# Market and Public Liquidity

*By* Patrick Bolton, Tano Santos and José A. Scheinkman[*]

As the record of Fed interventions over the past year, from December 2007 to December 2008, makes abundantly clear a foremost concern of monetary authorities in responding to the financial crisis has been to avoid a repeat of the great depression, and especially a repeat of the monetary contraction that Milton Friedman and Anna Schwartz (1963) have identified as the major cause of the 1930s depression. The Fed has shown tremendous resourcefulness and inventiveness in its liquidity injections, considerably widening the collateral eligible under the discount window and the term auction facility, and setting up new programs targeted at primary dealers, the commercial paper market and money market funds. At the same time it has stepped in to offer guarantees on assets held by some financial institutions (e.g. Citigroup) to avoid their bankruptcy (see David Enrich, Carrick Mollenkamp, Matthias Rieker, Damian Paletta, and Jon Hilsenrath, 2008).

This unprecedented intervention has had the intended effect of averting a major systemic financial meltdown and it has kept some critical financial institutions afloat. Yet, until now, banks have mostly responded by cutting new lending and hoarding liquidity, so that the ultimate goal of forestalling a credit crunch has not been achieved. For the most part banks also are still holding most of the toxic assets that have undermined the market's confidence in the soundness of the banking system. Moreover, the Fed has put its balance sheet at risk, increasing the assets it holds from $851 billion in the

summer of 2007 to $2.245 trillion at the end of 2008. Finally, the massive public liquidity injection has also had the effect of crowding out private liquidity and private capital as an alternative source of funding for banks.

These side effects of the public liquidity injection may undermine the effectiveness of public policy and may also impose substantial costs on the real economy. It is therefore important to explore with the benefit of hindsight whether less costly approaches to public liquidity injections aren't available. This is what we intend to do in this paper, by relying on the analytical framework we developed in Bolton, Santos and Scheinkman (2008) (BSS). The model we have developed in BSS is set up do address two issues that have been at the core of the current crisis. The first issue is the *originate-and-distribute* model of financial intermediation, what the underlying economic rationale for this model might be (if there is any), and how it might affect optimal liquidity provision. We propose a new explanation for origination and *contingent* distribution based on maturity shocks and the optimal allocation of long-term assets in the hands of long term investors. The second issue concerns the dynamics of liquidity crises and the optimal timing of public liquidity. At what point in a liquidity crisis is public liquidity most desirable?

Although, recent economic research provides a better understanding of the benefits of public intervention in credit markets during aggregate liquidity crises (Bengt Holmstrom and Jean Tirole, 1998 and Patrick Bolton and Howard Rosenthal, 2002) it does not touch on the issue of the optimal timing of liquidity in a dynamically unfolding liquidity crisis. Also, the monetary authorities did not have a blueprint they could rely on when the crisis broke out, and have essentially had to improvise their policy response as events unfolded.

The model in BSS only provides a most rudimentary dynamic structure, but it is sufficient to

* Bolton: Columbia University, NBER and CEPR, New York, NY, pb2208@columbia.edu. Santos: Columbia Business School, NBER and CEPR, New York, NY, USA, js1786@columbia.edu. Scheinkman: Princeton University and NBER, Princeton, NJ, joses@princeton.edu. We thank Doug Diamond for comments. This material is based on work supported by the National Science Foundation under award number SES-07-18407.

be able to frame the issue of the timing of public liquidity. We briefly outline the main building blocks of the model in the next section and in a subsequent section characterize equilibria using a numerical example. We then proceed to a discussion of the effects of public liquidity in our model.

Three main observations emerge from our analysis. First, lack of knowledge and opaqueness about asset-quality of institutions in need of liquidity, while it can facilitate liquidity trading (as Holmstrom and Tirole (2008) observe) also tends to induce inefficient liquidity provision by the market. Institutions who are faced with a liquidity shortage may trade assets for cash too soon in an effort to avoid future adverse selection problems, which undermine the liquidity of future secondary markets. By choosing to trade sooner these institutions forego a valuable option not to trade assets at fire-sale prices at all should their liquidity needs prove to be temporary.

