{\omega}) > 0$, and $\partial R^b/\partial\bar{\omega} > 0$ in a non-rationed equilibrium (Part 1 of Theorem 1). Then, it remains to show that $b_0^*u_b\partial R^b/\partial\bar{\omega} + (\partial\mathbb{U}/\partial\theta)(\partial\theta/\partial\bar{\omega}) > 0$. Using that $s_0^* = e_0 - b_0^*$, $\partial\theta/\partial\bar{\omega} = -(\theta/R^b)(\partial R^b/\partial\bar{\omega})$, and $U_s(\theta^*) = u_b(\theta^*)R^b(l_0^*, \bar{\omega}^*)$, we get that

$$b_0^*u_b\frac{\partial R^b}{\partial\bar{\omega}} + \frac{\partial\mathbb{U}}{\partial\theta}\frac{\partial\theta}{\partial\bar{\omega}} = \frac{b_0^*}{U_s}\left[U_su_b + \theta^*\frac{dU_s}{d\theta}u_b - \theta^*U_s\frac{du_b}{d\theta}\right] + e_0\frac{dU_s}{d\theta}\frac{\partial\theta}{\partial\bar{\omega}} > 0\,,$$

where the inequality follows from Lemma 1, where we established that $\mathbb{X} = U_su_b + \theta u_b(dU_s/d\theta) - \theta U_s(du_b/d\theta) > 0$; and from $dU_s/d\theta, \partial\theta/\partial\bar{\omega} < 0$.

Part 3: Show that the sign of $\partial\mathbb{U}/\partial\theta$ equals the sign of $\psi - \alpha(1+r)/(1+\alpha r)$. From Proposition 3 we know that when $\psi - \alpha(1+r)/(1+\alpha r) = 0$, then $\partial\mathbb{U}/\partial\theta = 0$. In addition, following the steps in the proof of Proposition 3, it is easy to show that $\psi - \alpha(1+r)/(1+\alpha r) > 0 \Leftrightarrow \partial\mathbb{U}/\partial\theta > 0$ and $\psi - \alpha(1+r)/(1+\alpha r) < 0 \Leftrightarrow \partial\mathbb{U}/\partial\theta < 0$.

Part 4: Characterize social optimum conditional on the sign of $\psi - \alpha(1+r)/(1+\alpha r)$. Suppose $\psi - \alpha(1+r)/(1+\alpha r) > 0$, then from Part 3 we have that $\partial\mathbb{U}/\partial\theta > 0$, and from Part 2, it follows that $d\bar{\omega}^{sp}(l_0^*)/dl_0 - d\bar{\omega}^{pe}(l_0^*)/dl_0 > 0$. Then, from Part 1 we conclude that $d\mathcal{L}^{sp}(l_0^*)/dl_0 < 0$, so the planner wants to reduce leverage relative to the private equilibrium allocation, l_0^*. By continuity of the derivative of the Lagrangian, the planner continues to reduce leverage until reaching the constrained efficient allocation l_0, so we conclude that $l_0 < l_0^*$. It follows that $d\mathcal{L}^{sp}/dl_0 \leq 0$ in $[l_0, l_0^*]$, so $d\bar{\omega}^{sp}/dl_0 \geq 0$ in $[l_0, l_0^*]$, with strict inequality at l_0^*. The Fundamental Theorem of Calculus imply that $\bar{\omega} < \bar{\omega}^*$. Finally, we show that $\theta > \theta^*$. Since $\partial\theta/\partial l_0 < 0$, then $\theta(l_0, \bar{\omega}^*) > \theta(l_0^*, \bar{\omega}^*) = \theta^*$, and since $\partial R^b/\partial\bar{\omega} > 0$, then $\theta = \theta(l_0, \bar{\omega}) > \theta(l_0, \bar{\omega}^*)$.

