Liquidity and Monetary Policy

Douglas Gale
Federal Reserve Bank of New York
and
New York University

March 12, 2010

Preliminary and incomplete
Please do not cite

1 Introduction

In the period leading up to the financial crisis of 2007-2008, financial institutions of all sorts increased their leverage in the wholesale markets, relying heavily on collateralized borrowing in the form of repurchase agreements ("repos") and issuance of asset backed commercial paper (ABCP). As the crisis approached, lenders became nervous and the significantly shortened the maturity of the loans they were prepared to make. When the crisis hit, the disappearance of short term-funding created severe problems for many financial institutions. Some large firms failed and further failures were only prevented by the intervention of the central banks.

Although the origin of the crisis may have been the US sub-prime mortgage market, the early stages of the crisis were manifested in the form of a liquidity crisis. At the end of July, two Bear Sterns funds filed for bankruptcy and a third suspended redemptions. More bad news followed and then, on August 7, BNP Paribas halted redemptions from three investment funds because it could not "fairly" calculate their NAV. Potential investors, mainly money-market mutual funds (MMF), declined to roll over their purchases of ABCP. Since many of the vehicles that comprised the parallel banking system (PBS) were sponsored by banks and/or had liquidity guarantees from banks, there was a fear that these assets would end up on bank balance sheets. This in turn raised concerns about counter-party risk amongst the banks and caused LIBOR to shoot upwards. The European Central Bank was forced to inject 95 billion Euros in overnight lending into the market in order to cope with the demand for liquidity. Two aspects of these events are notable. First, the events that triggered the collapse in the ABCP market were not themselves large. Second, the impact was felt by very different funds and institutions, some of which had nothing to do with sub-prime mortgages.

One famous victim of this "market freeze" was the British bank, Northern Rock. Originally a mutual organization known as a building society, it had converted to a bank in 1997 and by 2007 it had grown to be the fifth-largest mortgage lender in the UK. In order to achieve this market share, it had relied on securitization and wholesale funding rather than retail deposits. By mid-September the longer-term funding markets were closed for Northern Rock. When Lloyds TSB's offer to purchase the bank fell through, the Bank of England was forced to extend emergency assistance to Northern Rock directly. The announcement of the Bank's support for Northern Rock resulted in a run on the bank, which ended when the Government announced it would guarantee all Northern Rock's existing deposits. This was effectively the end of Northern Rock as an independent entity. Although Northern Rock would turn out to have its share of troubled assets, at the time of its near collapse, its problems seemed to be mainly liquidity related.

The failure of Bear Stearns in mid-March 2008 is another example of a market freeze.[1] As an intrinsic part of its business, Bear Stearns relied day-to-day on its ability to obtain short-term finance through secured borrowing. At this time, Bear was reported to be financing $85 billion of assets on the overnight market (Cohan, 2009). Beginning late Monday, March 10, 2008 and increasingly through that week, rumors spread about liquidity problems and eroded investor confidence in the firm. Even though Bear Stearns continued to have high quality collateral,[2] counterparties became less willing to enter into the normal funding arrangements with the firm. By the end of the week, counterparties were unwilling to make even secured funding available to the firm on customary terms. This unwillingness to fund on a secured basis placed enormous stress on Bear's liquidity. On Tuesday, March 11, the holding company liquidity pool declined from $18.1 billion to $11.5 billion. On Thursday, March 13, Bear Stearns' liquidity pool fell sharply and continued to fall on Friday. In the end, the market rumors about Bear Stearns' liquidity problems became self-fulfilling and led to the near failure of the firm. Bear Stearns was adequately capitalized at all times during the period from March 10 to March 17, up to and including the time of its agreement to be acquired by J.P. Morgan Chase. Even at the time of its sale, Bear Stearns' capital and its broker dealers' capital exceeded supervisory standards. In particular, the capital ratio of Bear Stearns was well in excess of the 10% level used by the Federal Reserve Board in its well-capitalized standard.

Understanding the reasons why liquidity dried up is obviously important if we want to avoid a repeat of the sub-prime crisis and to design a more stable and efficient financial system for the future. The financial crisis disrupted large parts of the financial system. The parallel or "shadow" banking system (PBS), consisting of SIVs, conduits, ABS, CDOs, etc., has virtually disappeared. Some estimates put the size of the global PBS in the neighborhood of 11 trillion dollars in 2007. Its collapse has sharply reduced the financial system's lending

[1] The discussion that follows is based on the Security and Exchange Commission's Chairman Christopher Cox's Letter to the Basel Committee in Support of New Guidance on Liquidity Management, available at: http://www.sec.gov/news/press/2008/2008-48.htm

[2] This high quality collateral consisted mainly of highly rated mortgage-backed assets which had low but not inconsequential credit risk by this time in the sub-prime crisis.

capacity. In spite of the well known problems with sub-prime asset backed securites (ABS), securitization has proved to be a valuable technique for spreading risk and and expanding sources of funding. Restoring the PBS and its demand for ABS is an important step in the reconstruction the financial system.

Reviving the PBS is also important as a way of getting ABS off the balance sheets of banks and central banks. Krishnamurthy and Vissing Jorgensen (2010) have estimated that most of the ABS previously held in the PBS have found their way onto the balance sheets of banks, assisted by the lending facilities extended by the central banks. More ABS are held by the Federal Reserve System and GSEs. If the central bank's balance sheet is to be restored to its normal size and the banks' credit channel is to be unblocked, these securities will have to find a home somewhere else. This will require an expansion of wholesale funding. Whether this will be possible without risking instability depends on the successful implementation of liquidity regulation.

The present paper focuses on what economic theory can tell us about the regulation of liquidity in the financial system. In Section 2 we describe a general equilibrium model of the financial system and characterize the efficient provision of liquidity. This benchmark model allows us to identify some of the causes of market failure and the central bank interventions that might be necessary to improve welfare. In particular, in Section 3 we see how the incompleteness of markets can lead to inefficient liquidity provision and, in some cases, market crises. In certain circumstances, market failures are relatively benign and can be rectified by requiring banks to hold adequate amounts of liquid assets. In other cases, more extensive interventions by the central bank are required. In Section 4 we discuss a variety of policies that might be to improve the allocation of liquidity in the financial system. Finally, in Section 5 we discuss a proposal to revive the PBS.

