Finance and Economics Discussion Series
Divisions of Research & Statistics and Monetary Affairs
Federal Reserve Board, Washington, D.C.

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

David M. Arseneau, David E. Rappoport, and Alexandros P. Vardoulakis

2017-033

Please cite this paper as:
Arseneau, David M., David E. Rappoport, and Alexandros P. Vardoulakis (2017). "Private and Public Liquidity Provision in Over-the-Counter Markets," Finance and Economics Discussion Series 2017-033. Washington: Board of Governors of the Federal Reserve System, https://doi.org/10.17016/FEDS.2017.033.

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

David M. Arseneau
Federal Reserve Board

David E. Rappoport
Federal Reserve Board

Alexandros P. Vardoulakis
Federal Reserve Board

March 29, 2017

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 because investors and firms fail to internalize how their actions affect liquidity in secondary markets. 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 are grateful to Regis Breton, Max Bruche (discussant), Francesca Carapella, Giovanni Favara, Zhiguo He (discussant), Nobu Kiyotaki, Michal Kowalik (discussant), Konstantin Milbradt, Cecilia Parlatore, Lasse H. Pedersen, Skander Van den Heuvel, Chris Waller, 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, 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 (Holmstrom and Tirole, 1998). Alternatively, public liquidity provision is justified when liquidity shortages arise, for example in 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 the public provision (or withdrawal) of liquidity when liquid assets are abundant and the prospects of fire sales unlikely? This question is particularly important in the aftermath of the global financial crisis, where unconventional monetary policies such as quantitative easing (QE)—implemented by central banks *well after* the onset of the crisis, at a point when liquidity shortages had moderated—are thought to have implications for market liquidity (Krishnamurthy and Vissing-Jorgensen, 2011).

To articulate our argument we develop a dynamic model of market-based financial intermediation which features a two-way interaction between primary credit markets and secondary OTC markets. On the one hand, long-term bonds issued by firms in the primary market are retraded in an OTC market, thus secondary market liquidity affects investors' supply of credit to firms.[1] On the other hand, 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. It is this trade-off between credit provision and liquidity provision in OTC markets that is the novel feature of our analysis.

The key financial friction of the model is the presence of search frictions in the *secondary* OTC market. The importance of search frictions for OTC markets is grounded in both the empirical evidence, which suggests that they are the main driver of illiquidity in OTC markets for bonds (Edwards et al., 2007, and Bao et al., 2011), and the large theoretical literature modeling OTC markets with search frictions (Duffie et al., 2005, Lagos and Rocheteau, 2009, He and Milbradt, 2014, Atkenson et al., 2015, and others). In addition, we model the interaction of firms and investors in the *primary* credit market following 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'

[1] We focus on credit provision through capital markets, which has become increasingly important for non-financial firms and households in the U.S. The fraction of credit that is 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. This evidence is obtained from the Financial Accounts of the United States and considers the fraction of commercial paper, municipal securities and loans, and corporate bonds in total credit market instruments for non-financial firms as well as the fraction of mortgages and consumer credit that is securitized for households.

funding.[2] In our framework, the trade and agency frictions interact to determine the credit supply and market liquidity in equilibrium.

Our 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. Funding is obtained 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 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. Thus, our concept of market liquidity is one of market thickness. The liquidity of the OTC market introduces a liquidity premium in firms' external financing and, thus, affects the supply of credit through a *liquidity premium channel* that operates through the cost of credit.

However, our model features an additional interaction between credit and OTC markets. The quantity of bonds issued reduces the liquid resources in the financial sector, which can be used to provide liquidity in secondary markets. 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 shift toward illiquid bonds. Thus, bond issuance affects secondary market liquidity through a *liquidity provision channel*.

The liquidity premium channel and the liquidity provision channel generate an interplay between primary credit markets and secondary OTC markets, as illustrated in Figure 1. These channels work in opposite directions and distort private decisions. Firms issue more debt exactly when the liquidity premium is low, which is the case when market liquidity is high. But, as investors hold more of this debt, they shift their portfolios away from liquid assets, thereby reducing secondary market liquidity. In our existence proof we establish that the two effects jointly determine the unique equilibrium in the primary and secondary markets. 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. Hence, our mechanism is different from models which feature an amplification between funding and market liquidity stemming from binding collateral constraints and limits to arbitrage

[2]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. This is expected to be a good description of markets that support credit intermediation.

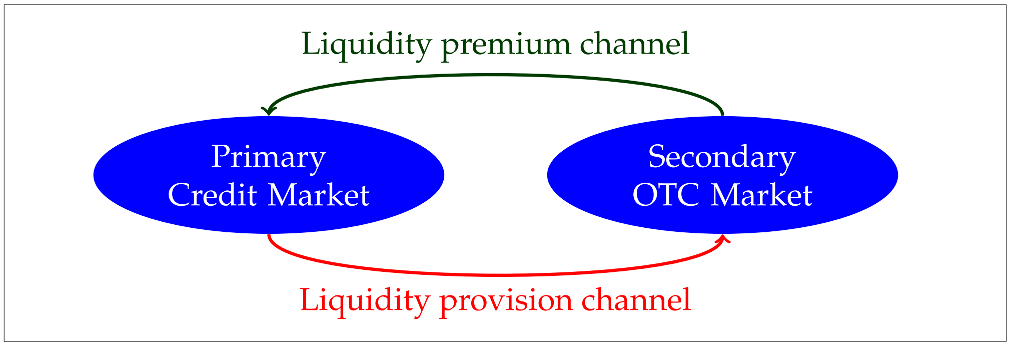


Figure 1: Feedback loop between primary and secondary market for corporate debt

(e.g., Brunnermeier and Pedersen, 2009, Gromb and Vayanos, 2002).

Moreover, the interaction between these two channels is a source of inefficiency. Search frictions and the dependence of market thickness on the abundance of liquid resources relative to the supply of bonds gives rise to a pecuniary externality for firms and a congestion externality for investors. The pecuniary externality works through the liquidity premium and reflects the fact that firms fail to internalize how their bond issuance affects their external financing costs. 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. 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 debt 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.

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.[3] These welfare gains allow the planner to reduce firms' funding costs and increase their profits despite having to operate on a smaller scale. In contrast, when private liquidity is inefficiently high a leverage subsidy combined with a storage tax are able to align the private and social incentives. Regardless of whether

[3]When private liquidity is inefficiently low (high), ex ante identical investors making the portfolio allocation decision are better off with higher (lower) liquidity because the gains (losses) to impatient investors outweigh the losses (gains) to patient investors.

private liquidity is inefficently high or low, optimal liquidity regulation allows the firm to internalize the pecuniary externality in a way that does not make investors worse off.

The interaction of the congestion and pecuniary externalities is critical for the efficacy of liquidity regulation. Indeed, the planner effectively exploits the congestion externality to create an additional surplus for investors that is then redistributed back to firms through more favorable funding costs. When the congestion externality is not present the planner has no means of achieving welfare-enhancing interventions. In this case, private liquidity coincides with its constrained-efficient level.

In addition to regulation, we also examine how the optimal management—provision or withdrawal—of public liquidity can alleviate trading frictions and improve economic efficiency. We focus on quantitative easing (QE) policies implemented by a central bank that uses liquid reserves to purchase less liquid assets from investors, yet quantitative tightening (QT) is expected to operate through the same mechanism. Through the lens of our model, any public policy that alters both public and private portfolios effectively shifts liquidity risk between the private and the public sector. This transfer of liquidity risk alters the liquidity premia which, in turn, influences savings and investment decisions in the real economy (see Stein, 2014, for a general discussion).

Our framework highlights the fact that public liquidity management 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. The difference in the two policies opens the door for them to coexist. Indeed, our analysis shows that either QE or QT should be supplemented with optimal regulation to generate even larger welfare gains. In this sense, liquidity regulation and liquidity management should be views as complements, not substitutes, in the policy toolkit.

Our analysis informs two related policy debates. On the one hand, despite their perceived efficacy the channels through which quantitative policies operate is still a matter of debate.[4] The two leading conceptual explanations are the signaling and the portfolio balance channels. The former considers that QE programs signal to investors that short-term rates will be lower in the future, which depresses long-term rates through the expectations channel. The latter considers that there is a downward slopping demand for the assets purchased by central banks, so these purchases contract the effective supply

[4]Gagnon et al. (2011) and Krishnamurthy and Vissing-Jorgensen (2011) present evidence of the efficacy of the program implemented by the Federal Reserve in reducing Treasury yields and mortgage rates. Similarly, Joyce et al. (2011) and Christensen and Rudebusch (2012) present evidence of the efficacy of the QE program implemented by the Bank of England.

of these assets, increasing their price and lowering their returns. This idea that can be traced back to Tobin (1969) can be rationalized if the market for reserves is segmented and the assets purchased by central banks are not perfectly substitutable (Christensen and Krogstrup, 2016). In contrast, our paper argues that the reason that the financial sector has a downward slopping demand for bonds (or an upward sloping supply of credit) is that changes in investors' portfolio composition affect the ease with which bonds can be retraded in secondary OTC markets.

On the other hand, the implementation of QE in the aftermath of the Great Recession has opened a debate about the optimal "exit strategy," i.e., what is the optimal strategy to unwind these quantitative policies. The Federal Open Market Committee (FOMC) has stated that it "is maintaining its existing policy of reinvesting principal payments from its holdings of agency debt and agency mortgage-backed securities [...] and of rolling over maturing Treasury securities [...], and it anticipates doing so until normalization of the level of the federal funds rate is well under way.[5]" Bernanke (2017) informally makes a case for such as strategy: given the uncertainty about the possible effects of a quantitative tightening (QT) program it seems prudent to increase the short-term policy rate in order to make room for monetary accommodation if needed. Our analysis provides an additional rationale for this strategy. Through the lens of our model, waiting to unwind the balance sheet until after interest rates have risen is adequate because optimal liquidity management calls for implementing QT to withdraw liquidity from OTC markets in this case.

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, 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 showing that, under the *same* financial frictions, private liquidity provision can be *either* suboptimally high or suboptimally low, depending on the conditions in the OTC market. This is an important result because these conditions are likely to vary over time, making the optimal liquidity management policy time varying.

Our paper is also related to the literature studying frictional OTC trade in financial

[5]See FOMC statement March 15, 2017. https://www.federalreserve.gov/newsevents/pressreleases/monetary20170315a.htm. The FOMC has laid out its plan for the exit strategy in the "Policy Normalization Principles and Plans" dated September 17, 2014. https://www.federalreserve.gov/newsevents/pressreleases/monetary20140917c.htm.

markets (Duffie et al., 2005, Lagos and Rocheteau, 2009, Geromichalos and Herrenbrueck, 2012, He and Milbradt, 2014, Atkenson et al., 2015). This literature has primarily focused on how search frictions matter for bid-ask spreads for assets traded in frictional markets. Some studies in this literature consider the role of endogenous market liquidity considering the cost of secondary market participation (Shi, 2015, Bruche and Segura, 2014, and Cui and Radde, 2015). Other studies consider the effect of monetary policy on trade frictions by changing agents' money holdings that are used in decentralized exchanges (Lagos and Wright, 2005, Lagos and Zhang, 2016, 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 market liquidity. Thus, we contribute to this literature by opening new avenues for research considering the interplay between liquidity provision and market liquidity in OTC markets.

Finally, our paper contributes to a recent literature that have studied the mechanism through which QE operates and its implications for the real economy. Gertler and Karadi (2011) study unconventional monetary policy in a new Keynesian model where financial intermediaries are credit constrained. The central bank, instead, is not credit constrained and thus can support credit extension when the intermediation capacity of private institutions is curtailed during episodes of stress. Farhi and Caballero (2013) show that QE can be effective when there is an excess demand for safe assets, as it substitutes risky with safe assets in private portfolios. However, exchanging long-term Treasury bonds with shorter maturity ones will not be effective in their model. Moreira and Savov (forthcoming) show a similar result in a model where the perceptions about the riskiness of assets used as collateral by private agents drives liquidity creation. Williamson (2012), within a new Monetarist model, argues that QE will affect the real economy only if it transfers credit risk from the private to the public sector. In contrast, our paper argues that QE affects the real economy by influencing the thickness of OTC markets, even when safe assets are abundant, there is no financial distress, or credit risk transfers between the private and public sector are not allowed.

The rest of the paper proceeds as follows. Section 2 presents our dynamic model of market-based financial intermediation and establishes the existence and uniqueness of the equilibrium. Section 3 describes the effect of secondary 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, and analytically describes the set of policy instruments that can implement the constrained efficient outcome. Section 5 analyses the effect of quantitative easing on secondary market liquidity and economic efficiency. Finally, section 6 concludes. All proofs are relegated to an Online Appendix.

2 A Dynamic 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 that 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 2).

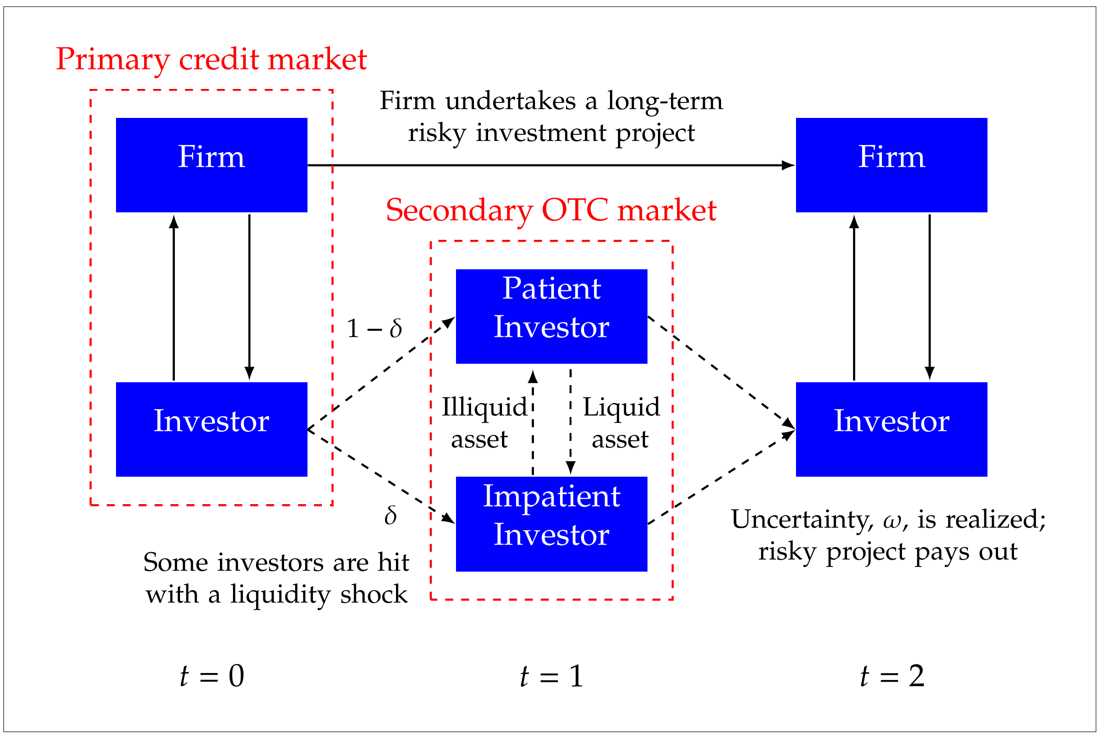


Figure 2: Timeline.

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 the return on their investment.[6] They operate a linear technology, which delivers $R^k\omega$ at $t = 2$, per unit of

[6] This is equivalent to maximize profits per unit of endowment, or the return on equity, as the endowment is fixed and is solely used for investment.

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, 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 the model. In order to produce, the firm must finance investment, denoted k_0, either through its own resources or by issuing financial contracts to investors. So, 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 will talk about entrepreneurs' projects 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 $t = 0$, and learn their preferences at $t = 1$. At $t = 1$ investors realize if they are *patient* or *impatient consumers*, a fraction $1 - \delta$ will turn out to be patient and a fraction δ impatient. Following Diamond and Dybvig (1983) the preference shocks are private information and are not contractible ex-ante.[7] 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 2). Patient investors that are able to acquire contracts in the OTC market realize an endogenous return Δ. But, as we show below this return equals an expression of exogenous parameters (see equation (8)).

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.[8]

Both the market for financial contracts, the primary market, and the OTC or secondary

[7] Financial intermediaries would be useful to improve allocations. Nevertheless, when retrading of financial claims is allowed in these models agents can self-insure against idiosyncratic liquidity risk (Jacklin 1987). But our secondary market is not Walrasian precluding the perfect self insurance. Moreover, we abstract from intermediaries to focus on the role of the OTC market.

[8] A long literature studies the effect of adverse selection in secondary trade as a source of illiquidity (Gorton and Pennacchi, 1990; Eisfeldt, 2004; Kurlat, 2013; Malherbe, 2014; Bigio, 2015). We have abstracted from informationally driven illiquidity to focus on the role of search frictions.

market for these contracts are described in detail below.[9]

In what follows we make the following assumptions.

Assumption 1 (Relative Returns) *The long-term return of the productive technology is larger than the cumulative two-period storage return and the return on storage plus the return on secondary markets, i.e.,* $(1+r)^2 < R^k$ *and* $(1+r)\Delta \leq R^k$*. 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) *Impatient investors discount future consumption at a higher rate than the return on storage,* $1/\beta - 1 \geq r$ *or* $\beta \leq 1/(1+r)$*, and the discounted expected return on firms' projects is larger than the return on storage* $1 + r \leq \beta R^k$.

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 hand, $R^k > (1+r)^2$, allows firms to offer a return that is higher than the cumulative two-period return on storage. On the other hand, $R^k \geq (1+r)\Delta$, allows firms to offer a return that is higher 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 ensures that there is no credit rationing in equilibrium. Assumption 3 makes impatient investors have a (weak) preference for current versus future consumption when the interest rate is r, $\beta \leq 1/(1+r)$, but not too impatient so that the return of entrepreneurs for an impatient investor is larger than the return on storage, $1 + r \leq \beta R^k$. 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

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 $t = 2$. 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

[9] Note that since $r > 0$ and since investors preferences have been assumed time separable and risk neutral, there was no loss of generality in abstracting away from consumption at $t = 0$ for investors, and consumption at $t = 1$ for patient investors.

assets.[10] 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 $\mathbb{E}_0 \max\left\{0, R^k \omega k_0 - Z b_0^d\right\}$. 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: $R^k \omega k_0 < Z b_0^d$. This condition defines a threshold productivity level, $\bar{\omega}$, such that the entrepreneur defaults when

$$\omega < \bar{\omega} = \frac{Z}{R^k} \frac{l_0 - 1}{l_0} \ . \tag{1}$$

The productivity threshold measures the credit risk of the financial contract, as it increases the firm's probability of default.[11]

For notational convenience, we define $G(\bar{\omega}) \equiv \int_0^{\bar{\omega}} \omega dF(\omega)$ and $\Gamma(\bar{\omega}) \equiv \bar{\omega}(1 - F(\bar{\omega})) + G(\bar{\omega})$. 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 profits per unit of endowment, can be expressed as $1/n_0\, \mathbb{E}_0 \max\{0, R^k \omega k_0 - Z b_0^d\} = [1 - \Gamma(\bar{\omega})]\, R^k l_0$.

Similarly, the expected gross return of holding a *single* bond to maturity R^b can be expressed as

$$R^b = \frac{1}{b_0} \left[\int_{\bar{\omega}}^{\infty} Z b_0^d dF(\omega) + (1 - \mu) \int_0^{\bar{\omega}} R^k \omega k_0 dF(\omega) \right] = \frac{l_0}{l_0 - 1} R^k \left[\Gamma(\bar{\omega}) - \mu G(\bar{\omega}) \right] \ , \tag{2}$$

which is a function of only leverage and the productivity threshold. Clearly R^b 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. Using this notation we can write down the firm's problem that defines the optimal contract, when the expected hold-to-maturity

[10] We consider deterministic monitoring rather than stochastic monitoring, which results in debt being the optimal contract. Krasa and Villamil (2000) derive the conditions under which deterministic monitoring occurs in equilibrium in costly enforcement models.

[11] Note that the productivity threshold is increasing in the spread between the promised return and the expected return on the entrepreneur investment, and it is increasing in the entrepreneur's leverage l_0.

return offered to investors equals R as

$$\max_{l_0,\bar{\omega}} [1 - \Gamma(\bar{\omega})] R^k l_0 \qquad \text{s.t.} \qquad R^b(l_0, \bar{\omega}) = R \,. \tag{3}$$

Note that this problem also defines the demand for credit in the primary market $b_0^d(R) = (l_0(R) - 1)n_0$, which is, as we show below, a strictly decreasing function of the expected hold-to-maturity return R. As it 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

The ex post heterogeneity introduced by the preference shock generates potential gains from trading financial contracts for consumption in the OTC market. Impatient investors want to exchange long-term bonds for consumption, as they would rather consume in period 1 than hold the bond to maturity until period 2 (Assumption 3). Patient investors are willing to take the other side of the trade if the return from buying bonds in the OTC market Δ is greater than the return on the storage technology $1 + r$.

To model the exchange in the OTC market we consider that each investor represents a large family of small *traders*. That is, each investor is comprised by a continuum of infinitesimal traders of mass e_0, where each trader has a portfolio restricted to one bond or q_1 units of consumption, which corresponds to the price of bonds in terms of consumption.[12] Traders are paired up according to a matching technology. Impatient investors send a mass of b_0^s traders to sell their bonds. Patient investors send a mass of $(1 + r)s_0/q_1$ traders to buy bonds in the OTC market. 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.

A key implication of the assumptions that trading is carried out by traders as opposed to investors and that traders meet only once with potential counterparties, is that the price in the OTC marker q_1 is independent of market thickness.[13] If we were to allow an effect on the secondary market price of market thickness there would be an additional mechanism

[12]That is, traders' portfolios are restricted as in Atkenson et al. (2015) or Bianchi and Bigio (2014). In our case, traders' portfolios are restricted to a single bond equivalent: one bond for sellers and q_1 units of consumption for buyers.