Similarly, if $\psi < \alpha(1+r)/(1+\alpha r)$, we conclude that $l_0 > l_0^*$, $\bar{\omega} > \bar{\omega}^*$, and $\theta < \theta^*$. ■

Proof of Proposition 5: First, we show that the proposed taxes implement the constrained efficient allocation. The firm's problem with taxes on storage and leverage can be written as $\max_{l_0,\bar{\omega}} [1 - \Gamma(\bar{\omega})]R^k l_0 - \tau^l l_0 + T^l$ subject to $(1-\tau^s)U_s = U_b = u_b R^b$. This last equation corresponds to the investors' break-even condition (IBEC) with a tax on storage, and it implicitly defines a function from leverage l_0 to risk $\bar{\omega}$. Since the firm takes as given secondary market liquidity and the tax on storage, it follows that the derivative of this function corresponds to $d\bar{\omega}^{pe}/dl_0$. Using this implicit function we can write the Lagrangian as a function of only l_0 and obtain the following optimality condition $[1-\Gamma(\bar{\omega})]R^k = \tau^l + \Gamma'(\bar{\omega})R^k l_0(d\bar{\omega}^{pe}/dl_0)$. Since in the constrained efficient equilibrium, optimality is given by equation (14) we have that $[1-\Gamma(\bar{\omega})] = \Gamma'(\bar{\omega})l_0(d\bar{\omega}^{sp}/dl_0)$. It follows that τ^l is given by equation (18). In addition, combining the original IBEC, the IBEC with taxes, and constraint (12) we derive the tax on storage shown in equation (18).

Now we sign the taxes conditional on the value of ψ relative to the intertemporal Hosios condition (15). If $\psi > \alpha(1+r)/(1+\alpha r)$ then from Proposition 4 the planner wants to increase secondary market liquidity so $\theta > \theta^*$. Thus, since $dU_s/d\theta < 0$, then the storage technology is subsidized, $\tau^s < 0$. To sign the tax on leverage we use that $\Gamma'(\bar{\omega})' > 0$ and the characterization of the sign of $d\bar{\omega}^{sp}/dl_0 - d\bar{\omega}^{pe}/dl_0$ conditional on the sign of $\psi - \alpha(1+r)/(1+\alpha r)$ derived in the proof of Proposition 4. In fact, in that proof we established that when $\psi > \alpha(1+r)/(1+\alpha r)$, $\partial\mathbb{U}/\partial\theta > 0$ and $d\bar{\omega}^{sp}/dl_0 - d\bar{\omega}^{pe}/dl_0 > 0$, so we conclude that the leverage is taxed when liquidity is under provided. Similarly, when $\psi < \alpha(1+r)/(1+\alpha r)$ we conclude that storage is taxed and leverage subsidized. Finally, when the intertemporal Hosios condition holds, $\psi = \alpha(1+r)/(1+\alpha r)$, $\theta = \theta^*$ and $d\bar{\omega}^{sp}/dl_0 = d\bar{\omega}^{pe}/dl_0$, so $\tau^s = \tau^l = 0$. ■

Proof of Proposition 6: The proof proceeds in two parts.

Part 1: Direct Effect on Market Liquidity. For a feasible QE program $\bar{b}_0$ and hold-to-maturity return R^b, firms' borrowing is given by $b_0(R^b) = n_0(l_0(R^b)-1)$, whereas investors' final bond holdings are given by $b_0(R^b) - \bar{b}_0$. Then from the budget constraint of entrepreneurs, $b_0(R^b) - \bar{b}_0 = n_0(l_0(R^b) - 1 - \bar{b}_0/n_0)$. In addition, from the investors' budget constraint, $s_0(R^b) = n_0(e_0/n_0 - (l_0(R^b)-1))$. Note that the size of the QE program does not affect the amount ultimately invested in storage, as the bonds the central bank purchases are offset with the reserves it takes from investors. Finally, from the CB's budget constraint we have that $\bar{s}_0 = \bar{b}_0$. Using the previous expressions we can express secondary market liquidity in terms of entrepreneurs leverage and QE, conditional on the interest on reserves relative to the return on the OTC market.