2 Optimal liquidity provision

In this section, we describe a general-equilibrium model of the financial system, characterizing the conditions under which market provision of liquidity is efficient. This model is a special case of the model developed in Allen and Gale (2004). The Allen-Gale model is itself an extension of the familiar Bryant-Diamond-Dybvig model. The two main innovations are the introduction of asset markets and aggregate uncertainty about asset returns and liquidity shocks. Asset markets allow banks to share risks and to share liquidity. The introduction of aggregate uncertainty about asset returns and liquidity shocks shifts the focus from sunspot phenomena, such as the panics in the Diamond and Dybvig (1983) model, to the analysis real shocks and their impact on risk sharing.

For the purpose of studying liquidity regulation, the model has a number of attractive features. (i) It incorporates sophisticated financial institutions in a general-equilibrium theory of markets. (ii) It endogenizes the costs of liquidating assets. (iii) It provides a robustness check on results obtained using simpler models. (iv) It allows one to study the relationship between liquidity provision and asset pricing. (v) It provides a foundation for the welfare analysis of financial regulation. (vi) Variants of the model can be used to explore financial

fragility, excess sensitivity, and sunspot phenomena.

2.1 Model primitives

Time is divided into three dates, indexed by $t = 0, 1, 2$. At each date, there is a single good, which can be used for consumption or investment.

There are two assets, a short-term asset and a long-term asset. The short-term asset is represented by a storage technology. One unit of the good invested at date t yields one unit of the good at date $t + 1$, for $t = 0, 1$. The long-term asset is a constant returns to scale technology that takes two periods to mature. One unit invested in the long asset at date 0 yields $R > 1$ units of the good at date 2. The asset structure represents the tradeoff between liquidity and returns. The short asset offers an earlier return but has a lower yield; the long asset offers a later return but has a higher yield.

Aggregate uncertainty is represented by two states of nature, indexed by $s = 1, 2$. The state is uncertain at date 0 and has a probability distribution $\boldsymbol{\pi} = (\pi(1), \pi(2))$, where $\pi(s) > 0$ is the probability of state s. At the beginning of date 1 the true state is publically observed.

At date 0 there are two ex ante types of consumer, indexed by $i = 1, 2$. A consumer's ex ante type is publically observable. We follow Diamond-Dybvig in representing liquidity preference in terms of uncertainty about the timing of consumption. Initially, consumers are uncertain whether they value consumption at date 1 or date 2. At the beginning of date 1, each consumer learns his ex post type: either he is an early consumer, in which case he only values consumption at date 1, or he is a late consumer, in which case he only values consumption at date 2. The probability of being an early or late consumer depends on the state of nature and the consumer's ex ante type. A consumer of type i has a probability $0 \leq \lambda_i(s) \leq 1$ of being an early consumer in state s. If a consumer of type i has a consumption bundle $\mathbf{x}_i : \{1, 2\} \to \mathbf{R}^2_+$ that promises $x_{it}(s)$ units of the good at date t in state s, then his utility is given by

$$U_i(\mathbf{x}_i) = \sum_{s=1,2} \{\lambda_i(s) u_i(x_{i1}(s)) + (1 - \lambda_{is}(s)) u_i(x_{i2}(s))\}.$$

We assume that $u_i(\cdot)$ has the usual properties.

Finally, each consumer is assumed to have an initial endowment of consisting of one unit of the good at date 0 and nothing at dates $t = 1, 2$.

A production plan for this economy is a four-tuple $\mathbf{y} = (y_0, y_1(1), y_1(2), y_2)$, where $y_0 \geq 0$ is the amount invested in the short asset at date 0, $y_1(s) \geq 0$ is the amount invested in the short asset at date 1 in state $s = 1, 2$, and $y_2 \geq 0$ is the amount invested in the long asset at date 0. An allocation consists of an ordered triple $(\mathbf{x}_1, \mathbf{x}_2, \mathbf{y})$, where $\mathbf{x}_i$ is a consumption bundle for $i = 1, 2$ and $\mathbf{y}$ is an investment plan. An allocation $(\mathbf{x}_1, \mathbf{x}_2, \mathbf{y})$ is incentive compatible if

$$x_{i1}(s) \leq x_{i2}(s), \text{ for } s = 1, 2 \text{ and } i = 1, 2. \tag{1}$$

The allocation $(\mathbf{x}_1, \mathbf{x}_2, \mathbf{y})$ is attainable if it is incentive compatible and satisfies the conditions

$$y_2 + y_0 = 2, \tag{2}$$

$$\sum_{i=1,2} \lambda_i(s)\, x_{i1}(s)\, y_1(s) = y_0, \text{ for } s = 1, 2, \tag{3}$$

and

$$\sum_{i=1,2} (1 - \lambda_i(s))\, x_{i2}(s) = y_1(s) + R y_2, \text{ for } s = 1, 2. \tag{4}$$

The first condition (2) simply says that the total amount invested in the short and long assets at date 0 must equal the total endowment of the two ex ante types. The second condition (3) requires that the total consumption of early consumers plus the investment in the short asset must equal the returns to the short asset in state s at date 1. Similarly, the third condition (4) requires that the total consumption of late consumers equals the returns of the long and short assets in state s at date 2.

An attainable allocation $(\mathbf{x}_1, \mathbf{x}_2, \mathbf{y})$ is incentive efficient if there does not exist an attainable allocation $(\mathbf{x}_1', \mathbf{x}_2', \mathbf{y}')$ such that $U_i(\mathbf{x}_i') \geq U(\mathbf{x}_i)$ for all i and $U_i(\mathbf{x}_i') > U(\mathbf{x}_i)$ for some i.

2.2 Decentralization

To decentralize an efficient allocation, we need markets and financial institutions. First, we assume there is a complete set of contingent commodity markets. Since all uncertainty is resolved at the beginning of date 1 (agents learn their types and the true state of nature is revealed) markets will be complete if there are markets at date 0 for the good at date 0 and for the good in state $s = 1, 2$ at date $t = 1, 2$. In other words, there are five contingent commodities. The assets can be interpreted in the usual way as claims to bundles of contingent commodities and priced accordingly. Let the good at date 0 be the numeraire and let $\mathbf{p} = (\mathbf{p}(1), \mathbf{p}(2))$ denote the contingent price vector, where $\mathbf{p}(s) = (p_1(s), p_2(s))$ is the vector of dated commodities in state s.