[13]See Bianchi and Biggio (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.

to rationalize the regulation and public provision of liquidity.

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 in each side of the market. And it is given by

$$m(A, B) = \min\left\{A, B, \nu A^{\alpha} B^{1-\alpha}\right\} , \tag{4}$$

with $0 < \nu$ a scaling constant and $0 < \alpha < 1$ the elasticity of the matching function 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 a concept of thickness 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, it is relatively easy for sellers to trade. But, at the same time, it is relatively hard for buyers to trade. 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

$$f(A, B) = \frac{m(A, B)}{A} \qquad \text{or} \qquad f(\theta) = m(1, \theta) , \tag{5}$$

and the *probability that a buy order is executed* is expressed as

$$p(A, B) = \frac{m(A, B)}{B} \qquad \text{or} \qquad p(\theta) = m(\theta^{-1}, 1) . \tag{6}$$

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

Once a buy order and a sell order are matched, the terms of trade are determined via a simple surplus sharing rule, determined by Nash bargaining and known by all agents. 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 receive an expected payoff βR in $t = 2$. Then, the surplus that accrues to an impatient investor is given by $S^I(q_1) = q_1 - \beta R$. Similarly, the value of a trading match to a buyer is the present value of the (expected) return on the bond, net of the price that needs to be paid for each bond in the secondary market, $S^P(q_1) = R/(1+r) - q_1$.

The price of the debt contract on the secondary market is determined by Nash bargaining, which maximizes the product of the respective surpluses, $\max_{q_1} (S^I(q_1))^{\psi} (S^P(q_1))^{1-\psi}$, where $\psi \in [0, 1]$ is the bargaining power of impatient investors.

The solution of the surplus splitting problem yields the following bond price in the secondary market

$$q_1 = R\left(\frac{\psi}{1+r} + (1-\psi)\beta\right) . \tag{7}$$

Note that $\psi = 1$ drives the price of the bond to the "ask" price, or the price that extracts full rent from the buyer, $q_1 = R/(1+r)$. By the same token, $\psi = 0$ drives the price of the bond to the "bid" price, or the price that extracts full rent from the seller, $q_1 = \beta R$. From equation (7) it follows that the return that patient investors make in the secondary market, per executed buy order, depends only on exogenous parameters and is given by

$$\Delta = \frac{R}{q_1} = \left(\frac{\psi}{1+r} + (1-\psi)\beta\right)^{-1} \geq 1 + r . \tag{8}$$

2.4 Investors and the Supply of Credit

As described above, investors are ex ante identical and are endowed with e_0 units of the consumption good. At $t = 0$ they can allocate their wealth across two assets: a storage technology s_0 and financial contracts b_0^s. Thus, their budget constraint is given by

$$s_0 + b_0^s = e_0 , \tag{9}$$

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, again paid in units of consumption. In this sense, storage is a liquid asset, as at any point in time it can be costlessly transformed into consumption. In contrast, bonds deliver an expected payoff R in $t = 2$ and are illiquid as an investor might be unable to turn her bond into consumption at $t = 1$ if her sell order does not find a match in the OTC market.

To describe the portfolio choice problem of investors, it is useful to first consider the optimal behavior of impatient and patient investors in $t = 1$ when they arrive in that period with a generic portfolio of storage and bonds (s_0, b_0^s).

2.4.1 Impatient Investors

By Assumption 3 at $t = 1$ impatient investors want to consume in the current period. 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

$$c_1^I = s_0(1+r) + f(\theta)q_1 b_0^s \ . \tag{10}$$

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

$$c_2^I = (1 - f(\theta))Rb_0^s \ , \tag{11}$$

with the utility derived from c_2^I discounted by β.

2.4.2 Patient Investors

Patient investors only value consumption in the final period and will be willing to place buy orders in the OTC market if there is a surplus to be made, i.e., if $q_1 \leq R/(1+r)$. The price determination in the OTC market guarantees that this is always the case $(1 + r \leq \Delta)$, 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 on average. So patient investors expect to increase their bond holding by $p(\theta)s_0(1+r)/q_1$ units. It follows that expected storage holdings at the end of $t = 1$, s_1^P, are 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 equals

$$c_2^P = (1 - p(\theta))s_0(1+r)^2 + \left[b_0^s + p(\theta)\frac{s_0(1+r)}{q_1}\right]R \ . \tag{12}$$

That is, the payout from consumption that was stored and not traded away in the OTC market plus the expected payout from bond holdings.

2.4.3 Optimal Portfolio Allocation

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 (9), and anticipating that optimal expected consumption of impatient and patient investors is given by equations (10)-(12).

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[14]

$$U_s(\theta) = \delta(1+r) + (1-\delta)\left[(1-p(\theta))(1+r)^2 + p(\theta)(1+r)\Delta\right] , \tag{13}$$

$$\text{and} \qquad U_b(R,\theta) = u_b(\theta)R , \tag{14}$$

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 make 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 and the expected loss due to the bond illiquidity $u_b(\theta)$.[15] By contrast, the utility of holding storage in period 0 only depends on market liquidity, through the probability that the return from providing liquidity can be realized $p(\theta)$.

Using these definitions, we can express the asset demand correspondence that maximizes the investors portfolio problem as

$$\begin{cases} b_0^s = e_0, & s_0 = 0 & \text{if } U_b > U_s \\ b_0^s \in [0, e_0], & s_0 = e_0 - b_0^s & \text{if } U_b = U_s \\ b_0^s = 0, & s_0 = e_0 & \text{if } U_b < U_s \end{cases}$$

That is, if the expected utility of holding bonds in period 0 is greater than the utility of holding storage in period 0—which incorporates the return that can be made by providing liquidity—then investors will demand only bonds in the initial period. On the contrary, if the expected utility of holding bonds is smaller than then expected benefit of holding storage in period 0, then investors will only hold storage in the initial period. Finally, if

[14] We are taking advantage of the result that Δ is a function of exogenous model parameters, but in general $\Delta(R, q_1) = R/q_1$.

[15] In equilibrium, this will introduce a dependence of the utility of holding a bond on the contract characteristics $(l_0, \bar{\omega})$, through the expected return on holding the bond to maturity R^b.

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 (9). In this case the *individual* credit supply is totally elastic when the expected hold-to-maturity return equals $U_s(\theta)/u_b(\theta)$.

Given our assumptions, in equilibrium, the portfolio allocation will be interior (i.e., $U_s = U_b$ with $s_0, b_0^s > 0$), thus we focus our analysis on this case.[16] For future reference we label this condition the *investors' break-even condition*.

$$U_s(\theta) = U_b(R, \theta) = u_b(\theta)R \ . \tag{15}$$

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 investors a hold-to-maturity return equal to $U_s(\theta)/u_b(\theta)$. Moreover, this condition describes the *aggregate* credit supply, $b_0^s(R)$. In fact, it is implicitly defined by $U_s(\theta(b_0^s, R)) = u_b(\theta(b_0^s, R))R$, where $\theta(b_0^s, R) = (1-\delta)(e_0 - b_0^s)(1+r)\Delta/(\delta b_0^s R)$. To be clear, our concept of aggregate credit supply is not just the sum of the individual investors' credit supply, but is one where the consistency of market thickness is taken into account, i.e., θ is a function of (b_0^s, R). As we show below, the aggregate credit supply is strictly increasing in the expected hold-to-maturity bond return.

2.5 Equilibrium

The equilibrium of the model is defined as follows.

Definition 1 (Private Equilibrium) *We say that $(l_0, \bar{\omega}, \theta, q_1)$ is a private equilibrium if and only if:*

1. *Given the outcome in the secondary market (θ, q_1), the financial contract is described by $(l_0, \bar{\omega})$ that maximizes entrepreneurs' return on investment subject to investors' break-even condition (15).*
2. *The credit market clears, i.e., $b_0^d(R) = b_0^s(R) \equiv b_0$, with $R = R^b(l_0, \bar{\omega})$ given by equation (2).*
3. *Market liquidity corresponds to $\theta = (1-\delta)(1+r)s_0/(q_1 \delta b_0^s)$.*
4. *q_1 is determined via the surplus sharing rule (7).*

[16] Note that the expected utility from investing in storage, U_s, is not smaller than the expected utility in financial autarky: $U_a = \delta(1+r) + (1-\delta)(1+r)^2$, since the return of buying a bond in the secondary market $\Delta \geq 1 + r$.

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

The equilibrium of the model is described by the entrepreneur's choice of leverage, l_0, and risk, $\bar{\omega}$, to maximize the payoff of the risky investment project. Entrepreneurs' profits are higher when l_0 is higher and when the promised payout is lower, that is, when $\bar{\omega}$ is lower. But firms are constrained in their choices of l_0 and $\bar{\omega}$ as they need to offer terms that make financial contracts attractive to investors: the investors' break-even condition. Firms are aware that when selling in the OTC market, investors obtain a price that depends on the contract characteristics and is determined via the sharing rule (equation 7). It follows that the firm's problem can be written as $\max_{l_0,\bar{\omega}} [1-\Gamma(\bar{\omega})]R^k l_0$, subject to the investors' break-even condition (15).

The entrepreneur's privately optimal choice of leverage trades off the marginal increase in profits from higher leverage against the marginal reduction in expected (hold-to-maturity) bond return for investors. Similarly, the privately optimal choice for the risk profile of the contract trades off the marginal increase in profits from lower risk against the marginal reduction in the expected (hold-to-maturity) bond return for investors.

These considerations can be summarized in the following optimality condition

$$\frac{1-\Gamma(\bar{\omega})}{\Gamma'(\bar{\omega})l_0} = -\frac{\partial R^b(l_0,\bar{\omega})/\partial l_0}{\partial R^b(l_0,\bar{\omega})/\partial\bar{\omega}} \,. \tag{16}$$

This equation, which describes the privately optimal debt contract, taken together with the investors' break-even condition (15), and the expressions that characterize the secondary market (θ, q_1) provide a complete description of the equilibrium of the model.

Finally, note that the expected hold-to-maturity return R^b, the price in the secondary market q_1, and secondary market liquidity θ all can be expressed as a function of the characteristics of the optimal financial contract $(l_0,\bar{\omega})$. In fact, the price is a function of the expected return on holding the bond to maturity, R^b, which depends on $(l_0,\bar{\omega})$; so we can write market liquidity as

$$\theta = \frac{(1-\delta)s_0(1+r)}{\delta b_0 q_1} = \frac{(1-\delta)(1+r)\Delta\left(e_0 - n_0(l_0-1)\right)}{\delta n_0(l_0-1)R^b(l_0,\bar{\omega})}. \tag{17}$$

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.*

The proof uses a fixed-point argument on a mapping that takes the expected bond return offered by firms and uses the equilibrium conditions to map it to the expected hold-to-maturity return required by investors. 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 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 slopping. In fact, the derivations in the proof of Theorem 1 allows to establish the following results that characterize the demand and supply for credit.

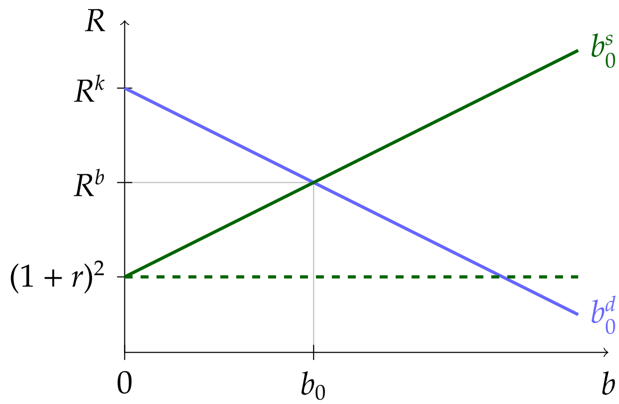
Corollary 1 (The Optimal Financial Contract and The Demand for Credit) *The characteristics of the optimal financial contract: leverage l_0 and risk $\bar{\omega}$, are strictly decreasing in the expected hold-to-maturity return R. That is, the demand for credit $b_0^d(R)$ is strictly decreasing.*

Corollary 1 follows from the characterization of the demand for credit in the proof of Theorem 1. 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, $b^s(R)$, 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 thickness need to drop so the return on storage increases and the expected loss from holding illiquid bonds decreases. Market thickness drops only if investors' portfolios

Figure 3: Aggregate Demand and Supply of Credit in the Primary Market.



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 generate an increasing aggregate supply of credit.

The previous results are useful to analyze the model as 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 3. This representation can be used to contrast our model with previous work. In the CSV literature it is typically assumed that aggregate credit supply is perfectly elastic at rate $(1 + r)^2$, e.g., Bernanke et al. 1999. This case is depicted by the green dashed line in Figure 3. In other models of OTC trade with search frictions, such as Duffie, Garleanu and Pederson (2005), where the trading probabilities and market thickness are functions of exogenous parameters, the aggregate credit supply will be totally elastic at some rate $R > (1 + r)^2$ and $R < R^k$.

3 OTC Trade, Bond Premia, and Credit Markets

This section defines the liquidity and default bond-premia and presents analytical characterizations for both. It then presents a frictionless benchmark that specifies the limiting cases where trade frictions in the OTC market become irrelevant. It continues to describe the relationship between the OTC market, i.e, the secondary market for bonds, and the credit market, i.e., the primary market for bonds. It concludes by presenting a numerical illustration of the model.

3.1 Analytical 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. Naturally, 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.[17] 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 benchmark rate, Z/R^{ℓ}. 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 given by $\Phi^t \equiv Z/R^{\ell} = \Phi^d\,\Phi^{\ell}$. 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

$$\Phi^d(\bar{\omega}) = \frac{\bar{\omega}}{\Gamma(\bar{\omega}) - \mu G(\bar{\omega})} \,. \tag{18}$$

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

[17] No arbitrage under perfectly liquid markets implies that trading a two-period bond should yield the same expected return for investors to rolling over one period safe investments, i.e. $\delta \cdot R^{\ell}/(1+r) + (1-\delta)\cdot R^{\ell} = \delta \cdot (1+r) + (1-\delta)\cdot(1+r)^2$.

is that leverage does not affect the default premium, as is the case in the benchmark CSV model, though leverage and risk are jointly determined in equilibrium. This is due to the fact that, for a fixed threshold level, $\bar{\omega}$, 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 (18)).

Moreover, from the definition of the liquidity premium and the investors' break-even condition (15), we have that

$$\Phi^\ell(\theta) = \frac{U_s(\theta)}{(1+r)^2 u_b(\theta)} \ . \tag{19}$$

That is, the liquidity premium equals the spread between the bond hold-to-maturity return, which equals the ratio between the expected utility from liquidity provision U_s and the expected utility loss due to bond illiquidity u_b, and the liquid two-period return $(1+r)^2$. Equation (19) provides an analytical characterization of the relationship between the liquidity premium and secondary market thickness θ, showing that the former is only a function of the latter. This analytical representation of the liquidity premium channel is key for our analysis. The following Lemma characterizes this channel.

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 (i.e., the liquidity premium) is inversely proportional to the amount of liquidity in secondary OTC markets. When secondary market liquidity, θ, is lower, investors require a higher liquidity premium, Φ^ℓ, or equivalently, a higher hold-to-maturity return, R^b. This relationship forms the basis for the liquidity premium channel shown by the arrow going from the secondary OTC market to the primary credit market in Figure 1. In our model, market liquidity determines the likelihood that investors' orders will be executed in an OTC trade. In particular, as the market becomes less liquid sell orders will be more difficult to execute ($f(\theta)$ decreases), and impatient investors will realize larger utility losses from the illiquidity of bonds ($u_b(\theta)$ decreases). By the same token, as liquidity declines buy orders are more likely to be executed ($p(\theta)$ increases), rising the expected private benefit from liquidity provision ($U_s(\theta)$ increases). Both of these channels lead to an increase in the expected hold-to-maturity bond return to keep investors indifferent between bonds and storage in period 0, that is, the liquidity premium increases.

3.2 A Frictionless Benchmark

The next proposition establishes the conditions under which trade in the OTC market is irrelevant, so that secondary market liquidity has no bearing on the equilibrium of the model.

Proposition 3 (Irrelevance of OTC Trade) *Under the following conditions, there is no liquidity premium, i.e., $\Phi^{\ell} = 1$, implying that the model collapses to the benchmark CSV model:*

1. *All investors are patient, so that $\delta = 0$; or*
2. *Impatient investors discount at rate $\beta = 1/(1 + r)$; or*
3. *Impatient investors are able to sell all their bonds in the secondary market at their reservation value, which is true when $\psi = 1$ and $e_0 \geq \bar{\bar{e}}_0$.*

The case in which $\delta = 0$ is straightforward. When all investors are patient, there is no need to trade in secondary markets; investors only care about the hold-to-maturity return. Liquidity is not priced in financial contracts and the model collapses to the standard CSV setup presented in, for example, Townsend (1979) and Bernanke and Gertler (1989).

The same result obtains for the second case, though for different reasons. When impatient investors discount future consumption at exactly the rate of return that comes from holding a unit of storage, so that $\beta = 1/(1 + r)$, they will be indifferent between consuming in the final or interim period. This indifference implies that there are no gains from OTC trade. In this case, the liquidity preference shock is immaterial and investors only consider the hold-to-maturity return when buying financial contracts in primary markets.

The third case considers the situation in which impatient investors can fully satisfy their liquidity needs in secondary markets, while there are no gains for patient investors from liquidity provision. That is, the terms of trade are set such that impatient investors extract the entire surplus, i.e., $\psi = 1$, and all their sell orders will be executed. Patient investors, then, earn only their opportunity cost from liquidity provision, i.e., $\Delta = (1 + r)^2$. In this case, as before, liquidity considerations will not factor in the lending decision of investors in primary markets. The condition that $f(\theta) = 1$ follows from $e_0 \geq \bar{\bar{e}}_0$. We derive this threshold for investors endowment in the proof of Proposition 3 in the Appendix. Intuitively, $f(\theta) = 1$, requires that there is enough storage at $t = 1$ that all sell orders can be satisfied, which requires that investors' endowment is sufficiently large.

It is worth mentioning that in principle there could be a fourth case where trade frictions are irrelevant: when patient investors can fully realize the gains from liquidity provision, i.e., $\psi = 0$ and $p(\theta) = 1$, while at the same time impatient investors are indifferent from holding bonds to maturity and trading in the OTC market, i.e., $\beta = 1/(1 + r)$. The first condition fixes the liquidity premium at $\beta^{-1}/(1 + r) \geq 1$. Thus, there will be no effect from market liquidity into the credit market. But trade frictions are irrelevant only when the liquidity premium collapses to 1, which is the case only when $\beta = 1/(1 + r)$. The latter condition is encompassed in case 2 in Proposition 3.[18]

3.3 Liquidity Premium and Liquidity Provision Channels

We now characterize the effects of frictional OTC secondary trade on primary credit markets. For the remainder of the paper, we consider only the cases in which trading frictions in the secondary market result in a non-negligible liquidity premium. That is, assume that (*i*) the probability of being an early consumer is positive, $\delta > 0$; (*ii*) impatient investors discount future consumption strictly more than what is implied by the storage rate, i.e., $\beta < 1/(1+r)$; and (*iii*) impatient investors cannot fully satisfy their liquidity needs in secondary markets, i.e., $\psi < 1$ or $e_0 < \bar{\bar{e}}_0$.

We are now ready to characterize the relationship between credit and OTC markets depicted in Figure 1. On the one hand, market thickness in the OTC market θ determines the liquidity premium that investors will require on illiquid bonds over liquid storage, $\Phi^{\ell}(\theta)$ (equation (19)). This is the *liquidity premium channel* that describes how market liquidity affects liquidity premia. Our model shows that the liquidity premium shapes the expected hold-to-maturity return R^b that firms need to offer investors and, thus, firms' demand for credit b_0^d. Therefore, market liquidity θ affects the equilibrium in the credit market (b_0, R^b).

On the other hand, the equilibrium in the credit market (b_0, R^b) will determine the secondary market liquidity, θ; this is the *liquidity provision channel* of credit markets into OTC markets. This channel is novel to the literature analyzing financial markets with trade frictions. In fact, to support the equilibrium level of bond issuance investors will have to hold those bonds in their portfolios and will reduce their holding of liquid storage, which is deployed in the OTC market to support market liquidity. In fact, equation (17) characterizes how OTC market thickness θ is a function of the volume and expected return of bonds in the credit market (b_0, R^b).

The liquidity premium effect and the liquidity provision effect work in opposite directions. Suppose there is an exogenous shock to market liquidity, then the liquidity

[18] Moreover, as we show in the proof of Proposition 1, Assumption 4 will rule out that liquidity can be arbitrarily low as to guarantee that patient investors can fully realize the gains from liquidity provision, i.e., $p(\theta) = 1$.

premium channel would increase the aggregate supply of credit. Firms respond to this shift in credit supply by issuing more bonds at lower expected (hold-to-maturity) returns. But the liquidity provision channel would yield a reduction in market liquidity as investors substitute liquid assets for bonds in their portfolio. This channel attenuates the initial increase in market liquidity. In our existence proof we establish that the two effects jointly determine the unique equilibrium in the primary and secondary markets. The direct effect dominates, while the indirect effect acts as an automatic stabilizer such that an improvement or a deterioration in market liquidity cannot perpetually increase or decrease bond issuance.

3.4 Comparative Statics

Now we describe the effect of the parameters that determine the demand and supply for credit for the equilibrium of the model. We begin by describing the effect of these parameters on the demand and supply for credit.