First, note that the number of sell orders is always equal to $A = \delta(b_0(R^b) - \bar{b}_0)$, as impatient investors will put all their bond holdings for sale in the OTC market. In the case when $\Delta > 1 + \bar{r}$, patient investors pledge all their liquid wealth to place buy orders in the OTC market so $B = (1-\delta)[(1+r)s_0(R^b) + (1+\bar{r})\bar{s}_0]/q_1$ and market liquidity $\theta(R^b)$ is given by (21), with $\partial\theta/\partial\bar{b}_0 = [(1-\delta)\Delta(1+\bar{r}) + \theta]\,/[\delta R^b(n_0(l_0(R^b)-1) - \bar{b}_0)] > 0$. In the case when $1+\bar{r} > \Delta$, patient investors place buy orders in the OTC market only using the liquid assets they hold after funding the reserves liquidated by impatient investors, so the number of buy orders $B = (1-\delta)((1+r)s_0(R^b) - \delta/(1-\delta)(1+\bar{r})\bar{s}_0)/q_1$ and market liquidity is given by $\theta(R^b) = ((1-\delta)\Delta((1+r)(e_0 - n_0(l_0(R^b)-1)) - \delta/(1-\delta)(1+\bar{r})\bar{b}_0))/(\delta R^b(n_0(l_0(R^b)-1) - \bar{b}_0))$, with $\partial\theta/\partial\bar{b}_0 = ((1-\delta)\Delta(1+r)[e_0 - n_0(l_0(R^b)-1)(1+\delta/(1-\delta)(1+\bar{r})/(1+r))])/(\delta R^b(n_0(l_0(R^b)-1) - \bar{b}_0)^2) > 0$, where the inequality follows from Assumption 4. Thus, in both cases $\partial\theta/\partial\bar{b}_0 > 0$.

Part 2: Equilibrium Effects. For a feasible QE program $\bar{b}_0$ we can express the equilibrium of the model as $J(R^b, \bar{b}_0) - R^b = 0$, where the function J is defined as in the proof of Theorem 1, using $\theta(R^b)$ derived in Part 1. By the Implicit Function Theorem, if the derivative of $J(R^b, \bar{b}_0) - R^b$ with respect to R^b is different than 0, then we can define $R^b(\bar{b}_0)$ and calculate its derivative. Recall that $J(R^b) = U_s(\theta(R^b))/u_b(\theta(R^b))$, then $dJ/dR^b = (J/\theta)\varepsilon_{\Phi^\ell,\theta}(d\theta/dR^b) < 0$, where $\varepsilon_{\Phi^\ell,\theta}$ the elasticity of the liquidity premium with respect to market liquidity. The inequality follows from $d\theta/dR^b > 0$ (equation (A.3)), Lemma 1, where we showed that the elasticity is negative, and Proposition 1, where we showed that the elasticity is different than zero.

Then, by the Implicit Function Theorem $dR^b/d\bar{b}_0 = -\left[\partial J/\partial R^b - 1\right]^{-1} \partial J/\partial\bar{b}_0$, i.e., the sign of $dR^b/d\bar{b}_0$ equals the sign of $\partial J/\partial\bar{b}_0$. Note that $\partial J/\partial\bar{b}_0 = dJ/d\bar{b}_0|_{R^b}$ and thus have the same sign. Moreover, $J = \Phi^\ell(1+r)^2$ so the sign of $dJ/d\bar{b}_0|_{R^b}$ and the sign of $d\Phi^\ell/d\bar{b}_0|_{R^b}$ are equal. Since U_s and u_b do not directly depend on the QE program, $\partial\Phi^\ell/\partial\bar{b}_0 = 0$, and $d\Phi^\ell/d\bar{b}_0|_{R^b} = \partial\Phi^\ell/\partial\bar{b}_0 + (\partial\Phi^\ell/\partial\theta)(d\theta/d\bar{b}_0|_{R^b}) = (\partial\Phi^\ell/\partial\theta)(d\theta/d\bar{b}_0|_{R^b}) < 0$, where the inequality follows from Lemma 1 and Part 1. So we conclude that $dR^b/d\bar{b}_0 < 0$. That is, QE lowers the equilibrium hold-to-maturity return, and thus, it increases firm leverage and investment. ■

Proof of Proposition 7: First, we want to show that QE only improved the constrained efficient allocation when $\psi > \alpha(1+r)/(1+\alpha r)$ and $R^b(l_0^{**}, \bar{\omega}^{**}) > (1+r)^2 + L(l_0^{**}, \bar{\omega}^{**})$. For the central bank, the constraints imposed by the period 0 budget constraint ($\bar{s}_0 = \bar{b}_0$), by the definition of secondary market liquidity (21), and by the pricing protocol in the secondary market (3), together define the function $\theta(l_0, \bar{\omega}, \bar{b}_0, \bar{r})$. Using this function, the investors' Pareto constraint, implicitly defines a function $\bar{\omega}^{cb} \equiv \bar{\omega}^{cb}(l_0, \bar{b}_0, \bar{r})$. We use this implicit function to write the lagrangian of the CB's problem as $\mathcal{L} = [1 - \Gamma(\bar{\omega}^{cb})] l_0 R^k - \gamma \left[(1+\bar{r})^2 + L(l_0, \bar{\omega}^{cb}) - R^b(l_0, \bar{\omega}^{cb})\right] - \nu(r - \bar{r}) + \eta \bar{b}_0$.