Markets for contingent commodities provide insurance against aggregate risk, but we also need intermediaries to provide insurance against idiosyncratic liquidity shocks to consumers. An intermediary takes deposits at date 0 in exchange for a complete contract $\mathbf{x}_i$ that promises $x_{it}(s)$ units of the good to a depositor of ex ante type i who withdraws at date t in state s. Because a depositor's ex post type is private information, we require the complete contract $\mathbf{x}_i$ to be incentive-compatible.

We assume that each intermediary caters to only one ex ante type of consumer. Otherwise, an intermediary could act as a central planner and internalize all gains from trade without needing to use markets. Since we want to investigate the use of markets to facilitate risk sharing and liquidity provision, we assume that transaction costs prevent this kind of super-intermediation. Thus, in equilibrium there will be two types of intermediaries, those catering to depositors of ex ante type 1 and those catering to depositors of ex ante type 2.

There is assumed to be free entry into the banking sector, so intermediaries earn zero profits. Competition among intermediaries forces them to maximize the expected utility of their depositors in equilibrium.

In addition to the restrictions on intermediaries' ability to intermediate among different types of consumers, we also restrict access to markets. In particular, we assume that intermediaries have access to all contingent and forward markets, whereas depositors do not have access to any of these markets, although they are allowed access to the storage technology (i.e., they can hold the short asset). These restrictions on market participation are needed to motivate the use of intermediaries. In fact, as Corn and Jacklin have short, if depositors have access to forward markets at date 1 they will arbitrage against the intermediaries. In equilibrium, it will be impossible for the intermediaries to provide any welfare gains from liquidity insurance.

An intermediary for consumers of type i chooses a complete, incentive compatible contract $\mathbf{x}_i$ to maximize the consumers' expected utility subject to a budget constraint that requires the value of the contract be less than or equal to the value of the deposit. Then the optimal contract $\mathbf{x}_i$ solves the decision problem

$$\begin{array}{ll} \max & U_i(\mathbf{x}_i) \\ \text{s.t.} & \mathbf{p} \cdot \mathbf{x}_i \leq 1. \end{array} \tag{5}$$

Although it is natural to think of intermediaries as holding assets, the existence of complete markets makes this unnecessary: whatever contingent commodities are needed to meet the intermediary's commitments can be purchased using the markets that exist at date 0. Someone must hold the assets, however, and in the Arrow-Debreu tradition we can assume that profit maximizing producers play this role. More precisely, a representative producer has access to a production set Y that is defined by

$$Y = \left\{ \mathbf{z} \in \mathbf{R}^5 : \mathbf{z} = (-y_0 - y_2, y_0 - y_1(1), y_0 - y_1(2), y_1(1) + Ry_2, y_1(2) + Ry_2), \ \exists \mathbf{y} \in \mathbf{R}^5_+ \right\}$$

and the producer chooses $\mathbf{z} \in \mathbf{Y}$ to maximize $\mathbf{p} \cdot \mathbf{z}$. In equilibrium, the producer earns zero profits and this implies that the no-arbitrage condition

$$\sup \{ \mathbf{p} \cdot \mathbf{z} : \mathbf{z} \in \mathbf{Y} \} = 0 \tag{6}$$

must be satisfied. This, implies that any production plan $\mathbf{y}$ satisfies

$$\sum_{s=1,2} \{ p_1(s)(y_0 - y_1(s)) + p_2(s)(y_1(s) + Ry_2) \} \leq y_0 + y_2.$$

We do not need to concern ourselves with the portfolio chosen by the bank since the completeness of markets assures us that any (undominated) portfolio is optimal.

Now we are ready to define an equilibrium of the economy: an equilibrium consists of an attainable allocation $(\mathbf{x}_1, \mathbf{x}_2, \mathbf{y})$ and a price vector $\mathbf{p}$ such that $\mathbf{x}_i$ is optimal (i.e., solves problem (5)) and satisfies the no-arbitrage condition (6).

This equilibrium shares the usual properties of the Arrow-Debreu competitive equilibrium:

Proposition 1 *Under a mild non-satiability condition, slightly stronger than the usual one because it has to take account of the fact that banks are restricted to incentive-compatible contracts, it can be shown that every equilibrium allocation is incentive efficient.*

In fact, in the special case studied here, the optimal contract satisfies the first-order condition

$$u_i'(x_{i1}(s)) \geq Ru_i'(x_{i2}(s)),$$

and equality holds if $y_1(s) > 0$. This implies that the incentive constraint $x_{i1}(s) \leq x_{i2}(s)$ is not binding, so the incentive-efficient allocation is actually a Pareto-efficient allocation. This stronger result does not hold in general, of course.

We can also show that equilibrium exists in general, but this requires a different notion of equilibrium. In particular, we cannot assume that every intermediary serving type i chooses the same contract $\mathbf{x}_i$. Instead, we have to allow for mixed allocations in which there is a distribution of contracts chosen by the intermediaries catering to each type i.

Proposition 2 *Under standard conditions, there exists a mixed equilibrium.*

The efficiency condition is straightforwardly extended to mixed equilibria.

2.3 Incomplete contracts and default

Although this provides an account of efficient liquidity provision, the intermediaries differ from the banks of the Bryant-Diamond-Dybvig model in one important respect: they offer complete, incentive-compatible contracts instead of standard deposit contracts. This is important because the contingency of these contracts rules out the need for default. In order to bring our model in line with standard practice in the banking literature, we need to introduce contracts with debt-like features. Allen and Gale (2004) allows for a general characterization of incomplete contracts, that is, contracts which are measurable with respect to a partition of states and allow for the possibility of default. Default is important both for realism and because it allows for potential welfare gains by increasing the contingency of the contract over and above the what is allowed by the incompleteness restrictions. For example, in the current environment with only two states, we could define incomplete contracts to satisfy the requirement that payoffs be independent of the state, that is,

$$x_{it}(s) = x_{it}(s'), \forall t = 1, 2, \ \forall s \neq s.$$

But if the bank cannot make the promised payments in one state, then it must liquidate the assets and pay them out immediately (i.e., at date 1) to early and late consumers, in which case

$$x_{i1}(s) = x_{i2}(s) = y_0 + \frac{p_2(s)}{p_1(s)} Ry_2.$$

With this change in the model, we have to change the definition of efficiency: an attainable banking allocation is defined to be an attained allocation as previously defined with

the additional requirement that the contracts are required to be incomplete. An attainable banking allocation is defined to be constrained efficient if there is no attainable banking allocation that Pareto dominates it. We define a banking equilibrium in an analogous way, adding the requirement that the banks must choose incomplete, incentive-compatible contracts. Then the following proposition shows that, in this sense, there is no market failure in equilibrium.

