Banks and Markets: Substitutes, Complements, or Both?*

BY MITCHELL BERLIN

In traditional banking arrangements, households hold their savings in the form of deposits at the bank, which makes loans to both firms and households and holds these loans to maturity. But in the United States, and to a lesser extent in other developed countries, markets have increasingly taken over the roles traditionally played by banks. The shift of financing activity from banks to financial markets, as well as their continued coexistence, raises a number of questions. In this article, Mitchell Berlin discusses some of these questions, such as: What factors determine the relative importance of banks and markets in a financial system in which the two types of finance coexist? Why do so many borrowers continue to use a mixture of bank loans and bonds? And perhaps most important: How does the mix of banks and market finance affect the real economy? That is, how much households save, how firms invest, and how fast the economy grows.

Banks play a central role in most developed financial systems. In traditional banking arrangements, households hold their savings in the form of deposits at the bank, which makes loans to both firms and households and holds these loans to maturity. But in the United States, and to a lesser extent in other developed countries, markets have increasingly taken over the roles traditionally played by banks. Since the 1980s, a larger share of firms' borrowing has shifted from bank loans to bonds (Figure 1). In addition, *securitized assets* — in which loans are packaged with other loans into marketable securities — have become an increasingly dominant channel for consumer finance in the U.S. (Figure 2) and in Europe (Figure 3).[1] While some breathless observers have predicted the ultimate decline of traditional banking altogether, most recognize that modern financial systems involve a mix of banks and markets. This is true even at the level of the individual firm. Firms with ready access to stock and bond markets continue to borrow from banks. And following the disruptions in the asset-backed securities market during the recent financial crisis, it no longer seems obvious that the consumer loan market will be so heavily dominated by securitized loans.

The shift of financing activity from banks to financial markets, as well as their continued coexistence, raises a number of questions. What factors determine the relative importance of banks and markets in a financial system in which the two types of finance coexist? Why do so many borrowers continue to use a mixture of bank loans and bonds? And perhaps most important: How does the mix of banks and market finance affect the real economy? That is, how much households save, how firms invest, and how fast the economy grows.

Mitchell Berlin is a vice president and economist in the Philadelphia Fed's Research Department. He is also head of the Banking and Financial Markets section.

This article is available free of charge at www.philadelphiafed.org/research-and-data/publications.

BANKS AND MARKETS BOTH PRODUCE INFORMATION, BUT DIFFERENTLY

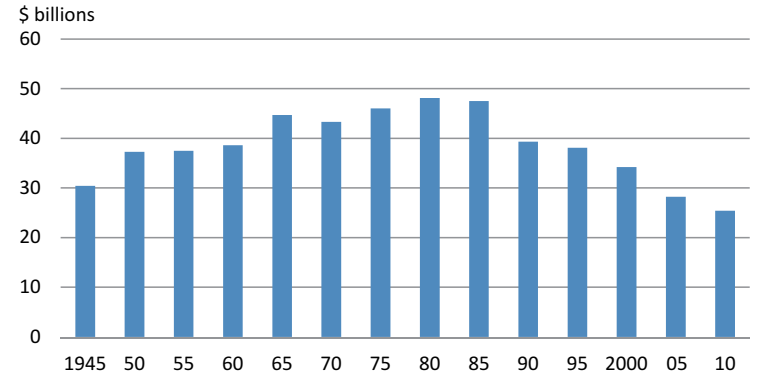
Before going further, we need to clarify some terms. I use the polar

*The views expressed here are those of the author and do not necessarily represent the views of the Federal Reserve Bank of Philadelphia or the Federal Reserve System.

[1] In their review article, Gary Gorton and Andrew Metrick explain how securitization works and discuss the underlying economics of securitization at length.