Proposition 4 (Aggregate Credit Supply Determinants) *Taking the expected hold-to-maturity return and bond issuance as given, investors require a higher liquidity premium, Φ^{ℓ}, and hence the market thickness θ is lower in the OTC market, when*

1. *(Preference shock) The probability of becoming impatient is higher, i.e., δ is higher; or*
2. *(Impatience) Impatient investors discount the future more heavily, i.e., β is lower; or*
3. *(Endowments) Investors have less to invest in storage, i.e., e_0 is lower.*

The proposition details how the parameters that describe investors' preferences (δ and β) and endowments (e_0) affect the supply of credit in the primary bond market, when firms' bond issuance is taken as given. As investors' preferences are more sensitive to liquidity risk (δ is higher or β is lower), the associated liquidity premium drives up the hold-to-maturity return that investors require to hold corporate debt, i.e., there is a contraction of the aggregate credit supply. On the other hand, when investors are poorer (e_0 is smaller) they reduce their holdings of liquid storage one-for-one conditional on buying the same number of financial contracts. Less liquid investors' portfolios reduce liquidity in secondary markets, and thus also drives up the required hold-to-maturity return through an increase in the liquidity premium (Lemma 1).

Proposition 5 (Credit Demand Determinants) *Taking the expected hold-to-maturity return offered to investors as given, credit demand is increasing in the firm's endowment n_0.*

Proposition 5 is a consequence of the fact that the optimal financial contract, which solves the firm's problem (3), specifies the optimal firm leverage l_0 as a function of the expected hold-to-maturity return R. In fact, as this return remains constant, so does leverage and thus firm's credit demand is given by $b^d(R) = (l_0(R)-1)n_0$, which an increasing function of n_0. This argument completes the proof.

The implications of the previous two results for the equilibrium of the model are summarized in the following proposition.

Proposition 6 (Equilibrium Comparative Statics) *In any of the following cases, the equilibrium expected return in credit markets increases. Thus, the leverage l_0 and risk $\bar{\omega}$ of the optimal contract and the default premium all decrease. Moreover, in the first three cases bond issuance decreases.*

1. *(Preference shock) The probability of becoming impatient is higher, i.e., δ is higher; or*
2. *(Impatience) Impatient investors discount the future more heavily, i.e., β is lower; or*
3. *(Investors' Endowments) Investors have less to invest in storage, i.e., e_0 is lower; or*
4. *(Firms' Endowments) Firms have more equity (i.e., n_0 is higher).*

This proposition presents the comparative statics in *equilibrium* for the parameters that describe preferences and endowments for investors and firms. For the first three cases, Proposition 4 establishes that an increase in δ or a decrease in β or e_0 will reduce the aggregate supply of credit and firms will see an increase in the liquidity premium of the bonds they issue. According to Proposition 6, firms adjust to this increase in expected bond return along two margins (recall that the debt contract is two-dimensional): they offer fewer and less risky contacts in the primary market. A reduction in the number of bonds issued in the primary market lowers the number of possible sell orders in the secondary market, boosting market thickness and attenuating the increase in the liquidity premium. In addition, the reduction of bonds' riskiness contributes to reduce firms' external financing cost by lowering the default premium. In equilibrium, thus, the total effect on the external financing premium is ambiguous and depends on whether the increase in the liquidity premium or the decrease in the default premium dominates.

The fourth case of Proposition 6 deserves special attention. In the benchmark CSV model, altering the firm's endowment of equity has no impact on expected return offered to investors, or the characteristics of the optimal contract. The reason is because in the frictionless benchmark there are no liquidity provision or liquidity premium channels at work. This result does not carry through in our framework, where these channels are present. As in the benchmark model—indeed, for exactly the same reason—there is no effect of an increase of firms' endowment on the optimal contract. But in our framework as credit demand and bond issuance increase, the liquidity provision channel reduces liquidity in the OTC market. This leads investors to reprice bonds' illiquidity: the liquidity premium channel. Thus, in our framework, the size of the corporate sector relative to the financial sector influences liquidity provision and liquidity premia. Moreover, our model is homogeneous of degree zero in (n_0, e_0). That is, increasing the size of the corporate sector n_0 and the financial sector e_0 in the same proportions have no effect on market thickness, the liquidity premium, or the characteristics of the optimal contract. The equilibrium in the credit market will be described by the same expected hold-to-maturity return and an increase in bond issuance commensurate to the increase in size of the two sectors.

3.5 A Numerical Illustration

We present a simple numerical illustration using the following parameter values. We set the endowment of firms at $n_0 = 0.2$ and the endowment of investors at $e_0 = 1$. Investors' preferences are described by a discount factor for impatient investors $\beta = 0.85$, while δ will take different values in $[0, 1]$ to illustrate the results established above. Firms' technology expected return is given by $R^k = 1.2$, whereas the return on storage is assumed to be $r = 0.01$. The parameters of the matching function in the OTC market are the scaling constant $\nu = 0.2$ and the elasticity of the matching function with respect to sell orders is $\alpha = 0.5$. The share of the surplus that accrues to impatient investors is $\psi = 1$. Idiosyncratic productivity shocks ω are distributed according to a log-normal distribution with mean equal 1 and variance equal to 0.25. Finally, monitoring costs are a share $\mu = 0.2$ of firms' revenue.

We begin with the frictionless benchmark, taking $\delta = 0$.[19] The equilibrium of the model is described by entrepreneurs' choice of leverage, l_0, and risk, $\bar{\omega}$, subject to the constraint imposed by the investors' break-even condition (15) and the consistency requirements for liquidity, θ, and price, q_1, in the OTC market. The characteristics of the optimal contract $(l_0, \bar{\omega})$ determine the hold-to-maturity return, R^b, and thus the secondary market price q_1. (Recall that the return on executed orders in secondary markets is pinned down by ψ, r, and β.) The optimal contract will determine the portfolio allocation of investors and thus secondary market liquidity θ (equation 17). Thus, we use the $(l_0, \bar{\omega})$-space to describe the optimal contract and the equilibrium of the model. Figure 4 depicts the firm's

[19] From Proposition 3 the frictionless benchmark is obtained if alternatively we set $\beta = 1/1.01$, or if $\psi = 1$ (as in our example) and e_0 is sufficiently high so $f(\theta) = 1$.

isoprofit curves in green,[20] and displays the investors' break-even condition by a red line. Firm's profits increase with leverage and decrease with risk, so isoprofit curves represent higher profits moving south-east in the figure. The private equilibrium in the frictionless benchmark economy is given by the tangency between the break-even condition and the isoprofit line shown by the solid black dot in Figure 4.

Figure 5 illustrates the case of an increase in the liquidity shock, δ, (i.e., case 1 of Propositions 4 and 6). As the probability of becoming impatient increases, investors require a higher liquidity premium to be compensated for liquidity risk (Proposition 4). In contrast, the firm's isoprofit lines for a given contract specified by $(l_0, \bar{\omega})$ are invariant to δ, thus the demand for credit is invariant to δ. Nevertheless, the firm adjusts the terms of the contract it offers in the primary market owing to the increase in the liquidity premium. In particular, the firm reduces its supply of primary debt, which partially compensates investors for the reduction in secondary market liquidity. The resulting equilibrium has a lower level of leverage and a less risky debt contract, as shown in Figure 5 (Proposition 6).

Finally, Figure 6 presents a decomposition of the total corporate premium Φ^t paid on the primary debt contract in terms of the default premium Φ^d and the liquidity premium Φ^ℓ. The figure shows that lower levels of leverage and risk due to increased liquidity demand result in lower total corporate bond premia. Naturally, the liquidity premium goes up, but the default premium decreases since the firm is offering a lower $\bar{\omega}$ (Proposition 6), and the latter effect dominates in this case.

4 Efficient Liquidity in OTC Markets

We consider a social planner that is constrained by the presence of search frictions and the structure of trade in the OTC market. Hence, our concept of efficiency is one of constrained efficiency, or second best.[21]

The planner chooses the optimal contract to maximize the profits of the firm while internalizing both the liquidity provision and liquidity premium channels. To formalize the planner's problem let $(l_0, \bar{\omega}, \theta, q_1)$ be allocations that describe the socially efficient outcome and let $(l_0^{pe}, \bar{\omega}^{pe}, \theta^{pe}, q_1^{pe})$ be the allocations in the private equilibrium described in section 3. Then, the planner's problem can be written as $\max_{\bar{\omega}, l_0, \theta, q_1} [1 - \Gamma(\bar{\omega})] R^k l_0$, subject

[20] Note that the shape of the isoprofit curves (increasing and concave) holds in general, as follows from the properties of the $\Gamma(\bar{\omega})$ function, and does not depend on the particular values used in our example.

[21] In the interest of space the analysis in sections 4 and 5 restricts attention to the more interesting case where $\theta \in (\underline{\theta}, \overline{\theta})$, so trading probabilities depend on the matching function (4) and are not pinned down by the minimum number of buy or sell orders.

to equations (7), (17), and

$$\mathbb{U}(l_0, \bar{\omega}, \theta, q_1) \geq \mathbb{U}(l_0^{pe}, \bar{\omega}^{pe}, \theta^{pe}, q_1^{pe}) \,. \tag{20}$$

Condition (20) says that the planner cannot choose equilibrium allocations that result in lower welfare for investors compared to the private equilibrium, whereas equations (7) and (17) force the planner to respect the determination of prices and liquidity, respectively, in the OTC market.[22] The social planning problem differs from the private equilibrium in two respects. First, the planner need not respect the investor's break-even condition (15), but cannot make investors worse off, i.e., needs to satisfy (20). Second, the planner internalizes how period 0 choices affect liquidity in the secondary market by explicitly considering (17) as a constraint, which, in contrast, is a condition of the private equilibrium.[23] In this manner, the planner internalizes both the liquidity provision and liquidity premium channels.

We substitute equations (7) and (17) in the planner's problem, and let λ be the multiplier on constraint (20), to obtain that the socially optimal choice of leverage is given by

$$[1 - \Gamma(\bar{\omega})]R^k = -\lambda \left[n_0(U_b - U_s) + b_0 u_b \frac{\partial R^b}{\partial l_0} + \frac{\partial \mathbb{U}}{\partial \theta} \frac{\partial \theta}{\partial l_0} \right] . \tag{21}$$

That is, the marginal increase in the firm's profits from additional leverage needs to be proportional to the marginal reduction in total expected utility for investors. The latter has three components: (*i*) the portfolio composition: as leverage increases investors need to re-allocate n_0 units from storage to bonds; (*ii*) the effect on the expected hold-to-maturity return R^b; and (*iii*) the effect through secondary market liquidity: as liquidity increases it becomes easier for impatient investors to sell their bonds, but it becomes more difficult for patient investors to buy bonds and earn the return Δ in the secondary market.

Similarly, the socially optimal choice for the risk profile of corporate debt is given by

$$l_0 \Gamma'(\bar{\omega}) R^k = \lambda \left[b_0 u_b \frac{\partial R^b}{\partial \bar{\omega}} + \frac{\partial \mathbb{U}}{\partial \theta} \frac{\partial \theta}{\partial \bar{\omega}} \right] . \tag{22}$$

That is, the marginal increase in the firm's profits from reducing risk need to be proportional to the marginal reduction in total expected utility for investors, which has

[22] We also considered a more general problem, as an alternative but not reported, where the planner can additionally determine the terms of trade in the secondary market and assigns Pareto weights on the two agents to maximize a social welfare function.

[23] Recall that investors, and thus firms, explicitly considered (7) in the private equilibrium as well, thus its explicit consideration does not modify the planner's problem relative to the private equilibrium, unless the planner can affect the terms of secondary trade.

two components: the effect on the hold-to-maturity return R^b and the effect through secondary market liquidity.

Taking a ratio of equations (21) and (22) gives

$$\frac{1-\Gamma(\bar{\omega})}{\Gamma'(\bar{\omega})l_0} = -\frac{n_0(U_b - U_s) + b_0 u_b \frac{\partial R^b}{\partial l_0} + \frac{\partial \mathbb{U}}{\partial \theta}\frac{\partial \theta}{\partial l_0}}{b_0 u_b \frac{\partial R^b}{\partial \bar{\omega}} + \frac{\partial \mathbb{U}}{\partial \theta}\frac{\partial \theta}{\partial \bar{\omega}}}. \tag{23}$$

This equation, together with the constraint on investors total expected utility (20), describes the socially optimal debt contract.[24] We are ready to establish the generic inefficiency of the private provision of liquidity.[25]

Proposition 7 (Generic Constrained Inefficiency of Liquidity Provision) *Consider a planner that designs an optimal financial contract, as described by (20), (23), (7) and (17). 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.*

Given Proposition 7, we can identify two distorted margins that drive a set of wedges between the private and socially efficient outcomes that are apparent from comparing the equilibrium conditions (15) and (16) to the social planner's counterparts (20) and (23). On the one hand, comparing equation (23) to equation (16), reveals the presence of a *pecuniary externality* that comes from the fact firms do not internalize how their bond issuance affects their funding cost. In fact, firms' funding cost are set by investors, who will be affected by bond issuance in two ways. First, additional bonds in their portfolios will change their expected utility if there is a difference between the utility they expect to receive when buying bonds or liquid assets, represented by the term $n_0(U_b - U_s)$. Second, additional bonds will affect market liquidity, which in turn is going to affect investors' expected utility, represented by the term $\partial \mathbb{U}/\partial \theta$.

The second distortion becomes apparent when comparing equation (15), the investors' break-even condition, to equation (20), the weak Pareto improvement constraint faced by the planner. Since $U_s = U_a + (1-\delta)(1+r)(\Delta - (1+r))p(\theta)$, we can rewrite equation (20) as $n_0(l_0 - 1)(U_b - U_s) = e_0(1-\delta)(1+r)(\Delta - (1+r))[p(\theta^{pe}) - p(\theta)]$. Written this way, the equation tells us that as long as the optimal level of public liquidity is different than the equilibrium level of private liquidity, i.e, $\theta^{pe} \neq \theta$, then the expected utility of holding

[24] The constraint will always be binding since the planner cares only about the firm, but this need not be the case if the planner maximizes aggregate social welfare. In that case the planner may want to split the aggregate gains according to some set of Pareto weights.

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

bonds and liquid assets will have to be different $U_b(\theta) \neq U_s(\theta)$. This reflects the presence of a *congestion externality* that comes from the fact that investors do not internalize how their portfolio choices affect market liquidity.

The following proposition summarizes the interaction between these two externalities.

Proposition 8 (Constrained Efficient Equilibrium) *The constrained efficient allocations can be characterized conditional on the model parameters (α, r, ψ) as follows:*

- *If $\psi(1 + \alpha r) > \alpha(1 + r)$ then secondary market liquidity generates a positive externality on investors ($\partial \mathbb{U}/\partial\theta > 0$); the planner implements a higher level of secondary market liquidity ($\theta > \theta^{pe}$); and the socially optimal financial contract is characterized by lower leverage, $l_0 < l_0^{pe}$, and less risk, $\bar{\omega} < \bar{\omega}^{pe}$.*

- *If $\psi(1 + \alpha r) < \alpha(1 + r)$ then secondary market liquidity generates a negative externality on investors ($\partial \mathbb{U}/\partial\theta < 0$); the planner implements a lower level of secondary market liquidity ($\theta < \theta^{pe}$); and the socially optimal financial contract is characterized by higher leverage, $l_0 > l_0^{pe}$, and more risk, $\bar{\omega} > \bar{\omega}^{pe}$.*

- *If $\psi(1 + \alpha r) = \alpha(1 + r)$ then there is no externality ($\partial \mathbb{U}/\partial\theta = 0$) and equilibrium is constrained efficient, i.e., $(l_0, \bar{\omega}, \theta) = (l_0^{pe}, \bar{\omega}^{pe}, \theta^{pe})$.*

To understand the intuition behind the proposition consider that the planner internalizes the pecuniary externality faced by firms and is also aware of the presence of the congestion externality faced by investors. As such, the planner faces a different trade-off between leverage and risk relative to an individual firm. Recall from the firms' problem (3) that profits increase whenever leverage l_0 is higher, or risk $\bar{\omega}$ is lower. Higher leverage implies a larger scope of the firm, whereas lower risk implies that a larger share of profits is retained by the firm. How can the planner increase the profitability of firms? Proposition 8 tells us that the answer depends on the parameters (α, r, ψ).

For example, consider the effect of an increase in liquidity on investors' welfare. On the one hand, increased liquidity generates ex ante welfare gains for impatient investors because it is easier to sell unwanted bonds in the secondary market. On the other hand, patient investors suffer as it becomes more difficult to earn a higher return by purchasing bonds at a discounted price. The gains to impatient investors outweigh the losses to patient investors, making investors ex ante better off, i.e., $\partial \mathbb{U}/\partial\theta > 0$ and we say there is a positive externality. This is the case when the model parameters satisfy $\psi(1+\alpha r) > \alpha(1+r)$. Intuitively, in this case the trade surplus that accrues to impatient investors is relatively large. Alternatively, the return on storage that patient investors receive if they fail to execute a secondary trade is sufficiently low, so that $r < (\psi - \alpha)/(\alpha - \alpha\psi)$.

In the case of the positive externality, in order to implement a higher level of market liquidity the planner is going to reduce bond issuance by firms, lowering leverage and scope of the firm. But, at the same time, the planner also reduces the risk of the financial contract, allowing firms to retain a larger share of profits in expectation. In this way the planner directs the firm to operate at a smaller scale, while at the same time paying lower financing costs. By restricting bond issuance and enhancing market liquidity, the cost of financing are much smaller than what a firm that does not internalize the externality would have expected, because the planner is able to redistribute the gains for investors back to firms in the form of even lower financing costs.[26]

In this way the planner effectively takes advantage of the congestion externality to create additional surplus for investors that is then redistributed back to firms by addressing the pecuniary externality. It is worth noting that the redistribution takes place through changes in the liquidity premium, as matter of fact, as we show below, our social planner problem can be implemented with no direct transfers from investors to firms.

By contrast, when there is a negative externality, i.e., $\partial \mathbb{U}/\partial\theta < 0$, a reduction in market liquidity generates gains to patient investors that outweigh the losses to impatient investors, making investors ex ante better off. This is the case when $\psi(1+\alpha r) > \alpha(1+r)$. In the case of a negative externality, in order to implement a lower level of market liquidity the planner is going to increase bond issuance by firms, increasing leverage and scope of the firm. However, at the same time the planner increases the risk of the financial contract, reducing the expected share of profits that the firm retains. In this way the planner allows the firm to operate at a larger scale, while at the same time paying higher cost of financing. But financing cost are lower than what a firm that does not internalize the externality would have expected at the new higher level of leverage, as the planner is able to redistribute the gains for investors owing to lower liquidity back to firms in the form of lower financing costs.

Finally, in the knife-edge case where $\psi(1+\alpha r) = \alpha(1+r)$ private liquidity is efficient so that at the margin an increase in liquidity generates gains for impatient investors that are perfectly offset by losses to patient investors, and the planner cannot exploit the congestion externality to improve upon the private equilibrium. This special case highlights the relationship of our result with the well-known Hosios condition.[27] But, note that in our

[26] It is interesting to note that by implementing higher secondary *market* liquidity, the planner in essence increases *funding* liquidity in the primary market by implementing a reduction in the liquidity premium and thus in the total bond premium.

[27] The parameter restriction is analogous to the Hosios (1990) rule that determines the efficient surplus split in search and matching models of the labor market. Arseneau and Chugh (2012) study the implications of inefficient surplus sharing for optimal labor taxation in a dynamic general equilibrium economy.

model the planner is not trying to correct the congestion externality, rather she aims to correct a pecuniary externality while exploiting the congestion externality.

4.1 Optimal Private Liquidity Regulation

To analyze optimal liquidity regulation we allow the planner access to a complete set of tax instruments. Specifically, we introduce a marginal tax τ^s on the return from storage in period 0 and a marginal tax τ^l on leverage (negative taxes correspond to subsidies).[28] With these tax instruments, the objective of investors becomes $\mathbb{U} = b_0 U_b + s_0 U_s (1 - \tau^s) + T^s$ and the objective of the firm changes to $[1 - \Gamma(\bar{\omega})]\, R^k l_0 - \tau^l \lambda^{pe} l_0 + T^l$. The taxes are funded in a lump-sum fashion on the same agents, thus $T^l = \tau^l \lambda^{pe} l_0$ and $T^s = \tau^s s_0 U_s$ in equilibrium. Also, in order to simplify the exposition note that we have normalized the tax on leverage by the Lagrange multiplier, $\lambda^{pe} > 0$, on the constraint faced by firms (i.e., the investors' break-even condition (15)).

Proposition 9 characterizes the optimal regulation of private liquidity provision.

Proposition 9 (Implementation of 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. The resulting optimal taxes on storage, τ^s, and leverage, τ^l, are given by:*

$$\tau^s = \frac{e_0}{b_0}\left(1 - \frac{U_s(\theta^{pe})}{U_s(\theta)}\right), \tag{24}$$

$$\tau^l = \frac{n_0 U_s u_b \frac{\partial R^b}{\partial \bar{\omega}} \tau^s + u_b \left[\frac{\partial R^b}{\partial l_0}\frac{\partial \theta}{\partial \bar{\omega}} - \frac{\partial R^b}{\partial \bar{\omega}}\frac{\partial \theta}{\partial l_0}\right]\frac{\partial \mathbb{U}}{\partial \theta}}{b_0 u_b \frac{\partial R^b}{\partial \bar{\omega}} + \frac{\partial \theta}{\partial \bar{\omega}}\frac{\partial \mathbb{U}}{\partial \theta}} \tag{25}$$

where the term in square brackets and the denominator in (25) are strictly positive.

The role for the tax on storage is to create a wedge in the investors' break-even condition (15), i.e., $U_b \neq U_s$ as long as $\theta^{pe} \neq \theta$. This allows the planner to implement the desired allocation without making investors worse off. Moreover, the role of the tax on leverage is to make the firm internalize the pecuniary externality of bond issuance on its funding costs.