Proposition 3 *Under a mild non-satiability condition, slightly stronger than the usual one because it has to take account of the fact that banks are restricted to incomplete, incentive-compatible contracts, it can be shown that every banking equilibrium allocation is constrained efficient.*

The importance of this result is that it shows that liquidity provision is efficient under laisser faire, even if incomplete contracts require default in some states of nature, as long as markets are complete. There is no market failure and the incidence of crises is socially optimal. In other words, there is no rationale for government intervention.

3 Incomplete markets

A crucial assumption underlying the results in the previous section is the existence of a complete set of contingent markets which allow intermediaries and banks to arrange at date 0 for the provision of liquidity at future dates. This is what we call ex ante provision of liquidity. In practice, markets are incomplete and liquidity must be obtained ex post, after the state of nature is realized.

3.1 Market provision of liquidity

To illustrate the provision of liquidity with incomplete markets we make the following changes in the model. First, we assume that there are no contingent commodity markets, but that intermediaries can hold assets to transfer wealth between dates. Secondly, we assume there is an asset market at date 1 that allows the long asset to be traded for the good. As before, we begin with the case of intermediaries that can offer depositors complete, incentive-compatible contracts.

In the absence of contingent commodity markets, intermediaries have to hold assets in order to provide for future liquidity needs. Let $\mathbf{y}_i$ denote the production plan chosen by an intermediary of type i. Let $\mathbf{q} = (q(1), q(2))$ denote the asset price vector, where $q(s)$ denotes the price of the long asset in terms of the good at date 1 in state s. The feasibility conditions are

$$y_{i0} + y_{i2} = 1,$$

$$\lambda_i(s)\, x_{i1}(s) + y_{i1}(s) = y_{i0} + q(s)\,(y_{i2} - z(s))$$

and

$$(1 - \lambda_i(s))\, x_{i2}(s) = y_{i1}(s) + Rz(s),$$

where $z(s)$ is the amount of the long asset held until the last date. The asset market allows interemediaries to collapse their sequential budget constraints into a single budget constraint at date 1. Using the date-0 budget constraint to write the investment in the long asset as $1 - y_{i0}$, we can write the date-1 budget constraint in present value terms as

$$\lambda_i(s)\, x_{i1}(s) + \frac{q(s)}{R}(1 - \lambda_i(s))\, x_{i2}(s) = y_{i0} + \frac{q(s)}{R}(1 - y_{i0}), \text{ for } s = 1, 2. \tag{7}$$

The intermediary of type i chooses a production plan y_{i0} and an incentive-compatiable contract $\mathbf{x}_i$ to maximize $U_i(\mathbf{x}_i)$ subject to the budget constraints (7). While this reduced form of the intermediary's decision problem suffices to determine the maximum utility, we also need the no-arbitrage conditions that come from the complete problem in order to characterize equilibrium:

$$q(s) < R \Longrightarrow y_{i1}(s) = 0, \text{ for } s = 1, 2. \tag{8}$$

In other words, there will be no storage unless the return on the short asset is equal to the return on the long asset. (The asset market will not clear unless $q(s) \leq R$ since someone must hold the long asset). An equilibrium for this economy with incomplete markets consists of an attainable allocation $(\mathbf{x}_1, \mathbf{y}_1, \mathbf{x}_2, \mathbf{y}_2)$ and a price vector $\mathbf{q}$ such that $(\mathbf{x}_i, \mathbf{y}_i)$ satisfies the no-arbitrage conditions (8) and maximizes $U_i(\mathbf{x}_i)$ subject to the budget constraints (7) for $i = 1, 2$.

To simplify, we assume the two ex ante types of consumers are mirror images. This requires that the two states are equally probable,

$$\pi(1) = \pi(2) = \frac{1}{2},$$

and the liquidity shocks have the same marginal distribution

$$\lambda_1(s) = \lambda_2(s') = \begin{cases} \lambda_H & \text{if } s = 1 \neq s' \\ \lambda_L & \text{if } s = 2 \neq s' \end{cases}.$$

Then there is no aggregate uncertainty since

$$\frac{1}{2}\sum_{i=1,2} \lambda_i(s) = \frac{1}{2}(\lambda_H + \lambda_L) = \bar{\lambda}, \text{ for } s = 1, 2.$$

Further, there exists a symmetric equilibrium in which each type of intermediary chooses the same initial portfolio y_0, the deposit contracts $\mathbf{x}_i$ are identical, modulo the permutation of states, and the asset price is independent of the state,

$$q(1) = q(2) = 1.$$

This equilibrium is generically inefficient because markets are incomplete. Moreover, it is constrained inefficient in the sense of Geanakoplos and Polemarchakis (1986). As an illustration, suppose that the two types of agents have the same power utility functions:

$$u_i(c) = \frac{1}{1-\rho} c^{1-\rho}.$$

Pareto efficiency requires that consumption be independent of the state, that is,

$$\mathbf{x}_i(1) = \mathbf{x}_i(2)$$

for each type i. However, if the coefficient of risk aversion ρ is greater than one, we can show that the optimal incentive compatible contract $\mathbf{x}_i$ satisfies

$$x_{i1}(s) > \frac{q(s)}{R} x_{i2}(s),$$

that is, the present value of consumption at date 1 is greater than the present value of consumption at date 2. Intuitively, the higher the coefficient of risk aversion, the more "insurance" the optimal contract provides to the early consumers. Given this inequality, an increase in the fraction of early consumers will raise the cost of of the deposit contract, other things being equal. Since the value of the intermediary's portfolio is constant across states, the budget constraint requires a compensating reduction in consumption at both dates. Thus, $\mathbf{x}_1(1) \ll \mathbf{x}_1(2)$ and $\mathbf{x}_2(1) \gg \mathbf{x}_2(2)$.

