

Monetary Policy as Financial-Stability Regulation

Jeremy C. Stein*
April 2010

Abstract

This paper develops a model that speaks to the goals and methods of financial-stability policies. There are three main points. First, from a normative perspective, the model defines the fundamental market failure to be addressed, namely that unregulated private money creation can lead to an externality in which intermediaries issue too much short-term debt and leave the system excessively vulnerable to costly financial crises. Second, it shows how in a simple economy where commercial banks are the only lenders, conventional monetary-policy tools such as open-market operations can be used to regulate this externality, while in more advanced economies it may be helpful to supplement monetary policy with other measures. Third, from a positive perspective, the model provides an account of how monetary policy can influence bank lending and real activity, even in a world where prices adjust frictionlessly and there are other transactions media besides bank-created money that are outside the control of the central bank.

* Department of Economics, Harvard University and NBER. I am grateful to Fan Zhang for research assistance, and to Effi Benmelech, Robin Greenwood, Sam Hanson, Arvind Krishnamurthy, David Scharfstein, Andrei Shleifer, Robert Vishny and seminar participants at Harvard and Northwestern for their comments. Thanks also to Mary Goodman, Sam Hanson, Matt Kabaker, Andrew Metrick, Charlie Nathanson, Larry Summers and Adi Sunderam for a series of early conversations that helped to shape the ideas in this paper.

I. Introduction

The modern literature on monetary policy takes the view that a central bank's primary mission is to achieve *price stability*.¹ Historically, however, a dominant concern for central bankers has been not just price stability, but also *financial stability*. Goodhart (1988) argues that the original motivation for creating central banks in many countries was to temper the financial crises associated with unregulated “free banking” regimes:

“In the nineteenth century, the advocates of free banking argued that the banking system could be trusted to operate effectively without external constraints or regulation....[But] experience suggested that competitive pressures in a milieu of limited information (and, thence, contagion risks) would lead to procyclical fluctuations punctuated by banking panics. It was this experience that led to the formation of noncompetitive, non-profit maximizing Central Banks.” (p. 77).

A related emphasis on crisis mitigation is evident in Bagehot's (1873) famous discussion of the lender-of-last-resort function.² And certainly, recent events have served to underscore the importance of the central bank's role in preserving financial stability.

In this paper, I develop a model that speaks to the goals and methods of central-bank financial-stability policies. The first step is to define the fundamental market failure that needs to be addressed. I begin with an unregulated banking system in which banks raise financing from households to invest in projects. Banks can raise this financing in the form of either short-term or long-term debt. Households are risk-neutral with respect to fluctuations in their consumption, but derive additional monetary services from holding any claim that is entirely riskless—with the notion being that riskless claims are easy to value and hence facilitate exchange among households. I show that banks can manufacture some amount of riskless private “money” of this sort, thereby lowering their

¹ See, e.g., Goodfriend (2007) for a recent articulation of this view.

² Tucker (2009) paraphrases Bagehot's (1873) dictum as follows: “to avert panic, central banks should lend early and freely (i.e., without limit) to solvent firms, against good collateral, and at ‘high rates’”.

financing costs. However, the only way for them to do so is by issuing short-term debt; no amount of long-term bank debt can ever be guaranteed to be risk-free.

The role for financial-stability policy arises because the private choices of unregulated banks with respect to money creation are not in general socially optimal. When banks issue cheaper short-term debt, they fully capture its social benefits, namely the monetary services it generates for households. However, they do not always fully internalize its costs. In an adverse “financial crisis” state of the world, the only way for banks to honor their short-term debts is by selling assets at fire-sale prices. I show that in equilibrium, the potential for such fire sales may give rise to a negative externality. Thus left to their own devices, unregulated banks may engage in excessive money creation, and may leave the financial system overly vulnerable to costly crises.

There are a variety of ways for a regulator to address this externality. One possibility is the use of conventional monetary-policy tools, i.e. open-market operations. To see how monetary policy might be of value, note that a crude approach to dealing with the externality would be for the regulator to just impose a cap on each bank’s total money creation. However, when the regulator is imperfectly informed about banks’ investment opportunities, he will not know where to set the cap, since it is desirable for banks with stronger investment opportunities to do more money creation. In this setting, the regulator can do better with a flexible “cap-and-trade” system in which banks are granted tradable permits, each of which allows them to do some amount of money creation.³ The market price of the permits reveals information about banks’ investment opportunities to the regulator, who can then adjust the cap accordingly—when the price of the permits

³ Kashyap and Stein (2004) suggest using an analogous cap-and-trade approach to implement time-varying bank capital requirements.

goes up, this suggests that banks in the aggregate have strong investment opportunities, and so the regulator should loosen the cap by putting more permits into the system.

All of this may sound a bit like science fiction; we don't observe cap-and-trade regulation of banks in the real world. However if banks' short-term liabilities are subject to reserve requirements, it turns out that monetary policy can be used as a mechanism for implementing the cap-and-trade approach. When the central bank injects reserves into the system, it effectively increases the number of permits for private money creation. And the nominal interest rate, which captures the cost of holding reserves, functions as the permit price. Thus open-market operations that adjust aggregate reserves in response to changes in short-term nominal rates can be used to achieve the cap-and-trade solution.

An interesting benchmark case is where reserve requirements apply to the money-like liabilities of *all* lenders in the economy. This allows the central bank to precisely control private money creation *with monetary policy alone*. While this case may roughly capture the situation facing central banks at an earlier period in history, it is less realistic as a description of modern advanced economies. Nowadays there are a range of short-term financial-intermediary liabilities that are not subject to reserve requirements, and yet may both: i) provide monetary services; and ii) create fire-sale externalities. For example, Gorton and Metrick (2009), and Gorton (2010a,b) argue that an important fraction of private money creation now takes place entirely outside of the formal banking sector, via the large volume of overnight repo finance in the "shadow banking" sector.

In this richer environment, conventional monetary policy is by itself generally not sufficient to rein in excessive money creation. Continuing with the above example, it may in addition be necessary to more directly regulate the volume of repo activity in the

shadow-banking sector. Thus the model helps to make clear the circumstances under which monetary policy needs to be supplemented with other measures. Moreover, it suggests that these other measures lie squarely in the central bank’s traditional domain, to the extent that they are all targeted at the fundamental externality associated with excessive private money creation. This is of interest in light of the ongoing debate over the appropriate mix of central-bank tools for achieving financial stability.⁴

In addition to its normative implications, the model is also relevant from a purely positive perspective. It provides a coherent account of how monetary policy “works”—i.e., of how open-market operations lead to changes in bank lending and output—in a simple environment that is arguably more realistic on some key dimensions than that found in other theories. In contrast to the textbook model, all prices are perfectly flexible. Moreover, I do not need to assume that the central bank has monopoly control over *all forms* of transactions media used by households. My model is unchanged if, for example, one introduces a set of non-reservable money-market-funds that provide the same monetary services to households as bank-created money.⁵ Indeed, I consider the limiting case where the interest-rate spread between money and bonds is fixed and unresponsive to their relative supplies. Monetary policy works in this case not by changing real interest rates, but through a pure quantity effect: a loosening of policy allows banks to finance themselves with more of the cheaper money, which encourages them to do more lending.

⁴ See, e.g. Adrian and Shin (2008), and Ashcraft, Garleanu and Pedersen (2010).

⁵ To be clear on the distinction: my model assumes that the central bank acts as a regulator, controlling those forms of private money creation that lead to negative externalities—in particular, short-term bank debt that finances risky long-term assets. However, unlike the textbook model, it does not require the central bank to control other, more benign forms of money creation, e.g., money-market-fund accounts backed by Treasury bills. See, e.g., Cochrane (1998), who argues that the latter feature is a major weakness of the textbook model.

The ideas in this paper connect to several strands of previous work. First, the basic model of fire sales that creates the rationale for policy intervention draws heavily on Shleifer and Vishny (1992, 1997).⁶ Second, the insight that banks create a valuable transactions medium by issuing low-risk claims is due to Gorton and Pennacchi (1990). Third, the notion that central bank reserves can be thought of as permits that allow banks to do more of a particular kind of cheap financing appears in Stein's (1998) elaboration of the bank lending channel of monetary policy transmission.⁷

And finally, in order to focus clearly on the financial-stability consequences of monetary policy, it helps to set aside any effects it might have on price stability. I do so by appealing to the fiscal theory of the price level, according to which the price level is determined not by the monetary base, but by total outstanding nominal government liabilities—i.e., by the sum of Treasury securities and the monetary base.⁸ As will become clear, this enables open-market operations that change the *mix* of Treasuries and bank reserves (while keeping their sum constant) to have real effects on bank investment and financing behavior, even in a world where all prices are perfectly flexible.

The rest of the paper is organized as follows. Section II develops the basic model of private money creation by banks. Section III compares banks' financing choices to the social planner's solution, and clarifies the conditions under which banks engage in excessive money creation. It also shows that a cap-and-trade approach to regulation can

⁶ On fire sales, see also Allen and Gale (2005), Brunnermeier and Pedersen (2009), Fostel and Geanakoplos (2008), Geanakoplos (2009), Gromb and Vayanos (2002), Morris and Shin (2004), Caballero and Simsek (2009), and Stein (2009).

