CHAPTER SIX

Monetary Policy with a Layered Payment System

Monika Piazzesi and Martin Schneider

This chapter considers the transmission of monetary policy in a modern monetary economy. It builds on Piazzesi and Schneider (2017), who introduce a model of a payment system with two layers. In an end user layer, households and firms pay for their purchases of goods and securities with inside money provided by financial intermediaries. Monetary policy operates in a different layer, the bank layer: the central bank controls the interest rate on an instrument that is traded almost exclusively by intermediaries. This is true whether the central bank targets an interbank rate such as the US federal funds rate or whether it sets the interest rate on reserves—households and firms directly hold neither interbank loans nor reserves.

The key departure from standard models of monetary policy is that different assets are not perfect substitutes. Rates of returns on different assets are therefore not equated in equilibrium. This feature shuts down the standard channel by which changes in a policy rate are transmitted to other interest rates. Under the standard transmission mechanism, the central bank is assumed to have direct control over interest rates that guide households' and firms' intertemporal decisions such as savings, investment, and price setting.

In a layered payment system, monetary policy instead works by changing banks' incentives to produce inside money. Inside

money is valued for its liquidity—it earns a convenience yield. It is not a perfect substitute for other assets since the convenience yield declines with the quantity of inside money that is available: end users tend to hold more money when it is cheaper to do so. In a layered payment system, reserves and interbank loans—which are essential for producing liquid inside money—similarly earn a convenience yield. Banks hold more of these assets if it is cheaper to do so. Equilibrium equates *total* returns on all assets: rate of return plus convenience yields.

Three important policy implications follow. First, interest rate policy tends to be weaker than what standard models predict. While the central bank can set the interest rate on, say, reserves, the total return on reserves also includes the convenience yield, which depends in turn on spending as well as the quantity of reserves. The convenience yield generates a wedge between the reserve rate and the rate on other, less liquid assets. This wedge moves around with the state of the economy and can change even if policy does not. Second, by affecting the convenience yield, the quantity of reserves can work as a separate policy tool. Finally, financial structure matters: the convenience yields on interbank loans and reserves depend on the role of these instruments in the payment system.

As a concrete example, this paper considers monetary tightening starting from a situation of abundant reserves, which describes the current regime in the United States as well as several other developed countries. In this regime, the spread between the interest rate on reserves and that on other short safe debt is negligible. Central banks currently discuss two possibilities. The first option is to raise the interest rate on reserves but keep the quantity of reserves sufficiently large so that reserves remain abundant, as the US Federal Reserve has done in recent years. Short safe interest rates would continue to rise together with the interest rate on reserves. The second option is for the central bank to further sell off securities for reserves, eventually shrinking the quantity of reserves to the

point where a spread between the interbank rate and the reserve rate emerges.

With a layered payment system, these two options for tightening have different effects on inflation and interest rates. An open market sale of securities changes the mix of collateral assets that banks use to back inside money. In particular, the policy reduces the quantity of the most liquid such asset: reserves. As long as reserves are abundant and the securities sold are similar in quality to reserves, the sale has no effect. A sufficiently large sale, however, makes liquidity scarce for banks. The interest rate on other short safe debt rises relative to the interest rate on reserves. The resulting spread makes it more costly to produce inside money and induces banks to issue less inside money.

In contrast, an increase in the interest rate on reserves increases the rate of return that banks earn on the safest form of collateral, namely short safe debt. In response, banks optimally increase their holdings of such collateral relative to the amount of inside money they issue. If the quantity of reserves available to the banking system is unchanged, banks optimally reduce their issuance of inside money. The contraction of money in this case works only through the money multiplier, not by changing the monetary base. Of course, it is possible to change the quantity of reserves as well. In fact, with a layered payment system, monetary policy can no longer be summarized by a single interest rate but must take into account the quantity of collateral available to banks.

Since the two policies affect different margins of banks' decision making, different features of financial structure matter for assessing their impact. The effects of an open market sale depend on liquidity management in the banking system as a whole. More efficient netting arrangements, for example, economize on reserves and imply that a larger open market sale is required to induce scarcity. Once reserves are scarce, the structure of the interbank market determines the banking system's demand for reserves. The effects of a

higher interest on reserves depend on the nature of other collateral that can be used to back inside money. Since a higher reserve rate reduces the money multiplier, it benefits the holders of nominal assets such as long-term fixed rate mortgages. If banks hold nominal assets, their balance sheets strengthen and they reduce their issuance of inside money by less, leading to a weaker policy impact.

The remainder of this chapter provides more details on the key mechanism. We first provide background on recent policy changes and their treatment in the literature, then sketch the model, and then discuss policy with a layered payment system. We focus throughout on the behavior of intermediaries. Details on how the mechanism can be embedded into macroeconomic models are included in related papers. Piazzesi and Schneider (2017) embed the banking sector described here into a flexible price model, whereas Piazzesi, Rogers, and Schneider (2018) consider a New Keynesian model with sticky prices. Lenel, Piazzesi, and Schneider (2018) provide evidence on the connection between bank balance sheets, portfolio risk, and interest rate spreads.