An optimal allocation for the CB needs to satisfy the following first-order conditions with respect to l_0, $\bar{b}_0$ and $\bar{r}$: $\partial\mathcal{L}/\partial l_0 = \left[1 - \Gamma(\bar{\omega}^{cb})\right] R^k - \Gamma'(\bar{\omega}^{cb}) l_0 R^k \left(\partial\bar{\omega}^{cb}/\partial l_0\right) - \gamma\left[\partial L/\partial l_0 - \partial R^b/\partial l_0 - \left(\partial R^b/\partial\bar{\omega} - \partial L/\partial\bar{\omega}\right)\left(\partial\bar{\omega}^{cb}/\partial l_0\right)\right] = 0$, $\partial\mathcal{L}/\partial\bar{b}_0 = -\Gamma'(\bar{\omega}^{cb}) l_0 R^k \left(\partial\bar{\omega}^{cb}/\partial\bar{b}_0\right) - \gamma\left[\partial L/\partial\bar{\omega} - \partial R^b/\partial\bar{\omega}\right]\left(\partial\bar{\omega}^{cb}/\partial\bar{b}_0\right) + \eta = 0$, and $\partial\mathcal{L}/\partial\bar{r} = -\Gamma'(\bar{\omega}^{cb}) l_0 R^k \left(\partial\bar{\omega}^{cb}/\partial\bar{r}\right) - \gamma\left[2(1+\bar{r}) + \left(\partial L/\partial\bar{\omega} - \partial R^b/\partial\bar{\omega}\right)\left(\partial\bar{\omega}^{cb}/\partial\bar{r}\right)\right] + \nu = 0$. To characterize the role of QE we evaluate these first-order conditions at the constrained efficient allocation $(l_0^{**}, \bar{\omega}^{**}, \theta^{**}, 0, r)$. By assumption, the no loss constraint is slack, so $\gamma = 0$. In addition, when $\bar{b}_0 = 0$, the CB Pareto improvement constraint corresponds to the Pareto improvement constraint of section 4, so, $\bar{\omega}^{cb}(l_0, 0, r) = \bar{\omega}^{sp}(l_0)$. Using these the first-order condition with respect to l_0 yields $[1 - \Gamma(\bar{\omega}^{sp})] R^k = \Gamma'(\bar{\omega}^{sp}) \partial\bar{\omega}^{sp}/\partial l_0$, which is satisfied at the constrained efficient allocation.

Similarly, the first-order condition with respect to $\bar{r}$ evaluated at the constrained efficient allocation gives $\Gamma'(\bar{\omega}^{cb}) l_0 R^k \left(\partial\bar{\omega}^{cb}/\partial\bar{r}\right) = \nu$. In addition, from the definition of $\bar{\omega}^{cb}$ it follows that $\partial\bar{\omega}^{cb}/\partial\bar{r} = \left[-(\partial\mathbb{U}/\partial\theta)(\partial\theta/\partial\bar{r}) - \bar{b}_0 (\partial U_{\bar{s}}/\partial\bar{r})\right] \Big/ \left[(\partial\mathbb{U}/\partial\theta)(\partial\theta/\partial\bar{\omega}) + (b_0^{**} - \bar{b}_0) u_b \left(\partial R^b/\partial\bar{\omega}\right)\right] = 0$. The last equality obtains evaluating at the constrained efficient allocation, where $\bar{b}_0 = 0$ and $\partial\theta/\partial\bar{r} = 0$. Then, we conclude that $\nu = 0$; that is, the constraint $\bar{r} \geq r$ is slack.