FIGURE 1

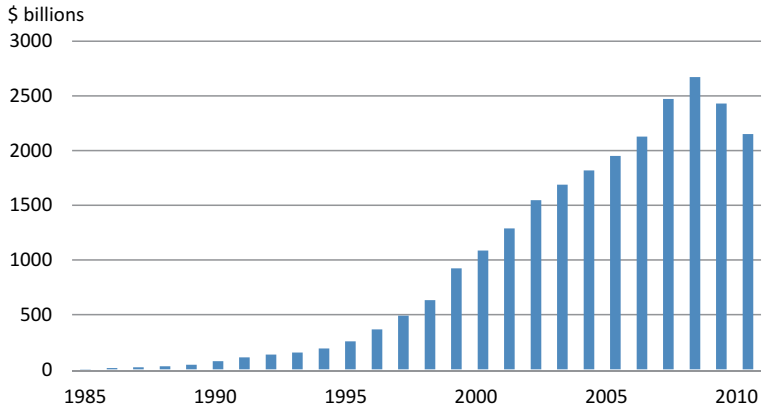
Bank Loans as a Share of Corporate Debt



Source: Flow of Funds, Federal Reserve Board. Total corporate debt is the sum of commercial paper, corporate bonds, and bank loans

FIGURE 2

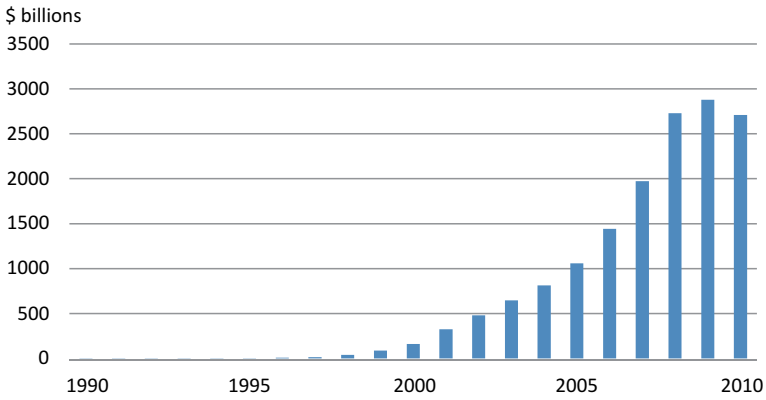
Total U.S. Asset-Backed Securities Outstanding



Source: Gorton and Metrick

FIGURE 3

Total European Asset-Backed Securities Outstanding



Source: Gorton and Metrick

terms *bank loans* and *bonds*, *banks* and *markets*, in order to simplify a complicated world. Intermediaries such as finance companies, insurance companies, and even some hedge funds may act much like commercial banks if they hold a large share of a firm's debt, even though they are not funded by deposits.[2] However, it will sometimes be important to think about banks more narrowly as deposit-taking firms. I use the term *bonds* to refer to widely held securities — including securitized loans — that may be held in households' portfolios but may also be held (and traded) by various types of intermediaries, including commercial banks. Thus, when a commercial bank originates credit card loans that are packaged into asset-backed securities and actively traded by the bank's trading subsidiary, I will classify these as market activities, not banking activities.

Banking economists have viewed banks as specialists in producing information about borrowers before the loan is made (*screening*) and monitoring their activities closely until the loan is repaid. For example, a banker will examine a borrowing firm's books to forecast future earnings growth, visit the firm's factory to examine the quality of the firm's receivables, and even talk to the firm's customers to make judgments about the firm's ability to pay. There is substantial empirical evidence for this view of banks, but the view that banks monitor firms while markets do not is too stark. Better said, banks and markets use different technologies for screening and monitoring borrowers.

Banks Monitor Firms Using Covenants. Business loans made by banks typically include covenants, a fundamental tool in bank lending. Broadly, covenants come in two varieties. Some covenants place direct

[2] Debt held by a small number of lenders is often called *private debt*.

restrictions on firm's activities, for example, restrictions on large new investments by the firm without the bank's approval. The second type requires the firm to maintain various measures of financial health and the ability to pay, for example, a minimum net worth ratio (the ratio of equity to total assets) or a minimum ratio of short-term to total assets.