BACKGROUND

In recent years, central banks around the world have massively expanded the supply of reserves. In many countries, monetary easing has made reserves *abundant*: there is no longer a spread between the interest rate on reserves and interest rates on short safe debt that provides the same payoff as reserves, such as overnight interbank loans. In a regime with abundant reserves (also referred to as a "floor" system), the central bank changes the stance of policy directly by altering the interest rate on reserves.

This new regime stands in sharp contrast to the typical policy environment before the global financial crisis when reserves were *scarce*: the spread between overnight and reserve rates was positive and large. For example, the US federal funds rate for overnight

interbank loans averaged 5.66 percent between 1955 and 2008, much higher than the reserve rate of zero. In a traditional scarce reserves regime (or "corridor system"), the central bank targets the overnight interbank rate and its trading desk adjusts the supply of reserves with this goal in mind.

It is important to distinguish between abundance of reserves and the so-called "zero lower bound" on interest rates. It is true that many central banks initially made reserves abundant at a time when the interest rate on reserves was in fact zero. The first data points about the new regime thus featured zero (or very low) reserve rates as well as overnight rates. This was only a coincidence, however: more recently, reserves have been abundant also at positive or negative levels of interest rates.

For example, in the wake of recent monetary tightening in the United States, the interest rate on reserves and the federal funds rate have been rising virtually in lockstep. Reserves thus remain abundant at rates that are substantially above zero. At the same time, the euro area experience has shown that there is no lower bound at zero. Indeed, in recent years, the negative Eonia (euro overnight index average) rate has been closely tracking the similarly negative ECB (European Central Bank) deposit facility rate.

Many central banks now face the question of whether reserves should remain abundant in the long run or whether they should reduce the supply of reserves so as to move back to scarcity. It is clear that the choice of regime affects how the payment system works. For example, banks manage liquidity differently when overnight borrowing is costly relative to reserves. What is less well understood is how the policy regime affects the transmission of monetary policy to the macroeconomy and, in particular, how monetary tightening affects inflation.

Standard models of monetary policy do not distinguish between scarce and abundant reserve regimes. Indeed, in the New Keynesian framework, the focus is on a single interest rate controlled by the

central bank that adopts the Taylor rule (Taylor 1993; see Woodford 2003 or Galí 2015 for textbook treatments). That interest rate also serves as a benchmark rate for all intertemporal decisions in the economy, such as investment and hiring. In other words, by controlling the policy rate, the central bank controls the entire yield curve as well as firms' cost of capital.

The key assumption underlying New Keynesian analysis is that financial markets treat all assets as perfect substitutes. In particular, there is nothing special about short safe debt such as reserves or overnight credit: much like other assets, short safe debt is valued only for its rate of return and not, say, for its liquidity or its quality as collateral. As a result, the transmission of monetary policy is very powerful and the stance of policy can be summarized by the policy rate alone. How the policy rate is changed in practice is a detail of the economy's "plumbing" that is not relevant for policy transmission.

Much recent analysis of the zero lower bound applies the same logic to a world in which nominal rates are bounded below by zero and prices are rigid, so that all real rates of return hit a lower bound. The bound on the nominal interest rate is often motivated by the presence of currency—the idea is that if nominal rates on reserves or deposits were negative, then banks or end users would exchange it one-for-one for currency. This approach to thinking about recent events cannot handle the fact that interest rates on both reserves and deposits have been negative. Moreover, it does not make a distinction between abundance of reserves and low interest rates.

A MODEL OF PAYMENTS IN A MODERN ECONOMY

The framework developed by Piazzesi and Schneider (2017) provides a different perspective on monetary policy transmission. It starts from two simple observations on fixed-income markets. First, central bank interest rate policy does not directly affect rates of

return on assets held by households and firms. Instead, whether the central bank targets an overnight interbank rate or varies the interest rate on reserves, it targets an instrument held directly only by banks or other intermediaries that provide payment services such as money market mutual funds.

The second observation is that interest rates on short safe instruments targeted by central banks are different from other rates that guide the decisions of households and firms. The short safe instruments that are held by nonbanks are typically "inside money" provided by banks or money market mutual funds and earn interest rates—such as deposit rates—that are lower than the short safe rates earned by banks on instruments such as short government bonds. The longer term instruments that are relevant for, say, investment decisions tend to earn higher interest rates than the short safe rates earned by banks.

Piazzesi and Schneider (2017) provide a simple model that is consistent with both observations. The model describes a layered payment system typical of modern economies: in an *end user layer*, households, firms, and institutional investors pay for goods and securities with inside money provided by intermediaries ("banks," for short). In the *bank layer*, banks handle end users' payment instructions and pay each other in central bank money. There is no currency in the model. The idea is that currency is used for a small share of transactions and is held mostly abroad and in the informal sector.