Next, we evaluate the derivative of the lagrangian with respect to $\bar{b}_0$ at the constrained efficient allocation to obtain $\partial\mathcal{L}/\partial\bar{b}_0 = -\Gamma'(\bar{\omega}^{**}) l_0^{**} R^k (\partial\bar{\omega}^{cb}/\partial\bar{b}_0) + \eta$. From the definition of $\bar{\omega}^{cb}$ it follows that $\partial\bar{\omega}^{cb}/\partial\bar{b}_0 = \left[u_b R^b - U_{\bar{s}} - (\partial\mathbb{U}/\partial\theta)\left(\partial\theta/\partial\bar{b}_0\right)\right] \Big/ \left[(\partial\mathbb{U}/\partial\theta)(\partial\theta/\partial\bar{\omega}) + (b_0^{**} - \bar{b}_0) u_b \left(\partial R^b/\partial\bar{\omega}\right)\right]$. At the constrained efficient allocation $U_{\bar{s}} = U_s = u_b R^b$ and $\bar{b}_0 = 0$. In addition, from Proposition 6 $\partial\theta/\partial\bar{b}_0 > 0$, from equation (9) $\partial\theta/\partial\bar{\omega} < 0$, and from Part 1 of the proof of Theorem 1 $\partial R^b/\partial\bar{\omega} > 0$. Moreover, from equation (21) it follows that $\partial\theta/\partial\bar{\omega} = -\theta/R^b \left(\partial R^b/\partial\bar{\omega}\right)$, so from Part 2 of the proof of Proposition 4 we conclude that $(\partial\mathbb{U}/\partial\theta)(\partial\theta/\partial\bar{\omega}) + b_0^{**} u_b \left(\partial R^b/\partial\bar{\omega}\right) > 0$. Then, if $\psi > \alpha(1+r)/(1+\alpha r)$ we have that $\partial\mathbb{U}/\partial\theta > 0$ and $\partial\mathcal{L}/\partial\bar{b}_0 > 0$. So bond purchases, or QE, improve upon the constrained efficient allocation.

To establish that if $\bar{b}_0 > 0$ then $\bar{r} > r$, we evaluate the first-order condition with respect to $\bar{r}$ when $\bar{b}_0 > 0$ and $\bar{r} = r$. Note that in this case where $1 + \bar{r} < \Delta$ we have that $\partial U_{\bar{s}}/\partial\bar{r} = \{\delta + (1-\delta)[p\Delta + (1-p)(1+r) + (\bar{r} - r)/(1-\delta)] + 1 + r\} > 0$ and $\partial\theta/\partial\bar{r} = (1-\delta)\Delta\bar{b}_0/(\delta[n_0(l_0 - 1) - \bar{b}_0]R^b) > 0$. Above we established that $(\partial\mathbb{U}/\partial\theta)(\partial\theta/\partial\bar{\omega}) + b_0^{**} u_b \left(\partial R^b/\partial\bar{\omega}\right) > 0$, so we conclude that $\partial\bar{\omega}^{cb}/\partial\bar{r} < 0$ and $\partial\mathcal{L}/\partial\bar{r} > 0$. Thus, the CB will set the interest on reserves strictly above the interest on storage, i.e., $\bar{r} > r$.

Following the previous steps it follows that if $\psi \leq \alpha(1+r)/(1+\alpha r)$ and $R^b(l_0^{**}, \bar{\omega}^{**}) > (1+r)^2 + L(l_0^{**}, \bar{\omega}^{**})$, then $\partial\mathcal{L}/\partial\bar{b}_0 < 0$. So QE worsens the constrained efficient allocation.

Finally, when the CB no loss constraint is violated at the constrained efficient allocation, i.e., $R^b(l_0^{**}, \bar{\omega}^{**}) > (1+r)^2 + L(l_0^{**}, \bar{\omega}^{**})$, then the CB cannot engage in bond purchases, so $\bar{b}_0 = 0$. ■

References

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

Allen, F. and D. Gale (1994), 'Limited market participation and volatility of asset prices', *American Economic Review* **84**(4), 933–955.

Allen, F. and D. Gale (2004), 'Financial intermediaries and markets', *Econometrica* **72**(4), 1023–1061.

Andrés, J., J. D. López-Salido and E. Nelson (2004), 'Tobin's imperfect asset substitution in optimizing general equilibrium', *Journal of Money, Credit and Banking* **36**(4), 665–690.