⁷ For early work on the bank lending channel see also Bernanke and Blinder (1988, 1992), Kashyap, Stein and Wilcox (1993), and Kashyap and Stein (2000).

⁸ The fiscal theory is developed in Leeper (1991), Sims (1994), Woodford (1995), and Cochrane (1998). My own adaptation of the theory is particularly indebted to Cochrane's exposition.

be useful when the social planner has imperfect information. Section IV demonstrates how the cap-and-trade approach can be implemented with open-market operations. Section V explores a number of other complementary policy tools; these include deposit insurance and a lender-of-last-resort function, as well as regulation of the shadow-banking sector. Section VI discusses how the model differs from other accounts of the monetary transmission mechanism. Conclusions are in Section VII.

II. A Model of Private Money Creation

The model features three sets of actors: households, banks, and “patient investors”. I begin by describing each of these groups, and then turn to the optimization problem faced by the banks.

A. Households

There are three dates, 0, 1, and 2. At time 0, households have an initial endowment of the one good in the economy. They can either consume this endowment at time 0, or invest some of it in financial assets and consume the proceeds from investment at time 2. They have linear preferences over consumption at these two dates. In addition to direct consumption, households also derive utility from monetary services. The key assumption is that monetary services can be provided by any privately-created claim on time-2 consumption, so long as that claim is *completely riskless*.⁹ Thus the utility of a representative household is given by:

$$U = C_0 + \beta E(C_2) + \gamma M, \quad (1)$$

⁹ This assumption is meant to capture, in a reduced-form way, the spirit of Gorton and Pennacchi (1990), and Dang, Gorton and Holmstrom (2009). These papers argue that information-insensitive securities are an attractive medium of exchange, because they eliminate the potential for adverse selection between transacting parties. Completely riskless securities are, by definition, entirely information-insensitive.

where M represents the household's time-0 holdings of privately-created "money".¹⁰ To be clear on the notational convention, when I say that a household has M units of money at time 0, I mean that it holds claims that are guaranteed to deliver M units of time-2 consumption.

Given their linear form, household preferences completely pin down two real rates. The first is the (gross) real return on risky "bonds" that pay off at time 2, which is given by $R^B = 1/\beta$. The second is the (gross) real return on riskless "money", which is given by $R^M = 1/(\beta + \gamma)$, where I am assuming that $\beta + \gamma < 1$. The latter follows from the observation that a household is always indifferent between having: i) $\beta + \gamma$ units of time-0 consumption; or ii) a riskless claim that promises one unit of time-2 consumption, since such a claim delivers β of utility from expected future consumption, along with an additional γ of utility in monetary services. The bottom line is that because riskless money offers households a convenience yield that risky bonds do not, in equilibrium it must have a lower rate of return.

The idea that money has a lower real return in equilibrium than bonds is standard in textbook models. But here, unlike in the textbook models, the return differential is *fixed and independent of the quantities of money and bonds*, thanks to the assumption of linear preferences on the part of households. This assumption is not necessary for anything that follows, and is easily relaxed. However, it serves to highlight a key novelty of my model: here changes in central bank policy work not by altering the real rates on

¹⁰ In a similar formulation, Krishnamurthy and Vissing-Jorgensen (2008) put the stock of Treasury securities directly into the representative agent's utility function. As one rationale for doing so, they cite the "surety" of Treasuries—i.e., the fact that Treasuries are riskless. Like I do, they posit that surety has an extra value above and beyond that which is captured in a standard asset-pricing model. See also Sidrauski (1967) for an early model with money in the utility function.

either type of claim, but rather by varying the proportions of each that banks use to finance themselves. In other words, looser monetary policy encourages banks to lend more by enabling them to tilt their capital structure towards cheap money financing, thereby lowering their weighted average cost of funds.

B. Banks

Households cannot invest their time-0 endowments directly in physical projects, because they do not have the monitoring expertise to do so. This investment must be undertaken by banks, who in turn issue financial claims—in the form of either riskless money or risky bonds—to households. There is a continuum of such banks, with total mass of one. Each bank faces the following investment opportunities. If an amount I is invested at time 0, and the good state prevails, which happens with probability p , total output at time 2 is given by the concave function $f(I) > I$. If instead the bad state prevails, total *expected* output at time 2 is $\lambda I \leq I$, and there is a positive probability that output collapses all the way to zero. In particular, in the bad state, output is either $\lambda I/q$ with probability q , or zero with probability $(1 - q)$.

At time 1, there is a public signal that reveals whether the good or bad state will be realized at time 2. At time 1 it is also possible for a bank to sell any fraction of its existing physical assets to a patient investor.¹¹ If a fraction Δ of the assets are sold, total proceeds to the bank are given by $\Delta k \lambda I$, where $0 \leq k \leq 1$, and the remaining unsold assets yield output at time 2 to the bank of $(1 - \Delta) \lambda I$. Thus k is a measure of the discount to expected value associated with a time-1 asset sale. A central feature of the model is that

¹¹ Since households only consume at time 0 and time 2, they do not consume the proceeds of any time-1 asset sales until time 2. One can think of these proceeds as being held on their behalf in a riskless escrow account in the interim.

k is endogenous, and depends on total asset sales by all banks in the economy. The equilibrium determination of k will be discussed shortly.

Other than their access to investment opportunities, banks have no initial endowments of their own, and hence have to raise the entire amount I externally. They can do so by issuing either short-term (maturing at time 1) or long-term (maturing at time 2) debt claims to households. Note that if they finance entirely with long-term debt, no amount of this debt can ever be riskless, irrespective of its seniority structure, since there is a positive probability of the assets yielding zero output at time 2 in the bad state. By contrast, short-term debt can be made riskless, as long as not too much is issued. This is because by forcing an asset sale upon seeing a bad signal at time 1, short-term creditors can escape early with a sure value equal to the proceeds from the sale.

Herein lies the central tradeoff: on the one hand, banks have an incentive to issue some short-term debt, because this debt can be made riskless—and hence by virtue of its money-ness, represents a cheap form of financing.¹² On the other hand, what keeps short-term debt safe is the bank's ability to sell assets to pay off short-term creditors in the bad state. As will become clear below, these sales of existing assets can lead to social costs that are not always fully internalized by individual banks when they pick their capital structures. As a result, there may be excessive private money creation by banks.

Suppose that a bank raises a fraction m of its total investment of I by issuing short-term debt. If this short-term debt can be made riskless, it will carry a rate of return of R^M , and the bank will owe its short-term creditors a repayment of $mIR^M \equiv M$. Can

¹² Other theories of short-term financing include Flannery (1986), Diamond (1991), and Stein (2005), who stress its signaling properties, and Diamond and Rajan (2001) who argue that short-term debt is a valuable disciplining device, particularly for financial intermediaries.

it meet this promise in the bad state by selling assets if necessary? From above, if it sells a fraction Δ of its assets, total proceeds are $\Delta k \lambda I$, so we require that:

$$\Delta k \lambda I = m I R^M, \text{ or } \Delta = \frac{m R^M}{k \lambda}. \quad (2)$$

Since it must be that $\Delta \leq 1$, there is an upper bound on private money creation given by:

$$m^{\max} = \frac{k \lambda}{R^M}. \quad (3)$$

Thus the potential for asset sales makes it possible for a bank to create riskless private money, by issuing short-term debt—so long as the amount issued is not too large.

Is it also the case that asset sales are an *unavoidable consequence* of money creation? One might think that since holding on to assets is positive-NPV relative to selling them at time 1, it might be possible for a bank to raise new funding at time 1 to pay off the departing short-term creditors, and thereby avoid forced sales. However, if one assumes that any new funding must be subordinated to existing long-term debt, such new funding may be blockaded by a severe debt overhang problem (Myers (1977)), given the low value of the assets in the bad state relative to the total face amount of already-issued debt.¹³ Thus under plausible circumstances, private money creation inevitably leads to some amount of asset sales.¹⁴

¹³ In particular, denoting the face value of the existing long-term debt by B , it must be that $M + B > I$, in order for the bank to have raised I at time 0 by issuing money and bonds. If the bank now wants to raise an amount M to pay off the short-term creditors in the bad state at time 1, it must do so by issuing new claims that are junior to the existing long-term debt. But given that they are junior, the value of these claims in the bad state is only $q(\lambda I/q - B)$. For q large enough (certainly for $q > \lambda$) the value of the new claims is necessarily less than M , so refinancing the short-term debt is impossible.

¹⁴ This line of argument leaves open the question of why the original long-term financing for the bank is in the form of senior debt, as opposed to say equity, or some other junior security that allows for new financing to come in on top of it. Following Hart and Moore (1995), it may be that this seniority of the long-term debt represents a valuable pre-commitment in the more likely good state of the world. For

Before moving on, it is worth fleshing out an issue of interpretation about the banks in the model. In the real world, banks do not invest in physical projects directly, but rather lend to firms who in turn do the project selection. Abstracting away from this extra layer of activity, as I do here, is tantamount to assuming that there are no contracting frictions between operating firms and banks, i.e. that firms can costlessly pledge all of their output to the banks. This then raises the question of whether it is appropriate to interpret what I label “banks” as really being financial intermediaries, as opposed to operating firms that borrow directly from households in the securities market.