A key feature of bank loan covenants is that they are set tightly and renegotiated frequently.[3] In their sample of bank loans, Ilia Dichev and Douglas Skinner examine two covenants frequently included in loan contracts and show that most firms maintain financial ratios just above the level that would put the firm in default; indeed, most firms are just in compliance when the contract is signed.[4]

The flip side of tight covenants is that it is easier for a single lender to renegotiate loan terms with a borrower than it is for widely dispersed bondholders. In his working paper, Michael Roberts found that loan contracts were renegotiated about once a year. For the most part, firms renegotiating contracts are not financially distressed, although Roberts and Amir Sufi found that covenant violations were most common in difficult economic environments. Over the life of the loan contract, the firm's business environment changes and contracts are adjusted to meet new realities — but only after the bank takes a close look into the firm's financial health.

Ease of renegotiation doesn't mean that every default is cured through renegotiation or that the terms on which loans are renegotiated are typically easy for the firm. Covenant violations lead to real constraints on the firm's behavior; the finding that the mass of firms are just in compliance provides indirect evidence that firms would be operating at lower liquidity or net worth levels if they were not constrained by covenant restrictions. More directly, Sufi finds that, following a covenant violation, both the used and unused portions of a firm's line of credit are typically reduced by between 15 and 25 percent, while Sudheer Chava and Roberts find that real investment declines by 13 percent.[5]

You might understand why a bank would like to keep a tight rein on borrowers; after all, a firm with high net worth and liquid assets is more likely to pay back the loan. But why would a firm accept such restrictions, and what types of firms would choose to use bank loans with tight covenants? From the firm's point of view, tight covenants may be attractive because the bank can profitably lend at a lower loan rate when the bank is better protected against loss. Without tight restrictions, many small firms and risky firms would simply find any outside funding to be too expensive. And we will examine in some detail the reasons why many larger firms will prefer to borrow using a mixture of bank loans and bonds; broadly, the reason is that a mix of bank loans and bonds often lowers the firm's total borrowing costs. But for low-risk firms that can afford the costs of borrowing on public debt markets, some mixture of short- and long-term bonds and internally generated funds may be preferable to the tight covenants and intrusive monitoring typical of bank lending.[6] Furthermore, much of the banking industry is regulated and regulatory costs are ultimately passed onto banks' customers, including borrowers. To avoid these costs, all firms have an incentive to limit their borrowing from banks.

Although I have focused here on covenants and renegotiation, researchers have also highlighted repeated lending between a single bank and borrower, a *lending relationship*, as a distinctive feature of bank lending. In a lending relationship, banks build up information about the borrowing firm over time. In addition, researchers have found evidence that banks use firms' deposit accounts as a mechanism for banks to monitor borrowing firms.[7]

A key feature of bank loan covenants is that they are set tightly and renegotiated frequently.

Markets Monitor by Aggregating Investors' Information. Nobel laureate Friedrich von Hayek first proposed the idea that market prices incorporate the information of market participants and, thus, provide guideposts to making economic decisions: buy, sell, invest. In financial markets, mutual fund managers, hedge fund managers, and other investors buy and sell stocks, bonds, and derivative securities based on their own research and the research

[3] My article with Loretta Mester formalizes this view of bank lending

[4] Dichev and Skinner focus on the current ratio (short-term debt over total assets) and net worth ratio because they are common and relatively standardized across loan contracts.

[5] These findings understate the constraints covenants impose on firms because they don't include the costs of the decisions the firm took to avoid breaching or renegotiating a contract. For example, firms may forgo a profitable investment in preference to seeking a change in its loan contract.

[6] However, even large firms that seldom borrow from banks retain backup lines of credit with banks to call on when financial markets are tight.

[7] Degryse and coauthors review the literature on lending relationships, and Loretta Mester and coauthors, among others, provide empirical evidence for the monitoring role of deposits.

of information specialists such as ratings agencies and industry analysts. Stated somewhat simplistically, based on their research, investors seek to buy securities that they believe will rise in price and to sell those securities that they believe will decline in price. Securities prices rise and fall accordingly.