Atkeson, A. G., A. L. Eisfeldt and P.-O. Weill (2015), 'Entry and Exit in OTC Derivatives Markets', *Econometrica* **83**(6), 2231–2292.

Bao, J., J. Pan and J. Wang (2011), 'The illiquidity of corporate bonds', *The Journal of Finance* **66**(3), 911–946.

Bernanke, B. and M. Gertler (1989), 'Agency costs, net worth and business fluctuations', *American Economic Review* **79**(1), 14–31.

Bernanke, B., M. Gertler and S. Gilchrist (1999), 'The financial accelerator in a quantitative business cycle framework', *in Handbook of Macroeconomics. Volume 1C, ed. John B. Taylor and Michael Woodford* pp. 1341–1393.

Bethune, Z., B. Sultanum and N. Trachter (2018), 'Asset issuance in Over-the-Counter markets', *working paper* .

Bianchi, J. and S. Bigio (2014), 'Banks, liquidity management and monetary policy', *NBER Working Paper No. 20490* .

Bigio, S. (2015), 'Endogenous liquidity and the business cycle', *American Economic Review, forthcoming* **105**(6), 1883–1927.

Bruche, M. and A. Segura (2017), 'Debt maturity and the liquidity of secondary debt markets', *Journal of Financial Economics* **124**(3), 599–613.

Brunnermeier, M. K. and L. H. Pedersen (2009), ‘Market liquidity and funding liquidity’, *Review of Financial Studies* **22**(6), 2201–2238.

Caballero, R. J. and E. Farhi (2013), A model of the safe asset mechanism (sam): Safety traps and economic policy, Working Paper 18737, National Bureau of Economic Research.

Carlstrom, C. T., T. S. Fuerst and M. Paustian (2017), ‘Targeting long rates in a model with segmented markets’, *American Economic Journal: Macroeconomics* **9**(1), 205–42.

Carpenter, S., J. Ihrig, E. Klee, D. Quinn and A. Boote (2015), ‘The Federal Reserve’s balance sheet and earnings: A primer and projections’, *International Journal of Central Banking* **11**(2), 237–284.

Chen, H., R. Cui, Z. He and K. Milbradt (forthcoming), ‘Quantifying liquidity and default risks of corporate bonds over the business cycle’, *Review of Financial Studies* .

Chen, H., V. Cúrdia and A. Ferrero (2012), ‘The macroeconomic effects of large-scale asset purchase programmes*’, *The Economic Journal* **122**(564), F289–F315.

Christensen, J. H. E. and G. D. Rudebusch (2012), ‘The Response of Interest Rates to US and UK Quantitative Easing’, *The Economic Journal* **122**(564), F385–F414.

Christensen, J. H. E. and S. Krogstrup (2016), Transmission of quantitative easing: The role of central bank reserves. Federal Reserve Bank of San Francisco Working Paper 2014-18.

Cui, W. and S. Radde (2019), ‘Search-based endogenous asset liquidity and the macroeconomy’, *Journal of the European Economic Association* .

Diamond, D. W. and P. H. Dybvig (1983), ‘Bank runs, deposit insurance, and liquidity’, *Journal of Political Economy* **91**(3), 401–419.

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

Edwards, A. K., L. E. Harris and M. S. Piwowar (2007), ‘Corporate bond markets transaction costs and transparency’, *The Journal of Finance* **62**(3), 1421–1451.

Eisfeldt, A. L. (2004), ‘Endogenous liquidity in asset markets’, *Journal of Finance* **59**(1), 1–30.

Farhi, E., M. Golosov and A. Tsyvinski (2009), ‘A theory of liquidity and regulation of financial intermediation’, *The Review of Economic Studies* **76**(3), 973–992.

Gagnon, J., M. Raskin, J. Remache, B. Sack et al. (2011), ‘The financial market effects of the federal reserve’s large-scale asset purchases’, *International Journal of Central Banking* **7**(1), 3–43.

Gale, D. and M. Hellwig (1985), ‘Incentive-compatible debt contracts: The one period problem’, *Review of Economic Studies* **52**(4), 647–663.