The equilibrium allocation is clearly not Pareto efficient but we can also show that it is not constrained efficient. Constrained inefficiency requires that we can make everyone better off ex ante by changing the choices at date 0 and allowing markets to clear at dates 1 and 2. Allen and Gale (2004) demonstrate that this can be achieved by regulating the amount of liquid assets held by both types of intermediary.

Theorem 4 *Asset prices are too high (too low) if relative risk aversion is less than (greater than) one. Welfare can be improved by imposing a maximum (minimum) on liquidity holdings.*

3.2 Arbitrage, fire sales, and asset price volatility

The preceding example shows that we do not need aggregate uncertainty to have inefficient provision of liquidity, but aggregate uncertainty makes the sources of inefficiency clearer. In the first place, selling assets in order to obtain liquidity is likely to be inefficient because of the risk of firesales. One of the properties of optimal insurance is that one pays the premium in states where the marginal utility of money is low and receives payment for damages in states where the marginal utility of money is high. If markets are complete, intermediaries can obtain liquidity in states where they need it and pay for this insurance in other states. When it is necessary to sell assets in order to obtain liquidity, there is a risk that one will end up selling assets in a "fire sale," that is, when prices are low, thus paying a high price for liquidity, the opposite of what optimal insurance would require. If several banks have to sell assets at the same time in order to obtain liquidity, the result may drive down prices, forcing the banks to sell even more assets, causing a vicious spiral that leads to large reductions in prices and capital losses for the selling banks. These losses do not represent a deadweight cost ex post because they are merely transfers from sellers to buyers; however, they do represent inefficient risk sharing ex ante, as we saw in the previous example, because the

seller is paying for liquidity in the state where his marginal utility of money is high rather than in the state where his marginal utility of money is low. The greater the fall in price, the greater the distortion.

This raises the question why the expected fall in prices does not increase the supply of liquidity. The reason is simply that holding liquidity is expensive–the opportunity cost is the foregone return on the long asset–and if the expected capital gains from buying underpriced assets must compensate for the cost of holding liquidity. If the probability of realizing these gains is small, the fall in price must be correspondingly large.

Allen and Gale (1998) study the provision of liquidity by arbitrageurs in a model with incomplete markets. With complete markets, liquidity is provided ex ante. There is no link between liquidity provision and asset sales. Hence asset prices are not affected by liquidity shocks. When markets are incomplete, liquidity provision requires asset price volatility. Arbitrage must be incomplete and asset price volatility will never be eliminated.

We can illustrate the relationship between the cost of liquidity and price volatility using a special case of the model. Suppose that consumers of ex ante type 1 are risk neutral and have no liquidity preference ($\lambda_1 = 0$) whereas consumers of ex ante type 2 have power utility functions with coefficient $\rho > 1$ and random liquidity parameters $\lambda_2(1) < \lambda_2(2)$. Consider an equilibrium $(\mathbf{x}_1, \mathbf{y}_1, \mathbf{x}_2, \mathbf{y}_2, \mathbf{q})$. In this case, we assume that $q(2) < q(1)$ and that the type-1 banks choose $y_{10} > 0$. They are holding liquidity in order to take advantage of the fire sale that occurs in state 2. However, the amount of the short asset held is too small to prevent the fall in asset prices. This is an equilibrium condition, since the first-order conditions for their optimum require that

$$\frac{\pi(1)}{q(1)} + \frac{\pi(2)}{q(2)} = 1.$$

Suppose that $q(s) < R$ for $s = 1, 2$, so that it is not optimal to hold the short asset between dates 1 and 2 in either state. Then the market-clearing condition becomes

$$\lambda_2(s)\, x_{21}(s) = y_{10} + y_{20}$$

for $s = 1, 2$, and subsituting this equation into the first-order conditions for optimality at date 1 gives us

$$\sum_{s=1,2} \pi(s) \left(\frac{y_{10} + y_{20}}{\lambda(s)} \right)^{\rho} (1 - q(s)) = 0,$$

or

$$\sum_{s=1,2} \pi(s)\, \lambda(s)^{-\rho} (1 - q(s)) = 0.$$

Then we have two equations in two unknowns.

Using these equations to solve for the equilibrium prices, we find that asset price volatility is large relative to the size of the liquidity shock when risk aversion is high. To illustrate this relationship, consider the following example. Let $R = 2$, $\lambda(1) = 0.5$, $\lambda(2) = 0.6$, $\pi(1) = 0.5$

and $\rho = 2$. Then the equations above become

$$\frac{0.5}{q(1)} + \frac{0.5}{q(2)} = 1$$
$$0.5(0.5)^{-2}(1 - q(1)) + 0.5(0.6)^{-2}(1 - q(2)) = 0.$$

Solving these equations, we obtain the price vector

$$\mathbf{q} = (1.2198, 0.847\,32).$$

Note that although the liquidity shock is 20%, ($\frac{0.6}{0.5} = 1.21$), the price change is 40% ($\frac{1.2198}{0.847\,32} = 1.4$). This turns out to be a general property of equilibrium.

Proposition 5 *Consider an economy in which* $u_1(c) = c$, $u_2(c) = \frac{1}{1-\rho}c^{1-\rho}$, $\lambda_1(s) = 0$ *for* $s = 1, 2$ *and* $0 < \lambda_2(s) < 1$ *for* $s = 1, 2$. *Let* $(\mathbf{x}_1, \mathbf{y}_1, \mathbf{x}_2, \mathbf{y}_2, \mathbf{q})$ *be an equilibrium and suppose that* $q(s) < R$ *for* $s = 1, 2$ *and* $y_{10} > 0$. *Then*

$$\log\left(\frac{q(2)}{q(1)}\right) = \rho \log\left(\frac{\lambda(1)}{\lambda(2)}\right).$$

Intuitively, the total consumption at date 1 is constant (see the market-clearing condition above), so as the liquidity shock goes up the per capita consumption goes down (and it goes down relative to the future consumption). For this adjustment to be optimal in equilibrium, prices have to go down by more than the change in relative in consumption because demand is relatively inelastic ($\rho > 1$).

The inefficiency implied by this relation is clear: whereas the first best requires that the risk neutral intermediaries absorb all the risk, the price volatility implied by this equation shows that most of it is borne by the risk averse types.

A fuller investigation of this example is contained in the appendix.