To create a meaningful distinction, suppose that any individual operating firm, once funded, always has some probability of immediate (i.e., before time 1) idiosyncratic failure, in which case it becomes public knowledge that its output will be zero in both the good and bad states. This risk of failure makes it impossible for an operating firm to ever issue riskless claims in any amount. Banks, on the other hand, represent highly diversified portfolios of such firm-level projects, and therefore their assets always have positive expected value as of time 1, as assumed above. The diversification associated with banks is thus a necessary condition for them to create riskless claims.¹⁵

C. Patient Investors

Patient investors (PIs) are another type of financial intermediary, and as such, any output that they produce reverts to the household sector at time 2. As a group, PIs are

example, it may prevent bank managers from using assets in place as collateral for negative-NPV empire-building investments. Thus, as in Hart and Moore, senior long-term debt is a double-edged sword: it serves to discipline wayward managers in the good state, but forces underinvestment (here, in the form of asset sales) in the bad state.

¹⁵ Thus, as in other models of intermediation, both pooling (i.e., diversification) and tranching (i.e., the issuance of properly structured senior securities) have roles to play in creating low-risk claims. See, e.g., Gorton and Pennacchi (1990), DeMarzo and Duffie (1999), and DeMarzo (2005).

endowed with total resources of W at time 1. For simplicity, I treat this endowment as exogenous in what follows, but one can also endogenize it by allowing the PIs to raise the W from the household sector at time 0 by issuing risky long-term claims. In this case, the PIs choose an optimal level of W at time 0 that equates the expected return on their time-1 investments to the cost of capital R^B . However, imposing this ex-ante breakeven condition does not affect the qualitative results of the model, so I leave it aside.¹⁶

In either formulation, what is crucial is that when time 1 rolls around and the state of the world is realized, W is fixed. Thus while it is fine to think of PIs as having full access to financial markets at time 0, they cannot go back and raise more at time 1 once they know the state of the world. In other words, W is an unconditional war chest, with the same amount available to PIs in the good and bad states.

PIs can do one of two things with their resources at time 1. First, they can invest in new, late-arriving real investment projects. Irrespective of the state of the world, an investment of K in such new projects at time 1 yields expected gross output of $g(K)$ at time 2, where $g(\cdot)$ is a concave function. Alternatively, PIs can absorb assets being sold by banks at time 1.

In the good state, there are no asset sales, so the PIs invest all of W in new projects, yielding $g(W)$.¹⁷ In the bad state, banks have to sell enough assets to repay short-term creditors the M they have promised them. Thus in equilibrium, PIs spend M on asset purchases, and invest only $(W - M)$ in new projects, yielding $g(W - M)$. For the PIs to be willing to allocate their endowment in this way, it must be that the marginal

¹⁶ One interesting feature that comes out of endogenizing W is that the optimal value of W is lower when the ex ante probability p of the good state is higher—since PIs earn a higher return when there are fire sales. This implies that the rarer is the crisis state, the more severe its consequences, all else equal.

¹⁷ I assume that $g'(W) > 1$ so that investing all of W in this way in the good state is optimal.

return on new projects is the same as the marginal return from buying existing assets from banks. This is what pins down the fire-sale discount k . In particular, we have that:

$$\frac{1}{k} = g'(W - M) \quad (4)$$

Equation (4) makes clear the real costs of fire sales, and hence of short-term debt financing by banks. The greater is M , and hence the more bank assets that the PIs have to absorb in the bad state at time 1, the less they have left over for investment in new projects. With scarce PI capital, the return on secondary-market arbitrage opportunities (buying up fire-sold assets) also becomes the hurdle rate for new investment, a point emphasized by Diamond and Rajan (2009a) and Shleifer and Vishny (2010).

D. The Bank's Optimization Problem

Let us now formulate the optimization problem for a bank that invests an amount I and finances it with some fraction $m \leq m^{\max}$ of money. The bank's expected net profits are given by:

$$\begin{aligned} \Pi = \{pf(I) + (1-p)\lambda I - IR^B\} + mI(R^B - R^M) - (1-p)zmIR^M = \\ \{pf(I) + (1-p)\lambda I - IR^B\} + \frac{M}{R^M}(R^B - R^M) - (1-p)zM \end{aligned} \quad (5)$$

where I have defined $z = \frac{(1-k)}{k}$ as the net rate of return on fire-sold assets. (Note that higher values of z correspond to larger fire-sale discounts, and $z = 0$ is the case where there is no discount.)

The three terms in (5) are easily interpreted. The first, $\{pf(I) + (1-p)\lambda I - IR^B\}$, is the net present value of investment assuming that investment is entirely financed at the higher bond-market rate—and hence that there is no need to ever sell assets. The second term, $mI(R^B - R^M)$, is the financing cost savings associated with using a fraction m of

money in the capital structure. And the last term, $(1-p)zmIR^M$, captures the expected fire sale losses associated with this riskier short-term capital structure.

Each bank picks privately-optimal values of m and I , or equivalently, of m and M . In what follows, I frame things the latter way, and compute the first-order conditions with respect to m and M , because this renders the analysis a little more transparent. In doing so, I assume that each bank is sufficiently small that they treat the fire-sale discount k as a fixed constant—i.e., they do not internalize the incremental impact of their capital-structure choices on the economy-wide fire-sale outcome. By contrast, when I examine the social planner's problem below, the key difference will be that the social planner takes into account the dependence of k on the capital structure of the banks.

Differentiating the objective function in (5) with respect to m , we have:

$$\frac{d\Pi}{dm} = I\{(R^B - R^M) - (1-p)zR^M\} \quad (6)$$

It follows that the bank is at a corner, setting $m = m^{\max}$, if $(R^B - R^M) > (1-p)zR^M$, i.e., if the equilibrium spread between bonds and money is sufficiently large. Alternatively, if the spread is smaller in equilibrium (that is, if $(R^B - R^M) = (1-p)zR^M$) then the bank chooses an interior value of m and we have that $\frac{d\Pi}{dm} = 0$.

Differentiating the objective function in (5) with respect to M yields:

$$\frac{d\Pi}{dM} = \{pf'(I) + (1-p)\lambda - R^B\} \left[\frac{dI}{dM} \right]_{Bank} + \frac{(R^B - R^M)}{R^M} - (1-p)z \quad (7)$$

Optimality for the bank requires that $\frac{d\Pi}{dM} = 0$. There are two ways that this can happen. First, the bank can be at an interior solution with respect to m , in which case

$(R^B - R^M) = (1-p)zR^M$, and it follows from setting (7) equal to zero that $pf'(I) + (1-p)\lambda = R^B$. Alternatively, the bank can be at a corner with $m = m^{\max}$, and $(R^B - R^M) > (1-p)zR^M$, in which case it follows from setting (7) equal to zero that $pf'(I) + (1-p)\lambda < R^B$. This reasoning leads to the following proposition.

Proposition 1: Define I^B as the optimal level of investment for a bank that finances itself exclusively in the long-term bond market: $pf'(I^B) + (1-p)\lambda - R^B = 0$. The solution to the bank's optimization problem involves two regions. In the low- M region (for $(R^B - R^M)$ relatively small) the bank chooses $m < m^{\max}$ and $I^* = I^B$. In the high- M region (for $(R^B - R^M)$ relatively large) the bank chooses $m = m^{\max}$ and $I^* > I^B$.

The important point to take away from the proposition is that in the low- M region, a bank's investment and financing choices are decoupled, while in the high- M region they are interdependent. This is because when $m < m^{\max}$, a bank's ability to tap low-cost money financing is not constrained by the amount of investment it does. By contrast, in the high- M region in which $m = m^{\max}$, a bank faces a binding collateral constraint—it can only issue more money if it increases the quantity of physical assets backing its debts. This is what ties investment and financing decisions together. If money financing is cheap enough that banks want to do a lot of it, and they begin to bump up against the collateral constraint, they will be induced to invest more so as to loosen the constraint.

III. Socially Excessive Money Creation: A Role for Regulation

The next step in the analysis is to identify the circumstances in which the process of private money creation described above involves an externality—i.e., when the level of money creation chosen by banks exceeds that preferred by a benevolent social planner.

A. The Social Planner's Problem

Given that all output of the banks and the PIs ultimately accrues to the household sector, the social planner seeks to maximize the utility of a representative household, as given by equation (1). If we think of the time-0 and time-1 endowments as exogenously fixed, it is easily shown that, disregarding constants, this utility is equivalent to:¹⁸

$$U = \{pf(I) + (1-p)\lambda I - IR^B\} + \frac{M}{R^M}(R^B - R^M) + E\{g(K) - K\} \quad (8)$$

Comparing this to the bank's expected profits in (5), we can see that the first two terms coincide. The difference is in the third term: the planner does not care about expected fire sale losses per se, because these only represent a transfer from the banks to the PIs. However, the planner does care about the net returns to investment by the PIs, as captured $E\{g(K) - K\}$.