Combining the insights of Proposition 9 with Proposition 8 above, it is easy to characterize the optimal tax system more specifically. When $\psi(1 + \alpha r) > \alpha(1 + r)$, the liquidity externality is positive so that the planner wants to implement higher liquidity relative to

[28] We consider tax instruments to correct the distorted margins, both other instruments such as leverage or portfolio restrictions could also be considered. See also Perotti and Suarez (2011), who propose Pigouvian taxation to address externalities from the under-provision of liquidity.

the private equilibrium, $\theta > \theta^{pe}$. Accordingly, the optimal regulation needs to be designed in a way that results in investors holding a more liquid portfolio. This can be achieved through a storage subsidy, so that $\tau^s < 0$. Moreover, the optimal regulation needs to be designed in a way that results in firms issuing fewer bonds in the primary market, which can be achieved through a tax on leverage, so that $\tau^l > 0$. By the same logic, when $\psi(1 + \alpha r) < \alpha(1 + r)$, the liquidity externality is negative and $\theta < \theta^{pe}$. The optimal tax system calls for a tax on storage, $\tau^s > 0$, and a leverage subsidy, $\tau^l < 0$. Only in the knife-edge case where $\psi(1 + \alpha r) = \alpha(1 + r)$ we have that $\tau^l = \tau^s = 0$.

4.2 A Numerical Illustration

We continue the numerical example in section 3.5. Recall that in this illustration, $\psi = 1$. Moreover, because the planner has the same objective as the firm, the isoprofits lines are the same in both problems. Figure 7 shows the planner's solution and the private equilibrium for two cases: $\delta = 0$ and $\delta = 0.1$. In a frictionless environment ($\delta = 0$), the planner's solution coincides with the private equilibrium (as we proved in Proposition 3). However, when there is a positive demand for liquidity, $\delta > 0$ and $\beta < 1/(1 + r)$, and secondary market liquidity is not sufficiently high to guarantee $f(\theta) = 1$, the planner chooses lower leverage and a less risky bond contract, i.e., lower l_0 and $\bar{\omega}$. The reason is because the planner internalizes the effect of the leverage decision on liquidity in the secondary market. This induces the planner to consider a steeper constraint compared to the breakeven condition considered by firms (where market liquidity is taken as given). As a result, the planner understands how lower leverage and risk improves borrowing terms on the margin, when the total social costs are taken into account.

Table 1 shows the change in equilibrium allocations between the private and planner's solutions for $\delta = 0.1$ as ψ moves from 1 to 0. Consistent with the analysis above, the planner's allocations can be replicated using appropriate tax instruments (subsidies if they are negative) on leverage and storage. For $\psi = 1$, the liquidity externality is positive implying that liquidity is suboptimally low in the private equilibrium. The planner would like to implement a tax on leverage to generate more liquidity in the secondary market (in this case all the surplus goes to sellers so the tax on storage is irrelevant). However, as the share of the gains from trade that accrues to impatient investors declines, the size of the liquidity externality shrinks. Hence, the planner is less aggressive in choosing the optimal combination of leverage tax and storage subsidy, i.e., both τ^l and τ^s shrink in absolute value. When the parameterization of ψ satisfies $\psi(1 + \alpha r) = \alpha(1 + r)$, the externality zeros out and the optimal tax system implies $\tau^l = \tau^s = 0$. For values of ψ below that point, the liquidity externality becomes negative, so that liquidity is over-provided in the private equilibrium. Accordingly, the sign of the optimal tax system flips so that leverage is subsidized, $\tau^l < 0$, and storage is taxed, $\tau^s > 0$.

5 Optimal Public Liquidity Management

In this section, we 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. Through the lens of our model, any public policy that alters both private and public portfolios effectively shifts liquidity risk between the private and the public sector. This shifting of liquidity risk alters the liquidity premia which, in turn, influences savings and investment decisions in the real economy. In practice, important policies that can alter public and private portfolios are quantitative easing or large scale asset purchases, as the ones implemented in the U.S., Europe, and Japan.

5.1 Quantitative Easing Policy

We model QE through direct purchases by the central bank of long-term illiquid assets (the financial contracts issued by firms and which are retraded by investors in OTC markets).[29] These purchases are financed by the issuance of short-term liquid liabilities, referred to as reserves, that offer a return that is at least as large as that offered by the storage technology. 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).

The key assumption we make to model QE is that the assets purchased by the central bank are relatively less liquid than reserves. 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. While it might be the case that these assets are highly liquid at or near origination, the evidence suggests that they become less liquid as they are *retraded* in venues that might be well represented by OTC markets.[30] This becomes less of a concern when considering QE programs in other jurisdictions, like Europe or Japan, where central

[29] Note that a pool of firms' contracts will have no credit risk, since there are no aggregate shocks.

[30] 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.

banks have purchased non-government guaranteed assets, which are perceived as less liquid than central bank's reserves.

At the beginning of the initial period, the central bank credibly commits to purchase a quantity $\bar{b}_0$ of bonds from investors and hold them to maturity. These bond purchases are financed through the issuance of $\bar{s}_0$ units of reserves that pay interest $\bar{r}$. In our model reserves are a perfect substitute for the storage technology from the point of view of investors, thus $\bar{r} \geq r$.[31] The central bank 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.[32]

Investors can freely trade reserves for consumption with the central bank at any point.[33] The central bank budget constraint in the initial period is simply

$$\bar{b}_0 = \bar{s}_0 \ . \tag{26}$$

In addition, we assume the central bank finances itself in period 1 with reserves only. This assumption prevents the central bank from injecting additional resources into the economy in the interim period. In order to keep its bond holdings, the central bank needs to roll over its outstanding reserves and pay interest on them in period 1. The central bank will have to borrow an amount equal to $(1 + \bar{r})\bar{s}_0$.[34] Finally, in period 2 the central bank receives the debt payout from the financial contract and expends $(1 + \bar{r})^2\bar{s}_0$ in interest and principal on outstanding reserves. It is assumed that the central bank allocates reserves evenly across investors who demand reserves in a given time period.

The central bank faces three constraints that, taken together, serve to limit the size of its QE program. First, we assume that the central bank 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 $\bar{\mu} > \mu$. Consequently, any positive effects of QE would not accrue from enhanced monitoring, but from its effect on liquidity premia. This

[31]Our results are qualitatively the same if we impose that $\bar{r} = r$. Nonetheless, as we will show in the next section, allowing the central bank to pay higher interest on reserves provides the central bank an additional tool to manage public liquidity.

[32]We have abstracted from trade frictions between the central bank and investors, as in practice the Federal Reserve announces in advance its intention to buy bonds and has readily available trading counterparties.

[33]This is isomorphic to a model in which trade in the Fed Funds market is frictionless. Another literature studies frictional trade in the Fed Funds market, see, for example, Afonso and Lagos (2015) or Bianchi and Bigio (2014).

[34]In practice, the long-term assets held by central banks pay interest in the interim period, and in an environment of low short-term interest rates these holdings will generate a positive net-interest income for the central bank. But for simplicity we abstract from these considerations. See, for instance, Carpenter et al. (2013) for estimates of net-interest income for the Federal Reserve.

implies that in expectation the central bank anticipates receiving $\bar{R}^b \bar{b}_0$ for its asset holdings, with $\bar{R}^b$ the expected hold-to-maturity return on financial contracts for the central bank, given by[35]

$$\bar{R}^b(l_0, \bar{\omega}) = \frac{l_0}{l_0 - 1} R^k \left[\Gamma(\bar{\omega}) - \bar{\mu} G(\bar{\omega})\right] = R^b(l_0, \bar{\omega}) - \frac{l_0}{l_0 - 1} R^k (\bar{\mu} - \mu) G(\bar{\omega}) \ .$$

Second, the central bank needs to fully finance its funding cost, i.e., the total interest on reserves, with its expected return on assets. That is,

$$\bar{R}^b \geq (1 + \bar{r})^2 \ . \tag{27}$$

Note that if the central bank buys a portfolio of bonds, it does not undertake any credit risk, as firms' returns are independent.

Finally, we assume that investors cannot be made worse off by QE, as we describe in section 5.3.

5.2 QE, Market Liquidity, and the Supply of Credit

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 payouts 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

$$c_1^I = (1 + r)s_0 + (1 + \bar{r})\bar{s}_0 + f(\theta) q_1 (b_0^s - \bar{b}_0) \ , \tag{28}$$

$$\text{and} \qquad c_2^I = (1 - f(\theta)) R (b_0^s - \bar{b}_0) \ . \tag{29}$$

Patient Investors. Patient investors only value consumption in the final period and, as a

[35] Note that the central bank in the model 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 central bank purchases bonds at a discount it will increase the expected return on asset purchases and relax constraint (27).

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 10 and 11—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 individual reserve holdings in the interim period for patient investors, $\bar{s}_1^P$, total $(1 + \bar{r})\bar{s}_0/(1 - \delta)$. All remaining units of consumption are placed into the lower yielding storage technology, so expected storage holdings at the end of $t = 1$, s_1^P, equal $s_1^P = (1 - p(\theta))\left[(1 + r)s_0 + (1 + \bar{r})\bar{s}_0\right] - (1 + \bar{r})\bar{s}_0/(1 - \delta)$, which is strictly positive from Assumption 4. It follows that expected consumption of patient investors equals

$$c_2^P = s_1^P(1 + r) + \frac{(1 + \bar{r})^2\bar{s}_0}{1 - \delta} + \left\{b_0^s - \bar{b}_0 + p(\theta)\frac{(1 + r)s_0 + (1 + \bar{r})\bar{s}_0}{q_1}\right\}R\ . \tag{30}$$

Using the optimal behavior of investors in period 1, summarized in equations (28)-(30), we can rewrite the expected lifetime utility as the portfolio weighted average of the utilities 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 (13) and (14), respectively. On the other hand, the expected utility of reserves is given by

$$U_{\bar{s}} = \delta(1 + \bar{r}) + (1 - \delta)(1 + \bar{r})\left[(1 - p(\theta))(1 + r) + \frac{\bar{r} - r}{1 - \delta} + p(\theta)\Delta\right]. \tag{31}$$

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 additional reserves bought in period 1.[36]

Finally, note that, when the equilibrium in the credit market is given by (b_0, R^b) , market liquidity corresponds to

$$\theta = \frac{(1-\delta)[(1+r)s_0 + (1+\bar{r})\bar{s}_0]}{\delta(b_0 - \bar{b}_0)q_1} = \frac{(1-\delta)\Delta\left[(1+r)\left(e_0 - n_0(l_0-1)\right) + (1+\bar{r})\bar{b}_0\right]}{\delta R^b\left(n_0(l_0-1) - \bar{b}_0\right)}. \tag{32}$$

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

Proposition 10 (The Real Effects of QE) *Quantitative easing, i.e., the size of the bond buying program, $\bar{b}_0$, increases secondary market liquidity θ and, hence, increases firm's investment.*

The intuition behind this result is straightforward. Each bond bought by the central bank 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. 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 publically provided liquid assets. The resulting reduction in the liquidity premium demanded by investors pushes down the cost of funding for firms, who respond by taking on higher leverage and 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 8 is presented from the perspective of a central bank 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 thickness.

5.3 Optimal Public Liquidity Management via Quantitative Policies

To understand the role of QE in the optimal policy mix, we consider a planner who wants to maximize firm profits, but is restricted by the central bank budget constraint, equation (26), and the financing constraint, equation (27). In addition, as with the planner

[36] If, $1+\bar{r} > \Delta$, patient investors will use their liquid wealth first to buy reserves, and then will use their remaining liquid wealth to place buy orders in the OTC market. Proceeding as above we can derive for patient investors s_1^P, c_2^P, and $U_{\bar{s}}$.

in Section 4, we assume the QE program cannot make investors worse off. To write this later constraint, let $\mathbb{U}(l_0, \bar{\omega}, \theta, \bar{b}_0, \bar{r})$ be the expected lifetime utility of investors when the equilibrium is described by $(l_0, \bar{\omega}, \theta)$, with the secondary market price given by (7), and the QE program described by $(\bar{b}_0, \bar{r})$. Similarly, let $\mathbb{U}(l_0^{pe}, \bar{\omega}^{pe}, \theta^{pe})$ be the expected lifetime utility of investors in the private equilibrium, when the secondary market price is given by (7). We refer to this planner that have access to QE policies as the central bank. Then, the central bank'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 equations (26), (27), (32), and

$$\mathbb{U}(l_0, \bar{\omega}, \theta, q_1, \bar{b}_0, \bar{r}) \geq \mathbb{U}(l_0^{pe}, \bar{\omega}^{pe}, \theta^{pe}, q_1^{pe}) \tag{33}$$

The following proposition characterizes the role of QE as part of the optimal policy mix.

Proposition 11 (QE as Part of the Optimal Policy Mix) *The optimal design of QE conditional on the model parameters (α, r, ψ) is described as follows:*

- *If $\psi(1 + \alpha r) > \alpha(1 + r)$, then QE improves upon the constrained efficient allocation, and the optimal QE program consists of a positive bond buying program, $\bar{b}_0 > 0$, coupled with an interest on reserves that is strictly greater than the return on storage, $\bar{r} > r$.*

- *If $\psi(1 + \alpha r) \leq \alpha(1 + r)$, then QE does not improve upon the constrained efficient allocation, and optimally the size of the QE program is $\bar{b}_0 = 0$.*

The QE program is effective because it affects market thickness, which in the presence of a liquidity externality, allows the central bank to increase the expected utility of investors and transfer those gains to firms. Intuitively, the QE program transfers illiquid bonds from investors, who value liquidity, to the central bank, who does not value liquidity because it is a long-term investor and is not subject to runs.

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.

However, the proposition shows that this technological improvement can only be realized when there are social gains from managing liquidity. Indeed, when the liquidity externality does not exist, i.e., the knife edge case where $\psi(1 + \alpha r) = \alpha(1 + r)$, there is no role for the management of public liquidity. When the externality is negative (so that liquidity is suboptimally high in the private equilibrium), the central bank wants to remove liquidity from the secondary market. As discussed at the end of the previous subsection, this could, in principle, be done through a QT program, whereby the central banks reduces the size of its balance sheet by selling bonds to investors in exchange for reserves.

There are a few additional points worth mentioning. First, 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.

Second, 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). We examine this in more detail in the quantitative exercise below.

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 central bank disintermediates the bond market. That is, the optimal policy is for the central bank to buy all the bonds offered by the firm and offer the corresponding amount of reserves to investors, paying $\bar{r} = r$. Doing so would allow the central bank to replicate the frictionless benchmark of section 3.2. However, as mentioned above, in our model the size of the QE program is limited by the constraints faced by the central bank.

5.4 A Numerical Illustration

Table 2 extends our numerical example to study the optimal public liquidity management, as implemented through QE. The table shows the changes in allocations relative to the private equilibrium for three different economies. The first column shows the decentralization of the socially efficient outcomes achieved through optimal liquidity regulation (implemented with the leverage tax, τ^l, and storage subsidy, τ^s) , but without QE. The second column shows the effects of QE in absence of liquidity regulation. Finally, the third column shows the case in which liquidity management coexists with liquidity regulation. All cases assume the parameterization $\alpha = 0.5$, $\psi = 0.9$, and $r = 0.01$. We choose this parameterization because it puts the model in a region of the parameter space where QE is effective, as per proposition 11. In addition, we assume that $\bar{\mu} = 0.3$, which is 50 percent higher than the baseline value of $\mu = 0.2$.

The first column (which, for reference, corresponds to a point half way between the results shown in the first and second columns of table 1) shows that in absence of QE, the efficient allocation is decentralized with a leverage tax, $\tau^l = 0.21$, and a subsidy for storage, $\tau^s = -0.04$. By raising liquidity in the secondary market, and hence depressing the liquidity premium, the resulting reduction in funding costs raises profits by 14 basis points relative to the private equilibrium, leaving the utility of investors unchanged. The second column presents results where we shut down liquidity regulation, but allow the planner access to a QE program. Even when we shut down the tax system, so that $\tau^l = \tau^s = 0$, the planner can use QE to achieve an even greater increase in firm profitability without harming investors. With QE the planner can achieve a similar outcome in terms of liquidity, without tax instruments. Finally, the last column of the table shows that QE, by itself, is not a panacea. A planner can do even better by implementing optimal liquidity management through QE in conjunction with liquidity regulation. The way to interpret this last result is that although QE improves the intermediation technology in the economy, it does nothing to remove the underlying distortions, arising from the pecuniary and congestion externalities discussed above.

Figure 8 shows how the gains to the firm vary with ψ for different levels of the efficiency of the central bank monitoring technology. The thick lines show the case for $\bar{\mu} = 0.3$ assuming QE in conjunction with the optimal tax system (the thick solid line) and, alternatively, assuming QE alone with no supporting tax system (the thick dashed line). The thin solid and dashed lines correspond to the same information when the monitoring cost is lower, so that $\bar{\mu} = 0.2$. Finally, the thin dotted line shows the gains to the firm from optimal tax policy alone in absence of QE. There are four things to take from the figure. First, QE is always more effective when combined with the optimal tax policy (the solid lines are always above the dashed line for the same monitoring cost assumption). Second, the effectiveness of QE is limited by the parameterization of ψ (the dashed lines are downward sloping), so that as the gains from trade that accrue to impatient investors decline, QE becomes less effective. Third, the effectiveness of QE depends importantly on the quality of the central bank's monitoring technology (the thick lines are below the thin ones, so the worse the technology, the less effective is QE). Finally, there are parts of the parameter space in which QE is ineffective to the point at which a planner would strictly prefer optimal taxation to QE (the regions in which the thick and thin dashed lines lie below the thin dotted line).

6 Conclusion

We show that trade frictions in OTC markets provide a rationale for the regulation and public management of market liquidity. In our model, investors face liquidity risk and need to allocate their limited liquid resources between liquidity provision and illiquid long-term bonds. Bonds provide credit to productive firms, so there is a trade-off between liquidity and credit provision. Trade friction together with this trade-off result in an inefficiency because investors and firms fail to internalize how their actions affect liquidity in the secondary market. A novel aspect of our model 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 and show how the socially efficient equilibrium can be decentralized with two tax instruments. Finally, we show how both the provision (as in QE) as well as the withdrawal of public liquidity can enhance welfare.

Our model suggests a set of testable predictions for the relationship between the availability of short-term liquid assets and liquidity premia. While there is only a single OTC market in our setup, in practice there are many, potentially segmented secondary markets (see, for example, Vayanos and Wang, 2007 and Vayanos and Weill, 2008). Given that central bank reserves can be used to settle transactions across most markets, we expect that quantitative easing should have an effect on the liquidity premia of not just those illiquid assets directly purchased by central banks, but *all* securities traded in OTC markets where participants' portfolios are affected by the policy (see Christensen and Krogstrup (2016) for related empirical support). Our model also suggests that QE will be more effective when the assets purchased by the Central Bank are more illiquid. In this sense, our mechanism provides an explanation for the fact that the Federal Reserve's strategy of purchasing relative illiquid mortgage-backed securities in the aftermath of the financial crisis might have been more effective than exchanging long-term Treasury bonds with shorter maturity ones as argued in Krishnamurthy and Vissing-Jorgensen (2011).

There are a number of directions for future work. First of all, our paper opens up new avenues for research on optimal liquidity provision when financial intermediation is conducted in markets with OTC characteristics. 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 of our model would be to jointly consider bank- and bond-financing and study the interaction of these two sources of funding for the real economy, as well as the spillovers from bank (liquidity) regulation on market liquidity (bank credit provision).

References

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

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

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

Atkeson, A. G., Eisfeldt, A. L. and Weill, P.-O. (2015), 'Entry and exit in otc derivatives markets', *Econometrica* **83**(6), 2231–2292.

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

Bernanke, B. (2017), Shrinking the fed's balance sheet. Blog. Brookings. https://www.brookings.edu/blog/ben-bernanke/2017/01/26/shrinking-the-feds-balance-sheet/.

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

Bernanke, B., Gertler, M. and Gilchrist, S. (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.

Bianchi, J. and Bigio, S. (2014), 'Banks, liquidity management and monetary policy', *Federal Reserve Bank of Minneapolis, Research Department Staff Report 503* .

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

Bruche, M. and Segura, A. (2014), 'Debt maturity and the liquidity of secondary debt markets', *working paper* .

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

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

Carpenter, S., J. Ihrig, E. Klee, D. Quinn and Boote, A. (2013), The Federal Reserve's Balance Sheet and Earnings: A primer and projections, Finance and Economics Discussion Series 2013-01, Federal Reserve Board.

Christensen, J. H. E. and Krogstrup, S. (2016*a*), A portfolio model of quantitative easing. Federal Reserve Bank of San Francisco Working Paper 2016- 12.

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

Christensen, J. H. E. and Rudebusch, G. D. (2012), 'The response of interest rates to us and uk quantitative easing*', *The Economic Journal* **122**(564), F385–F414.
URL: *http://dx.doi.org/10.1111/j.1468-0297.2012.02554.x*

Cui, W. and Radde, S. (2014), 'Search-based endogenous illiquidity and the macroeconomy'.

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

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

Edwards, A. K., Harris, L. E. and Piwowar, M. S. (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.

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

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

Geanakoplos, J. and Polemarchakis, H. (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 Herrenbrueck, L. (2012), 'Monetary policy, asset prices, and liquidity in over-the-counter markets', *Working Papers, University of California, Department of Economics* .

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

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

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

Hart, O. and Zingales, L. (2015), 'Liquidity and inefficient investment', *Journal of the European Economic Association, forthcoming* .

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

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

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

Jacklin, C. J. (1987), Demand deposits, trading restrictions, and risk sharing, *in* E. C. Prescott and N. Wallace, eds, 'Contractual Arrangements for Intertemporal Trade'.

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

Krasa, S. and Villamil, A. P. (2000), 'Optimal contracts when enforcement is a decision variable', *Econometrica* **68**(1), 119–134.