But how does all this buying and selling affect firms' real decisions? The market for corporate control is one channel. For example, Alon Brav and coauthors have recently examined the role of hedge funds in the market for corporate control between 2001 and 2006. Some hedge funds specialize in buying up the securities (stocks or bonds) of underperforming firms and using their financial stake to put pressure on the firms' managers or to get rid of current management. The fund's investors gain if a firm's performance improves and its stock or bonds increase in value. Indeed, Brav and coauthors find that just the announcement of a hedge fund's intent to play an active role increases a firm's stock price, on average, and these gains are not reversed. Alex Edmans and coauthors provide evidence that a decline in a firm's stock price significantly increases the likelihood of a takeover attempt.

A second channel is the direct effect of market prices on management decisions. A growing body of evidence shows that managers' investment decisions are affected by the firm's stock price. Furthermore, managers' investment decisions appear to be improved when stock prices are more informative.[8]

How Does the Shift from Banks to Markets Affect Information Production? To date, researchers have only started to think about the implications for the larger economy of changes in the information environment when activity shifts from banks to markets.[9] For example, Christine Parlour and Guillaume Plantin demonstrate that the option to securitize assets may inefficiently reduce banks' information production about borrowers' financial health. This happens if the bank has large cost savings from shifting assets off its books. To see why, imagine the bank did produce information about borrowers, hoping to sell well-performing loans at a higher price by certifying when a loan is healthy. But rational buyers will be suspicious of the bank's claims and demand a large discount in the fear that the bank was selling them a *lemon*, a troubled loan being passed off as a healthy one. Thus, producing information about the firm will not be profitable for the bank. When the gains from moving loans off the bank's books are large, the loans will be sold, but only at a price so low that the bank can't profitably produce information. And since no information is produced, neither the bank nor the buyer of the loan knows whether the loan is healthy or troubled.[10]

From another perspective, Vincent Glode and coauthors argue that markets may also provide incentives to produce too much information. They show that fund managers who make their living trading securities produce information to gain a bargaining advantage over other traders. In their model, much of the research simply affects the distribution of gains between traders: what one trader (and his investors) gain and another trader (and his investors) lose. The information doesn't increase the total profits shared by investors, only the distribution of these profits.

To date, researchers have only started to think about the implications for the larger economy of changes in the information environment when activity shifts from banks to markets.

Insights such as these are a starting point for developing a deeper understanding of how incentives to produce information change with an evolving mix of financial activities carried out through banks and through markets.

COMPETITION UNDERMINES INTERMEDIATION

Competition in financial markets increased dramatically in the last quarter of the 20th century, in significant part due to deregulation, with banks facing increased competition on both sides of their balance sheets.[11] (See *Deregulation and Competition*.) Smaller and riskier firms that could only have borrowed from banks in the past could now borrow directly on bond and stock markets. One indicator of this trend is the decline in the age of firms going public. Between 1970 and 2000, the

[8] See Yaron Leitner's *Business Review* article for an accessible account of the theory and evidence on the effects of market prices on managerial decisions.

[9] Arnoud Boot and Anjan Thakor's and Fenghua Song and Thakor's articles are notable exceptions. Both articles contain models in which banks and markets coexist. My distinction between close monitoring and aggregating information follows theirs.

[10] More formally, Parlour and Plantin show that when the gains from selling are large, the only equilibrium is a *pooling* equilibrium without information production.

[11] I focus here on competition from financial markets rather than competition between banks. More competition between banks has much the same effect as competition from financial markets.

Deregulation and Competition

On the asset side, the deregulation of underwriting fees in 1973 and commercial bank entry into investment banking made it cheaper for firms, especially smaller firms and riskier firms, to gain access to public debt and stock markets.* The securitization of mortgages was largely the result of the collapse of the savings and loan industry in the 1980s, which was primarily driven by the deregulation of deposit rates in 1980. This technology was then adapted to a wide range of loans, providing access to securities markets to a whole new range of borrowers, mainly households. Finally, the dismantling of barriers first to intrastate and then to interstate banking increased competition between banks for borrowers' business. On the liability side, competition from money market funds — beginning in the 1970s — increased households' access to financial markets. While money market funds did not develop strictly because of deregulation, they were an innovation that was largely driven by regulatory arbitrage; money market funds could hold commercial paper without the capital requirements that were first imposed by regulators on banks in the 1980s.