The social planner's first-order condition with respect to M is therefore given by:

$$\frac{dU}{dM} = \{pf'(I) + (1-p)\lambda - R^B\} \left[\frac{dI}{dM} \right]_{Planner} + \frac{(R^B - R^M)}{R^M} + E\{(g'(K) - 1) \frac{dK}{dM}\} = 0. \quad (9)$$

¹⁸ In particular, households have some fixed time-0 and time-1 endowments of X_1 , and X_2 , respectively, and they have invested a non-contingent amount W of their time-1 endowment with the PIs.

From equation (4) above, we know that in the bad state, which occurs with probability $(1 - p)$, we have $g'(K) = \frac{1}{k}$, and $\frac{dK}{dM} = -1$. In the good state, by contrast, $\frac{dK}{dM} = 0$. This observation implies that we can re-write (9) as:

$$\frac{dU}{dM} = \{pf'(I) + (1-p)\lambda - R^B\} \left[\frac{dI}{dM} \right]_{Planner} + \frac{(R^B - R^M)}{R^M} - (1-p)z = 0. \quad (10)$$

Comparing this expression to its counterpart (7) in the bank-optimization case, note that the third terms are now identical—they are given by $(1 - p)z$ in both cases. This captures the fact that although the social planner does not care about fire-sale discounts directly (again, these are just a transfer), when it comes to picking an optimal value of M , the social planner nevertheless acts as if he would like to reduce these discounts. This is because greater fire-sale discounts are associated with reduced real investment by the PIs, and the planner does care about this real investment. Thus when we compare (7) and (10) we see one place where social and private incentives are perfectly aligned: when choosing an optimal capital structure, one consideration for banks is that they would like to avoid fire-sale discounts, all else equal. This leads them to internalize the planner's desire to maintain real investment by the PIs.

So where is the externality? The only divergence between social and private incentives shows up in the first term of equations (7) and (10). In particular, this

divergence arises out of the fact that, in the high- M region, $\left[\frac{dI}{dM} \right]_{Planner} > \left[\frac{dI}{dM} \right]_{Bank}$, or

alternatively, that $\left[\frac{dM}{dI} \right]_{Planner} < \left[\frac{dM}{dI} \right]_{Bank}$. To see this, note that in the high- M region,

we have $M = k\lambda I$. Thus from the perspective of an individual bank that takes the fire-

sale discount k as exogenous, $\frac{dM}{dI} = k\lambda$. However from the perspective of a social planner who recognizes the dependence of k on the total amount of money issued by all banks, $\frac{dM}{dI} = \frac{k\lambda}{(1 - I\lambda \frac{dk}{dM})}$. And as can be seen from (4), $\frac{dk}{dM} < 0$: the more money is created in the aggregate, the lower are the prices of fire-sold assets.

The externality can be understood as follows. When a given bank raises its own investment and money creation, it takes into account the fact that, in the bad state this will force it to sell more assets at a discount in order to pay off its own short-term claimants—this is the third term in (7) and (10), given by $(1 - p)z$. However, what the bank fails to internalize is that its greater level of money creation, by reducing the equilibrium value of k , *effectively lowers the collateral value of all other bank's assets*. As aggregate M goes up, other banks can no longer individually generate as much M for a given level of I , since greater fire-sale discounts imply that the resale value of their assets is reduced.¹⁹ This is why $\left[\frac{dM}{dI} \right]_{Planner} < \left[\frac{dM}{dI} \right]_{Bank}$. Importantly, this externality only comes into play in the high- M region, when banks' money creation is at a corner ($m = m^{\max}$) and constrained by the availability of collateral. If, by contrast, banks are only creating an interior amount of money, a marginal reduction in the value of their collateral is of no consequence.

The following proposition summarizes the analysis.

¹⁹ This effect is reminiscent of Shleifer and Vishny's (1992) concept of industry debt capacity.

Proposition 2: Denote the private and socially optimal values of M by M^* and M^{**} respectively. In the low- M region, $M^* = M^{**}$. In the high- M region, $M^* > M^{**}$.

Thus banks may create a socially excessive amount of money, but this happens only if the spread between money and bonds ($R^B - R^M$) is high enough. If the spread is so low that any individual bank choose an interior value of money creation $m < m^{\max}$, there is no divergence between private and social incentives.

Example 1: Pick these functional forms and parameter values: $f(I) = \psi \log(I) + I$, $g(K) = \theta \log(K)$, $R^B = 1.04$; $R^M = 1.02$; $\psi = 3.5$; $\theta = 150$; $\lambda = 1$; $W = 200$; and $p = 0.95$. For these values, the private optimum is in the high- M region, and involves banks choosing $M^* = 79.1$ and $I^* = 98.1$, with an associated rate of return on fire-sale assets of $z = 24.1\%$ ($k = 0.806$). By contrast, in the social optimum, the planner chooses $M^{**} = 76.8$ and $I^{**} = 93.5$, leading to a rate of return on fire-sale assets of $z = 21.8\%$ ($k = 0.821$).

Figure 1 expands on Example 1, keeping all of the other parameter values the same as above, but allowing R^M to vary between 1.00 and 1.04, thereby causing the bond-money spread ($R^B - R^M$) to vary between zero and four percent. As can be seen, for low values of the spread, the private and socially optimal values of M and I coincide. But as the spread widens, these values diverge further and further from one another.

The result that there is no externality in the low- M region when $m < m^{\max}$ is dependent on the assumption that, when the PIs invest in real projects, they can capture all the social surplus associated with these projects. If instead one assumes that this

surplus is only partially pledgeable (due to financial contracting frictions, or positive spillovers associated with investment), then private money creation is always excessive from a social perspective, irrespective of parameter values. In particular, suppose that the social return to an investment project financed by a PI is still given by $g(K)$, but that only $\phi g(K)$ can be pledged to the PI. In this case, the equilibrium determination of k in (4) is altered so that $\frac{1}{k} = \phi g'(W - M)$. In other words, a given amount of underinvestment by the PIs is now associated with a smaller fire sale discount. Hence a bank's aversion to fire sales no longer leads it to fully internalize the social costs of underinvestment in real projects. This variant of the model is briefly explored in the appendix.

B. A “Cap-and-Trade” Approach to Bank Liquidity Regulation

The analysis thus far makes clear that in some cases banks will choose to create more money than is socially optimal, thereby inflicting inefficiently high levels of fire sales on the economy. This suggests a role for regulation. In the full-information case, in which the regulator observes all the relevant parameters of the model, the social optimum can be easily implemented with a cap on money creation: each bank can simply be prohibited from issuing more short-term claims than the desired level of M^{**} .

However, if the regulator is imperfectly informed, it becomes more challenging to set the cap appropriately. Consider a situation in which banks know the productivity of their investment opportunities—i.e., they know what the function $f(I)$ looks like—but the regulator does not. As can be seen from equation (10), the value of M^{**} depends on, among other things, the marginal product of investment $f'(I)$. Intuitively, it makes sense to allow banks to create more cheap money financing when they have better

investment opportunities. Thus without knowledge of the value of $f'(I)$, it is impossible for the regulator to target the socially optimal level of money creation with a simple cap.

One way for the regulator to generate the required information is through a system of cap-and-trade. In particular, each bank can be granted permits that allow it to issue some amount of money; by picking the aggregate quantity of permits, the regulator can, as before, effectively target the total amount of money M in the economy. Moreover, if the permits can be traded among banks, their price (per unit of money creation allowed) will equal the shadow value of the constraint to the banks, as given by

$\frac{d\Pi}{dM}$.²⁰ And as can be seen from equation (7), conditional on the regulator knowing the other parameters of the model, observing $\frac{d\Pi}{dM}$ allows him to infer the value of $f'(I)$.

It follows from this reasoning that the regulator can implement the M^{**} solution by making the permits tradable, and by targeting the appropriate price for these permits—namely the price at which $\frac{dU}{dM} = 0$. With a little bit of algebra based on equations (7) and (10), this can be shown to imply:

Proposition 3: A regulator who is imperfectly informed about the nature of bank lending opportunities can implement the desired level of money M^{**} with a system of

²⁰ Note that since the banks in the model are all identical, the volume of trade in the permits is zero. Nevertheless, there is a unique equilibrium price, given by the common shadow value of the constraint.

tradable permits for money creation. This involves adjusting the number of permits such that their market clearing price P is given by:²¹

$$\begin{aligned}
 P &= \left\{ \frac{(R^B - R^M)}{R^M} - (1-p)z \right\} \left\{ 1 - \frac{\left[\frac{dI}{dM} \right]_{Bank}}{\left[\frac{dI}{dM} \right]_{Planner}} \right\} \\
 &= \left\{ \frac{(R^B - R^M)}{R^M} - (1-p)z \right\} \left\{ \frac{-\lambda Idk / dM}{(1 - \lambda Idk / dM)} \right\}.
 \end{aligned} \tag{11}$$

Example 2: Keep everything the same as in Example 1: $f(I) = \psi \log(I) + I$, $g(K) = \theta \log(K)$, $R^B = 1.04$; $R^M = 1.02$; $\psi = 3.5$; $\theta = 150$; $\lambda = 1$; $W = 200$; and $p = 0.95$. At the social optimum of $M^{**} = 76.8$, the price of permits is $P = 0.00336$. Now suppose there is a positive productivity shock, and ψ rises to 4.0. If the cap is not adjusted, the price of permits spikes to $P = 0.00873$. However, this price increase reveals the new value of ψ to the regulator, who can increase the number of permits in the system, raising the quantity of money in the system to its new optimal value of $M^{**} = 81.3$. At this new optimum, the price of permits is given by $P = 0.00262$.