The banking sector in the model adds value because it provides a technology for payments. End users like inside money ("deposits," for short) because it provides liquidity that is convenient for trading goods and securities. The liquidity benefit of deposits explains why end users are willing to hold them even at low interest rates. In order to provide liquidity benefits, banks face financial frictions. First, end users require that deposits be backed by sufficient collateral—that is, banks hold assets of sufficient quality. Second,

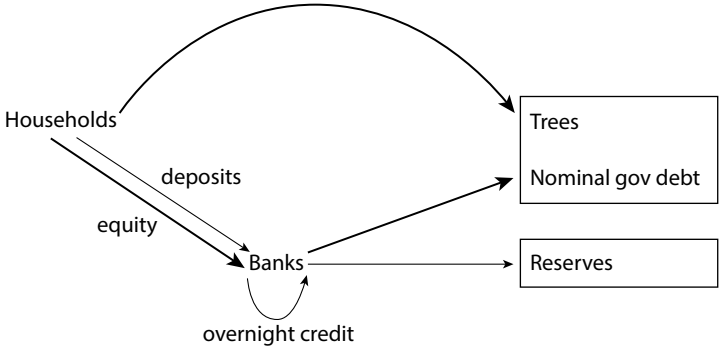

FIGURE 6.1.1. Ownership of Assets with a Layered Payment System

end users' payment instructions generate deposit flows between banks, which require liquidity management. Banks can handle deposit flows by holding reserves or borrowing in the interbank market.

The basic structure of asset ownership is illustrated in figure 6.1.1. Households can invest in three different assets that are in nonzero net supply: reserves, nominal government debt, and "trees" that are other claims on output. Households can invest in assets either directly or through banks. Banks are competitive firms that issue deposits and equity. Since households own bank shares and banks maximize shareholder value, households also decide banks' capital structure and investment. In particular, they decide to what extent trees and debt are held through the banking system as opposed to directly. At the same time, reserves can be held only via banks, reflecting the current regulatory environment in most countries. There is also an interbank market where banks make short-term loans to each other.

In a frictionless world, all asset positions—indicated by arrows in figure 6.1.1—would be indeterminate, all assets would be perfect substitutes, and all rates of return would be equal, at least after risk adjustment. To make predictions about positions, we thus

introduce financial frictions. Assets that help overcome frictions are then valuable beyond simply their rate of return: they convey liquidity or collateral benefits that decline with the quantity of assets available. The presence of liquidity and collateral benefits accounts for spreads between different interest rates. Since benefits decline with quantity, trade-offs between spreads and quantities generate determinate optimal positions.

The first friction is that it is costly to exchange assets for goods, and this cost is smaller for deposits. Since deposits thus facilitate the purchase of goods, households value them for their liquidity and not only for their rate of return. As a result, the model predicts a spread between deposits and other assets such as bank equity that reflects the liquidity benefit of deposits. We emphasize that this spread is always positive: even when reserves are abundant so reserves are no longer more liquid for banks than other short debt, it remains the case that deposits are more liquid for households than any other debt. This property of the model is important to match the behavior of spreads in the data.

COSTLY LEVERAGE

The second friction is that bank leverage is costly. In particular, there is a resource cost per unit of debt of the bank that is decreasing in the *collateral ratio* of the bank, defined as weighted assets divided by debt (deposits plus interbank borrowing), or the inverse of leverage. The interpretation is that a bank has to make some effort to convince depositors (or regulators) that it is sound. This is less costly if the bank has lower leverage or its assets are of higher quality (that is, they receive a higher weight in the collateral ratio). It is not necessary for these leverage costs to be large. What matters is that they decline with the collateral ratio and tend to zero as the collateral ratio increases—a bank that issues no debt has an infinite collateral ratio and pays no leverage cost.

Costly leverage leads to a number of key predictions. First, there is an optimal collateral ratio that trades off the spread between deposits and bank equity against the cost of leverage. Bank deposits are a cheap source of financing for banks because they provide a liquidity benefit to households and hence earn a lower rate of return than bank equity. For banks with low overall levels of debt, it is therefore cheaper to issue deposits than to raise equity. Since the first dollar of debt implies no leverage cost, it is always optimal to issue *some* deposits. The more deposits a bank issues, the lower its collateral ratio, which in turn increases its leverage cost. At the optimal collateral ratio, the marginal leverage cost is equated to the spread between the deposit rate and the bank's cost of capital, its rate of return on equity.

Costly leverage further implies that banks value assets as collateral and not only for their rate of return. When a bank invests a dollar in short safe debt, say, it knows that this not only contributes to its return on assets but also increases its collateral ratio and hence lowers its leverage cost. Banks are thus willing to hold short safe debt even if its rate of return is below the rate of return on bank equity. The model thus predicts a spread between safe short-term debt held by intermediaries that issue deposits and assets held directly by households that are valued only for their payoffs.

Moreover, the model predicts that the spread between short safe debt and other assets should be higher when bank balance sheets are weaker in the sense that collateral ratios are lower. Banks with low collateral ratios are willing to pay more for collateral. They bid up the price of short safe debt, which lowers the interest rates. This mechanism fits well with recent evidence: as the economy entered the financial crisis, bank balance sheets were weak and interest rates on short safe debt were low.