* In addition to these regulatory changes, Michael Milkin's recognition that portfolios of junk bonds would yield predictable returns expanded high-risk firms' access to public debt markets. His discovery may be thought of as a "technological" advance in financial markets.

median age of a firm undertaking an *initial public offering* — selling stock to the public for the first time — declined from around 40 years to five years, with the most dramatic decline in the 1970s following the deregulation of underwriting fees.[12] Household borrowers also gained access to securities markets via securitized mortgages and credit card loans; these assets were increasingly moved off banks' balance sheets. On banks' liability side, depositors could now choose to invest their savings in securities through a wide range of intermediaries that held securities instead of loans, for example, mutual funds or hedge funds.

As a general rule, competition lowers fees and increases the variety and availability of financial services. But some of the distinctive services provided by banks depend on cross-subsidies among bank customers. Cross-subsidization is feasible only when banks have market power over their customers.

Firms' Access to Markets Undermines Lending Relationships. Financial economists have found convincing evidence that firms in a long-term lending relationship with a bank are less likely to be required to post collateral and less likely to be denied loans. In essence, banks make loans to young firms and risky firms that are profitable only if the firm sticks with the bank and pays higher than purely competitive loan rates in the future. So, in a bank loan portfolio, the profits from older and safer firms subsidize the loans to younger and riskier firms.

This works only as long as the bank has some market power over older and safer borrowers. If it is easy and cheap for a firm to go public and to sell securities, the bank can't charge the firm a high loan rate or maintain its accustomed level of control over the firm's activities, and the scope for such cross-subsidies decreases. Supporting this view of the decline in banks' market power over firms with access to public markets, Carola Schenone shows that the rate a firm pays on its bank loan declines when the firm goes public.

In addition to losing older and safer borrowers to bond markets, banks' more limited ability to cross-subsidize across borrowers means that bank loans to younger and riskier borrowers become increasingly *arm's length*, in the language of the banking literature. Essentially, this terms means that the bank screens the borrower when it makes the loan but does not renegotiate loan terms or provide temporarily concessionary rates if the firm is in trouble. In turn, younger and riskier firms find that borrowing exclusively from a bank becomes relatively less attractive compared to selling bonds.

Banks Provide Less Liquidity When Households Have Access to Financial Markets. One of the traditional roles of banks is to allow households to put their money in checking or savings accounts and allow them to withdraw their money on demand. In their classic article, Douglas Diamond and Philip Dybvig demonstrate how a bank can do so even while holding a portfolio of mainly illiquid assets (e.g., loans), which have a higher yield than liquid assets such as cash. Diamond and Dybvig assume that investors have no alternative to putting their funds in the bank, a relatively accurate picture of the real world until the 1980s. But what happens when some households have the alternative of investing directly in securities markets? In his follow-up article, Diamond explicitly considers the effect of households' ac-

[12] These numbers are from the article by Jason Fink and coauthors. The median age increased to 12 years by 2006, suggesting that market participants reacted to the excesses of the Internet boom of the late 1990s by demanding more seasoning before a firm could go public.

[13] In his model, he views households as if they were trading for themselves, but you can just as easily think of them as customers who can shift their savings from a deposit to a mutual fund or a hedge fund and have a manager trade on their behalf.

cess to financial markets.[13]

In Diamond and Dybvig's model, the feasibility of the banking arrangement depends on a cross-subsidy among depositors. Some households find that they need funds right away — they face a *liquidity shock* — while others have no immediate need for funds. As long as households have no alternative to the bank, the bank can promise households access to their funds on demand with only a small penalty. But this is only possible if households that don't need their funds will accept a lower rate than they could get in the market; that is, they are subsidizing the households that withdraw funds.