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

Kurlat, P. (2013), 'Lemons markets and the transmission of aggregate shocks', *American Economic Review* **103**(4), 1463–1489.

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

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

Lagos, R. and Zhang, S. (2016), Turnover liquidity and the transmission of monetary policy. Sloan Foundation Economics Research Paper Forthcoming. Available at SSRN: https://ssrn.com/abstract=2773782.

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

Malherbe, F. (2014), 'Self-fulfilling liquidity dry-ups', *Journal of Finance* **69**(2), 947–970.

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

Moreira, A. and Savov, A. (forthcoming), 'The macroeconomics of shadow banking', *Journal of Finance* .

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

Stein, J. (2014), 'Incorporating financial stability considerations into a monetary policy framework'. Speech at the 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 Wang, T. (2007), 'Search and endogenous concentration of liquidity in asset markets', *Journal of Economic Theory* **136**(1), 66–104.

Vayanos, D. and Weill, P.-O. (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.

Tables and Figures

Table 1: Planning outcomes and Implementation

ψ	1.0	0.8	0.6	0.4	0.2	0.0
% change in l_0	-8.62%	-5.03%	-1.63%	1.72%	5.13%	8.63%
% change in $\bar{\omega}$	-5.27%	-3.06%	-0.99%	1.04%	3.08%	5.17%
% change in θ	62.01%	27.75%	7.44%	-6.70%	-17.42%	-26.03%
% change in Π	0.23%	0.07%	0.01%	0.01%	0.06%	0.16%
% change in $\mathbb{U}$	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
τ^l	0.27%	0.15%	0.05%	-0.05%	-0.13%	-0.21%
τ^s	0.00%	-0.05%	-0.03%	0.04%	0.14%	0.27%

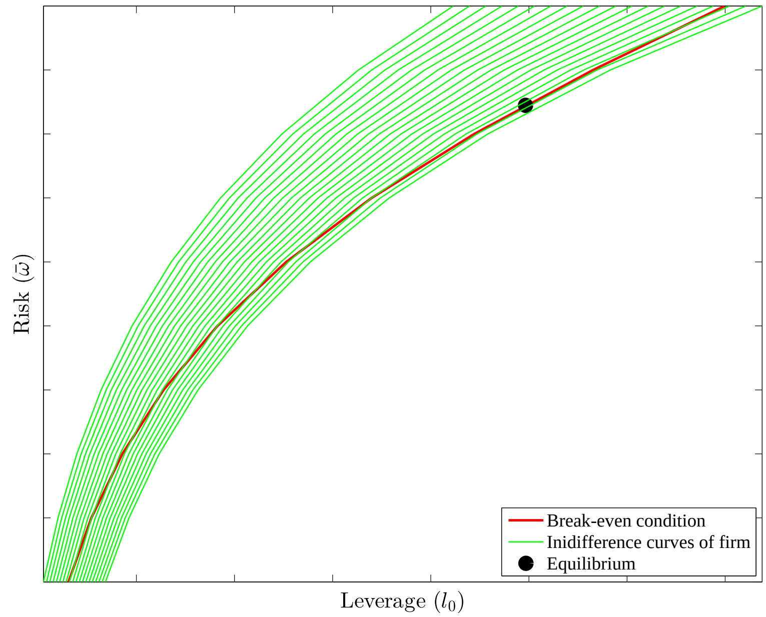
Note: Percentages correspond to deviations with respect to the private equilibrium for variables: leverage (l_0), risk ($\bar{\omega}$), market liquidity (θ), firms' profits (Π), and investors' utility ($\mathbb{U}$); and to the level of the optimal taxes on leverage (τ^l) and storage (τ^s). Negative values for taxes corresponds to subsidies. For details see section 4.2.

Table 2: Outcomes with Quantitative Easing

	Constrained Efficient Allocations	Quantitative Easing with $\tau^s = \tau^l = 0$	Quantitative Easing with τ^s, τ^l Chosen Optimally
% change in l_0	-6.78%	1.68%	-3.05%
% change in $\bar{\omega}$	-4.13%	0.72%	-2.35%
% change in θ	42.19%	43.37%	167.72%
% change in Π	0.14%	0.42%	0.98%
% change in $\mathbb{U}$	0.00%	0.00%	0.00%
$\bar{r}$		1.16%	1.10%
$\bar{s}_0$		0.09	0.18
τ^l	0.21%		0.17%
τ^s	-0.04%		-0.05%

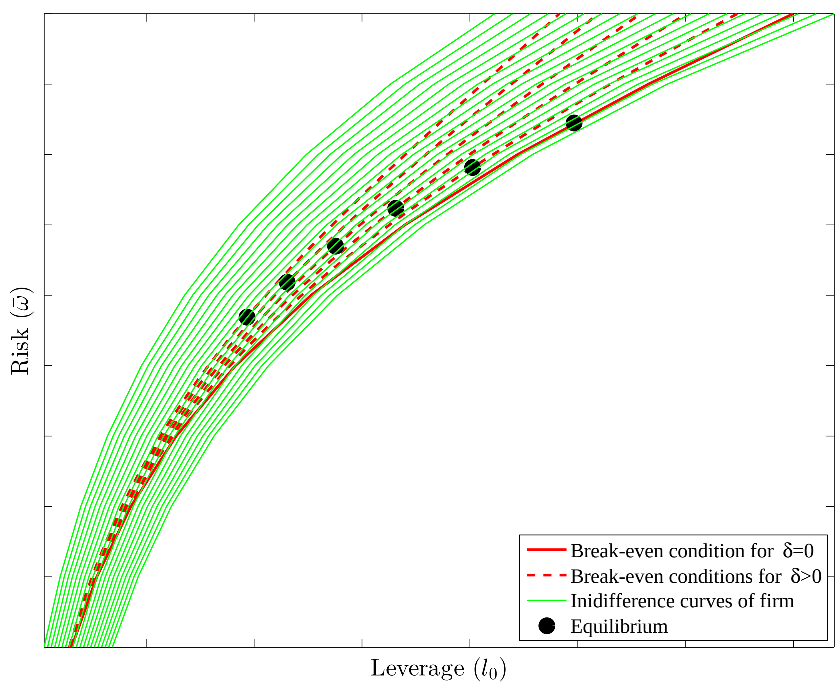
Note: Percentages correspond to deviations with respect to the private equilibrium for variables: leverage (l_0), risk ($\bar{\omega}$), market liquidity (θ), firms' profits (Π), and investors' utility ($\mathbb{U}$); and to the level of: tax on leverage (τ^l), tax on storage (τ^s), and interest rate on reserves ($\bar{r}$). Values for reserves ($\bar{s}_0$) are in levels. Negative values for taxes corresponds to subsidies. For details see section 5.4.

Figure 4: Equilibrium in the Frictionless Benchmark



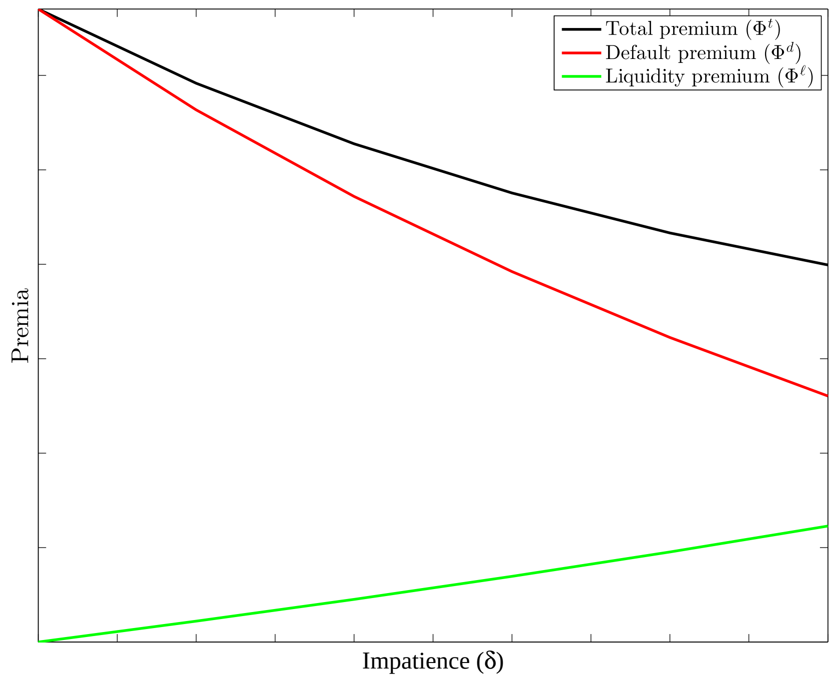
Note: For details see section 3.5.

Figure 5: Comparative Statics on δ.



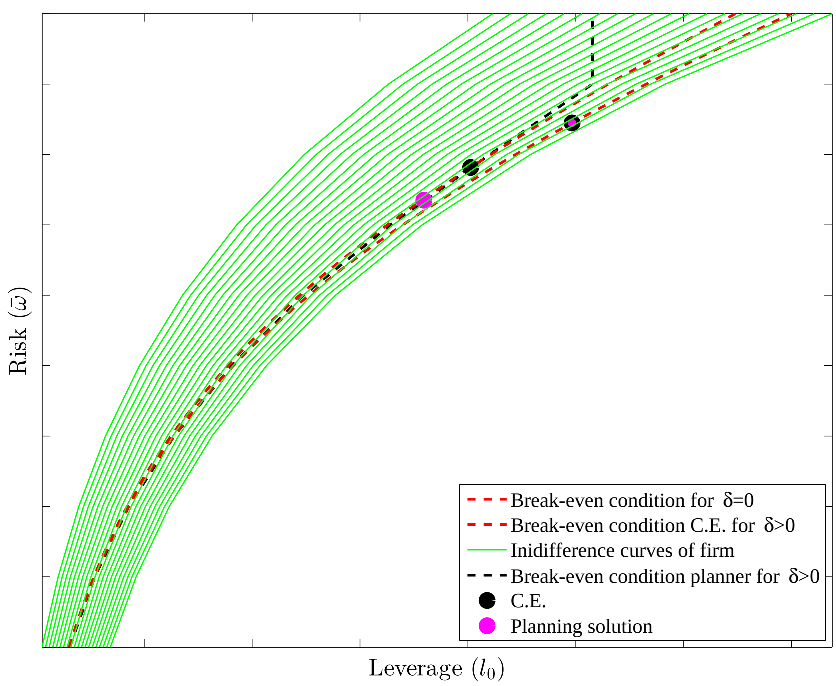
Note: δ take values in $\{0, 0.1, \ldots, 0.5\}$. See section 3.5.

Figure 6: Bond Premia Decomposition



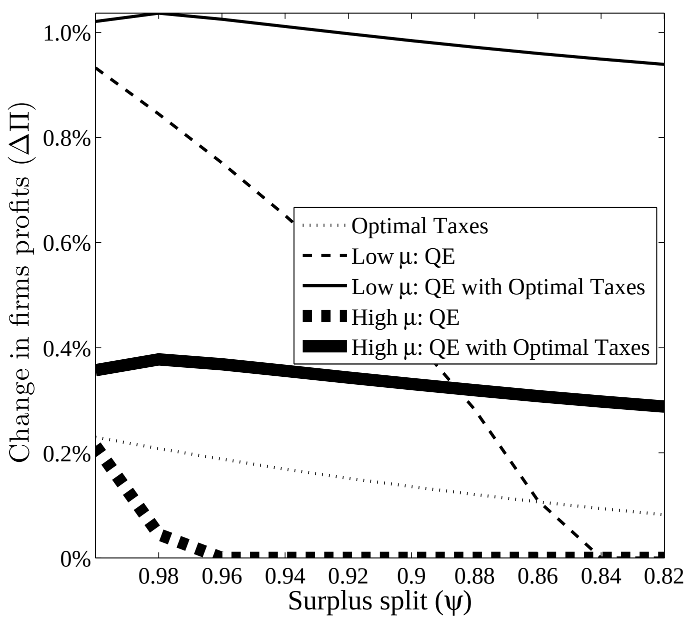
Note: For details see section 3.5.

Figure 7: Constrained Efficient Equilibrium



Note: For details see section 4.2.

Figure 8: Effect of Quantitative Easing



Note: For details see section 5.4.

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

David M. Arseneau
Federal Reserve Board

David E. Rappoport
Federal Reserve Board

Alexandros P. Vardoulakis
Federal Reserve Board

Online Appendix

Proofs and Derivations

Proof of Theorem 1: *We need to show that there is a unique equilibrium, and that in this equilibrium credit is not rationed. For that, first, we establish that the privately optimal contract is an interior solution to the firm's optimization problem (Part 1). Then, we establish existence of equilibria (Parts 2). Finally, we establish uniqueness (Part 3).*

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* $\bar{\omega} > 0$

$$\begin{gathered}
\Gamma(\bar{\omega}) > 0\,, \qquad 1-\Gamma(\bar{\omega}) = \mathbb{P}(\omega \geq \bar{\omega})\mathbb{E}[\omega-\bar{\omega}|\omega\geq\bar{\omega}] > 0 \\
1 > \Gamma'(\bar{\omega}) = 1-F(\bar{\omega}) > 0\,, \qquad \Gamma''(\bar{\omega}) = -dF(\bar{\omega}) < 0 \\
0 < G(\bar{\omega}) < 1\,, \qquad \mu G(\bar{\omega}) < G(\bar{\omega}) < \Gamma(\bar{\omega}) \\
G'(\bar{\omega}) = \bar{\omega}dF(\bar{\omega}) > 0\,, \qquad G''(\bar{\omega}) = dF(\bar{\omega}) + \bar{\omega}\frac{d(dF(\bar{\omega}))}{d\bar{\omega}} \\
\lim_{\bar{\omega}\to 0}\Gamma(\bar{\omega}) = 0\,, \qquad \lim_{\bar{\omega}\to\infty}\Gamma(\bar{\omega}) = \bar{\omega}\mathbb{P}(\omega\geq\bar{\omega}) + \mathbb{P}(\omega<\bar{\omega})\mathbb{E}[\omega|\omega<\bar{\omega}] = 1 \\
\lim_{\bar{\omega}\to 0}G(\bar{\omega}) = 0 \qquad \textit{and} \qquad \lim_{\bar{\omega}\to\infty}G(\bar{\omega}) = 1\,.
\end{gathered} \tag{A.1}$$

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,*

$$\Gamma'(\bar{\omega}) - \mu G'(\bar{\omega}) = (1-F(\bar{\omega}))\left(1-\mu\frac{\bar{\omega}dF(\bar{\omega})}{1-F(\bar{\omega})}\right) \quad \begin{cases} > 0 & \textit{if } \bar{\omega} < \bar{\bar{\omega}} \\ = 0 & \textit{if } \bar{\omega} = \bar{\bar{\omega}} \\ < 0 & \textit{if } \bar{\omega} > \bar{\bar{\omega}} \end{cases}.$$

The value of the firm, $[1-\Gamma(\bar{\omega})]\,R^k l_0$*, is increasing in leverage,* l_0*, and decreasing in risk,* $\bar{\omega}$*. In addition, if investors' expected (hold-to-maturity) return is* $R \in [(1+r)^2, R^k]$*, then* $R = R^b(l_0, \bar{\omega})$

imply that

$$l_0^{ier}(\bar{\omega}) = \frac{R}{R - R^k[\Gamma(\bar{\omega}) - \mu G(\bar{\omega})]} \ . \tag{A.2}$$

Since $\Gamma(\bar{\omega}) - \mu G(\bar{\omega})$ *attains a maximum at* $\bar{\bar{\omega}}$

$$l_0^{ier}(\bar{\omega}) \leq \frac{R}{R - R^k[\Gamma(\bar{\bar{\omega}}) - \mu G(\bar{\bar{\omega}})]} \leq \frac{(1+r)^2}{(1+r)^2 - R^k[\Gamma(\bar{\bar{\omega}}) - \mu G(\bar{\bar{\omega}})]} \equiv \bar{\bar{l}}_0$$

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* $1 \leq l_0 \leq \bar{\bar{l}}_0$.

To establish the properties of the optimal contract, it will be useful to consider the following contract problem

$$\max_{l_0, \bar{\omega}} [1 - \Gamma(\bar{\omega})] R^k l_0 \tag{A.3}$$

$$s.t. \qquad R^b(l_0, \bar{\omega}) = R \ , \qquad 1 \leq l_0 \leq \bar{\bar{l}}_0 \qquad and \qquad 0 \leq \bar{\omega} \leq \bar{\bar{\omega}} \ .$$

Note that since the firm's objective is continuous the maximum is achieved in the closed set defined by the constraints. Now we want to establish that the maximum is interior, i.e. $0 < \bar{\omega} < \bar{\bar{\omega}}$ *or equivalently* $0 < l_0 < \bar{\bar{l}}_0$*. We write the Lagrangian for this problem as*

$$\mathcal{L} = [1 - \Gamma(\bar{\omega})] R^k l_0 - \lambda \left[R^b(l_0, \bar{\omega}) - R\right] - \check{\eta}_l[1 - l_0] - \hat{\eta}_l[l_0 - \bar{\bar{l}}_0] + \check{\eta}_\omega \bar{\omega} - \hat{\eta}_\omega[\bar{\omega} - \bar{\bar{\omega}}]$$

Then, the FOC are

$$(l_0) \qquad 0 = [1 - \Gamma(\bar{\omega})] R^k - \lambda \frac{\partial R^b}{\partial l_0} + \check{\eta}_l - \hat{\eta}_l$$

$$(\bar{\omega}) \qquad 0 = -\Gamma'(\bar{\omega}) R^k l_0 - \lambda \frac{\partial R^b}{\partial \bar{\omega}} + \check{\eta}_\omega - \hat{\eta}_\omega$$

with

$$\frac{\partial R^b}{\partial l_0} = -\frac{R^b}{l_0(l_0 - 1)} < 0 \qquad and \qquad \frac{\partial R^b}{\partial \bar{\omega}} = \frac{R^b[\Gamma'(\bar{\omega}) - \mu G'(\bar{\omega})]}{\Gamma(\bar{\omega}) - \mu G(\bar{\omega})} \ . \tag{A.4}$$

Suppose now $\bar{\omega} = 0$ *and* $l_0 = 1$ *then* $\check{\eta}_l, \check{\eta}_\omega > 0$*. Note that* $l_0^{ier}(0) = 1$*, then, from the FOC*

$$0 < \check{\eta}_l = -R^k - \lambda \frac{R^b}{l_0(l_0 - 1)} < 0 \ ,$$

which is a contradiction. So we conclude that $\bar{\omega} > 0$ *and* $l_0 > 1$*. Similarly, if* $\bar{\omega} = \bar{\bar{\omega}}$ *and* $l_0 = \bar{\bar{l}}_0$ *from the FOC*

$$0 < \hat{\eta}_\omega = -\Gamma'(\bar{\bar{\omega}}) R^k \bar{\bar{l}}_0 - \lambda \frac{R^b[\Gamma'(\bar{\bar{\omega}}) - \mu G'(\bar{\bar{\omega}})]}{\Gamma(\bar{\bar{\omega}}) - \mu G(\bar{\bar{\omega}})} < 0 \ ,$$

which is a contradiction. So we conclude that $\bar{\omega} < \bar{\bar{\omega}}$ *and* $l_0 < \bar{\bar{l}}_0$.

Thus, we conclude that there exist a solution to the modified contract problem (A.3) and this *solution is interior.*

Part 2. Existence of equilibria.