Second, if the monetary authorities wrongly time their injection of liquidity they risk crowding out liquidity that may be available in the market, outside the banking sector (mainly in hedge funds, pension funds and sovereign wealth funds). At the same time, if liquidity is injected in the form of a collateralized lending facility, public liquidity will undermine financial institutions' incentives to obtain liquidity outside the banking sector by selling (problem) assets for cash.

Third, public liquidity injections while alleviating the liquidity needs of sound institutions, may also provide a life-line to holders of worthless assets. Unfortunately, the monetary authorities may not have the knowledge required to be able to optimally time their liquidity injections and to be able to discriminate between sound and worthless institutions. It may thus be desirable to give the authorities greater powers to monitor the financial system and the financial institutions that may one day have to rely on its liquidity facility.

## The BSS model

In BSS we consider a model with two types of investors, long-run (LR) and short-run (SR) investors. The latter have expertise in identifying valuable risky projects that typically ma-

ture early, while the former invest in higher return long-duration assets. Assets that mature early are risky and expose their holders to both maturity and return risk. The other assets are riskless. There are gains from trade between LR and SR investors when the risky asset matures late. In this case SR investors prefer to sell the asset to LR investors as long as the price is at least as high as the future value of the asset's returns discounted at their higher discount factor. If SRs anticipate to be able to sell in these contingencies they are more willing to invest in the risky asset. Similarly, if LRs anticipate to be able to buy risky assets at marked down prices in these events they are willing to hold more cash. In sum, there is a natural complementarity between LR and SR investors. SRs sell assets in states where they value them the least and LRs provide cash when SRs value cash the most.

In a frictionless financial system it is efficient for SRs to rely on this source of outside liquidity. This mechanism allows SRs to originate a larger volume of valuable assets and to distribute them to the highest-value holders. However, in reality there are at least two frictions that may disrupt this financing model. First, the originator may have private information about the underlying value of the asset. Second, both sides of the market must coordinate their portfolio composition decisions and the timing of their trades to generate maximum gains from trade. Indeed, the secondary market for assets can completely dry up at any moment if SRs expect LRs not to carry much cash, or if LRs expect to be able to purchase these assets at even more marked down prices in the future.

More formally, the BSS model allows for four periods. At date 0, the representative LR chooses the amount $M$ to hold in cash and the amount $(\kappa - M)$ of his endowment $\kappa$ to invest in a long-term decreasing returns-to-scale project that yields a return $\varphi(\kappa - M) > \kappa - M$ at date 3. Similarly, a representative SR chooses the fraction $(1 - m)$ of her unit endowment to invest in an $i.i.d$ risky project that she originates and that can be scaled up to at most one unit; the remainder $m$ is held in cash. Both LR and SR investors are assumed to be risk neutral. They differ only in their time preferences, with SR investors discounting date 3 consumption with discount factor $\delta < 1$ but not LR investors.

Risky projects are likely to mature early: they pay an amount $\widetilde{\rho}_t$ at either dates $t = 1, 2, 3$, where $\widetilde{\rho}_t \in \{0, \rho\}$ and $\rho > 1$. At date 1 risky assets yield $\rho$ with probability $\lambda$, and with probability $(1 - \lambda)$ they only yield a positive return at either dates 2 or 3. The date 1 shock to cash-flows is an aggregate publicly observable shock. Subsequent cash-flow shocks, however are *i.i.d.* idiosyncratic shocks: (i) the asset matures with probability $\theta$ at date 2 and with probability $(1 - \theta)$ at date 3; (ii) when it matures it yields $\widetilde{\rho}_t = \rho$ with probability $\eta$ and $\widetilde{\rho}_t = 0$ with probability $(1 - \eta)$. Only the holders of a risky asset are able to observe the realization of the idiosyncratic shock. This informational asymmetry introduces a key friction in the secondary market for risky assets at date 2. It features prominently in the BSS model to capture the increased concern about the quality of assets held by financial institutions as the current crisis unfolded and as reflected by the widening of spreads.