As long as households are concerned that they may need their funds at short notice — and as long as only a fraction of households need to withdraw funds at any time — this arrangement is attractive to all households. Most households would prefer to avoid being penalized whenever a pressing need for funds arises, and they would be willing to give up some return for this assurance. You can think of the bank as a type of insurance company that provides insurance against liquidity shocks.[14]

Things change when some households have direct access to securities markets. Since it is unrealistic to think that a bank can really tell why a depositor needs to make a withdrawal, the deposit rate has to be the same for all households. This means two things: (i) any subsidy paid to households that withdraw funds to, say, make a mortgage payment must also be paid to those who withdraw their funds to trade in the market; and (ii) only households without access to securities markets can be the source of the subsidy. So as more customers have easy access to securities markets, the interest rate the bank can offer to households with immediate liquidity needs decreases, and the liquidity insurance offered by the bank becomes less valuable. In turn, even more activity shifts from banks to markets.

BANKS AND MARKETS ARE ALSO COMPLEMENTARY

Competition tends to make bank services less unique and to shift activities from banks to markets. But this doesn't mean that the banking sector will shrink until banks become niche providers, serving only very small firms and the most cautious and unsophisticated households. First, greater competition doesn't mean that market power disappears completely. Furthermore, not all of the services provided by banks depend on monopoly power and cross-subsidies. Perhaps most important, bank loans and bonds are *complements*.[15]

Competition tends to make bank services less unique and to shift activities from banks to markets.

Firms' Optimal Financing Mix Includes Bank Debt.[16] Since the 1960s, financial economists have made a huge effort to understand firms' capital structure, that is, how much equity and how much debt were chosen by firms and why. Beginning in the 1990s, theorists began to think more carefully about the composition of firms' debt, e.g., short-term versus long-term debt, bank debt versus public debt. More recently, empirical financial economists have explored the structure of debt contracts in much more detail.

Consider a firm that is large enough to borrow in bond markets; in principle, at least, the firm could avoid borrowing from a bank altogether and thus avoid the bank's monitoring. Indeed, the firm would gain maximum flexibility by selling long-term bonds, let's say 30-year bonds. But would this be the cheapest way for the firm to borrow? Sensible bondholders will be concerned that a lot can change in 30 years. The firm's markets may dry up, or new managers with a taste for high risk or costly empire building may replace current management. The firm may have to pay quite a high rate of interest to convince bondholders to accept these types of risks, or there might not be a rate high enough to convince them.

One possible alternative for the firm is to split its borrowings into short-term debt (commercial paper) and long-term bonds. In this case, the firm will have to prove that its finances are healthy by paying off its short-term debt on a regular basis. And if bondholders are no longer convinced that the firm's prospects are good, short-term investors can pull the plug and the firm will be forced to scramble for funds. Thus, short-term debt may serve as a disciplinary device that, in turn, facilitates borrowing for a longer term. While this debt structure is feasible for low-risk firms with an impeccable reputation, it poses problems for riskier firms.[17]

[14] In addition to providing customers with more liquidity, the bank also changes the mix of investments in the economy. Specifically, the bank holds a portfolio with a larger fraction of illiquid (but high-yielding) investments than individuals could hold in their own portfolios. Without the bank, individuals would have to hold lots of low-yield liquid investments (cash in mattresses) to self-insure against liquidity shocks.

[15] Broadly, two products are complements when the cost of producing (or using) one good lowers the cost of producing (or using) the other.

[16] The theoretical description in this section includes insights from articles by Eric Berglof and Ernst Ludwig von Thadden and by Cheol Park.

[17] For very low-risk firms, the disciplinary role of short-term debt is probably a secondary matter. For such firms, short-term borrowings are simply a convenient way to finance working capital.