3.3 Limited market participation

[To be completed]

3.4 Microprudential regulation

We have seen that, under certain circumstances, a regulated increase in the holding of liquid assets can lead to a Pareto-improvement. In our specific example, when the coefficient of relative risk aversion is greater than one, an increase in holdings of liquid assets, which causes an increase in the asset price, will increase ex ante expected utility. More generally, when markets are incomplete, the work of Geanakoplos and Polemarchakis tells us that equilibrium are generically constrained inefficient. So we should not expect the market provision of liquidity to be constrained efficient. There will exist *some* intervention by the regulator that

is welfare increasing. Identifying this policy may not be easy and implementing it may be impractical, but the possibility of welfare improving regulation exists.

What conclusions can be drawn from the preceding discussion? Certainly, holding liquid assets entails an opportunity cost in the form of the higher rate of return that could have been earned on less liquid assets. When markets are incomplete, one bank's holding of liquid assets creates a pecuniary externality not internalized by bank itself. The harder question to answer is whether the level of liquidity in the system is likely to be too high or too low. Intuition may suggest that each bank will hold too little, since it bears the cost but does not capture all of the benefit. But in a general equilibrium system, we cannot assume that there is a benefit from more liquidity, however intuitive this may sound. No doubt there will be conditions under which there is too little liquidity, but we can also imagine cases where there is too much.

Even if we willing to assume that the level of liquidity is too small most of the time, the theoretical examples reviewed do not suggest that minimum liquidity ratios would solve the problem. We have seen examples where requiring banks to hold more liquid assets may raise the future asset price or reduce asset-price volatility, *but only if the banks are free to use their liquidity to buy assets*. More generally, liquidity that banks are forced to hold come what may is of no value in providing liquidity for the system as a whole. This is just another example of *Goodhart's taxi*. The beneficial effects of having "cash in the market," as Allen and Gale call it, are limited to cash that can be used whenever the opportunity arises, rather than cash that banks are forced to hold as a result of liquidity requirements.

There are several alternatives to liquidity ratios. One is the payment of interest on liquid balances or reserves, essentially a Pigovian subsidy on liquidity. This will encourage the accumulation of liquid balances that can then be run down when profitable opportunities arise. Another, is the creation of liquidity facilities by the central bank, essentially an extension of its role as lender of last resort. A third would be the introduction of some sort of liquidity insurance, in effect, an attempt to replicate the complete markets of the first model we looked at. How well these alternatives will work is beyond the scope of the present discussion, but they all deserve further study.

4 Macroprudential regulation

Macro-prudential regulation raises a different set of issues from micro-prudential regulation. We argued above that there may be an inadequate level of liquidity in the financial system when markets are incomplete; but this conclusion is consistent with the view of Goodfriend and King (1988) that, *under normal conditions*, financial markets allocate liquidity efficiently as long as the aggregate level of liquidity in the system is adequate. In a crisis, on the other hand, providing adequate liquidity in the aggregate may not be enough. When the stability of the entire financial system is threatened, there is a tendency for liquidity to "dry up." There is a flight to quality, an unwillingness to lend at any but the shortest maturities, and banks may be unwilling to lend to one another. There may be enough liquidity in the system, but the markets are frozen. In this case, the central bank must step in as the lender of first

resort and replace the frozen markets.

During the crisis, central banks responded by creating a variety of different facilities to provide liquidity to the financial system, going beyond the normal lender of last resort function. The Federal Reserve System encouraged banks to make use of the discount window and, when this proved insufficient, set up a term auction facility (TAF) to allow banks to access liquidity without the stigma of the discount window. Subsequently, the FRS went on to set up a variety of facilities that provided loans to banks and non-banks and to assist borrowing by issuing guarantees. Instead of requiring treasuries or agency debt as collateral, a wide range of collateral was accepted. The Bank of England and the ECB took similar steps.

These were emergency measures and in some cases were taken quite unwillingly. Here, for example, are the views of the Governor of the Bank of England in a letter to the chairman of the House of Commons Treasury Committee (King, 2007).

> "Third, the moral hazard inherent in the provision of ex post insurance to institutions that have engaged in risky or reckless lending is no abstract concept. The risks of the potential maturity transformation undertaken by off-balance sheet vehicles were not fully priced. The increase in maturity transformation implied by a change in the effective liquidity in the markets for asset-backed securities was identified as a risk by a wide range of official publications, including the Bank of England's Financial Stability Report, over several years. If central banks underwrite any maturity transformation that threatens to damage the economy as a whole, it encourages the view that as long as a bank takes the same sort of risks that other banks are taking then it is more likely that their liquidity problems will be insured ex post by the central bank. The provision of large liquidity facilities penalises those financial institutions that sat out the dance, encourages herd behaviour and increases the intensity of future crises."

By contrast, others were claiming that providing lliquidity is costless for the central bank—and only the central bank—so liquidity should be provided on demand.

> "Liquidity is a public good. It can be managed privately (by hoarding inherently liquid assets), but it would be socially inefficient for private banks and other financial institutions to hold liquid assets on their balance sheets in amounts sufficient to tide them over when markets become disorderly. They are meant to intermediate short maturity liabilities into long maturity assets and (normally) liquid liabilities into illiquid assets. Since central banks can create unquestioned liquidity at the drop of a hat, in any amount and at zero cost, they should be the liquidity providers of last resort, both as lender of last resort and as market maker of last resort. There is no moral hazard as long as central banks provide the liquidity against properly priced collateral, which is in addition subject to the usual 'liquidity haircuts' on this fair valuation. The private provision of the public good of emergency liquidity is wasteful. It's as simple as that."

(Willem Buiter as quoted by Cao and Illing, 2010).

The problem of moral hazard as outlined by the Governor is a puzzling one. What exactly is cost of providing liquidity to the banks in question? Cao and Illing (2010) have described a model in which banks engage in *real* maturity transformation, borrowing short in order to invest in assets that have a random payoff date. When output arrives late, the consumers suffer. Bailouts in the form of monetary injections depreciate the nominal claims of the depositors and allow the banks to avoid default and gain profits in the form of delayed output at the expense of their depositors. Anticipating a bailout, banks will engage in even more of this naughty transformation. This story conforms fairly closely to the paranoid vision of the Governor, but it does not look like what has been happening over the last few years.