We assume that there is a unit mass of both LR and SR investors and we assume that the law of large numbers applies, so that $\theta$ is also the proportion of risky assets that matures at date 2 and $\eta$ is the proportion of risky assets that pay off $\rho$.

For later reference, we refer to $\omega_{1\rho}$ as the state in which all risky assets yield $\rho$ at date 1 and $\omega_{1L}$ as the state in which all risky assets are known to mature at dates 2 or 3. Similarly, we refer to $\omega_{2\rho}$, $\omega_{20}$, $\omega_{2L}$, $\omega_{3\rho}$, and $\omega_{30}$ as states idiosyncratic to a risky project, which refer to respectively a payoff of $\rho$ or 0 at date 2, the event that the asset matures at date 3, and the payoff of $\rho$ or 0 at date 3.

### An Example

Most of our analysis can be illustrated with the help of the following example: $\lambda = .85$, $\eta = .4$, $\rho = 1.13$, $\kappa = .2$, $\delta = .192$, $\varphi(x) = x^\gamma$, $\gamma = .4$. The only free parameter is $\theta$, which we allow to vary between 0 and $\overline{\theta} = .4834$. This free parameter plays a central role in the analysis as it is simultaneously a measure of the expected maturity of the asset and of the informational rent of the originators of the asset. As $\theta$ increases the risky asset is more attractive to SR investors, since the probability $(\lambda + (1 - \lambda)\theta)$ that the asset matures before date 3 is then higher. It is

straightforward to verify that for any value $\theta \leq \overline{\theta}$, SR investors prefer to only hold cash under autarchy.

Note also that in this example $\varphi'(\kappa) > 1$ so that LR investors must be able to purchase risky assets in secondary markets at marked down prices to compensate for the opportunity cost of holding cash. In other words, in this example equilibrium secondary market prices must be *cash-in-the-market* prices, a term first coined by Franklin Allen and Douglas Gale (1998).

### Equilibrium

BSS solve for symmetric, competitive, rational expectations equilibria in which LR and SR investors choose their optimal portfolio and asset trades taking prices as given. They solve for two types of equilibria, an *immediate-trading equilibrium* in which secondary markets are active only at date 1, and a *delayed-trading equilibrium* in which secondary markets are active only at date 2.

The immediate-trading equilibrium exists for all $\theta \in [0, \overline{\theta}]$ and is such that $M_i^* > 0$, $m_i^* \geq 0$, and

$$q^*(\omega_{1L}) = Q^*(\omega_{1L}) = 1 - m_i^*,$$

where $q^*(\omega_{1L})$ and $Q^*(\omega_{1L})$ respectively denote the SR asset supply and LR asset demand in the event $\omega_{1L}$ where risky assets do not mature at date 1.

This equilibrium is supported by *on-the-equilibrium-path* market-clearing prices such that:

$$P_{1i}^* = \frac{M_i^*}{1 - m_i^*} = \frac{1 - \lambda\rho}{1 - \lambda},$$

and *off-the-equilibrium-path* prices at date 2, $P_{2i}^*$, such that neither SRs nor LRs have an incentive to trade at date 2. SRs prefer to sell assets at date 1 for a price $P_{1i}^*$ rather than wait to trade at date 2 if necessary at price $P_{2i}^*$ if the following condition holds:

$$P_{1i}^* \geq \theta\eta\rho + (1 - \theta\eta) P_{2i}^*.$$

As for LRs, they also prefer to trade at date 1 if their off-the-equilibrium-path beliefs are such that they expect to buy only *lemons*, so that the conditional expected payoff of the risky asset at date 2 is $\mathbf{E}[\widetilde{\rho}_3 | \mathcal{F}] = 0$, where $\mathcal{F}$ is the information set of LRs at date 2.