The example suggests that, in the face of productivity shocks, it is optimal for the regulator to actively lean against incipient changes in the price of permits. When a positive shock pushes the price of permits up, the regulator should increase the supply of permits, thereby driving their price back down. In fact, optimality in this setting requires

²¹ In adjusting the quantity of permits, the regulator is looking for a fixed point. Suppose he picks an initial value of M . If this is in fact the social optimum, the market price of permits will be given by (11), which the regulator can calculate based on his knowledge of M and the other observable parameters of the model. If the observed market price in fact turns out to be higher than the target value given by (11), the regulator increases M , and vice-versa. The social optimum is achieved at that value of M where the target price in (11) equals the observed market price.

the supply response to be sufficiently strong that the equilibrium price of permits actually falls as productivity rises.

IV. Implementing the Cap-and-Trade Approach with Monetary Policy

The cap-and-trade approach to bank regulation outlined above may seem alien—it does not have any direct counterpart in the real world. However, I now argue that the cap-and-trade approach can be implemented with something that looks very much like conventional monetary policy—with open-market operations in which the central bank adjusts the quantity of nominal reserves in the banking system. In this setting, reserves play the role of permits for money creation, given the existence of a binding reserve requirement. And the nominal interest rate corresponds to the price of the permits.

In drawing this analogy, one important caveat is that I have so far been working in an entirely real economy. In order to introduce a central bank and a role for monetary policy, I need to bring in a set of nominally-denominated government liabilities, and then pin down the price level. To do so, I rely on the fiscal theory of the price level (Leeper (1991), Sims (1994), Woodford (1995), Cochrane (1998)). In particular, the government is assumed to issue two types of nominal liabilities: Treasury bills, and bank reserves. According to the fiscal theory, the *sum* of these two nominal liabilities is what is relevant for determining the price level. And given the sum, the *composition* of these liabilities is effectively a real variable, since only reserves—and not Treasury bills—can be used to satisfy reserve requirements. Thus holding fixed total government liabilities, when there are more reserves in the system, banks are able to create more money in real terms, i.e. to

finance a greater fraction of their operations with short-term debt. Hence reserves correspond exactly to the concept of regulatory permits in the purely real model.²²

To operationalize the fiscal theory, I assume that the government anticipates real tax revenues of T at time 2, and that the value of T is exogenously fixed. At time 0, the government has total nominal liabilities outstanding of l_0 , composed of Treasury bills b_0 , and bank reserves r_0 . Thus $l_0 = b_0 + r_0$. The time-0 price level P_0 , is then determined by the requirement that the real value of the government's obligations must equal the present value of its future tax revenues:

$$\frac{l_0}{P_0} = \frac{T}{R^M} \quad (12)$$

Two points are worth noting here. First, the relevant real discount rate for the government is R^M , given that its obligations are riskless in this setting. In other words, when households own Treasury bills, they derive the same monetary services from these bills that they do from privately-created bank money, so the return on Treasury bills is equal to R^M . Second, in order to keep real tax revenues fixed at T as the composition of government liabilities varies, I need to assume that the government rebates any seignorage revenues derived from the issuance of non-interest bearing reserves in a lump-sum fashion to the household sector.²³

²² Since the price level is pinned down by fiscal considerations in this model, the goal of achieving price stability cannot be the central bank's job. Rather, the central bank is effectively left in the role of financial-stability regulator.

²³ Without this assumption, the composition of government liabilities would influence real tax revenues. In particular, as the government issued more non-interest-bearing reserves and fewer interest-bearing bills, its effective tax revenues would go up through a seignorage mechanism. The assumption can be loosely motivated by the idea that the government has some kind of social compact with its citizens that prevent it from letting total tax revenues—no matter how they are raised—go above T .

Again, the key distinction between Treasury bills and bank reserves is that only the latter can be used to satisfy reserve requirements. In particular, any bank wishing to issue a dollar of short-term debt must hold ρ dollars of reserves, where ρ is the fractional reserve requirement. Hence the *net* amount of short-term debt financing made possible by one dollar of reserves is $(1 - \rho)/\rho$ dollars.²⁴ It follows that in real terms, the total amount of M that can be created by the banking sector is now given by:

$$M = \frac{(1-\rho)r_0}{\rho P_0} = \frac{(1-\rho)T}{\rho R^M} \frac{r_0}{l_0} \quad (13)$$

This expression makes it clear that the ratio of r_0 to l_0 —namely, the *composition* of the government’s nominal liabilities—is a real variable, and in particular is the means by which the government can target total real money creation by banks. An open-market operation that increases the supply of reserves and reduces the supply of Treasury bills is isomorphic to an increase in the regulatory limit on M in the all-real cap-and-trade version of the model.

Moreover, as noted above, the analog to the price of permits is the current setting is the nominal interest rate. This is because when banks want to create money, they are forced to hold non-interest bearing reserves, and the nominal interest rate represents the opportunity cost of doing so.

Denoting the nominal interest rate by i , one can express the time-2 price level as:

$$P_2 = \frac{P_0(1+i)}{R^M}. \quad (14)$$

²⁴ As an example, suppose $\rho = .10$. In this case, with one dollar of reserves, a bank is allowed to raise 10 dollars of short-term debt. But given that it must hold the reserves as an asset, only 9 of these dollars represent *net* financing that is available to fund new loans.

Now suppose a bank wishes to increase its net issuance of real M by one unit at time 0, which will increase its real time-2 profits by $\frac{d\Pi}{dM}$. To do so, the bank must increase net nominal M by P_0 units at time 0, which requires it to hold $\rho P_0/(1 - \rho)$ of nominal reserves. To finance these reserve holdings, it must pay $\rho i P_0/(1 - \rho)$ of nominal financing costs at time 2. The real time-2 value of these financing costs is therefore $\rho i P_0/(1 - \rho)P_2$ or, using equation (13), $\rho i R^M/(1 - \rho)(1 + i)$. For a bank to be at an indifference point, it must be that these real costs are equal to $\frac{d\Pi}{dM}$. Thus it follows that the nominal interest rate is given by:

$$\frac{i}{(1+i)} = \frac{(1-\rho)}{\rho R^M} \frac{d\Pi}{dM}. \quad (15)$$

Example 3: Keep everything the same as in Example 1: $f(I) = \psi \log(I) + I$, $g(K) = \theta \log(K)$, $R^B = 1.04$; $R^M = 1.02$; $\psi = 3.5$; $\theta = 150$; $\lambda = 1$; $W = 200$; and $p = 0.95$. We saw previously that at the social optimum of $M^{**} = 76.8$, $\frac{d\Pi}{dM} = 0.00336$. With a fractional reserve requirement of $\rho = 0.10$, if the social optimum is implemented with monetary policy, the resulting nominal interest rate is given by $i = 3.05\%$. (Since the nominal rate exceeds the riskless real rate of 2.0%, the implied rate of inflation between time 0 and time 2 is 1.05%.) If we keep all else the same but set $R^M = 1.01$, the new social optimum involves $M^{**} = 81.4$, which is implemented with a nominal interest rate of $i = 6.35\%$. Intuitively, as the spread between money and bonds increases, banks have a stronger desire to create private money. So the nominal interest rate, which is equivalent to the value of a permit for money creation, must rise as well.

V. Other Policy Tools

A. Deposit Insurance and Lender of Last Resort

In the version of the model above, the only way for banks to pay off their short-term creditors in the crisis state is by fire-selling their assets, and the only role for policy is to control the amount of short-term debt that is created *ex ante*. An alternative approach would be for the government to try to stem the amount of socially costly fire sales that occur for a *given* amount of short-term bank debt. This could be done either with either deposit insurance, or a lender-of-last resort policy.

Unlike in the classic framework of Diamond and Dybvig (1983), such policies are not costless to the government in equilibrium, because here, in the crisis state, there is a probability $(1 - q)$ that the banks' assets will turn out to be entirely worthless. So there is always a chance that taxpayers will be left on the hook. If taxpayer-financed bailouts create deadweight losses, the overall optimum set of policies may have the realistic feature that: i) some fraction of banks' money-like claims are insured by the government; ii) the remainder are uninsured, and hence still subject to fire-sale risk; and iii) as before, it makes sense for the regulator to control the total quantity of bank-created money.