Let $\mathcal{J} : C \to \mathbb{R}$, *with* $C = [(1+r)^2, R^k]$. *For* $R \in C$, $\mathcal{J}(R)$ *is defined as follows. Given R define* $(l_0(R), \bar{\omega}(R))$ *as the solution to the optimization problem (A.3). Note that the solution to the contract problem is feasible, as* $\bar{\bar{l}}_0 < e_0/n_0 + 1$ *from Assumption 4.*

Use $(l_0(R), \bar{\omega}(R))$ *to calculate* $b_0(R) = n_0(l_0(R) - 1)$ *and* $s_0(R) = e_0 - b_0(R)$ *and* $\theta(R)$ *as*

$$\theta(R) = \frac{(1-\delta)s_0(R)(1+r)\Delta}{\delta b_0(R) R} .$$

Then,

$$\mathcal{J}(R) = \frac{U_s(\theta(R))}{u_b(\theta(R))}$$

Intuitively, for any hold-to-maturity two-period return R, the function $\mathcal{J}(R)$ *gives the hold-to-maturity return that makes investors indifferent between liquid storage and illiquid two-period bonds, given that firms optimally choose the contract given R and that the level of secondary market liquidity is consistent with the investors portfolios that support the optimal firms' bond issuance* $b_0(R)$. *It follows that a fix point of* $\mathcal{J}$ *constitute a private equilibrium.*

Now we want to show that $\mathcal{J}$ *is a continuous single valued function and* $\mathcal{J}(C) \subset C$, *so* $\mathcal{J}$ *has a fixed point* $R = \mathcal{J}(R)$, *which constitute a non-rationing equilibrium.*

First, we show that $\mathcal{J}$ *is a single valued function. For that it suffices to show that the optimal contract as a function of R 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, is not convex, so we need to rule out that the firm's indifference curves and the investors' expected return condition,* $R = R^b(l_0, \bar{\omega})$, *intersect more than once.*

On the one hand, the firm's indifference curves are described by

$$l_0^{ic}(\bar{\omega}) = \frac{L}{R^k[1 - \Gamma(\bar{\omega})]} ,$$

where L is a constant that describes the level of profits at the indifference curve. Then,

$$\frac{dl_0^{ic}}{d\bar{\omega}} = \frac{L\,\Gamma'(\bar{\omega})}{R^k[1-\Gamma(\bar{\omega})]^2} .$$

On the other hand, differentiating (A.2) we get

$$\frac{dl_0^{ier}}{d\bar{\omega}} = \frac{RR^k[\Gamma'(\bar{\omega}) - \mu G'(\bar{\omega})]}{(R - R^k[\Gamma(\bar{\omega}) - \mu G(\bar{\omega})])^2} .$$

At the optimal contract, these two curves intersect. We use the condition that the two curves intersect to express L in terms R. In fact, $l_0^{ic} = l_0^{ier}$ *imply*

$$L = \frac{RR^k[1 - \Gamma(\bar{\omega})]}{R - R^k[\Gamma(\bar{\omega}) - \mu G(\bar{\omega})]} .$$

Moreover, these two curves have the same slope, i.e., $dl_0^{ic}/d\bar{\omega} = dl_0^{ibec}/d\bar{\omega}$*, if and only if*

$$\frac{\Gamma'(\bar{\omega})}{1-\Gamma(\bar{\omega})} = \frac{R^k[\Gamma'(\bar{\omega}) - \mu G'(\bar{\omega})]}{R - R^k[\Gamma(\bar{\omega}) - \mu G(\bar{\omega})]}$$

$$\frac{R^k - R}{\mu R^k} = G(\bar{\omega}) + \frac{[1-\Gamma(\bar{\omega})]G'(\bar{\omega})}{\Gamma'(\bar{\omega})} \equiv \mathcal{H}(\bar{\omega}) \; . \tag{A.5}$$

The function $\mathcal{H}(\bar{\omega})$ *is a strictly increasing function of* $\bar{\omega}$*. In fact,*

$$\mathcal{H}'(\bar{\omega}) = \frac{[1-\Gamma(\bar{\omega})]}{\Gamma'(\bar{\omega})^2}\left[\Gamma'(\bar{\omega})G''(\bar{\omega}) - G'(\bar{\omega})\Gamma''(\bar{\omega})\right] = \frac{[1-\Gamma(\bar{\omega})]}{\Gamma'(\bar{\omega})^2}\frac{d(\bar{\omega}h(\bar{\omega}))}{d\bar{\omega}}(1-F(\bar{\omega}))^2 > 0 \; .$$

Then we conclude that there is only one solution to the firms' maximization problem. Therefore, $(l_0(R), \bar{\omega}(R))$ *are single valued functions and so are* $b_0(R)$*,* $s_0(R)$*,* $\theta(R)$*, and* $\mathcal{J}(R)$*.*

It follows from above that $\mathcal{J}(R)$ *is continuous. In fact,* $\mathcal{H}(\bar{\omega})$ *is a continuous function with* $\mathcal{H}'(\bar{\omega}) > 0$*, so by the Implicit Function Theorem* $\bar{\omega}(R)$ *is a continuous strictly decreasing function in* $\mathcal{C}$*, i.e.,* $\bar{\omega}'(R) < 0$*. That is, the risk of the optimal contract is decreasing in the expected hold-to-maturity return offered to investors. Then, from the investors' expected return condition (A.2)* $l_0(R)$ *is a continuous function in* $\mathcal{C}$*, given that* $R = R^b > R^k[\Gamma(\bar{\omega}) - \mu G(\bar{\omega})]$*. It follows, that* $b_0(R)$ *and* $s_0(R)$ *are continuous in* $\mathcal{C}$*, and thus, that* $\theta(R)$ *and* $\mathcal{J}(R)$ *are continuous in* $\mathcal{C}$*.*

Now we show that $\mathcal{J}(R) \geq (1+r)^2$ *and* $\mathcal{J}(R) \leq R^k$*. From* $(1+r) \leq \Delta \leq \beta^{-1}$ *we have that*

$$\delta(1+r) + (1-\delta)(1+r)^2 \leq U_s(\theta(R)) \leq \delta(1+r) + (1-\delta)(1+r)\Delta$$

$$\text{and} \qquad \delta\beta + 1 - \delta \leq u_b(\theta(R)) \leq \delta\Delta^{-1} + 1 - \delta \; .$$

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,

$$\mathcal{J}(R) \leq \frac{\delta(1+r) + (1-\delta)(1+r)\Delta}{\delta\beta + 1 - \delta} \leq R^k \; .$$

On the other hand, since $\Delta \geq 1 + r$ *we have that*

$$\mathcal{J}(R) \geq \frac{\delta(1+r) + (1-\delta)(1+r)^2}{\delta\Delta^{-1} + 1 - \delta} \geq (1+r)^2 \; .$$

Part 3. Uniqueness: Show that $\mathcal{J}(R)$ *is decreasing in* $\mathcal{C}$*.*

Differentiating we obtain

$$\frac{d\mathcal{J}(R)}{dR} = \mathcal{J}(R)\left[\frac{1}{U_s(\theta(R))}\frac{dU_s(\theta(R))}{d\theta} - \frac{1}{u_b(\theta(R))}\frac{du_b(\theta(\bar{\omega}))}{d\theta}\right]\frac{d\theta(R)}{dR} \; .$$

To sign this derivative note that

$$\frac{dU_s}{d\theta} = (1-\delta)(1+r)p'(\theta)\,[\Delta - (1+r)] \leq 0\,,$$
$$\text{and} \qquad \frac{du_b}{d\theta} = \delta f'(\theta)\left[\Delta^{-1} - \beta\right] \geq 0\,. \tag{A.6}$$

where the inequalities follow from $p'(\theta) \leq 0$, $f'(\theta) \geq 0$, *and* $(1+r) \leq \Delta \leq \beta^{-1}$. *Thus, the term in square brackets is negative.*

We are left to show that $d\theta/dR > 0$. *For that note that*

$$\frac{d\theta}{dR} = \frac{\theta}{s_0}\frac{ds_0}{dR} - \frac{\theta}{b_0}\frac{db_0}{dR} - \frac{\theta}{R}\,.$$

In addition, from above we had that $\bar{\omega}'(R) < 0$ *and*

$$\frac{dl_0}{dR} = \frac{RR^k[\Gamma'(\bar{\omega}(R)) - \mu G'(\bar{\omega}(R))]\bar{\omega}'(R)}{(R - R^k[\Gamma(\bar{\omega}(R)) - \mu G(\bar{\omega}(R))])^2} < 0\,.$$

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

$$\frac{ds_0}{dR} = -\frac{db_0}{dR}\,, \qquad \frac{db_0}{dR} = n_0\frac{dl_0}{dR}\,, \qquad \text{and} \qquad \frac{dR^b}{dR} = 1 = \frac{\partial R^b}{\partial l_0}\frac{dl_0}{dR} + \frac{\partial R^b}{\partial \bar{\omega}}\frac{d\bar{\omega}}{dR}\,.$$

Then

$$\frac{R}{\theta}\frac{d\theta}{dR} = -\frac{dl_0}{dR}\left[\frac{\left(R - R^k[\Gamma(\bar{\omega}(R)) - \mu G(\bar{\omega}(R))]\right)^2}{R^k[\Gamma(\bar{\omega}(R)) - \mu G(\bar{\omega}(R))]} + \frac{R(e_0+n_0)}{l_0(e_0 - n_0(l_0-1))}\right] > 0\,. \tag{A.7}$$

Where we used that in an interior contract $l_0 < e_0/n_0 + 1$. ■

Proof of Proposition 1: *The aggregate credit supply is determined by the investors' break-even condition (15), from where it follows that*

$$\frac{dU_s}{d\theta}\left[\frac{\partial\theta}{\partial b_0^s}\frac{db_0^s}{dR} + \frac{\partial\theta}{\partial R}\right] = \frac{du_b}{d\theta}\left[\frac{\partial\theta}{\partial b_0^s}\frac{db_0^s}{dR} + \frac{\partial\theta}{\partial R}\right]R + u_b\,.$$

Then,

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

From the definition of market thickness (17)

$$\frac{\partial\theta}{\partial b_0^s} = -\frac{\theta}{b_0^s}\frac{e_0}{(e_0 - b_0^s)e_0} < 0 \qquad and \qquad \frac{\partial\theta}{\partial R} = -\frac{\theta}{R} ,$$

where the inequality follows from Assumption 4. This inequality and equation (A.6) imply that the denominator in the previous expression is non-negative. 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 $\theta > \overline{\theta}$ and $\psi = 1$, or $\theta < \underline{\theta}$ and $\psi = 0$. The first case corresponds to case 3 in Proposition 3, i.e., one of the conditions that make OTC trade irrelevant, which violates the assumption that $\psi < 1$ or $e_0 < \bar{\bar{e}}_0$. The second case corresponds to the case where the liquidity premium is fixed at $\beta^{-1}/(1+r) \geq 1$. But in this case $\theta < \underline{\theta}$ imply that

$$\frac{(1-\delta)(1+r)(e_0 - b_0)\beta^{-1}}{\delta b_0 R^k} \leq \frac{(1-\delta)(1+r)(e_0 - b_0)\beta^{-1}}{\delta b_0 R} < \min\{1, \nu^{\frac{1}{\alpha}}\} .$$

But then, rearranging and using that $b_0 \leq n_0(\bar{\bar{l}}_0 - 1)$ we get

$$e_0 \leq \left[\frac{\delta\beta R^k \min\{1, \nu^{\frac{1}{\alpha}}\}}{(1-\delta)(1+r)} + 1\right] b_0 \leq \left[\frac{\delta\beta R^k \min\{1, \nu^{\frac{1}{\alpha}}\}}{(1-\delta)(1+r)} + 1\right] n_0(\bar{\bar{l}}_0 - 1) ,$$

which is in contradiction with 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 is positive, i.e.,

$$\frac{\theta}{u_b}\frac{du_b}{d\theta} - \frac{\theta}{U_s}\frac{dU_s}{d\theta} < 1 ,$$

which follows from Lemma 1, i.e., $\left|\varepsilon_{\Phi^\ell,\theta}\right| < 1$. ■

Proof of Proposition 2: *Taking derivative wrt to $\bar{\omega}$ in equation (18) yields*

$$\frac{d\Phi^d(\bar{\omega})}{d\bar{\omega}} = \frac{1}{\Gamma(\bar{\omega}) - \mu G(\bar{\omega})} - \frac{\bar{\omega}[\Gamma'(\bar{\omega}) - \mu G'(\bar{\omega})]}{[\Gamma(\bar{\omega}) - \mu G(\bar{\omega})]^2} > 0$$

$$\Leftrightarrow \qquad \Gamma(\bar{\omega}) - \mu G(\bar{\omega}) - \bar{\omega}[\Gamma'(\bar{\omega}) - \mu G'(\bar{\omega})] > 0$$

$$\Leftrightarrow \qquad (1-\mu)G(\bar{\omega}) + \bar{\omega}\mu G'(\bar{\omega}) > 0 .$$

Where we used that $\Gamma(\bar{\omega}) = \bar{\omega}\Gamma'(\bar{\omega}) + G(\bar{\omega})$, and where the last inequality follows from $1 - \mu > 0$, $G(\bar{\omega}) \geq 0$ and $G'(\bar{\omega}) = \bar{\omega}dF(\bar{\omega}) > 0$, for any $\bar{\omega} > 0$. ■

Proof of Proposition 3: *When there is no need to compensate investors for liquidity risk, there is no liquidity premium, i.e., $\Phi^\ell(\theta) = 1$ and $R^b = (1+r)^2$. In other words, the expected return* *from lending to entrepreneurs is equal to the outside option of holding storage for two periods. Note that we can rewrite the previous condition as* $k_0 R^k \left[\Gamma(\bar{\omega}) - \mu G(\bar{\omega})\right] = (k_0 - n_0)(1+r)^2$, *which is the break-even condition in the benchmark costly state verification model. In addition, note that entrepreneurs' profits do not depend directly on secondary market liquidity. We proceed by showing that* $\Phi^\ell(\theta) = 1$ *under the three alternative condition stated in Proposition 3.*

Condition 1: $\delta = 0$. *This implies that secondary market liquidity* $\theta \to \infty$, *hence* $p(\theta) = 0$. *Setting* $\delta = 0$ *and* $p(\theta) = 0$ *yields* $\Phi^\ell(\theta) = 1$.

Condition 2: $\beta = (1+r)^{-1}$ *then* $\Delta = 1 + r$ *and simple substitution yields* $\Phi^\ell(\theta) = 1$.

Condition 3: $\psi = 1$ *and* $f(\theta) = 1$. *Simple substitution yields* $\Phi^\ell(\theta) = 1$. *We want to find* $\bar{e}_0$ *such that* $f(\theta) = 1$ *if* $e_0 \geq \bar{e}_0$.

$$f(\theta) = 1 \quad \Leftrightarrow \quad m(A,B) = A \quad \Leftrightarrow \quad \min\{\nu A^\alpha B^{1-\alpha}, B\} \geq A \quad \Leftrightarrow \quad \theta \geq \max\{\nu^{-1/(1-\alpha)}, 1\}$$

From (A.7) we have that $\theta \geq \theta\left((1+r)^2\right)$ *and since* $\psi = 1$ *we have that* $\Delta = 1 + r$, *so* $\theta \geq (1-\delta)\bar{\bar{s}}_0 / \left(\delta \bar{\bar{b}}_0\right)$, *where* $\bar{\bar{s}}_0 = e_0 - \bar{\bar{b}}_0$ *and* $\bar{\bar{b}}_0 = n_0\left(\bar{\bar{l}}_0 - 1\right)$ *with* $\bar{\bar{l}}_0$ *the upper bound on firm leverage defined in the proof of Theorem 1. We impose that* $(1-\delta)\bar{\bar{s}}_0 / \left(\delta \bar{\bar{b}}_0\right) \geq \max\left\{\nu^{-1/(1-\alpha)}, 1\right\}$ *to obtain a lower bound for the endowment of investors such that* $f(\theta) = 1$.

$$(1-\delta)\left(e_0 - n_0\left(\bar{\bar{l}}_0 - 1\right)\right) \geq \max\left\{1, \nu^{-1/(1-\alpha)}\right\} \delta n_0 \left(\bar{\bar{l}}_0 - 1\right)$$

$$e_0 \geq \bar{\bar{e}}_0 = \frac{n_0}{1-\delta}\left(\bar{\bar{l}}_0 - 1\right)\left(1 + \delta \max\left\{0, \nu^{-1/(1-\alpha)} - 1\right\}\right)$$

■

Proof of Lemma 1: *We want to show that the derivative of the liquidity premium wrt 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*

$$\frac{dU_s(\theta)}{d\theta} = (1-\delta)(1+r)\left[\Delta - (1+r)\right]\frac{dp(\theta)}{d\theta} \leq 0 \,,$$

$$\text{and} \qquad \frac{du_b(\theta)}{d\theta} = \delta\left[\Delta^{-1} - \beta\right]\frac{df(\theta)}{d\theta} \geq 0 \,,$$

where the inequalities follow from $\beta \leq 1/(1+r)$, *equations (5) and (6), and that the matching function* $m(A,B)$ *is increasing in both arguments. From equation (19) we have that*

$$\frac{d\Phi^\ell(\theta)}{d\theta} = \Phi^\ell(\theta)\left[\frac{1}{U_s(\theta)}\frac{dU_s(\theta)}{d\theta} - \frac{1}{u_b(\theta)}\frac{du_b(\theta)}{d\theta}\right] \leq 0 \,, \tag{A.8}$$

where the inequality follows from the previously established inequalities.

Regarding the second part of the Lemma, the elasticity of the liquidity premium, Φ^ℓ, *with* *respect to the secondary market liquidity, θ, is written, using equation (A.8), as:*

$$\varepsilon_{\Phi^\ell,\theta} = \frac{\theta}{\Phi^\ell}\frac{d\Phi^\ell}{d\theta} = \left[\frac{\theta}{U_s(\theta)}\frac{dU_s(\theta)}{d\theta} - \frac{\theta}{u_b(\theta)}\frac{du_b(\theta)}{d\theta}\right], \tag{A.9}$$

Then $\left|\varepsilon_{\Phi^\ell,\theta}\right| < 1$ requires:

$$-\left[\frac{\theta}{U_s(\theta)}\frac{dU_s(\theta)}{d\theta} - \frac{\theta}{u_b(\theta)}\frac{du_b(\theta)}{d\theta}\right] < 1$$

$$\Leftrightarrow \quad U_s(\theta)u_b(\theta) + \theta\frac{dU_s(\theta)}{d\theta}u_b(\theta) - \theta U_s(\theta)\frac{du_b}{d\theta} > 0 \tag{A.10}$$

First, lets consider the case where $\theta \in \left(\underline{\theta}, \overline{\theta}\right)$. In this case, $f(\theta) = \nu\theta^{1-\alpha}$ and $p(\theta) = \nu\theta^{-\alpha}$. Thus, $\theta(df(\theta)/d\theta) = (1-\alpha)f(\theta)$ and $\theta(dp(\theta)/d\theta) = -\alpha p(\theta)$. Then,

$$\frac{dU_s(\theta)}{d\theta}\theta = -\alpha U_s(\theta) + \alpha(1+r)[\delta + (1-\delta)(1+r)] \leq 0$$
$$\frac{du_b(\theta)}{d\theta}\theta = (1-\alpha)u_b(\theta) - (1-\alpha)[\beta\delta + (1-\delta)] \geq 0 .$$

Then,

$$\begin{aligned} & U_s(\theta)u_b(\theta) + \theta\frac{dU_s(\theta)}{d\theta}u_b(\theta) - \theta U_s(\theta)\frac{du_b}{d\theta} \\ &= U_s u_b + u_b\left\{-\alpha U_s + \alpha(1+r)[\delta + (1-\delta)(1+r)]\right\} - U_s\left\{(1-\alpha)u_b - (1-\alpha)[\beta\delta + (1-\delta)]\right\} \\ &= \alpha u_b(\theta)(1+r)[\delta + (1-\delta)(1+r)] + (1-\alpha)U_s(\theta)[\beta\delta + (1-\delta)] > 0 . \end{aligned}$$

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 $dp(\theta)/d\theta = dU_s(\theta)/d\theta = 0$. Want to show that $u_b(\theta) - \theta(du_b(\theta)/d\theta) > 0$. From above $du_b(\theta)/d\theta = \delta[\Delta^{-1} - \beta]$. Then,

$$u_b(\theta) - \theta\frac{du_b(\theta)}{d\theta} = \delta\beta + (1-\delta) > 0 .$$

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 $U_s(\theta) + \theta(dU_s(\theta)/d\theta) > 0$. From above, $\theta(dU_s(\theta)/d\theta) = -\theta^{-1}(1-\delta)(1+r)[\Delta - (1+r)]$. Then,

$$U_s(\theta) + \theta\frac{dU_s(\theta)}{d\theta} = \delta(1+r) + (1-\delta)(1+r)^2 > 0 . \qquad \blacksquare$$

Proof of Proposition 4: *For this proof we consider the liquidity premium a function of both market thickness, θ, and model parameters δ and β, and consider market thickness as a function of R, which we will take as given, and model parameters δ, β, and e_0. That is, we can write the* *liquidity premium as $\Phi^\ell(\theta, \delta, \beta)$.*

Case 1: Effect of δ. Want to show that

$$\left.\frac{d\Phi^\ell}{d\delta}\right|_R = \frac{\partial \Phi^\ell}{\partial \delta} + \frac{\partial \Phi^\ell}{\partial \theta}\left.\frac{d\theta}{d\delta}\right|_R > 0 \, .$$

From the definition of secondary market liquidity, given in equation (17) we have that

$$\left.\frac{d\theta}{d\delta}\right|_R = -\frac{\theta}{\delta(1-\delta)} \, .$$

Using this expression we get

$$\left.\frac{d\Phi^\ell}{d\delta}\right|_R = \frac{\partial \Phi^\ell}{\partial \delta} - \frac{\Phi^\ell}{\delta(1-\delta)}\varepsilon_{\Phi^\ell,\theta} \, ,$$

where $\varepsilon_{\Phi^\ell,\theta}$ is the elasticity of the liquidity premium with respect to secondary market liquidity, which is negative (Lemma 1), therefore, the second term is positive.

It is left to show that $\partial\Phi^\ell/\partial\delta > 0$. For that we compute the derivatives of $U_s(\theta, \delta, \beta)$ and $u_b(\theta, \delta, \beta)$ with respect to δ.

$$\frac{\partial U_s}{\partial \delta} = 1 + r - (1+r)\left[(1-p(\theta))(1+r) + p(\theta)\Delta\right] \quad \text{and} \quad \frac{\partial u_b}{\partial \delta} = \left[f(\theta)\Delta^{-1} + (1-f(\theta))\beta\right] - 1 \, .$$

Then, from equation (19) we have that

$$\frac{\partial \Phi^\ell}{\partial \delta} = \Phi^\ell\left[\frac{1}{U_s}\frac{\partial U_s}{\partial \delta} - \frac{1}{u_b}\frac{\partial u_b}{\partial \delta}\right] ,$$

which is strictly greater than zero if and only if $u_b(\partial U_s/\partial\delta) > U_s(\partial u_b/\partial\delta)$

$$\Leftrightarrow \quad \frac{\partial U_s}{\partial \delta}\left[\delta\frac{\partial u_b}{\partial \delta} + 1\right] > \left[\delta\frac{\partial U_s}{\partial \delta} + 1 + r - \frac{\partial U_s}{\partial \delta}\right]\frac{\partial u_b}{\partial \delta} \quad \Leftrightarrow \quad \frac{\partial U_s}{\partial \delta} > \left[1 + r - \frac{\partial U_s}{\partial \delta}\right]\frac{\partial u_b}{\partial \delta}$$

$$\Leftrightarrow \quad 1 + r - (1+r)\left[(1-p)(1+r) + p\Delta\right] > (1+r)\left[(1-p)(1+r) + p\Delta\right]\left\{\left[f\Delta^{-1} + (1-f)\beta\right] - 1\right\}$$

$$\Leftrightarrow \quad 1 > \left[(1-p)(1+r) + p\Delta\right]\left[f\Delta^{-1} + (1-f)\beta\right] \, .$$

It is easy to check that after distributing terms in the previous expression the four remaining terms, are a weighted average of terms strictly smaller than 1, with the weights given by the product of probabilities f and p adding up to 1. In fact, $\beta < 1/(1+r)$ imply that $\beta(1+r) < 1$, $\Delta^{-1}(1+r) < 1$, and $\Delta\beta < 1$.