The equilibrium portfolio policies $[m_i^*, M_i^*]$ are obtained by solving respectively the SR and LR optimization problems at date 0 given the equilibrium price $P_{1i}^*$. An SR's payoff function at date 0 is linear in $m$ and given by

$$\pi_i(m) = m + (1 - m)\left[\lambda\rho + (1 - \lambda)P_{1i}^*\right],$$

so that SRs are indifferent between any cash holding $m \in [0, 1]$ if $P_{1i}^* = \frac{1 - \lambda\rho}{1 - \lambda}$. Similarly, an LR investor's payoff function at date 0 is given by:

$$\Pi_i(M) = \varphi(\kappa - M) + \lambda M + (1 - \lambda)\frac{\eta\rho}{P_{1i}^*}M,$$

so that $M_i^*$ is given by

$$\varphi'(\kappa - M_i^*) = \lambda + (1 - \lambda)^2\frac{\eta\rho}{1 - \lambda\rho}.$$

Then, setting $m_i^* = m$ such that

$$\frac{M_i^*}{1 - m} = \frac{1 - \lambda\rho}{1 - \lambda}$$

completes the characterization of the immediate-trading equilibrium.

A *delayed-trading equilibrium* may also exist for a subset $\theta \in [0, \tilde{\theta}]$, where $\tilde{\theta} = .4628 < \overline{\theta} = .4834$. This equilibrium is such that $m_d^* \in [0, 1)$, $M_d^* \in (0, \kappa)$, and

$$q^*(\omega_{20}) = q^*(\omega_{2L}) = (1 - m_d^*)$$

$$Q_{2d}^* = (1 - \theta\eta)(1 - m_d^*).$$

Here $q^*(\omega_{20})$ and $q^*(\omega_{2L})$ respectively denote the SR asset supplies in the (idiosyncratic) event $\omega_{20}$ when the risky asset is known to be worthless to SR and the event $\omega_{2L}$ when the risky asset is known to mature at date 3. $Q_{2d}^*$ denotes the LR asset demand at date 2.

This equilibrium is supported by *on-the-equilibrium-path* market-clearing prices such that:

$$(1) \qquad P_{2d}^* = \frac{M_d^*}{(1 - \theta\eta)(1 - m_d^*)}.$$

Under delayed trading, the total cash in the market is $M_d^*$ and the total supply of risky assets is given by the fraction of SRs who want to trade

$(1 - \theta\eta)$ times the total amount of assets they each have available to trade $(1 - m_d^*)$. Equilibrium prices are then simply given by the aggregate cash-to-asset ratio at date 2.

The equilibrium portfolio policies $[m_d^*, M_d^*]$ are obtained by solving the SR and LR optimization problems at date 0 under the assumption that trade takes place only at date 2. The SRs' payoff function at date 0, $\pi_d(m)$, is again linear in $m$ and is given by:

$$m + (1 - m)\left[\lambda\rho + (1 - \lambda)(\theta\eta\rho + (1 - \eta\theta)P_{2d}^*)\right].$$

The LR's payoff function at date 0, $\Pi_d(M)$, is given by:

$$\varphi(\kappa - M) + \lambda M + (1 - \lambda)\left(\frac{(1 - \theta)\eta\rho}{(1 - \theta\eta)P_{2d}^*}\right)M,$$

so that

$$(2) \quad \varphi'(\kappa - M_d^*) = \lambda + (1 - \lambda)\frac{(1 - \theta)\eta\rho}{(1 - \theta\eta)P_{2d}^*}.$$

For $\theta \in [0, \widehat{\theta})$ with $\widehat{\theta} = .4196$ the equilibrium is such that $m_d^* > 0$ and the equilibrium price is such that SRs' are indifferent between any $m \in [0, 1]$ :

$$(3) \qquad P_{2d}^* = (1 - \lambda)\rho + \frac{1 - \rho}{1 - \theta\eta}.$$

Equations (1), (2), and (3) determine $m_d^*$, $M_d^*$, and $P_{2d}^*$. For $\theta \in [\widehat{\theta}, \tilde{\theta}]$ the equilibrium is such that $m_d^* = 0$, $M_d^*$ is given by

$$\varphi'(\kappa - M_d^*) = \lambda + (1 - \lambda)\frac{(1 - \theta)\eta\rho}{M_d^*},$$

and $P_{2d}^*$ immediately obtains from (1). To show that $P_{2d}^*$ is in fact an equilibrium it remains to check that $P_{2d}^* \geq \delta\eta\rho$, so that the SRs in state $\omega_{2L}$ do have an incentive to supply the risky asset rather than carrying it to date 3. It can be checked that this is indeed the case whenever $\theta \in [0, \tilde{\theta}]$, where $\tilde{\theta} = .4628$.