The liquidity facilities set up by the central banks allow banks to borrow reserves if they deposit enough collateral. If the central bank is simply substituting one asset, either reserves or government debt, for another, it is not clear that any cost is imposed on the system. Some economists fret about the inflationary implications of expanding the central bank's balance sheet, but so far these fears have not been realized. In any case, new techniques of monetary policy, such as the payment of interest on reserves, are being used to sterilize the effects of these liquidity injections.

Of course, once the central bank is seen to make funding available in emergencies, it will encourage the banks to use short-term lending as a source of finance. If there are limits to the central bank's ability to provide liquidity, there may be a threat of dislocation in the financial system if these limits are breached. It is easy to see why a central banker might fear the practical difficulties of providing so much liquidity. But if there is an argument for limiting the extent of maturity transformation, it is probably best achieved through more intelligent and aggressive liquidity regulation, either by setting explicit limits on the asset-liability maturity gap or by linking capital requirements to the asset-liability maturity gap.

If King is too paranoid, perhaps Buiter is too cavalier. In journalistic discussions, any action by the central bank tends to be characterized as a "bailout" without much consideration for who benefits and who pays the cost. It is an interesting and often over-looked question whether liquidity provision necessarily constitutes a "bailouts." On the one hand, if the central bank is lending against the security of (somewhat) risky assets, there will be some states of nature in which it makes a loss. On the other hand, if the central bank demands riskless collateral—prior to the collapse of Norther Rock, the Bank of England was only willing to lend against gilts—then it is not supplementing the market provision of liquidity in any meaningful sense. A bank that can put up riskless assets as collateral should not have any problem obtaining as much liquidity as it needs from the market. Of course, it is all a matter of degree. In circumstances where the market seems excessively or unreasonably risk averse, the central bank may feel that it is reasonable to adopt a more flexible definition of liquidity provision, as the various central banks did beginning in the autumn of 2007.

A simple example may make these distinctions clearer. As usual, we assume there are three dates, indexed by $t = 0, 1, 2$, with a single good that can be used for consumption or

investment at each date. There are two assets, a short-term safe asset and a long-term risky asset. The short asset is represented by a storage technology: one unit of the good invested at date t yields one unit of the good at date $t+1$. The long asset produces a random return of $\tilde{R}$ units of the good at date 2 for each unit of the good invested at date 0. We assume that

$$\tilde{R} = \begin{cases} R_H & \text{w. prob. } 1-\varepsilon \\ R_L & \text{w. prob. } \varepsilon \end{cases}$$

where $R_H > R_L > 0$ and $0 < \varepsilon < 1$ are constants.

There is a large number of risk averse consumers, each of whom has an endowment of one unit of the good at date 0 and values consumption at date 2 only. Their risk preferences are represented by a vNM utility function having the form $u(c) = \frac{1}{1-\rho} c^{1-\rho}$, where $\tilde{\rho}$ is a random variable defined by

$$\tilde{\rho} = \begin{cases} \rho_H & \text{w. prob. } \delta \\ \rho_L & \text{w. prob. } 1-\delta \end{cases}$$

where $\rho_H > \rho_L > 0$ and $0 < \delta < 1$. We assume that $\tilde{R}$ and $\tilde{\rho}$ are independent. The information structure is illustrated in Figure 1 below.

—Figure 1 about here—

The planner's problem is to choose an amount y to invest in storage so as to maximize the expected utility of the typical consumer.

$$\begin{aligned} &\max \mathbf{E}\left[u\left(y+(1-y)\tilde{R}\right)\right] = \\ &\quad \delta\left\{\frac{1}{1-\rho_H}\left((1-\varepsilon)\left(y+R_H(1-y)\right)^{1-\rho_H}+\varepsilon\left(y+R_L(1-y)\right)^{1-\rho_H}\right)\right\}+ \\ &\qquad (1-\delta)\left\{\frac{1}{1-\rho_L}\left((1-\varepsilon)\left(y+R_H(1-y)\right)^{1-\rho_L}+\varepsilon\left(y+R_L(1-y)\right)^{1-\rho_L}\right)\right\}. \end{aligned}$$

Suppose, for example, that $R_H = 2$, $R_L = 0.5$, $\varepsilon = 0.01$, $\rho_L = 0$, $\rho_H = 20$ and $\delta = 0.01$. The objective function achieves a maximum at $y = 0.22046$.

The question is: How can we implement this allocation if the banks use short-term finance, for example, issuing one period bonds that have to be rolled over at the intermediate date? The problem is that in the event that risk aversion is high at the intermediate date, it will be impossible to roll over the debt. More precisely, the maximum value of the debt that can be issued will be much less in the high risk aversion state than in the low risk aversion state. If we use the parameters from the numerical example, we can see that the debt capacity, the maximum amount that can be borrowed at date 1 with high risk aversion is

$$\frac{\left(0.99\left(0.22046+(1-0.22046)\,2\right)^{-20}2+0.01\left(0.22046+(1-0.22046)\,0.5\right)^{-20}0.5\right)}{\left(0.99\left(0.22046+(1-0.22046)\,2\right)^{-20}+0.01\left(0.22046+(1-0.22046)\,0.5\right)^{-20}\right)} = 0.5$$

compared to

$$\frac{(0.99\,(2) + 0.01\,(0.5))}{(0.99 + 0.01)} = 1.985$$

with low risk aversion. At the initial date, it is necessary to issue debt with a high face value 1.985 in order to finance the investment in the portfolio consisting of 0.22046 units of the short asset and 0.77954 units of the long asset. But this will ensure a default in the bad state.

Decentralization of this allocation presents a problem because, when risk aversion is very high, investors do not want to hold even slighty risky assets. What can the central bank do to implement the planner's problem? The central bank can accommodate the flight to quality by selling government bonds to individuals and lending the revenue directly to the banks. The banks are able to pay off their debts and the previous holders of the debt use this repayment to purchase government bonds from the central bank. In the final period, the firms will have to repay their loans to the central bank; or, the central bank will seize the collateral put up by the firms in the form of claims on the output. In the event that the returns are high, the banks can repay the central bank and use this money to redeem the government bonds. If the returns are low, on the other hand, neither banks will be unable to repay their loans to the central bank and the government will be forced to bail out the banking system.