Geanakoplos, J. and H. Polemarchakis (1986), ‘Existence, regularity, and constrained suboptimality of competitive allocations when the asset market is incomplete’, *In W. Heller, R. Starr, and D. Starrett (eds.), Essays in Honor of Kenneth Arrow, Vol. 3. Cambridge University Press* pp. 65–95.

Geromichalos, A. and L. Herrenbrueck (2012), ‘Monetary policy, asset prices, and liquidity in over-the-counter markets’, *Working Papers, University of California, Department of Economics* .

Geromichalos, A. and L. Herrenbrueck (2016), ‘The strategic determination of the supply of liquid assets’, *Working Papers 161, University of California, Department of Economics* .

Gertler, M. and P. Karadi (2011), ‘A model of unconventional monetary policy’, *Journal of Monetary Economics* **58**(1), 17–34.

Gorton, G. B. and G. Pennacchi (1990), ‘Financial intermediaries and liquidity creation’, *Journal of Finance* **45**(1), 49–71.

Gromb, D. and D. Vayanos (2002), ‘Equilibrium and welfare in markets with financial constrained arbitrageurs’, *Journal of Financial Economics* **66**(2-3), 361–407.

Hart, O. and L. Zingales (2015), ‘Liquidity and inefficient investment’, *Journal of the European Economic Association, forthcoming* **13**(5), 737–769.

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

He, Z. and P. Kondor (2016), ‘Inefficient investment waves’, *Econometrica* **84**(2), 735–780.

Holmström, B. and J. Tirole (1998), ‘Private and public supply of liquidity’, *The Journal of Political Economy* **106**(1), 1–40.

Hosios, A. J. (1990), ‘On the efficiency of matching and related models of search and unemployment’, *The Review of Economic Studies* **57**(2), 279–298.

Jacklin, C. J. (1987), Demand deposits, trading restrictions, and risk sharing, *in* E. C.Prescott and N.Wallace, eds, ‘Contractual Arrangements for Intertemporal Trade’, Minneapolis: Univ. Minnesota Press, pp. 26–47.

Joyce, M., A. Lasaosa, I. Stevens, M. Tong et al. (2011), ‘The financial market impact of quantitative easing in the united kingdom’, *International Journal of Central Banking* **7**(3), 113–161.

Krishnamurthy, A. and A. Vissing-Jorgensen (2011), ‘The effects of quantitative easing on interest rates: channels and implications for policy’, *Brookings Papers on Economic Activity* .

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

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

Lagos, R. and S. Zhang (forthcoming), ‘Turnover liquidity and the transmission of monetary policy’, *American Economics Review* .

Lebeau, L. (2019), Payment frictions and participation in over-the-counter markets. mimeo.

Lester, B., G. Rocheteau and P.-O. Weill (2015), ‘Competing for Order Flow in OTC Markets’, *Journal of Money, Credit and Banking* **47**(S2), 77–126.

Lorenzoni, G. (2008), ‘Inefficient credit booms’, *Review of Economic Studies* **75**(3), 809–833.

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

Moen, E. R. (1997), ‘Competitive search equilibrium’, *Journal of Political Economy* **105**(2), 385–411.

Moreira, A. and A. Savov (2017), ‘The macroeconomics of shadow banking’, *The Journal of Finance* **72**(6), 2381–2432.

Rogerson, R., R. Shimer and R. Wright (2005), ‘Search-theoretic models of the labor supply market: A survey’, *Journal of Economic Literature* **43**(4), 959–988.

Shi, S. (1997), ‘A Divisible Search Model of Fiat Money’, *Econometrica* **65**(1), 75–102.

Shi, S. (2015), ‘Liquidity, assets and business cycles’, *Journal of Monetary Economics* **70**, 116 – 132.

Stein, J. (2014), ‘Incorporating financial stability considerations into a monetary policy framework’, *Speech at International Research Forum on Monetary Policy, Washington, D.C* .

Tobin, J. (1969), ‘A general equilibrium approach to monetary theory’, *Journal of Money, Credit and Banking* **1**(1), 15–29.

Townsend, R. M. (1979), ‘Optimal contracts and competitive markets with costly state verification’, *Journal of Economic Theory* **21**(2), pp. 265–293.

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

Williamson, S. D. (2012), ‘Liquidity, Monetary Policy, and the Financial Crisis: A New Monetarist Approach’, *American Economic Review* **102**(6), 2570–2605.