The *off-the-equilibrium-path* prices at date 1, $P_{1d}^*$, must also be such that neither SRs nor LRs have an incentive to trade at date 1, or:

$$P_{1d}^* \leq \theta\eta\rho + (1 - \theta\eta)P_{2d}^* \text{ and } \frac{\eta\rho}{P_{1d}^*} \leq \frac{(1 - \theta)\eta\rho}{(1 - \theta\eta)P_{2d}^*}.$$

The first inequality ensures that SRs are better off trading at date 2, and the second that LRs also prefer to trade at date 2. If LRs trade at date 1 their net return is given by the total expected return of the risky asset at date 1, $\eta\rho$, divided by the price of the asset $P_{1d}^*$. Similarly at date 2 the conditional expected return on the asset is $(1-\theta)\eta\rho/(1-\theta\eta)$, as SRs don't trade with probability $\theta\eta$ and are expected to trade lemons with probability $\theta(1-\eta)$ at date 2. It is straightforward to verify that for our parameter values it is always possible to find a price $P_{1d}^*$ that satisfies these inequalities.

For $\theta = .35$ the immediate and delayed trading equilibrium are such that $(M_i^*, m_i^*) = (.0169, .9358)$ and $(M_d^*, m_d^*) = (.0540, .4860)$, respectively. Moreover, although both equilibria are *interim efficient*, it can be shown that the delayed trading equilibrium (weakly) *Pareto dominates* the immediate trading equilibrium ex-ante. Indeed, for this parametrization $\pi_i^* = \pi_d^* = 1$ and $\Pi_d^* = .5317 > .5258 = \Pi_i^*$. The intuition for these results is as follows. In our framework efficiency gains occur whenever more risky projects are implemented and the amount of liquidity carried by both SRs and LRs is lowered. But SRs do not implement risky projects in autarky; they only do so if enough outside liquidity is supplied to absorb potential sales in either dates 1 or 2. That is, more risky projects require more outside liquidity. In the delayed trading equilibrium there is a larger fraction of risky projects undertaken and overall liquidity is lower than in the immediate trading equilibrium, though there is more outside liquidity. The efficiency gains associated with the implementation of more risky projects overwhelms the efficiency losses associated with the increase in outside liquidity. In the delayed trading equilibrium LRs only need to acquire the assets of SRs that are in states $\omega_{2L}$ and $\omega_{20}$ and SRs retain the "upside" of the risky asset. In contrast, in the immediate trading equilibrium LRs have to absorb the full measure of all risky projects and this requires more outside liquidity per unit of risky project undertaken which entails a loss of efficiency.

Although delaying trading at date 2 is ex-ante efficient, the delayed trading equilibrium may fail to exist due to adverse selection problems at that date. This occurs whenever the candidate price at date 2, obtained as above, is such that $P_{2d}^* < \delta\eta\rho$. In this case SRs' who have assets that mature at date 3 prefer to hold onto those assets rather than trade at highly dilutive prices at date 2, and this leads to a market breakdown. In the example the delayed trading equilibrium fails to exist for $\theta \in (.4628, .4834]$.

## The Timing of Public Liquidity

In the BSS model market liquidity takes the form of inside liquidity carried by SRs and outside liquidity provided by LRs. We now consider the effect of anticipated public liquidity injection on market liquidity.

There is a welfare-improving role for public liquidity in the BSS model in situations when the delayed trading equilibrium fails to exist. In such situations, the monetary or fiscal authorities could intervene by providing price support in the secondary market at date 2 and thus restore existence of the delayed-trading equilibrium. Another, related welfare improving intervention is to ensure that the economy coordinates on the efficient equilibrium by providing price support at date 2, so as to put a price floor on *off-the-equilibrium prices* at date 2 and thus ensure that an immediate-trading equilibrium cannot exist. These anticipated forms of public liquidity provision help induce an efficient amount and mix of market liquidity provision at date zero.