Anyway, if the use of the risky asset is efficient, it seems that a government bailout will implement the efficient allocation. There is another problem, however. The anticipation of a government bailout will change the investment decision at date 0. Even though the probability of a low return in the high risk aversion state is very improbable ($p = 0.001$), depending on the size of the haircut, the "bailout" component of the central bank's liquidity facility may cause the risky asset dominate the safe asset. For example, if the government guarantees the payoff to the risky asset is at least 60 in the bad state, the optimum for the planner's problem is $y = 0$. So the anticipated "bailout" has the expected effect of increasing the size of the investment that must be bailed out.

We have so far assumed that it makes no difference whether the central bank is providing real or nominal liquidity. If the central bank wants to maintain price stability, it may not matter much whether we model the provision of liquidity in real or monetary terms. The argument that liquidity is costless is based on the notion that the central bank can "print money" and sooner or later this will produce inflation. Of course, the inflation tax is far from costless although there are circumstances in which it will not have any distortionary effect. Allen, Carletti and Gale (2009) have investigated the role of the interbank market and central bank intervention in the bond market in achieving an efficient allocation of risk through the financial system. Allen, Carletti and Gale (2010) conducts a similar investigation in a monetary economy in which bank deposits are denominated in terms of money and the central bank can provide liquidity by printing money. In that case, it is possible to achieve the first best through changes in the price level that adjust the real value of deposits in response to real shocks.

We have also seen that excessive asset price volatility arises when markets are incomplete

and that this volatility may be exacerbated by incomplete market participation. Whereas some authors (e.g., Taylor) have argued that loose monetary policy played a role in creating asset price bubbles, monetary policy may have a more beneficial role in preventing fire sales.

There is also an efficiency question lurking behind the moral hazard argument. Is it more efficient to use longer term finance if it is more expensive? Suppose that banks have two sources of funds, one long-term and expensive, the other short-term and less expensive. If the roll-over of debt at the same rate of interest is assured, the bank will prefer to use short-term debt. The fact that the two sources of funds both exist suggests that there is some difference between them. Either the interest rate at which the debt can be rolled over is uncertain or the possibility of rolling over the debt is uncertain. If the probability of rolling over the short term debt and making a profit is sufficiently high, the bank may prefer to choose short-term funding even if it involves some risk of illiquidity. Whether it is more efficient to limit the use of short-term finance in this case requires a cost-benefit analysis that so far has not been made.

5 The parallel banking system

Despite the problems caused by the sub-prime crisis, the securitization model has proved itself as an efficient and stable technology for sharing risk and expanding access to credit. A functioning securitization market is essential to restore the capacity of the financial system to provide credit for business. The PBS needs to be rebuilt but it needs to be a different kind of banking system, one based on limited purpose financial companies (LPFC) performing the function of a financial utility warehousing high quality assets rather than a hedge fund or SIV taking large risks in pursuit of high rates of return on equity.[3]

Why are such utility banks needed? since the early 1990s, banks have earned much higher rates of return on equity (ROE) than the traditional utility bank could achieve. They did by taking on much higher, by engaging in proprietary trading, and by securitizing loans in order to increase the velocity of capital. The increasing complexity of banks has also added to their opacity. Capital is expensive for banks because of the combination of risk and opacity. There is a need for Narrow Banks to provide a lower but safer return from a more transparent business model. These Narrow Banks are also essential for the functioning of the casino-like universal banks, because they provide a demand for high-quality assets that do not require the expertise or expensive capital that universal banks can provide. By off-loading high-quality assets to the PBS, universal banks can hold high-risk, high-return and opaque assets and concentrate on information intensive activities such as origination and securitization. In the financial ecosystem, Narrow Banks are somewhere between MMFs, on the one hand, and traditional banks and investment banks (or universal banks), on the other.

Another reason for wanting to revive the PBS is to allow the central banks to exit from their current role as providers of liquidity directly to banks and market participants. In the

[3] This section is based on a presentation by Gordian Knot to the Liquidity Working Group at the Federal Reserve Bank of New York on October 9, 2009.

last few years, the FRS has instituted a number of facilities to supplement the ABCP market and other sources of liquidity. Most of these lending facilities involve collateral in the form of securities. The danger of providing these services in the longer run is that it will undermine the private sector's provision of credit. What would be lost is the private sector's presumably superior ability to assess and price risk. As Goodfriend and King (1988) persuasively argued, the Fed does not have a comparative advantage in providing credit except to the extent that its control of the money supply makes it easy to obtain loanable funds. This suggests that the efficiency of the financial system requires the Fed to find an exit from its current role and that requires a re-invigorated PBS.

The basic building block of the PBS is a Narrow Bank (LPFC) that holds securities originated by traditional banks and finances them with a combination of capital (equity) and short- and medium-term financing from MMFs, securities lenders, pension and insurance companies, and underexposed banks. The requirements to be a Narrow Bank are those of a utility company:

- The target returns to capital providers will be low, but stable,
- The activities that the Narrow Bank can engage in (e.g., the type of securities held) must be fully defined by a charter that is transparent to capital and debt investors and to regulators.
- The capital structure must be responsive to (i) the credit quality of the portfolio; (ii) the maturity of the assets; (iii) the diversification of the portfolio; and (iv) the asset-liability maturity gap. For example, the capital requirements could be based on a Basel-type formula that incorporates measures of credit quality, maturity, diversification, and asset-liability gap.
- No off-balance-sheet exposures or contingent risk exposures would be allowed.
- There would be no proprietary trading and most assets would be held to maturity.
- Interest rate and currency exposure would be fully hedged.
- Risk management would be consistent with "best practice;" in particular, asset quality evaluation would be based on in-house research and would not rely on rating agencies.
- **Most importantly, the Narrow Bank would be regulated by the FRS and would have access to central bank liquidity facilities.**

The key ingredient in this proposal is the highlighted requirement for access to liquidity from the FRS. The extension of liquidity to institutions that are not traditionally recognized as banks will be resisted by many who see any lending, even as a last resort, as a "bailout." To the extent that the Fed is accepting risky collateral, it may incur losses in the future; however, to the extent that the Fed is responding to a liquidity crisis and accepts only "good security," in Bagehot's phrase, the extension of liquidity may be costless or even profitable.

If the Fed's intervention is successful and profitable, it may become acceptable to have the monetary authorities providing such facilities in on a wider basis, but only as a last resort.