To see this explicitly, consider a case where the deadweight costs of taxation take the following form: there is no cost to raising any amount less than L to pay for a bailout, but it is infinitely costly to raise anything more than L . It follows that the amount of government-insured money that can be created, M^l , is bounded by $M^l \leq L$, and it will in fact always be optimal to set $M^l = L$. Note too that if the government offers insurance on some amount of bank deposits, it will have to put in place a rule to prevent banks from selling *all* of their assets in a crisis state to satisfy the demands of uninsured depositors;

otherwise banks will create just as much uninsured money as before, and the deposit insurer will always be left holding an empty shell in the crisis state. A simple version of such a rule—which can effectively be thought of as a ban on fraudulent conveyance—is a requirement that the fraction of assets sold in a crisis, Δ , not exceed the relative proportion of uninsured deposits. Thus the requirement that goes along with insurance is

that $\Delta \leq \frac{M^U}{M^U + M^I}$, where M^U is the quantity of uninsured money created by the bank.

It follows that the *total* amount of money—insured plus uninsured—that can be created must satisfy the same collateral constraint as before: $M = M^U + M^I \leq k\lambda I$. The only thing that is changed is the determination of the fire-sale discount k . Since insured depositors are protected and do not need to demand repayment at time 1, only uninsured deposits give rise to fire sales. Thus k is now given by:

$$\frac{1}{k} = g'(W - M^U) = g'(W - M + L) \quad (16)$$

In other words, the outcome in a world with limited deposit insurance is equivalent to that in a world with no deposit insurance, but where the wealth of the PIs is augmented from W to $(W + L)$. A given amount of total money creation now causes less fire-sale damage, and as a result, more money can be created in equilibrium.

Equation (16) also makes clear the close connection between deposit insurance and a lender-of-last-resort function. Given that the government can never put itself in a position to lose more than L , an alternative to deposit insurance would be for it to leave all deposits uninsured, but to commit to step in and invest L alongside the PIs in the event of a fire sale. This would have exactly the same effect—it would reduce the fire-sale discount per equation (16), and thereby allow for more total money creation.

The bottom line is that one can add deposit insurance to the model in such a way as to make it more realistic, without changing any of its qualitative properties. The optimal policy mix will involve limited use of deposit insurance or equivalently, limited use of a lender-of-last-resort function. Banks will continue to issue uninsured money-like claims alongside insured deposits, and hence will continue to create some degree of fire-sale risk. Thus as before, there will continue to be a motive for regulating the creation of these uninsured short-term claims.

B. Regulating the Shadow Banking System

The model also assumes that all private money is manufactured by commercial banks that are subject to reserve requirements. Hence private money creation can be directly controlled by open-market operations. While this may be an adequate representation of an earlier period in history, it omits an important form of money creation in the modern economy. As Gorton and Metrick (2009) and Gorton (2010a,b) argue, private money—in precisely the sense meant here—is also created by the unregulated shadow banking system, via the large volume of repo finance that is collateralized by securitized loan pools of one form or another. Gorton (2010a) writes:

“Banking *means* creating short-term trading or transactions securities backed by longer term assets. Checking accounts (demand deposits) are the leading example of such securities....Before the crisis trillions of dollars were traded in the repo market....Repo and checks are both forms of money....There have always been difficulties creating private money (like demand deposits) and this time around was no different.”

This line of argument suggests that, to more accurately capture the modern financial system in the framework of the model, it makes sense to think of there being two distinct sets of bank-like intermediaries: a group of “commercial banks”, who are as described above, and who are subject to reserve requirements on the money they create,

and a group of “shadow banks” who are otherwise identical to the commercial banks but whose liabilities are not subject to regulatory control of any sort, and who therefore are free to choose the privately-optimal level of money creation.

It is easy to see that in this setting, monetary policy still works, in the sense that there continues to be an operative bank lending channel: contractions in reserves still affect nominal interest rates, as well as the lending behavior of the commercial banks. And monetary policy retains some ability to limit the overall externality associated with excessive money creation, since it can rein in the amount of money created by the commercial banks. But this regulatory ability is now impaired, because there is no control over the quantity of money created by the shadow banks.

Clearly, it would be better for commercial banks and shadow banks to be regulated in a more symmetric fashion. If, for some reason, the liabilities of shadow banks cannot be subjected to reserve requirements, an alternative approach might be to impose a regime of “haircut” requirements on their investments. In particular, the central bank could specify the maximum fraction of private money—that is, repo financing—that could be issued against a given amount of collateralizable assets. Moreover, just as the optimal quantity of bank-created money M^{**} varies with economic conditions, optimal haircuts would respond to these conditions as well. The appendix provides a brief analysis of haircut regulation. It turns out that while such regulation is indeed useful, it is less efficient than direct control of the quantity of privately-created money.

C. Government Debt Maturity

As we have seen, the magnitude of the externality associated with private money creation is related to the bond-money spread ($R^B - R^M$): when the spread widens, the

wedge between the social and private returns to money creation goes up. Thus an alternative way to moderate the externality would be to compress the spread. In the current version of the model this is impossible—given the assumption of linear preferences, the spread is exogenously fixed and insensitive to asset supplies.

However, if one changes the model so that the monetary services enjoyed by households are a concave function of the total amount of money available—i.e., so that there is diminishing marginal utility to holding money—then it becomes possible for the government to act on the bond-money spread. For example, since short-term Treasury bills are riskless, they can provide the same monetary services as short-term bank debt. Hence an increase in the supply of Treasury bills will, in this modified setting, reduce the bond-money spread.

One appeal of dealing with the externality in this fashion is that unlike some other regulatory approaches, it does not invite evasion. For example, a possible drawback with the sort of haircut regulation described above is that private actors may try to get around limits on their ability to use short-term debt by using various forms of hidden borrowing, e.g., by embedding the borrowing in an opaque derivative contract. In contrast, when the relative cost of short-term borrowing goes up—because the market has been saturated with riskless short-term claims—the incentive to create private money is blunted.

In Greenwood, Hanson and Stein (2010), we use this observation as the point of departure for a normative theory of government debt maturity. We argue that the government should choose a shorter debt maturity—and in particular, should issue more riskless T-bills—than it otherwise might, in an active effort to crowd out the short-term debt of financial intermediaries. The argument is based on a principle of comparative

advantage. On the one hand, tilting its issuance towards short-term debt is not without cost for the government, since with stochastic interest rates this increases the variability of future interest payments and ultimately disrupts efforts to smooth tax rates over time. On the other hand, short-term government debt, unlike the short-term debt of financial intermediaries, does not create fire-sale risk. To the extent that the fire-sale externality is more costly to the economy at the margin than the disruption of tax smoothing, it can make sense for the government to take on a bigger role in providing the short-term riskless claims that the economy demands.

Of course, precisely because of tax-smoothing considerations, it will not generally be optimal for the government to tilt so strongly towards short-maturity issuance as to entirely eliminate the bond-money spread in equilibrium. Rather, optimal behavior by the government on this dimension will typically involve leaving the spread only partially compressed. So while government debt maturity may be one helpful tool in addressing the problem of excessive private money creation, it is not a panacea, and it is unlikely to eliminate the usefulness of the other tools discussed above.

VI. A Distinctive Account of the Monetary Transmission Mechanism

Much of the discussion above has focused on the normative implications of the model. But the model is also of interest as a positive account of the monetary transmission mechanism. Three of its properties are particularly noteworthy in this regard. First, monetary policy has real effects even though all prices are perfectly flexible. Second, monetary policy works entirely through a quantitative effect on bank lending. That is, the real rates on both money and bonds are fixed and independent of the

stance of policy; an easing of policy impacts bank lending only because it enables banks to use more of the former, relatively cheaper, funding source. This is a pure version of the bank lending channel, and as such helps to explain how monetary policy can have important real effects even when it does not move long-term open-market interest rates by much, or when firm investment does not appear to be very responsive to such open-market rates.

Third, the model has the property that the central bank does not lose control of the monetary transmission mechanism when other, non-reservable forms of money are introduced. Consider what happens if there is, in addition to the risky production technology already in the model, a completely safe storage technology. Claims to this technology are riskless, and hence circulate as an alternative transactions medium alongside bank-created money, bearing the same gross interest rate of R^M . They are also not subject to reserve requirements. (To be a bit more concrete, one can interpret these claims as money-market-fund deposits backed by, say, Treasury bills.) Even if the volume of these claims is large, nothing in the model changes. All real rates are already pinned down by the linearity of household preferences, and are therefore unaffected by the total quantity of money in circulation.

By contrast, the standard textbook account of the monetary transmission mechanism depends crucially on the hard-to-swallow premise that the central bank has monopoly control over the transactions medium used by households. For if it cannot control the aggregate quantity of money, it has no lever over real interest rates, which is the key mechanism in the textbook model. Thus for the textbook story, the existence of

non-reservable money market funds is a big problem, a point emphasized by, e.g., Cochrane (1998).

In the model of this paper, the central bank's ability to influence real outcomes derives not from its control over the total quantity of transactions-facilitating claims available to households, but rather from the fact that it is the unique provider of permits that allow banks to issue short-term debt and hence finance themselves more cheaply. Simply put, only central-bank-provided reserves can be used to satisfy the reserve requirements that constrain short-term debt issuance by banks. This "permits" aspect of monetary policy is also emphasized in Stein (1998), though the model in that paper differs significantly on other dimensions.