Note that either forms of intervention are market-making interventions similar to those initially envisaged under TARP, that aim to support outside liquidity by facilitating the *transfer* of (troubled) assets from SRs to LRs. Thus, our analysis suggests that rather than the government playing a role of lender of last resort it should play a role of *market-maker of last resort*. By inducing SRs to obtain liquidity through asset sales, the government makes optimal use of market liquidity and helps maintain the efficiency of origination and distribution of risky assets under the delayed-trading equilibrium. To the extent that monetary authorities may not be legally authorized to play such a market-making role of last resort, fiscal authorities need to intervene in this capacity.

On the other hand, public liquidity provision through collateralized lending has the perverse effect of encouraging hoarding and crowding out outside liquidity, thus undermining the efficient

distribution of risky assets originated by SRs. More precisely in our model such an intervention has the effect of raising $\delta$ and thus encouraging SRs to inefficiently hold risky assets until they mature at date 3. As a result the delayed trading equilibrium may disappear, and we are left with the immediate trading equilibrium in which SRs carry inefficient amounts of inside liquidity. Another unintended effect of central banks broadening of collateralized lending, by accepting a larger set of securities, is that it worsens the lemons problem in secondary markets, as only the worst assets, those that cannot serve as collateral, will be kept on the books of financial institutions. This may help explain why LIBOR spreads increased following some recent interventions by central banks.

Our analysis, thus highlights an important concern with Fed interventions over the past year that other commentators have also emphasized: namely that it does not do much more than provide a life-line to financial institutions. It does not induce them to engage in new lending. On the contrary, it encourages zombie lending by helping banks maintain non-performing assets on their balance sheet. What is more, it transfers a potentially major asset risk to the Fed.

In the BSS model SRs *originate* more risky projects if the delayed equilibrium is being played - that is if they can *distribute* these assets when it is efficient to do so. Tobias Adrian and Hyun Song Shin (2009) document that in the run up to the current crisis, an increased fraction of credit originated by financial intermediaries has been distributed. Until this option to distribute is restored, it is unlikely that financial intermediaries will originate enough new loans to alleviate the credit crunch. The option to distribute can only be restored after cleaning up the intermediaries' balance sheets. This cleaning up can only be delayed by the availability of collateralized lending by the monetary authorities.

An even more efficient intervention could be envisaged if the authorities were able to identify institutions in states $\omega_{20}$ and $\omega_{2L}$. In that case, liquidity could be granted to the sound SRs who need liquidity and not to the worthless SRs in state $\omega_{20}$. To be able to pull this off, however, the monetary authorities would need a much more detailed knowledge of financial institutions' assets and liabilities than they currently have. In sum the efficient provision of public liquidity requires detailed knowledge by monetary authorities to be able to time the intervention optimally and to be able to sort solvency from liquidity problems.

### References

**Adrian, Tobias and Hyun Song Shin** (2009) "Money, Liquidity and Monetary Policy," *American Economic Review Papers and Proceedings*, this issue.

**Allen, Franklin and Douglas Gale** (1998) "Optimal Financial Crises," *Journal of Finance*, 53, 1245-1284.

**Bolton, Patrick, and Howard Rosenthal** (2002) "Political Intervention in Debt Contracts, "*Journal of Political Economy*, 110, 1103-34.

**Bolton, Patrick, Tano Santos and Jose Scheinkman** (2008) "Outside and Inside Liquidity, "mimeo, Columbia University

**Enrich, David, Carrick Mollenkamp, Matthias Rieker, Damian Paletta and Jon Hilsenrath** (2008) "US Agrees to Rescue Struggling Citigroup,"Wall Street Journal, November 24th, 2008

**Friedman, Milton and Anna Schwartz** (1963) *Monetary History of the United States, 1867-1960*

**Holmstron, Bengt and Jean Tirole** (1998) "Private and Public Supply of Liquidity," *Journal of Political Economy,* vol. 106 no. 1, 1-40.

**Holmstrom, Bengt and Jean Tirole** (2008) "Inside and Outside Liquidity," Wicksell Lectures, IDEI, Toulouse.