A final prediction of the model that follows from costly leverage is market segmentation: in equilibrium, short safe bonds are held only by banks and not directly by households. Given the spread

between short bonds and other assets, banks that receive collateral benefits are happy to hold short safe bonds even at low rates of return. In contrast, households perceive these bonds to be too expensive and prefer not to hold them directly—they prefer other assets such as bank equity. This prediction is in line with observed positions: most short paper in the US economy is held by intermediaries, whereas the short-term assets held by households are predominantly deposits.

Summing up, we have now described the positions indicated by dark arrows in figure 6.1.1. Bank deposits are produced because they provide liquidity to households. Since leverage is costly, banks also issue equity and buy assets as collateral. Banks therefore buy short safe debt. They can buy short safe debt as well as trees. How many trees they buy in equilibrium is not essential for the argument here. In the full model in Piazzesi and Schneider (2017), banks compete for trees with other asset management firms and the allocation of trees depends on who has the better technology to invest in certain classes of trees.

LIQUIDITY MANAGEMENT

The final element of the model is bank liquidity management, represented by the light arrows at the bottom of figure 6.1.1. In order to capture the idea that reserves provide liquidity to banks, we assume that banks face liquidity shocks. The idea is that banks may experience deposit outflows as part of their provision of payment instruments: sometimes customers' payment instructions require wire transfers to other banks. We further assume that such transfers are subject to a liquidity constraint: banks must either wire reserves they already hold or access the overnight credit market to obtain additional reserves.

Banks thus manage liquidity with two tools: reserves and overnight credit. Since overnight credit is subject to the same leverage

cost as deposits, it is always better to first exhaust reserves before turning to the overnight market. When do they have to borrow? We assume that liquidity shocks are proportional to deposits. The likelihood that banks must borrow overnight is then captured by a second key balance sheet ratio: the *liquidity ratio*, defined as their reserves divided by deposits. If a bank's liquidity ratio is low, then there is a good chance that deposit outflows require borrowing. At the other extreme, if the bank were to invest only in reserves, then it could for sure withstand any liquidity shock without borrowing.

What is the optimal liquidity ratio chosen by banks? It is determined again by trading off a spread against a nonpecuniary benefit. Banks value reserves for their liquidity: holding reserves allows them to avoid costly overnight borrowing in the event that they experience a large deposit outflow. This liquidity benefit is decreasing in the liquidity ratio: if the liquidity ratio is higher, the chance that banks must turn to costly borrowing declines. Under the reasonable assumption that liquidity shocks are bounded—that is, there is a largest outflow that is lower than 100 percent of deposits—there will be some threshold ratio such that the liquidity benefit reaches zero. If all conceivable outflows can be covered with existing reserves, any further dollar of reserves does not convey a liquidity benefit.

We therefore have two cases for the choice of liquidity ratio. If there is a positive spread between the interest rate on short safe debt and the interest rate on reserves, then there exists an optimal liquidity ratio such that the marginal liquidity benefit of the last dollar of reserves added to the balance sheet is exactly equal to the spread. We note that this optimal liquidity ratio depends on the bank's collateral ratio: a better collateralized bank finds it cheaper to borrow in the interbank market, and therefore perceives a lower marginal liquidity benefit of reserves. In the case with a positive spread, banks borrow in the interbank market if they experience a large enough liquidity shock. The second case occurs when the

spread between the short rate and the reserve rate is zero. In this case, the bank chooses a liquidity ratio above the threshold where it never has to borrow and would be willing to hold additional reserves at zero spreads.

The optimal liquidity management by banks thus generates a demand for reserves. At high interest rates, banks choose high collateral ratios that make it cheap to borrow overnight from other banks. In this case, banks choose low liquidity ratios. At lower rates, banks choose lower collateral ratios at which it would be expensive to borrow overnight from other banks. In this case, banks choose higher liquidity ratios. The resulting inverse relationship between interest rates and demand for liquidity is much like a conventional "money demand function" for banks.

TWO EQUILIBRIA

Liquidity management and costly leverage together determine the optimal choice of each bank, which determines the liquidity ratio and collateral ratio. A key property of the model is that although we allow for heterogeneous banks that make a rich set of balance sheet choices, these banks choose the same liquidity ratio and collateral ratio, so that we can describe the behavior of the aggregate banking sector with these ratios. An equilibrium of the model requires optimal savings and portfolio choices by households, optimal balance sheet choices by banks, and market-clearing for goods and assets. This equilibrium is fully characterized by the behavior of the two aggregate banking ratios.

The equilibrium liquidity ratio determines whether the economy is in a regime with abundant or scarce reserves. In the regime with abundant reserves, banks have a high enough liquidity ratio that they do not have to borrow from each other. The equilibrium fed funds rate and the reserve rate are the same and all activity on the overnight interbank credit market ceases. After several rounds

of quantitative easing, the United States has been in the abundant reserve regime after the financial crisis. Before the crisis, the United States was in a scarce reserves regime, where banks choose a low liquidity ratio and need to borrow from other banks if they are hit by large liquidity shocks. As a result, reserves are valued for their convenience, which creates a positive spread between the fed funds rate and the reserve rate.