VII. Conclusions

The basic message of this paper can be summarized as follows. Banks and other financial intermediaries like to fund themselves with short-term debt. With sufficient collateral backing it, this short-term debt can be made into riskless money, which, because of the transactions services it generates, represents a cheap source of finance for banks. While society benefits from this private money creation, banks' private incentives lead them to overdo it, since they do not fully internalize the fire-sales costs that necessarily come with their maturity-transformation activities. The externality associated with excessive private money creation provides the fundamental rationale for financial-stability regulation, and arguably, for the existence of central banks.

In a sufficiently simple institutional environment, the externality can be addressed with conventional monetary policy, complemented by either deposit insurance or a

lender-of-last-resort facility. Indeed, this is one interpretation of what central banks have done for much of their history. In a more realistic modern-day setting, where a substantial shadow-banking sector exists alongside traditional commercial banks, other tools, such as haircut regulation, may also be necessary. If so, central banks should not be reluctant to deploy these additional tools—to the extent that they do so in an effort to contain excessive private money creation, they can be said to be pursuing one of their traditional core missions in a more comprehensive and effective manner.

Appendix

A. A Variant of the Model With Imperfect Pledgeability

As noted in the text, the result that there is no externality in the low- M region when $m < m^{\max}$ is dependent on the assumption that, when the PIs invest in real projects, they capture all the social surplus associated with these projects. An alternative approach is to assume that the social return to a project financed by a PI is still given by $g(K)$, but that only $\varphi g(K)$ can be pledged to the PI. In this case, the equilibrium determination of k in (4) is altered so that $\frac{1}{k} = \varphi g'(W - M)$.

Equations (7) and (10), for the bank's and the social planner's first-order conditions with respect to M , still hold as stated. For convenience, they are repeated here as (A.1) and (A.2) respectively.

$$\frac{d\Pi}{dM} = \{pf'(I) + (1-p)\lambda - R^B\} \left[\frac{dI}{dM} \right]_{Bank} + \frac{(R^B - R^M)}{R^M} - (1-p)z = 0 \quad (\text{A.1})$$

$$\frac{dU}{dM} = \{pf'(I) + (1-p)\lambda - R^B\} \left[\frac{dI}{dM} \right]_{Planner} + \frac{(R^B - R^M)}{R^M} + E\{(g'(K) - 1) \frac{dK}{dM}\} = 0. \quad (\text{A.2})$$

However, one can no longer simplify (A.2) by replacing $E\{(g'(K) - 1) \frac{dK}{dM}\}$ with $-(1-p)z$. Instead, (A.2) can be rewritten as:

$$\begin{aligned} \frac{dU}{dM} &= \{pf'(I) + (1-p)\lambda - R^B\} \left[\frac{dI}{dM} \right]_{Planner} + \frac{(R^B - R^M)}{R^M} - (1-p)z \\ &- (1-p) \frac{(1-\varphi)}{\varphi} g'(W - M) = 0 \end{aligned} \quad (\text{A.3})$$

It follows that even when $I = I^B$, so that the first term in both (A.1) and (A.3) is zero, there is still a wedge of $(1-p) \frac{(1-\varphi)}{\varphi} g'(W - M)$ between $\frac{d\Pi}{dM}$ and $\frac{dU}{dM}$, and

hence a role for regulation. Thus even in the low-M region where $m < m^{\max}$ and $I = I^B$, the optimal price of permits P will now be strictly positive, and given by $(1-p)\frac{(1-\varphi)}{\varphi}g'(W-M^{**})$. Alternatively, in the monetary-policy implementation of the optimum, the nominal interest rate will now be strictly positive for all parameter values.

Figure 2 illustrates, plotting private and socially-optimal values of M and I as R^M varies. Everything is identical to Figure 1, except that now $\varphi = .90$. As can be seen, the privately-optimal value of M is now always greater than the socially-optimal value, even in the region where both banks and the social planner choose $I = I^B$.

Another noteworthy feature of this version of the model is that it implies different comparative statics than the baseline model with respect to the ex ante probability of a financial crisis, as captured by $(1-p)$. Here, if we are in the low-M region, an increase in $(1-p)$ raises the desired value of the permit price P , or equivalently, the nominal interest rate. By contrast, in the baseline model with perfect pledgeability, equation (11) says that an increase in $(1-p)$ lowers the desired permit price. Intuitively, the difference is that in the baseline version of the model, banks do a better job of internalizing the social costs of fire sales. Indeed when the risk of a fire sale goes up, banks become sufficiently more cautious about using short-term debt that they become better aligned with the social planner, which in turn implies that there is less need to rein them in by raising permit prices/interest rates. However, with imperfect pledgeability, there is an effect in the opposite direction, since banks tend to underweight the social costs of fire sales even when the collateral constraint is not binding.

B. Haircut Regulation

To see the case for haircut regulation most transparently, consider the imperfect-pledgeability version of the model described just above. Suppose that we are in a “shadow-banking” economy where all else is the same as before, with one exception: it is impossible to regulate the absolute quantity of privately-created money M directly—say because shadow banks cannot be subjected to reserve requirements—but it is possible to impose a cap $m^{cap} < m^{\max}$ on the *fraction* of investment that is money-financed. Finally, assume just for the moment that shadow-bank investment is not a choice variable, but is exogenously fixed at I^f .

Given imperfect pledgeability, it will always be desirable for the social planner to control the level of M , even when $m < m^{\max}$, for the reasons described just above. And note that with fixed investment, it is trivial to implement any desired money target M^{**} with haircut regulation, simply by setting $m^{cap} = \frac{M^{**}}{I^f R^M}$. In other words, with fixed investment, haircut regulation is exactly equivalent to controlling the aggregate quantity of money.

However, when investment is a choice variable, this equivalence breaks down. Now haircut regulation, while still useful, is a second-best means of intervention as compared to controlling the aggregate quantity of money. This is because the social costs of fire sales are a function of M , so this is the item that the planner would ideally like to control. And trying to do this indirectly, by picking a value of m^{cap} , will now have the undesired side-effect of encouraging shadow banks to raise their investment above the optimal level of I^b . (I am assuming that we are in the low- M region of the parameter

space, so that absent haircut regulation, shadow banks would choose $I = I^b$.) Intuitively, haircut regulation always gives shadow banks the option to create more cheap money financing at the margin so long as they are willing to raise the level of investment.

This can be seen formally by considering the first-order condition with respect to M for a shadow bank facing binding haircut regulation:²⁵

$$\frac{d\Pi}{dM} = \frac{\{pf'(I) + (1-p)\lambda - R^B\}}{m^{cap} R^M} + \frac{(R^B - R^M)}{R^M} - (1-p)z = 0 \quad (\text{A.4})$$

It follows that it is impossible to use haircut regulation to implement the optimum in (A.3) that corresponds to the case where the planner controls M . For if (A.3) is satisfied with $I = I^b$, it must be that $\frac{(R^B - R^M)}{R^M} - (1-p)z = (1-p)\frac{(1-\varphi)}{\varphi}g'(W - M) > 0$.

But then for (A.4) to be satisfied, i.e., for the shadow bank to be optimizing given the haircut constraint, we require $pf'(I) + (1-p)\lambda - R^B < 0$, which means that $I > I^b$.

²⁵ Note that when there is haircut regulation, a shadow bank is free to choose its desired level of M , which is equivalent to choosing its level of investment I .

References

- Adrian, Tobias, and Hyun Song Shin, 2008, Financial intermediaries, financial stability and monetary policy, in *Maintaining Stability in a Changing Financial System*, Federal Reserve Bank of Kansas City, 287-334.
- Allen, Franklin, and Douglas Gale, 2005, From cash-in-the-market pricing to financial fragility, *Journal of the European Economic Association* 3, 535-546.
- Ashcraft, Adam, Nicolae Garleanu, and Lasse Heje Pedersen, 2010, Two monetary tools: Interest rates and haircuts, working paper.
- Bagehot, Walter, 1873, *Lombard Street: A Description of the Money Market* (Charles Scribner's Sons, New York).
- Bernanke, Ben S., and Alan Blinder, 1988, Credit, money, and aggregate demand, *American Economic Review* 78, 435-39.
- Bernanke, Ben S., and Alan Blinder, 1992, The federal funds rate and the channels of monetary transmission, *American Economic Review* 82, 901-21.
- Brunnermeier, Markus K., and Lasse H. Pedersen, 2009, Market liquidity and funding liquidity, *Review of Financial Studies*, 22, 2201-2238.
- Caballero, Ricardo J., and Alp Simsek, 2009, Fire sales in a model of complexity, working paper.
- Cochrane, John, 1998, A frictionless model of U.S. inflation, *NBER Macro Annual*, 323-384.
- Dang, Tri Vi, Gary B. Gorton, and Bengt Holmstrom, 2009, Opacity and the optimality of debt for liquidity provision, working paper.
- DeMarzo, Peter M., 2005, The pooling and tranching of securities: A model of informed intermediation, *Review of Financial Studies* 18, 1-35.
- DeMarzo, Peter M., and Darrell Duffie, 1999, A liquidity-based model of security design, *Econometrica* 67, 65-99.
- Diamond, Douglas W., 1991, Debt maturity structure and liquidity risk, *Quarterly Journal of Economics* 106, 709-737.
- Diamond, Douglas W. and Philip Dybvig, 1983, Bank runs, deposit insurance and liquidity, *Journal of Political Economy* 91, 401-419.