Case 2: Effect of β. Want to show that

$$\left.\frac{d\Phi^\ell}{d\beta}\right|_R = \frac{\partial \Phi^\ell}{\partial \beta} + \frac{\partial \Phi^\ell}{\partial \theta}\left.\frac{\partial \theta}{\partial \beta}\right|_R < 0 \, .$$

From the definition of secondary market liquidity, given in equation (17) we have that

$$\left.\frac{\partial\theta}{\partial\beta}\right|_R = -\frac{\theta}{q_1}\frac{\partial q_1}{\partial\beta} = -\theta(1-\psi)\Delta \ .$$

Thus,

$$\left.\frac{d\Phi^\ell}{d\beta}\right|_R = \frac{\partial\Phi^\ell}{\partial\beta} - (1-\psi)\Delta\theta\frac{\partial\Phi^\ell}{\partial\theta} \ .$$

But the sign of the right-hand side term is ambiguous. The reason is that a higher β, on one hand, reduces the preference for liquidity by impatient households, i.e., $\partial\Phi^\ell/\partial\beta < 0$. But, on the other hand, it increases the secondary market price, q_1, which pushes market thickness θ down and liquidity premia up. This second force, represented by the second term depends crucially on the bargaining power of impatient investors: the lower their bargaining power the more important the effect of their valuation, i.e., β, will be on the price.

To show that the term in the right-hand side is negative, we use that

$$\frac{\partial\Phi^\ell}{\partial\beta} = \Phi^\ell\left[\frac{1}{U_s}\frac{\partial U_s}{\partial\beta} - \frac{1}{u_b}\frac{\partial u_b}{\partial\beta}\right] ,$$

with the derivatives of $U_s(\theta,\delta,\beta)$ and $u_b(\theta,\delta,\beta)$ with respect to β given by

$$\frac{\partial U_s}{\partial\beta} = -(1-\delta)(1+r)p(\theta)\Delta^2(1-\psi) < 0 ,$$

$$\text{and} \qquad \frac{\partial u_b}{\partial\beta} = \delta[f(\theta)(1-\psi) + 1 - f(\theta)] = \delta(1-\psi f(\theta)) > 0 ,$$

where the inequalities follow from our assumption about δ, ψ, and $f(\theta)$. Then, $\partial\Phi^\ell/\partial\beta < 0$. So we need to show that

$$(1-\psi)\Delta\theta\left[u_b\frac{\partial U_s}{\partial\theta} - U_s\frac{\partial u_b}{\partial\theta}\right] - u_b\frac{\partial U_s}{\partial\beta} + U_s\frac{\partial u_b}{\partial\beta} > 0 \ .$$

Using the expressions derived above for these derivatives the previous expression equals

$$u_b(1-\psi)\Delta\alpha(1+r)\left\{[\Delta-(1+r)]\theta\left.\frac{\partial p}{\partial\theta}\right|_{b_0} + p\Delta\right\} + U_s\delta(1-\psi)\left\{f - [1-\Delta\beta]\theta\left.\frac{\partial f}{\partial\theta}\right|_{b_0}\right\} + U_s\delta(1-f) > 0 ,$$

where the inequality follows from the fact that the terms in curly brackets are positive. In fact, when $\theta \in (\underline{\theta}, \overline{\theta})$, then $\theta(\partial p/\partial\theta)|_{b_0} = -\alpha p$ and $\theta(\partial f/\partial\theta)|_{b_0} = (1-\alpha)f$ so we have

$$[\Delta-(1+r)]\theta\left.\frac{\partial p}{\partial\theta}\right|_{b_0} + p\Delta = (1-\alpha)p\Delta + \alpha p(1+r) > 0 ,$$

$$\text{and} \qquad f - [1-\Delta\beta]\theta \left.\frac{\partial f}{\partial \theta}\right|_{b_0} = \alpha f + (1-\alpha) f \Delta\beta > 0\,.$$

When $\theta < \underline{\theta}$, then $(\partial p/\partial\theta)|_{b_0} = 0$ and $\theta(\partial f/\partial\theta)|_{b_0} = f$ so we have

$$[\Delta - (1+r)]\theta \left.\frac{\partial p}{\partial \theta}\right|_{b_0} + p\Delta = p\Delta > 0 \qquad \text{and} \qquad f - [1-\Delta\beta]\theta \left.\frac{\partial f}{\partial \theta}\right|_{b_0} = f\Delta\beta > 0\,.$$

Finally, when $\theta > \overline{\theta}$, then $\theta(\partial p/\partial\theta)|_{b_0} = -p$ and $(\partial f/\partial\theta)|_{b_0} = 0$ so we have

$$[\Delta - (1+r)]\theta \left.\frac{\partial p}{\partial \theta}\right|_{b_0} + p\Delta = (1+r)p > 0 \qquad \text{and} \qquad f - [1-\Delta\beta]\theta \left.\frac{\partial f}{\partial \theta}\right|_{b_0} = f > 0\,.$$

Case 3: Effect of e_0. Want to show that $d\Phi^\ell/de_0|_R < 0$. Note that investors' endowment e_0 affects liquidity premium Φ^ℓ only through its effect on secondary market liquidity θ. In particular, it has an effect only through $s_0 = e_0 - b_0$ given that we consider bond issuance as given. Thus,

$$\left.\frac{d\Phi^\ell}{de_0}\right|_R = \frac{\partial \Phi^\ell}{\partial \theta}\frac{\partial \theta}{\partial s_0}\frac{ds_0}{de_0} = \frac{\partial \Phi^\ell}{\partial \theta}\frac{\theta}{s_0} < 0\,,$$

where the inequality follows from Lemma 1 and Proposition 1. ■

Proof of Proposition 6: *From the proof of Theorem 1 we know that the equilibrium is described by the fixed point of the function $\mathcal{J}$, i.e, $\mathcal{J}(R) = R$. Then, considering a generic model parameter ϱ we can express the equilibrium of the model as $\mathcal{J}(R,\varrho) - R = 0$. By the Implicit Function Theorem, if the derivative of the previous expression with respect to R is different than 0, then we can define $R(\varrho)$ and calculate its derivative from the previous expression. Recall that $\mathcal{J}(R) = U_s(\theta(R))/u_b(\theta(R))$, then $d\mathcal{J}/dR = (\mathcal{J}/\theta)\varepsilon_{\Phi^\ell,\theta}(d\theta/dR) < 0$, where $\varepsilon_{\Phi^\ell,\theta}$ the elasticity of the liquidity premium with respect to market thickness. The inequality follows from $d\theta/dR > 0$ (equation (A.7)), 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

$$\frac{dR}{d\varrho} = -\left[\frac{\partial \mathcal{J}}{\partial R} - 1\right]^{-1}\frac{\partial \mathcal{J}}{\partial \varrho}\,.$$

From where we conclude that the sign of $dR/d\varrho$ equals the sign of $\partial\mathcal{J}/\partial\varrho$. Note that $\partial\mathcal{J}/\partial\varrho = d\mathcal{J}/d\varrho|_R$ and thus have the same sign. Moreover, $\mathcal{J} = \Phi^\ell(1+r)^2$ so the sign of $d\mathcal{J}/d\varrho|_R$ and the sign of $d\Phi^\ell/d\varrho|_R$ are equal, with the latter established in Proposition 4. Now we consider the effect of each of model parameters.
Case 1: Comparative Statics on δ. From Proposition 4 $d\Phi^\ell/d\delta|_R > 0$, so we conclude that $dR/d\delta > 0$.
Case 2: Comparative Statics on β. From Proposition 4 $d\Phi^\ell/d\beta|_R > 0$, so we conclude that $dR/d\beta > 0$.
Case 3: Comparative Statics on e_0. From Proposition 4 $d\Phi^\ell/de_0|_R > 0$, so we conclude that

$dR/de_0 > 0$.
Case 4: Comparative Statics on n_0. From Proposition 5 $db_0/dn_0 > 0$. In this case as the demand for credit increases, and investors' portfolio become more illiquid, investors need to receive a higher compensation to buy those bonds in equilibrium, i.e., as bond issuance increase the expected hold-to-maturity return increases (Proposition 1). So we conclude that $dR/dn_0 > 0$.

In all cases as the expected hold-to-maturity return increases, the characteristics of the optimal contract decrease, i.e., leverage l_0 and risk $\bar{\omega}$ decrease, and thus, the default premium increases. Finally, in the first three cases given credit demand remains fixed, as the expected return increases bond issuance decreases, whereas in the fourth case the total amount of bond issuance depends on the relative strength of two effect. The increase in bond issuance from the increase in firms' endowments and the decrease in bond issuance from the decrease in firms' leverage. ■

Proof of Proposition 7: *We want to show that if the private equilibrium is constrained efficient, then $(\alpha, \psi, r) \in \emptyset$, a set of measure zero.*

Suppose $(l_0^{pe}, \bar{\omega}^{pe}, \theta^{pe}, q_1^{pe})$, the private equilibrium, is constrained efficient. Since $(l_0^{pe}, \bar{\omega}^{pe}, \theta^{pe}, q_1^{pe})$ is a private equilibrium the investor break-even condition (15) holds, i.e., $U_s = U_b$, and from equation (16) it must be that

$$\frac{1-\Gamma(\bar{\omega}^{pe})}{l_0^{pe}\Gamma'(\bar{\omega}^{pe})} = -\frac{\partial U_b/\partial l_0}{\partial U_b/\partial \bar{\omega}} \, .$$

On the other hand, since $(l_0^{pe}, \bar{\omega}^{pe}, \theta^{pe}, q_1^{pe})$ is constrained efficient, from equation (23) it must be that

$$\frac{[1-\Gamma(\bar{\omega}^{pe})]}{l_0^{pe}\Gamma'(\bar{\omega}^{pe})} = -\frac{n_0(U_b - U_s) + b_0^{pe}\frac{\partial U_b}{\partial l_0} + \frac{\partial \mathbb{U}}{\partial \theta}\frac{\partial \theta}{\partial l_0}}{b_0^{pe}\frac{\partial U_b}{\partial \bar{\omega}} + \frac{\partial \mathbb{U}}{\partial \theta}\frac{\partial \theta}{\partial \bar{\omega}}} \, .$$

Using that $U_s = U_b$, then

$$\frac{b_0^{pe}\frac{\partial U_b}{\partial l_0} + \frac{\partial \mathbb{U}}{\partial \theta}\frac{\partial \theta}{\partial l_0}}{b_0^{pe}\frac{\partial U_b}{\partial \bar{\omega}} + \frac{\partial \mathbb{U}}{\partial \theta}\frac{\partial \theta}{\partial \bar{\omega}}} = \frac{\frac{\partial U_b}{\partial l_0}}{\frac{\partial U_b}{\partial \bar{\omega}}} \, ,$$

which is the case iff

$$\frac{\partial \mathbb{U}}{\partial \theta}\left[\frac{\partial U_b}{\partial \bar{\omega}}\frac{\partial \theta}{\partial l_0} - \frac{\partial U_b}{\partial l_0}\frac{\partial \theta}{\partial \bar{\omega}}\right] = 0 \, . \tag{A.11}$$

Note that,

$$\frac{\partial U_b}{\partial \bar{\omega}}\frac{\partial \theta}{\partial l_0} - \frac{\partial U_b}{\partial l_0}\frac{\partial \theta}{\partial \bar{\omega}} < 0 \, , \tag{A.12}$$

since

$$\frac{\partial U_b}{\partial l_0} = -\frac{U_b}{l_0(l_0-1)} < 0 \qquad \text{and} \qquad \frac{\partial U_b}{\partial \bar{\omega}} = \frac{U_b[\Gamma'(\bar{\omega}) - \mu G'(\bar{\omega})]}{\Gamma(\bar{\omega}) - \mu G(\bar{\omega})} > 0 \, , \tag{A.13}$$

where the last inequality follows from Theorem 1, and $\partial\theta/\partial l_0, \partial\theta/\partial\bar{\omega} < 0$ from equation (17).

Then, A.11 holds iff $\partial\mathbb{U}/\partial\theta = 0$, which is the case iff

$$s_0^{pe}\frac{\partial U_s}{\partial\theta} + b_0^{pe}\frac{\partial U_b}{\partial\theta} = 0$$

$$s_0^{pe}(1-\delta)(1+r)[\Delta-(1+r)]p'(\theta^{pe}) + b_0^{pe}\delta[\Delta^{-1}-\beta]f'(\theta^{pe})R^b = 0$$

$$p(\theta^{pe})\frac{\alpha}{\theta^{pe}}s_0^{pe}(1-\delta)(1+r)[\Delta-(1+r)] = f(\theta^{pe})\frac{1-\alpha}{\theta^{pe}}b_0^{pe}\delta[\Delta^{-1}-\beta]R^b$$

$$\alpha s_0^{pe}(1-\delta)(1+r)[\Delta-(1+r)] = \theta^{pe}(1-\alpha)b_0^{pe}\delta[\Delta^{-1}-\beta]R^b \ .$$

But from equation (17) $\theta^{pe} = (1-\delta)(1+r)\Delta s_0^{pe}/(\delta b_0^{pe}R^b)$, then

$$\alpha[\Delta-(1+r)] = (1-\alpha)\Delta[\Delta^{-1}-\beta] = (1-\alpha)[1-\Delta\beta]$$

$$\Leftrightarrow \qquad \Delta[\alpha+(1-\alpha)\beta] = 1+\alpha r \qquad \Leftrightarrow \qquad \frac{\psi}{1+r} + (1-\psi)\beta = \frac{\alpha+(1-\alpha)\beta}{1+\alpha r}$$

$$\Leftrightarrow \qquad \psi = \frac{\alpha(1-\beta(1+r))}{1+\alpha r}\frac{1+r}{(1-\beta(1+r))} \qquad \Leftrightarrow \qquad \psi(1+\alpha r) = \alpha(1+r) \ . \tag{A.14}$$

The set of (α,ψ,r) satisfying (A.14) is, thus, of measure zero. ■

Proof of Proposition 8:

Part 1. The sign of the externality determines the socially optimal level of secondary market liquidity.

Let $\mathcal{L}$ be the Lagrangian of the planner's problem, which is given

$$\mathcal{L} = [1-\Gamma(\bar{\omega})]R^k l_0 - \lambda[\mathbb{U}^{pe} - s_0U_s - b_0U_b] \ ,$$

Fully differentiating and evaluating at the private equilibrium allocation $(l_0^{pe}, \bar{\omega}^{pe}, \theta^{pe})$ we have

$$d\mathcal{L}(l_0^{pe}, \bar{\omega}^{pe}, \theta^{pe}) = \lambda\frac{\partial\mathbb{U}}{\partial\theta}d\theta \ ,$$

where we have substituted the optimality conditions in the private equilibrium. Thus, the planner, who internalizes the effect of liquidity on the investor's utility, would like to increase liquidity in secondary markets when the externality is positive, i.e., $\partial\mathbb{U}/\partial\theta > 0$, and decrease liquidity if the externality is negative, i.e., $\partial\mathbb{U}/\partial\theta < 0$.

Part 2. Show that the sign of the externality depends on the relationship between the parameters (α, r, ψ).

Want to show that

$$\psi(1+\alpha r) > \alpha(1+r) \qquad \Leftrightarrow \qquad \frac{\partial\mathbb{U}}{\partial\theta} > 0 \ .$$

In fact,

$$\psi(1+\alpha r) > \alpha(1+r) \qquad \Leftrightarrow \qquad \Delta > \frac{\alpha[\Delta-(1+r)]}{(1-\alpha)[\Delta^{-1}-\beta]}$$

$$\Leftrightarrow \qquad \theta > \frac{\alpha s_0(1-\delta)(1+r)[\Delta-(1+r)]}{(1-\alpha)b_0\delta[\Delta^{-1}-\beta]R^b}$$

$$\Leftrightarrow \qquad b_0\frac{\partial U_b}{\partial\theta}+s_0\frac{\partial U_s}{\partial\theta} > 0 \qquad \Leftrightarrow \qquad \frac{\partial \mathbb{U}}{\partial\theta} > 0.$$

Part 3. Characterization of the efficient contract.

Let $\bar{\omega}^{pi}(l_0)$ be the function implicitly defined by the Pareto improvement constraint in the planner's problem (20). Using the Implicit Function Theorem we have that

$$\frac{d\bar{\omega}^{pi}}{dl_0} = -\frac{\frac{\partial\mathbb{U}}{\partial l_0}+\frac{\partial\mathbb{U}}{\partial\theta}\frac{\partial\theta}{\partial l_0}}{\frac{\partial\mathbb{U}}{\partial\bar{\omega}}+\frac{\partial\mathbb{U}}{\partial\theta}\frac{\partial\theta}{\partial\bar{\omega}}}\ .$$

Similarly, using the notation introduced in the proof of Theorem 1, let $\bar{\omega}^{ier}(l_0)$ denotes the function implicitly defined by the investors' expected return condition $R^b(l_0,\bar{\omega}) = U_s(\theta)/u_b(\theta)$ for $\bar{\omega} < \bar{\bar{\omega}}$, we have that

$$\frac{d\bar{\omega}^{ier}}{dl_0} = -\frac{\frac{\partial R^b}{\partial l_0}}{\frac{\partial R^b}{\partial\bar{\omega}}} = -\frac{\frac{\partial U_b}{\partial l_0}}{\frac{\partial U_b}{\partial\bar{\omega}}}\ . \tag{A.15}$$

Note that the private equilibrium is a feasible point of the pareto improvement constraint, so $\bar{\omega}^{pi}(l_0^{pe}) = \bar{\omega}^{ier}(l_0^{pe})$. Moreover, note that

$$\frac{d\bar{\omega}^{pi}(l_0^{pe})}{dl_0}-\frac{d\bar{\omega}^{ier}(l_0^{pe})}{dl_0} = \frac{\frac{\partial\mathbb{U}}{\partial\theta}\left[\frac{\partial\theta}{\partial\bar{\omega}}\frac{\partial U_b}{\partial l_0}-\frac{\partial\theta}{\partial l_0}\frac{\partial U_b}{\partial\bar{\omega}}\right]}{\frac{\partial U_b}{\partial\bar{\omega}}\left[b_0^{pe}\frac{\partial U_b}{\partial\bar{\omega}}+\frac{\partial\mathbb{U}}{\partial\theta}\frac{\partial\theta}{\partial\bar{\omega}}\right]}\ ,$$

where all the derivatives on the RHS are evaluated at $(l_0^{pe},\bar{\omega}^{pe},\theta^{pe})$, and we used that

$$\frac{\partial\mathbb{U}(l_0^{pe},\bar{\omega}^{pe},\theta^{pe})}{\partial l_0} = n_0(U_b(l_0^{pe},\bar{\omega}^{pe},\theta^{pe})-U_s(\theta^{pe}))+b_0^{pe}\frac{\partial U_b(l_0^{pe},\bar{\omega}^{pe},\theta^{pe})}{\partial l_0} = b_0^{pe}\frac{\partial U_b(l_0^{pe},\bar{\omega}^{pe},\theta^{pe})}{\partial l_0}\ .$$

Note that

$$b_0^{pe}\frac{\partial U_b}{\partial\bar{\omega}}+\frac{\partial\mathbb{U}}{\partial\theta}\frac{\partial\theta}{\partial\bar{\omega}} = b_0^{pe}\frac{u_bR^b[\Gamma'-\mu G']}{\Gamma-\mu G}+\left[s_0^{pe}\frac{dU_s}{d\theta}+b_0^{pe}\frac{du_b}{d\theta}R^b\right]\frac{\partial\theta}{\partial\bar{\omega}}\ .$$

And using that $s_0^{pe} = e_0-b_0^{pe}$, $R^b(l_0^{pe},\bar{\omega}^{pe}) = U_s(\theta^{pe})/u_b(\theta^{pe})$ and

$$\frac{\partial\theta}{\partial\bar{\omega}} = -\frac{\theta[\Gamma'-\mu G']}{\Gamma-\mu G}\ ,$$

we obtain that

$$b_0^{pe}\frac{\partial U_b}{\partial \bar{\omega}} + \frac{\partial \mathbb{U}}{\partial \theta}\frac{\partial \theta}{\partial \bar{\omega}} = \frac{b_0^{pe}[\Gamma' - \mu G']R^b}{U_s[\Gamma - \mu G]}\left\{U_s u_b + \theta^{pe}\frac{dU_s}{d\theta}u_b - \theta^{pe}U_s\frac{du_b}{d\theta}\right\} + e_0\frac{dU_s}{d\theta}\frac{\partial \theta}{\partial \bar{\omega}} > 0\,, \quad \text{(A.16)}$$

where the inequality follows from (A.10) and $\partial U_s/\partial\theta,\ \partial\theta/\partial\bar{\omega} < 0$. It follows from the previous inequality and equations (A.12) and (A.13) that

$$\frac{d\bar{\omega}^{pi}(l_0^{pe})}{dl_0} - \frac{d\bar{\omega}^{ier}(l_0^{pe})}{dl_0} > 0 \qquad \Leftrightarrow \qquad \frac{\partial \mathbb{U}}{\partial \theta} > 0\,.$$

Then, if $\psi(1+\alpha r) > \alpha(1+r)$, from Part 2, $\partial\mathbb{U}/\partial\theta > 0$, and, thus,

$$\frac{d\bar{\omega}^{pi}(l_0^{pe})}{dl_0} > \frac{d\bar{\omega}^{ier}(l_0^{pe})}{dl_0} > 0\,,$$

where the last inequality follows from equation (A.15). That means there are points, along the weak Pareto improving constraint, that are feasible for the planner where $(l_0,\bar{\omega}) << (l_0^{pe},\bar{\omega}^{pe})$ and the firm achieves higher profits, so the planner will choose an allocation with lower leverage and risk. (Note that this imply that the planer will set a higher secondary market liquidity: $\theta > \theta^{pe}$. In fact, lower leverage and risk require a lower equilibrium interest rate in the credit market which yields a higher level of market liquidity).