TIGHTENING IN THE ABUNDANT RESERVE REGIME

There are two options for monetary policy to tighten. The first option is an open market sale of securities which changes the mix of collateral assets that banks use to back inside money. In particular, the policy reduces the quantity of reserves. As long as reserves are still abundant after the sale and the securities sold are similar in quality to reserves, the sale has no effect. A sufficiently large sale (for example, as in a large unwind of the Fed's portfolio), however, makes liquidity scarce for banks. The model predicts that the interest rate on other short safe debt rises relative to the interest rate on reserves. The resulting spread makes it more costly to produce inside money and induces banks to issue less inside money.

The second option is an increase in the interest rate on reserves. This policy increases the rate of return that banks earn on the safest form of collateral, namely short safe debt. In response, banks optimally increase their holdings of such collateral relative to the amount of inside money they issue. If the quantity of reserves available to the banking system is unchanged, banks optimally reduce their issuance of inside money. The contraction of money in this case works only through the money multiplier, not by changing the monetary base.

While both policies can lead to higher interest rates, they have opposite effects on the liquidity of the banking system. An open

market sale reduces the aggregate liquidity ratio, whereas a higher reserve rate increases the liquidity ratio. Since these policies affect bank balance sheets differently, their effectiveness depends on financial structure in different ways. A key question for the effectiveness of an open market sale is how banks manage to operate with fewer reserves. If banks' liquidity management is highly efficient, an open market sale may not be able to increase interest rates because it may not be substantial enough to reach scarcity. For example, in an economy with few banks, many transactions are netted on the books of the same bank. Moreover, more efficient netting between financial institutions also reduces the need for reserves.

An important aspect for the effectiveness of changes in the reserve rate is the nature of other collateral that banks can use to back inside money. Since a higher reserve rate reduces the money multiplier, the policy tends to benefit the holders of nominal assets such as long-term fixed rate mortgages. If banks hold many nominal assets, their balance sheets strengthen in response to a higher reserve rate and they can reduce their issuance of inside money by less, leading to a weaker policy impact.

References

Galí, Jordi. 2015. *Monetary Policy, Inflation, and the Business Cycle*. Princeton, NJ: Princeton University Press.

Lenel, Moritz, Monika Piazzesi, and Martin Schneider. 2018. "The Short Rate Disconnect in a Monetary Economy." Working paper, Stanford University, Stanford, CA.

Piazzesi, Monika, Ciaran Rogers, and Martin Schneider. 2018. "Money and Banking in a New Keynesian Model." Working paper, Stanford University, Stanford, CA.

Piazzesi, Monika, and Martin Schneider. 2017. "Payments, Credit, and Asset Prices." Working paper, Stanford University, Stanford, CA.

Taylor, John B. 1993. "Discretion versus Policy Rules in Practice." Carnegie-Rochester Conference Series on Public Policy 39: 195–214.

Woodford, Michael. 2003. *Interest and Prices: Foundations of a Theory of Monetary Policy*. Princeton, NJ: Princeton University Press.

DISCUSSANT REMARKS

Oleg Itskhoki

This is an extremely meaty paper. It took me quite a while to work through it, and it was absolutely worthwhile. There were a lot of dimensions to the paper. I'm going to discuss just a few of them, but the paper definitely contains a lot more.

The idea of the paper is to build a detailed micro-model of how the monetary transmission works through the banking system, and then study the macroeconomic implications of this transmission mechanism. One can then use it to ask the question of the optimal aggregate liquidity management by the government. The model combines quite a few ingredients, and each of them appears simple and intuitive in partial equilibrium. Yet the paper's main contribution is in having all these individual ingredients work nicely together in general equilibrium. And it is very impressive how Monika and Martin can characterize the equilibrium outcomes in a tractable way.

The model is an endowment economy. The macro variables to be determined in equilibrium are consumption, inflation, and asset prices. The reason why consumption is not equal to output is because there are real collateral costs, and so consumption can differ across equilibria depending on how large are the aggregate collateral costs in the economy.

There are three types of agents: the households (consumers), the banks, and the government. There are multiple assets in the economy (deposits, reserves, short-term debt, bank equity, stock markets, etc.) and the paper characterizes equilibrium prices of all assets. Typically, one can characterize the prices of all assets either when they are all equivalent or when the markets are complete. Neither is the case in the present paper, and it is very impressive how all asset prices admit closed-form characterization in this rich equilibrium environment.

There are two types of frictions in the payment system: (a) a liquidity constraint (cash-in-advance) on both households and banks and (b) a collateral constraint (costly leverage) on both banks and the government. Therefore, there are a total of four constraints on three types of agents, and all of them are consequential. It is the intersection of these four constraints that creates the interesting equilibrium outcomes.

In the limiting case, the economy becomes a frictionless neoclassical economy where the constraints don't bind. I will refer to it as a Friedman-rule economy. The Friedman rule is more complex in this economy. It not only guarantees that money and bonds have the same rates of return, but in fact that all assets have the same returns, and agents are indifferent about their portfolio choices, and no constraints are binding. Away from the Friedman rule, the constraints are binding and the asset prices are different for different assets, and the allocation is not first best.