- Diamond, Douglas W., and Raghuram G. Rajan, 2001, Liquidity risk, liquidity creation, and financial fragility: A theory of banking, *Journal of Political Economy* 109, 287-327.
- Diamond, Douglas W., and Raghuram G. Rajan, 2009a, Fear of fire sales and the credit freeze, working paper.
- Diamond, Douglas W., and Raghuram G. Rajan, 2009b, Illiquidity and interest rate policy, working paper.
- Flannery, Mark J., 1986, Asymmetric information and risky debt maturity choice, *Journal of Finance* 41, 19-37.
- Fostel, Ana, and John Geanakoplos, 2008, Leverage cycles and the anxious economy, *American Economic Review* 98, 1211-1244.
- Geanakoplos, John, 2009, The leverage cycle, working paper.
- Goodfriend, Marvin, 2007, How the world achieved consensus on monetary policy, *Journal of Economic Perspectives* 21, 47-68.
- Goodhart, Charles, 1988, *The Evolution of Central Banks* (MIT Press, Cambridge MA).
- Gorton, Gary B., 2010a, Questions and answers about the financial crisis, working paper.
- Gorton, Gary B., 2010b, *Slapped by the Invisible Hand: The Panic of 2007* (Oxford University Press).
- Gorton, Gary B., and Andrew Metrick, 2009, Securitized banking and the run on repo, working paper.
- Gorton, Gary B., and George Pennacchi, 1990, Financial intermediaries and liquidity creation, *Journal of Finance* 45, 49-72.
- Greenwood, Robin, Samuel Hanson, and Jeremy C. Stein, 2010, A comparative advantage approach to government debt maturity, work in progress.
- Gromb, Denis, and Dimitri Vayanos, 2002, Equilibrium and welfare in markets with financially constrained arbitrageurs, *Journal of Financial Economics* 66, 361-407.
- Hart, Oliver, and John Moore, 1995, Debt and seniority: An analysis of the role of hard claims in constraining management, *American Economic Review* 85, 567-585.
- Kashyap, Anil K, and Jeremy C. Stein, 2000, What do a million observations on banks say about the transmission of monetary policy?, *American Economic Review* 90, 407-28.

- Kashyap, Anil K, and Jeremy C. Stein 2004, Cyclical implications of the Basel-II capital standards, *Federal Reserve Bank of Chicago Economic Perspectives* 28 Q1, 18-31.
- Kashyap, Anil K, Jeremy C. Stein, and David Wilcox, 1993, Monetary policy and credit conditions: Evidence from the composition of external finance”, *American Economic Review* 83, 78-98.
- Krishnamurthy, Arvind, and Annette Vissing-Jorgensen, 2008, The aggregate demand for Treasury debt, working paper.
- Leeper, Eric, 1991, Equilibria under active and passive monetary policies, *Journal of Monetary Economics* 27, 129-147.
- Morris, Stephen, and Hyun Song Shin, 2004, Liquidity black holes, *Review of Finance* 8, 1-18.
- Myers, Stewart C., 1977, Determinants of corporate borrowing, *Journal of Financial Economics* 5, 147-175.
- Shleifer, Andrei, and Robert W. Vishny, 1992, Liquidation values and debt capacity: A market equilibrium approach, *Journal of Finance* 47.
- Shleifer, Andrei, and Robert W. Vishny, 1997, The limits of arbitrage, *Journal of Finance* 52, 35-55.
- Shleifer, Andrei, and Robert W. Vishny, 2010, Unstable banking, *Journal of Financial Economics*, forthcoming.
- Sidrauski, Miguel, 1967, Rational choice and patterns of growth in a monetary economy, *American Economic Review* 57, 534-544.
- Sims, Christopher A., 1994, A simple model for the determination of the price level and the interaction of monetary and fiscal policy, *Economic Theory* 4, 381-399.
- Stein, Jeremy C., 1998, An adverse-selection model of bank asset and liability management with implications for the transmission of monetary policy, *RAND Journal of Economics* 29, 466-86.
- Stein, Jeremy C., 2005, Why are most funds open-end? Competition and the limits of arbitrage, *Quarterly Journal of Economics* 120, 247-272.
- Stein, Jeremy C., 2009, Presidential address: Sophisticated investors and market efficiency, *Journal of Finance* 64, 1517-1548.

Tucker, Paul, 2009, The repertoire of official sector interventions in the financial system: Last resort lending, market-making, and capital, remarks at the Bank of Japan 2009 International Conference, May 27-28.

Woodford, Michael, 1995, Price level determinacy without control of a monetary aggregate, *Carnegie-Rochester Conference Series on Public Policy* 43, 1-46.

Figure 1
Private and Socially Optimal Outcomes Versus the Money-Bond Spread

The figure plots private and socially optimal values of money creation M and investment I as a function of R^M . Functional forms and parameter values are as follows: $f(I) = \psi \log(I) + I$; $g(K) = \theta \log(K)$; $R^B = 1.04$; $\psi = 3.5$; $\theta = 150$; $\lambda = 1$; $W = 200$; and $p = 0.95$. R^M varies between 1.0 and 1.04.

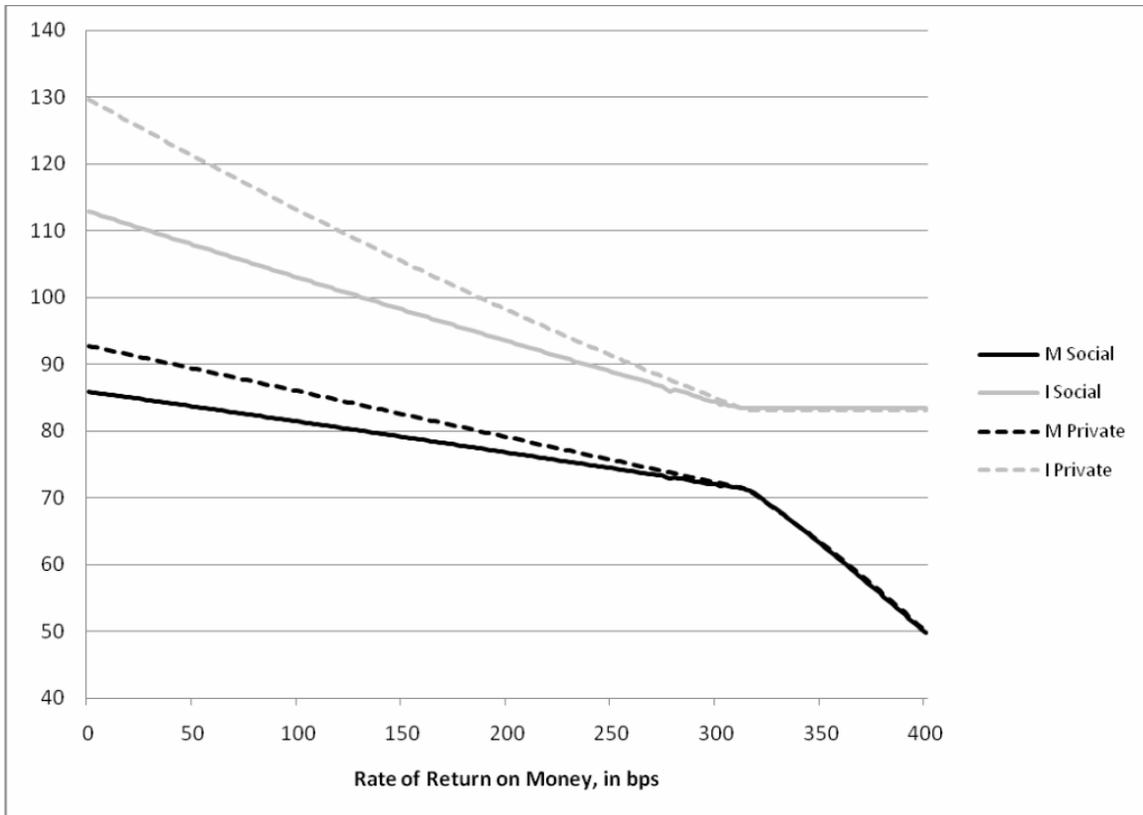


Figure 2
Private and Socially Optimal Outcomes Versus the Money-Bond Spread:
The Case of Imperfect Pledgeability

The figure plots private and socially optimal values of money creation M and investment I as a function of R^M . Functional forms and parameter values are as follows: $f(I) = \psi \log(I) + I$; $g(K) = \theta \log(K)$; $R^B = 1.04$; $\psi = 3.5$; $\theta = 150$; $\lambda = 1$; $W = 200$; and $p = 0.95$. R^M varies between 1.0 and 1.04. In contrast to Figure 1, only a fraction $\phi g(K)$ of the output from time-1 projects can be pledged to private investors, with $\phi = .90$.