Similarly, if $\psi(1+\alpha r) < \alpha(1+r)$, from Part 2, $\partial\mathbb{U}/\partial\theta < 0$, so

$$0 < \frac{d\bar{\omega}^{pi}(l_0^{pe})}{dl_0} < \frac{d\bar{\omega}^{ier}(l_0^{pe})}{dl_0}\,.$$

That means there are points that are feasible for the planner where $(l_0,\bar{\omega}) >> (l_0^{pe},\bar{\omega}^{pe})$ and firms enjoy higher profits, so the planner will choose an allocation with higher leverage and risk, and lower market liquidity $\theta < \theta^{pe}$. ■

Proof of Proposition 9:

Part 1. Deriving the tax instruments.

The firm's problem with taxes on storage and leverage can be written as

$$\max_{l_0,\bar{\omega}}\ [1-\Gamma(\bar{\omega})]R^k l_0 - \tau^l\lambda^{pe}l_0 + T^l$$

subject to

$$U_b = (1-\tau^s)U_s\,. \quad \text{(A.17)}$$

We write the Lagrangian for this problem as

$$\mathcal{L} = [1-\Gamma(\bar{\omega})]R^k l_0 - \tau^l\lambda^{pe}l_0 + T^l - \lambda^{pe}[(1-\tau^s)U_s - U_b]\,.$$

Then, the optimality conditions are

$$[1-\Gamma(\bar{\omega})]R^k = \tau^l \lambda^{pe} - \lambda^{pe}\frac{\partial U_b}{\partial l_0} \ ,$$

$$\textit{and} \qquad \Gamma'(\bar{\omega})R^k l_0 = \lambda^{pe}\frac{\partial U_b}{\partial \bar{\omega}} \tag{A.18}$$

Note that the FOC for $(\bar{\omega})$, equation (A.18), together with equation (A.13) ensures that $\lambda^{pe} > 0$, which is not necessarily the case with equality constraints. And the optimal contract is described by

$$\frac{1-\Gamma(\bar{\omega})}{l_0\Gamma'(\bar{\omega})} = -\frac{\frac{\partial U_b}{\partial l_0} - \tau^l}{\frac{\partial U_b}{\partial \bar{\omega}}} \ . \tag{A.19}$$

Equating the previous expression and equation (23), and using that $U_b - U_s = -\tau^s U_s$, we derive the tax on leverage:

$$\tau^l = \frac{n_0 U_s \frac{\partial U_b}{\partial \bar{\omega}}\tau^s + \left[\frac{\partial U_b}{\partial l_0}\frac{\partial \theta}{\partial \bar{\omega}} - \frac{\partial U_b}{\partial \bar{\omega}}\frac{\partial \theta}{\partial l_0}\right]\frac{\partial \mathbb{U}}{\partial \theta}}{b_0\frac{\partial U_b}{\partial \bar{\omega}} + \frac{\partial \theta}{\partial \bar{\omega}}\frac{\partial \mathbb{U}}{\partial \theta}} \ .$$

The term in square brackets is positive from equation (A.12). In addition, the denominator is positive from equation (A.16).

On the other hand, the break-even condition of investors with a tax on storage was given by equation (A.17). Combining it with constraint (20) we derive the tax on storage:

$$\tau^s = \frac{e_0}{b_0}\left(1 - \frac{U_s(\theta^{pe})}{U_s(\theta)}\right) \ .$$

Part 2. Signing the tax on storage.

If $\psi(1+\alpha r) > \alpha(1+r)$ then from Proposition 8 the planner wants to increase secondary market liquidity so $\theta > \theta^{pe}$. Thus, the storage technology is subsidized: $\tau^s \leq 0$. In fact, the tax on storage is strictly negative from equation (24) if $\psi < 1$ and is zero if $\psi = 0$.

On the contrary, if $\psi(1+\alpha r) < \alpha(1+r)$, then the externality is negative, the planner wants to reduce secondary market liquidity, and, therefore, $\tau^s > 0$.

Part 3. Signing the tax on leverage.

We start by describing the feasible allocations for a firm that chooses the optimal contract and faces the optimal tax on storage, and the efficient level of secondary market liquidity. That is, τ^s is given by equation (24) and θ is the one that the planner would choose optimally. In this case we have

$$(1-\tau^s)U_s(\theta) = \left(1 - \frac{e_0}{b_0}\frac{U_s(\theta) - U_s(\theta^{pe})}{U_s(\theta)}\right)U_s(\theta) = \frac{b_0 U_b(l_0^{pe}, \bar{\omega}^{pe}, \theta^{pe}) + s_0 U_s(\theta^{pe}) - s_0 U_s(\theta)}{b_0}$$

where we used that in the private equilibrium $U_s(\theta^{pe}) = U_b(l_0^{pe}, \bar{\omega}^{pe}, \theta^{pe})$, and $b_0^{pe} + s_0^{pe} = e_0$.

Lets consider first the case when $\psi(1+\alpha r) > \alpha(1+r)$. In this case $\partial\mathbb{U}/\partial\theta > 0$ and $\theta > \theta^{pe}$, *then*

$$b_0 U_b(l_0^{pe}, \bar{\omega}^{pe}, \theta^{pe}) + s_0 U_s(\theta^{pe}) < b_0 U_b(l_0^{pe}, \bar{\omega}^{pe}, \theta) + s_0 U_s(\theta) \ .$$

So we conclude that

$$(1-\tau^s)U_s(\theta) < U_b(l_0^{pe}, \bar{\omega}^{pe}, \theta) \ .$$

Since $\partial U_b/\partial\bar{\omega} > 0$, *given the leverage* l_0^{pe} *a feasible level of risk will be lower than the risk in the private equilibrium* $\bar{\omega}^{pe}$. *So the investor's break-even condition with the optimal tax and the efficient level of liquidity will lie below the investor's break-even condition in the private problem. Moreover, from the mapping* $\bar{\omega}^{ier}(l_0)$ *the slope of the investor's break-even condition at* l_0^{pe}, *which has the same expression regardless of the tax, will be flatter.*

The firm, then, if it were to face this constraint without a tax on leverage will choose a higher leverage, at odds with the planner optimal prescriptions. The planner then will distort the firm's decision to disincentivize the use of leverage by levying a tax on leverage. One way to see this is that the planner will introduce a distortion such that the distorted isoprofit lines are flatter in the $(l_0, \bar{\omega})$*-space.*

Let $\Pi^\tau = [1-\Gamma(\bar{\omega})]R^k l_0 - \tau^l \lambda^{pe} l_0 + T^l$, *and denote by* $\bar{\omega}^{\Pi^\tau}(l_0)$ *the function that for any* l_0 *gives the associated risk level* $\bar{\omega}$ *along the taxed firm isoprofit line. Then, the Implicit Function Theorem implies that*

$$\frac{d\bar{\omega}^{\Pi^\tau}}{dl_0} = \frac{[1-\Gamma(\bar{\omega})]R^k - \tau^l \lambda^{pe}}{\Gamma'(\bar{\omega})R^k l_0} \ ,$$

so a flatter slope requires a positive τ^l. ■

Using the same reasoning we conclude that if $\psi(1+\alpha r) < \alpha(1+r)$, *then* $\tau^l < 0$.

Proof of Proposition 10: *In the presence of quantitative easing, firms' borrowing is given by* b_0, *whereas investors' final bond holdings are given by* $b_0 - \bar{b}_0$. *Then from the budget constraint of entrepreneurs we have that* $k_0 = n_0 + b_0$, *so investors' lending can be written in terms of entrepreneurs leverage and QE as* $n_0(l_0 - 1 - \bar{b}_0/n_0)$. *On the other hand, from the investors' budget constraint,* $b_0 - \bar{b}_0 + s_0 + \bar{s}_0 = e_0$, *so we can express the amount invested in the storage technology in terms of entrepreneurs leverage as* $s_0 = n_0(e_0/n_0 - (l_0 - 1))$. *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 central bank'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. Note that the number of sell orders is always equal to $A = \delta(b_0 - \bar{b}_0)$, *as impatient investors will put all their bond holdings for sale in the OTC market.*

If $\Delta > 1 + \bar{r}$ *patient investors pledge all their liquid assets to place buy orders in the OTC market so the number of buy orders* $B = (1-\delta)[(1+r)s_0 + (1+\bar{r})\bar{s}_0]/q_1$ *and market liquidity is given by*

$$\theta = \frac{(1-\delta)[(1+r)s_0 + (1+\bar{r})\bar{s}_0]}{\delta(b_0 - \bar{b}_0)q_1} = \frac{(1-\delta)\Delta\left[(1+r)\left(e_0 - n_0(l_0-1)\right) + (1+\bar{r})\bar{b}_0\right]}{\delta R^b\left(n_0(l_0-1) - \bar{b}_0\right)} \ . \tag{A.20}$$

Then,

$$\frac{\partial\theta}{\partial\bar{b}_0} = \frac{(1-\delta)\Delta(1+\bar{r})}{\delta R^b\left(n_0(l_0-1)-\bar{b}_0\right)} + \frac{(1-\delta)\Delta\left[(1+r)\left(e_0-n_0(l_0-1)\right)+(1+\bar{r})\bar{b}_0\right]}{\delta R^b\left(n_0(l_0-1)-\bar{b}_0\right)^2} > 0\,. \qquad \text{(A.21)}$$

On the other hand, 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 $B=(1-\delta)[(1+r)s_0-\delta/(1-\delta)(1+\bar{r})\bar{s}_0]/q_1$ and market liquidity is given by

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

Then,

$$\begin{aligned}\frac{\partial\theta}{\partial\bar{b}_0} &= -\frac{\Delta(1+\bar{r})}{R^b\left(n_0(l_0-1)-\bar{b}_0\right)} + \frac{(1-\delta)\Delta\left[(1+r)\left(e_0-n_0(l_0-1)\right)-\frac{\delta}{1-\delta}(1+\bar{r})\bar{b}_0\right]}{\delta R^b\left(n_0(l_0-1)-\bar{b}_0\right)^2} \\ &= \frac{(1-\delta)\Delta(1+r)\left[e_0-n_0(l_0-1)\left(1+\frac{\delta}{1-\delta}\frac{1+\bar{r}}{1+r}\right)\right]}{\delta R^b\left(n_0(l_0-1)-\bar{b}_0\right)^2} > 0\,.\end{aligned}$$

where the inequality follows from Assumption 4. Then, $\partial\theta/\partial\bar{b}_0>0$. ■

Proof of Proposition 11: *We want to show that a planner that has access to QE as an additional policy tool will only use it when $\psi(1+\alpha r)>\alpha(1+r)$. Let $(l_0^{sp},\bar{\omega}^{sp},\theta^{sp})$ be the allocations chosen by the social planner studied in section 4 and denote by λ^{sp} the lagrange multiplier on the constraint faced by this planner (20).*

Let $\mathcal{L}$ be the Lagrangian of the central bank, which can be written as

$$\mathcal{L} = [1-\Gamma(\bar{\omega})]\,R^k l_0 - \lambda\left[\mathbb{U}^{pe}-\mathbb{U}(l_0,\bar{\omega},\theta(l_0,\bar{\omega},\bar{b}_0,\bar{r}),\bar{b}_0,\bar{r})\right] - \gamma\left[(1+\bar{r})^2-\bar{R}^b\right] - \nu[r-\bar{r}] + \eta\bar{b}_0\,,$$

where we are considering the constraint imposed by the definition of secondary market liquidity (17) writing $\theta(l_0,\bar{\omega},\bar{b}_0,\bar{r})$ and where we have already substituted in $\bar{s}_0=\bar{b}_0$. An optimal allocation for this planner needs to satisfy the following FOCs:

$$(l_0) \qquad 0 = \frac{\partial\mathcal{L}}{\partial l_0} = [1-\Gamma(\bar{\omega})]\,R^k + \lambda\left[\frac{\partial\mathbb{U}}{\partial l_0}+\frac{\partial\mathbb{U}}{\partial\theta}\frac{\partial\theta}{\partial l_0}\right] + \gamma\frac{\partial\bar{R}^b}{\partial l_0}$$

$$(\bar{\omega}) \qquad 0 = \frac{\partial\mathcal{L}}{\partial\bar{\omega}} = -\Gamma'(\bar{\omega})R^k l_0 + \lambda\left[\frac{\partial\mathbb{U}}{\partial\bar{\omega}}+\frac{\partial\mathbb{U}}{\partial\theta}\frac{\partial\theta}{\partial\bar{\omega}}\right] + \gamma\frac{\partial\bar{R}^b}{\partial\bar{\omega}}$$

$$(\bar{b}_0) \qquad 0 = \frac{\partial\mathcal{L}}{\partial\bar{b}_0} = \lambda\left[\frac{\partial\mathbb{U}}{\partial\bar{b}_0}+\frac{\partial\mathbb{U}}{\partial\theta}\frac{\partial\theta}{\partial\bar{b}_0}\right] + \eta$$

$$(\bar{r}) \qquad\qquad 0 = \frac{\partial \mathcal{L}}{\partial \bar{r}} = \lambda \left[\frac{\partial \mathbb{U}}{\partial \bar{r}} + \frac{\partial \mathbb{U}}{\partial \theta} \frac{\partial \theta}{\partial \bar{r}} \right] - 2\gamma(1+\bar{r}) + \nu$$

The next step is to evaluate the FOCs at the constrained efficient allocation (without QE), i.e., $\left(l_0^{sp}, \bar{\omega}^{sp}, \theta^{sp}, 0, r\right)$. *If* $\bar{R}^b(l_0^{sp}, \bar{\omega}^{sp}) \leq (1+r)^2$ *the central bank cannot implement QE without violating its funding constraint (27). So we consider that we are in the interesting case where* $\bar{R}^b(l_0^{sp}, \bar{\omega}^{sp}) > (1+r)^2$ *and the central bank has some scope to offer a higher return on reserves relative to the storage technology. In this case the multiplier of this constraint at* $\left(l_0^{sp}, \bar{\omega}^{sp}, \theta^{sp}, 0, r\right)$ *equals zero, i.e.,* $\gamma = 0$. *Moreover, note that at* $\bar{b}_0 = 0$, *investors' expected utility* $\mathbb{U}$ *has the same functional form as in the case of the planner studied in section 4. Similarly, at* $\bar{b}_0 = 0$ *secondary market liquidity* θ, *equation (A.20), is the same function of choice variables as in the case without QE, equation (17). So we conclude that the FOCs wrt leverage* l_0 *and risk* $\bar{\omega}$ *are satisfied at* $\left(l_0^{sp}, \bar{\omega}^{sp}, \theta^{sp}, 0, r\right)$. *(In fact, we can use either FOC to obtain that* $\lambda = \lambda^{sp}$, *from where the other FOC follows.)*

Next, note that

$$\frac{\partial \mathbb{U}}{\partial \bar{r}} = \bar{s}_0 \frac{\partial U_{\bar{s}}}{\partial \bar{r}} = \bar{b}_0 \frac{\partial U_{\bar{s}}}{\partial \bar{r}} \qquad \Rightarrow \qquad \frac{\partial \mathbb{U}\left(l_0^{sp}, \bar{\omega}^{sp}, \theta^{sp}, 0, r\right)}{\partial \bar{r}} = 0\,.$$

And given that $(1+\bar{r}) = (1+r) < \Delta$ *from equation (A.20) we have that*

$$\frac{\partial \theta}{\partial \bar{r}} = \frac{(1-\delta)\Delta \bar{b}_0}{\delta R^b \left(n_0(l_0-1) - \bar{b}_0\right)} \qquad \Rightarrow \qquad \frac{\partial \theta\left(l_0^{sp}, \bar{\omega}^{sp}, \theta^{sp}, 0, r\right)}{\partial \bar{r}} = 0\,.$$

So the FOC wrt interest on reserves $\bar{r}$ *is trivially satisfied, with* $\nu = 0$.

Finally, we need to evaluate the FOC wrt $\bar{b}_0$ *at* $\left(l_0^{sp}, \bar{\omega}^{sp}, \theta^{sp}, 0, r\right)$. *From this condition it follows that*

$$\frac{\partial \mathbb{U}}{\partial \bar{b}_0} + \frac{\partial \mathbb{U}}{\partial \theta} \frac{\partial \theta}{\partial \bar{b}_0} < 0 \qquad \Rightarrow \qquad \eta > 0 \quad \text{and} \quad \bar{b}_0 = 0\,.$$

To sign $\partial \mathbb{U}/\partial \bar{b}_0 + (\partial \mathbb{U}/\partial \theta)(\partial \theta/\partial \bar{b}_0)$ *we proceed to compute these derivatives and evaluate at* $\left(l_0^{sp}, \bar{\omega}^{sp}, \theta^{sp}, 0, r\right)$. *By definition*

$$\mathbb{U}(l_0, \bar{\omega}, \theta, \bar{b}_0, \bar{r}) = [e_0 - n_0(l_0-1)]U_s + \bar{b}_0 U_{\bar{s}} + [n_0(l_0-1) - \bar{b}_0]U_b\,.$$

Then,

$$\frac{\partial \mathbb{U}\left(l_0^{sp}, \bar{\omega}^{sp}, \theta^{sp}, 0, r\right)}{\partial \bar{b}_0} = U_{\bar{s}}\left(\theta^{sp}, r\right) - U_b\left(l_0^{sp}, \bar{\omega}^{sp}, \theta^{sp}\right) = U_s\left(\theta^{sp}\right) - U_b\left(l_0^{sp}, \bar{\omega}^{sp}, \theta^{sp}\right)\,,$$

where we used that if interest on reserves are equal to the return on the storage technology then $U_{\bar{s}}\left(\theta^{sp}, r\right) = U_s\left(\theta^{sp}\right)$, *from equation (31). In addition, from the conditions that describe the* *planner's allocations we have that*

$$s_0^{sp} U_s(\theta^{sp}) + b_0^{sp} U_b\left(l_0^{sp}, \bar{\omega}^{sp}, \theta^{sp}\right) = s_0^{pe} U_s(\theta^{pe}) + b_0^{pe} U_b\left(l_0^{pe}, \bar{\omega}^{pe}, \theta^{pe}\right) = e_0 U_s(\theta^{pe})$$

$$\Rightarrow \qquad U_b\left(l_0^{sp}, \bar{\omega}^{sp}, \theta^{sp}\right) - U_s\left(\theta^{sp}\right) = \frac{e_0\left[U_s\left(\theta^{pe}\right) - U_s\left(\theta^{sp}\right)\right]}{b_0^{sp}} \ . \tag{A.22}$$

Then, from the characterization of the constrained efficient allocation (Proposition 8) we have that if $\psi(1+\alpha r) < \alpha(1+r)$, *then* $\theta^{sp} < \theta^{pe}$ *and* $\partial\mathbb{U}/\partial\theta < 0$. *So we conclude that* $\partial\mathbb{U}/\partial\bar{b}_0 < 0$. *In addition, from Proposition 10 we have that* $\partial\theta/\partial\bar{b}_0 > 0$, *thus* $\partial\mathbb{U}/\partial\bar{b}_0 + (\partial\mathbb{U}/\partial\theta)(\partial\theta/\partial\bar{b}_0) < 0$. *Thus, it must be that* $\eta > 0$ *and the optimal QE designs calls for not buying bonds, i.e.,* $\bar{b}_0 = 0$.

Alternatively, when $\psi(1+\alpha r) < \alpha(1+r)$ *from Proposition 8 we have that* $\theta^{sp} > \theta^{pe}$ *and* $\partial\mathbb{U}/\partial\theta > 0$, *thus* $\partial\mathbb{U}/\partial\bar{b}_0 > 0$ *and we conclude that* $\partial\mathbb{U}/\partial\bar{b}_0 + (\partial\mathbb{U}/\partial\theta)(\partial\theta/\partial\bar{b}_0) > 0$. *Therefore,* $\partial\mathcal{L}/\partial\bar{b}_0 > 0$, *i.e., the central bank will want to increase the size of the bond buying program when it is zero and we conclude that* $\bar{b}_0 > 0$, *improving upon the constrained efficient allocation.*

We are left to establish that at the allocation implemented by the central bank, where $\bar{b}_0 > 0$, $\bar{r} > r$. *For that we consider evaluate the FOC wrt* $\bar{r}$ *when* $\bar{r} = r$. *Note that in this case where* $1 + \bar{r} < \Delta$ *we have that*

$$\frac{\partial\mathbb{U}}{\partial\bar{r}} = \bar{s}_0 \frac{\partial U_{\bar{s}}}{\partial\bar{r}} = \bar{b}_0\left\{\delta + (1-\delta)\left[p\Delta + (1-p)(1+r) + \frac{\bar{r}-r}{1-\delta}\right] + 1 + r\right\} > 0 \ .$$

In addition,

$$\frac{\partial\theta}{\partial\bar{r}} = \frac{(1-\delta)\Delta\bar{b}_0}{\delta[n_0(l_0-1) - \bar{b}_0]R^b} > 0 \ .$$

So we conclude that

$$\frac{\partial\mathcal{L}(l_0, \bar{\omega}, \theta, \bar{b}_0, r)}{\partial\bar{r}} = \lambda\left[\frac{\partial\mathbb{U}}{\partial\bar{r}} + \frac{\partial\mathbb{U}}{\partial\theta}\frac{\partial\theta}{\partial\bar{r}}\right] + \nu > 0 \ .$$

Recall that we assumed that $\bar{R}^b(l_0^{sp}, \bar{\omega}^{sp}) > (1+r)^2$. *So we conclude that the central bank will set the interest on reserves strictly higher that the interest on storage, i.e.,* $\bar{r} > r$. ■