This is how the model works. First, consider the liquidity constraint on the household sector. In order to consume, the households need liquidity, and this liquidity is inside money, or deposits. In a conventional model, one needs to hold money, the reserves, in order to buy consumption. Here, one does not need money issued by the government to buy consumption. Instead, one uses deposits to finance consumption expenditure in each period. Therefore, the households need the deposits in order to pay for their consumption; hence, they are willing to take lower interest rates on their deposits relative to other assets, when the liquidity constraint is binding. As a result, the interest rate on deposits is lower than on the other assets, when the constraint is binding. In turn, this allows the banks to make profits using this gap in interest rates.

The banks pay back these profits to their stakeholders, the households. To summarize, the first sign that the liquidity constraints are binding is that the rate on deposits is lower than the interest rate on other assets, and the banks are making profits off the deposits.

The second friction is also a liquidity constraint, but now on the banks. The banks issue deposits and sometimes they need to quickly liquidate their positions. If people want their money back, the banks need to be able to pay them back on short notice. This is modeled as a random liquidity shock on the banks. In order to pay back, the banks need to have either reserves, which is the outside money issued by the government, or they can borrow in the interbank market. Because these funds are useful for the bank when liquidity constraints bind, the banks are willing to take lower interest rates on holding the reserves. If they hold bonds, they're also willing to hold them at a lower interest rate, as long as bonds provide liquidity (collateral) services.

This is not enough, however. The model also requires that the banks find it costly to hold too much of the reserves, as reserves are expensive and the banks need leverage to make profits. The banks face an exogenous real cost of leverage. As a result, there exists a collateral ratio—the ratio of liquid assets to liabilities—and the collateral costs decrease with the collateral ratio. If the banks have a lot of collateral, the leverage costs are low, but this means that the banks are not making the differential returns on assets and liabilities, and this creates a trade-off. The leverage costs are real costs, and they reduce the amount of output left for consumption in the economy. To summarize, the banks want to hold the reserves to relax both the collateral constraint and the liquidity constraint. However, holding reserves is costly, if the returns on reserves are low relative to the returns on other assets. The Friedman rule increases the quantity of reserves in the economy, making them cheap and abundant, increasing the return on reserves, until both the liquidity and the collateral constraints of the banks are fully relaxed and the economy approaches the first best.

What are the other assets? One can hold short-term bonds and the equity of the banks. But they are less effective as means of dealing with collateral and liquidity constraints, while reserves are

most effective in relaxing both. Finally, why can't one get the high return on reserves? In principle, the government could provide a lot of reserves at no cost. But the problem is that the government also faces costs of leverage, and the leverage for the government is modelled as the size of the transactions in the economy relative to the size of the balance sheet of the government. Hence, the government also does not like to have a big balance sheet, and this is why it offers a limited quantity of reserves, making them scarce and expensive, and driving low the return on reserves. This is the reason why the Friedman rule is not achieved in this economy.

Taking into account that increasing the balance sheets of both banks and the government is costly, there exists an internal solution for the optimal quantity of liquidity in the economy, and the constraints are binding in equilibrium. In the internal solution, there is a differential return on different types of assets. The government can choose the interest rates on reserves, and it can choose the growth rate of the outside money (the reserves), and it can also choose the composition of its balance sheet, subject to the constraints. Therefore, the government has three choice variables, with the goal of maximizing welfare in the economy—that is, minimizing the aggregate collateral costs to both banks and the government. In turn, the banks choose the collateral ratio and the liquidity ratio to maximize the value of their shareholders, taking as given the actions of the other banks.

A natural question then is: What is the optimal thing for the government to do? The government would always want to minimize its balance sheet and simultaneously to relax both of the constraints for the banks. While these may seem to be conflicting goals, in principle, it is possible to achieve both by simultaneously issuing lots of reserves and saving a comparable amount in the form of private bonds. In other words, the government could create lots of liquidity and simultaneously save in other assets to reduce its leverage costs, by expanding its balance sheet. Hence, in order to make

the problem interesting, the paper must impose an upper bound on how much the government can save in private assets—or, in other words, on the size of the balance sheet of the government, which is feasible without recurring to government leverage. Empirically, it is an interesting question why the governments shy away from large balance sheets and do not want to provide more of the liquidity services.

I will next turn to my comments on the modeling approach in the paper. There are two related trade-offs the paper must confront. The first is how detailed versus concise the model must be. The paper opts for a very detailed model of the banking system, with a lot of details that are often ignored in the macro literature. It is very impressive how far the authors can go with such a detailed model. The natural question, however, is which details are absolutely essential and which ones may be dispensed with in the future, when we incorporate these mechanisms into the workhorse macro framework. Does there exist a concise version of this model, which maintains the main mechanisms and trade-offs but which we can easily wrap into a full macroeconomic model with production and other features? Or, perhaps, there are circumstances when all these micro details do not matter and we can default to the baseline model (e.g., in "normal times"), and there are circumstances when these features become first-order for the macroeconomic outcomes and need to be modeled in full detail (e.g., in "crisis times").

Having "complained" that the model is perhaps too detailed, my second comment is that it is arguably not detailed enough. Luckily, the format of the discussion allows me to not be fully coherent, and just explore different directions. Indeed, the second trade-off the paper faces is between having some ad hoc constraints in the model and fully micro-founding them. It's not clear that is feasible at all given the state of the literature. The main ad hoc constraint the model relies on is the collateral constraint or, rather, the costs of leverage, on both the banks and the government. Literally speaking,

when there is not enough collateral, the banks and the government need to burn resources. Of course, it is a parable for something. For example, in a partial equilibrium of the banking sector, it is perhaps not very consequential, and indeed intuitive. There are likely real costs for the banks of not being collateralized enough, and so one can just say that effectively the banks lose resources if that happens. But once we go to the general equilibrium, a specific model of such costs becomes consequential. Indeed, one needs to know whether the resources are burnt proportionally in every period or instead in certain infrequent states of the world, e.g., when there is a crisis. This, in turn, is likely consequential for the macroeconomic outcomes such as inflation and aggregate consumption. Do we have a sense of robustness concerning for which macro outcomes the stylized nature of the constraints is consequential and for which it is not?

Furthermore, and perhaps more important, once we go to a model with ad hoc constraints, studying optimal policy becomes very tricky. Indeed, the authors acknowledge this by carefully avoiding making strong policy prescriptions based on the model. Yet, of course, the questions of the optimal aggregate liquidity management are of the highest applied interest, and it is hard to avoid thinking about this issue in the context of this paper. Unfortunately, answering these questions without knowing more about the particular micro nature of the collateral costs is difficult. In particular, we do not know how specific policies may affect the collateral cost functions for the banks and the government, whether they would keep them unchanged or may alter them in some fundamental way. This is a version of the Lucas critique in the context of this model. And this is the main reason why the whole literature should think harder about the deeper micro foundations behind the liquidity and collateral constraints that are commonly adopted in the more positive work, which hence needs to avoid making strong normative recommendations.

It is also interesting to know the nature of the liquidity shock for the banking system. The banks need to have liquidity in certain states of the world. In partial equilibrium, this is very natural as a description of the environment for the banks. But once you start thinking about general equilibrium, you start wondering: With which aggregate shocks do the micro-level liquidity shocks interact, contributing to the cycle? Are these liquidity shocks something that could, in principle, be effectively diversified, so that the advances of technology would make these shocks less consequential for the macroeconomy? Or is it something about the aggregate state of the economy which does not allow it to effectively diversify the liquidity shocks, no matter what the market structure and technology used in the interbank market? Are these shocks rare, correlated events like the one that happened in 2007, and hence there is no effective way of avoiding them? If yes, how frequently do we expect to see such shocks in equilibrium?

A quick additional remark is about the welfare objective in the economy, which is exclusively to minimize the collateral costs, as it is an endowment economy. Of course, the natural next step is to extend the environment to a production economy, where sticky-price and/or financial constraints result in endogenous cyclical output fluctuations, which are then reinforced by the constraints in the banking system. The authors have a companion project where they do just that, and this is an important continuation to this research agenda.

I have three remaining comments, related to the empirical verification of the model's mechanism. First, the basic fact about the world is that deposit rates are low relative to other rates of return in the economy, even after controlling for the associated risk. One can go in at least two different ways about interpreting this fact. The way this paper interprets the fact is to say that liquidity constraints are binding, and hence return differentials. Thus, since we observe return differentials in the data, it must be that liquidity constraints

are really binding and we must take this mechanism seriously. An alternative interpretation of the data is that of the market paper: perhaps the low deposit rates reflect the local monopoly power of the banks over retail customers. Is it possible to separate empirically the low rates on deposits due to market power versus those that are due to liquidity constraints? Perhaps the amount of market power changes slowly, at low frequencies (even though the recent crisis was followed by a wave of consolidation in the banking sector), and hence much of the cyclical fluctuation in the deposit spread is due to liquidity constraints. A further possibility is that the observed deposit spreads are due to some form of interaction between market power and financial frictions. Then it is interesting to know how the technological improvements in the high-tech financial sector may wipe out the market power of the conventional banks and what the implications are of such changes for the cyclical analysis.

My last two comments are about the more direct ways one can look at the data to get some empirical validation of the model's assumptions. I think there are two salient predictions of the model. The first salient implication is that the households don't want to hold the stock market other than the banking system, while the banking system will hold all of the non-banking equity. The banking equity offers high returns for the households, while the rest of the stock market is not particularly useful for the households, as it does not allow them to relax any of their liquidity needs. This is, of course, very stylized, and should not be taken literally. But one can ask a more nuanced question in the data. Specifically, in periods when liquidity constraints tighten, is it true that the expected equity returns for the banks are higher than equity returns on the rest of the stock market, and would an investor with deep pockets be able to take advantage of this?

The second salient implication of the model is about the cross section of countries with different institutions and, hence, arguably different reduced-form leverage cost functions. In some countries,

leverage is very costly for the government, while in others governments run very large balance sheets with a lot of leverage. For some governments, it is very easy to borrow, and for other governments it is much costlier. Such variation in the leverage costs should translate into different choices of collateral and liquidity ratios in the private sector, through the endogenous mechanisms of the model. This, in turn, should translate into different macroeconomic outcomes. Can we look at the cross section of countries where governments have a differential ability to increase their balance sheets and see whether this indeed translates into different equilibrium outcomes in the banking system, as predicted by the model? Perhaps there exists anecdotal evidence of such effects.

This is a rich and insightful paper, and I look forward to the new developments in this exciting research agenda!

GENERAL DISCUSSION

MARTIN SCHNEIDER: I think we develop different types of models for different purposes. If we wanted to perform a full-on welfare analysis, more detail may be useful. The leverage cost in our model is basically a cost of breaking promises. It's effectively about expected bankruptcy cost—the probability of bankruptcy, how the legal system works during bankruptcy, and what investors typically recover. Adding more details about bankruptcy would make this aspect of our model more explicit.

We also have worked on a version of our model that has a more elaborate macro side. The key implication of our approach for macro modeling is that monetary policy targets the interest rate on an instrument that incorporates a convenience yield. This means that there is no direct transmission from the policy target to the marginal rate of substitution of households. Moreover, the wedge between the policy rate and the interest rate that matters for households depends on the current condition of the banking system. The map between this interest rate spread and variables that describe the banking system can be captured by a reduced form mapping. We can add details and other features to our model. The current version is the simplest way to derive the result that financial structure matters for the transmission of monetary policy.

And the other ideas you had I think were great. We'll follow up on that.

ARVIND KRISHNAMURTHY: In this model, you have interest on reserves, but yet there is a convenience yield on liquid assets and the quantity of liquidity matters. This world then appears different than the one that Bob Hall was telling us about where interest on reserves appeared sufficient to describe the monetary equilibrium. I was wondering if you could say more about that and

clarify the difference between your model of the world and Bob's model of this morning.

MARTIN SCHNEIDER: I'll say two sentences about ours and then pass it on to Bob. So here the convenience of the reserves in our setting comes from two pieces. There is a liquidity benefit to the banks, and that is something that can shrink to zero. That is what happens in an abundant reserve regime in the liquidity trap. In this situation, T-bills and reserves become perfect substitutes for banks. However, because banks face a leverage cost, which is an additional financial friction, reserves have a convenience yield even when they are abundant because they are useful as collateral for the banks. The same is true for T-bills and other fixed-income instruments that banks hold. In other countries where banks can hold equity, equity values will also reflect a collateral benefit. In this sense, there is still convenience yield even in a liquidity trap. That's an important part of our theory.

ROBERT HALL: Reserves, as you point out, form the basis of the bottom of a transaction system. Ultimately, you need to transfer reserves to meet a financial obligation. On the other hand, the influential instrument that sets the interest rate is the RRP (reverse repurchase agreement).

[Note: Since the conference, the Fed has stopped using the RRP to set interest rates and reverted to using the interest rate on bank reserves.]

MARTIN SCHNEIDER: I consider the existence of the reverse repo program of the Fed, and the fact that currently the federal funds rate and the reserve rate are not exactly the same, but the fed funds rate is a bit lower, peculiarities of the payment system in the United States, where there are payment providers and other entities that use reserves but don't get interest on them. This goes back to US regulatory history. In other countries, it's not like that. In other countries, our model applies more cleanly in the

sense that there are two regimes. There is an interbank market. There is the rate on interbank borrowing, and the central bank targets that rate. In the regime with abundant reserves, the interbank market essentially shuts down, and then one can think of equality between the reserve rate and the interbank rate.

In the United States, there are additional wiggles, which means that the fed funds market did not shut down completely, and then there was a reverse repo program introduced to make these other payment instrument providers more like banks, even though legally they're not.

From the perspective of our model, this is a feature that is left out. But I think for the main point of how policy transmission works across these two regimes, that is a detail. Adding more detail meant to put even finer tubes in the plumbing can be interesting, but I don't think it changes the overall spirit of the analysis.

JEFF LACKER: This idea of the RRP rate setting interest rates for the market has come up a couple times. It's important to clarify what Martin was saying. The Fed sets an interest rate on excess reserves. Any bank can borrow from anyone and put the resulting funds it obtains at the Federal Reserve at the interest rate on excess reserves. So the marginal cost of doing that will determine the spread between what they're willing to pay on, say, deposits or fed funds borrowed and the interest rate on excess reserves. When the Fed goes out and borrows money via RRPs, it auctions off that transaction. Where's the rate going to be? Well, it's going to be set in the market at a rate commensurate, risk-adjusted, with the rates banks are willing to pay and other intermediaries are willing to pay on those funds. Everything else is sort of just little wiggles and complications.

MARTIN SCHNEIDER: The way that we think about it in our model is that right now we're in a regime where all short paper is basically

the same and moves in lockstep as the interest rate on reserves is changed. The last couple of years showed that when policy changes, as long as we stay in this regime, that's sort of how it works. There are small differences, but that's the main theme. The liquidity trap remains in the bank layer, not elsewhere in the economy, and everything moves together.