Let's take a firm with a significant chance of default. In fact, let's consider a firm that is unable to pay off its short-term creditors because of financial difficulties. Crucially, a firm facing financial problems is often worth a lot more alive than dead; simply auctioning off the firm's assets inside or outside bankruptcy proceedings would fetch a lower price than the firm is worth as a going concern. With the prospect of a looming default, bondholders will have a powerful incentive to agree if the firm's managers propose the following deal: Accept new claims that pay less than the original contractual return but not much more than they would have received by auctioning off the firm's assets and sharing the proceeds. As long as all creditors have the same priority, that is, each of the firm's creditors has a pro rata claim on the firm's assets in the event of default, they would unanimously agree to this deal.

While a restructuring to avoid default is often better for both bondholders and managers, it is not hard to see that the possibility of renegotiation undermines the threat to impose default, so short-term debt doesn't have as much disciplinary power as it appeared on first sight. Most worrisome, if managers know that bondholders will renegotiate, they may take more risks or build empires, and we are back where we started: high borrowing costs.[18]

Short-Term Bank Debt Is a Hard Claim. The threat to impose default can be made more believable if the short-term creditor has priority over other creditors because the creditor with priority captures more than its pro rata share of the auction value of the firm's assets. (For these purposes, having a collateralized claim serves much the same purpose.) Since it gets a disproportionately higher payoff in default, the creditor with priority will be a hard bargainer; economists would say that he or she holds a *hard claim*. Even if the threat to impose default is never actually carried out, discipline is improved because the firm's managers know that default will be painful. And even though the short-term creditor gains at the expense of other creditors in negotiations, the firm's long-term creditors benefit from the discipline imposed by the hard claim.[19] Furthermore, negotiations will be more efficient if the holder of the short-term claim has the capacity to closely examine the firm's financial condition. While holding a hard claim is valuable to discipline managers, liquidating fundamentally sound firms or mistakenly relaxing contract terms for genuinely troubled firms makes everyone worse off.[20]

In summary, risky firms with access to bond markets will borrow through a mixture of subordinated long-term debt and higher priority short-term debt (or debt with stringent covenants) held by a creditor that can monitor the firm closely.[21] That is, for all but the safest firms, bank debt is part of the debt mix that reduces borrowing costs. For that matter, a firm's access to bond markets may depend on the role played by the bank.

Evidence for the Value of Hard Claims. Recent research has provided support for the role of hard claims in risky firms' debt structure. Mark Carey and Michael Gordy examine a large sample of firms that entered bankruptcy and ask which firms have a larger *recovery value*, that is, which firms ultimately paid creditors the largest amount per dollar invested when the firm's assets were liquidated. Carey and Gordy find that recovery values are higher for firms with a higher share of bank debt and that other factors are of secondary importance. In their terminology, banks are "grim reapers." Banks discover financial troubles early and their interventions prevent managers from imposing greater losses on creditors.[22]

Joshua Rauh and Amir Sufi examine a sample of *fallen angels*, firms that experience a dramatic drop in their credit rating from investment grade

While holding a hard claim is valuable to discipline managers, liquidating fundamentally sound firms or mistakenly relaxing contract terms for genuinely troubled firms makes everyone worse off.

[18] Making it very hard to renegotiate would improve discipline. But the option to renegotiate is also valuable for firms in risky environments. Furthermore, if the firm can choose to enter Chapter 11 bankruptcy proceedings, the bondholders get bargaining plus bankruptcy lawyers!

[19] Long-term creditors will also insist on receiving an interest rate that compensates them for the likelihood that the short-term creditor does too well at their expense in contract negotiations.

[20] Note that while my account focuses on the disciplinary role of short-term debt, longer-term debt with strict covenants and with priority over other long-term creditors has a similar disciplinary effect.

[21] Subordinated bondholders receive a payoff only after other debt holders have been paid in full. Thus, subordinated debt holders have higher priority than stockholders but lower priority than other bondholders.

[22] This evidence doesn't imply that bankruptcy was an efficient outcome, only that the threat to impose default was effective and that the decision to liquidate was informed, in the sense that creditors tended to gain after liquidation.

to junk status, which means that their risk of default increases substantially. They find that these firms were originally funded primarily by unsecured debt and equity, but after the collapse in their credit rating, they shifted toward a mix of secured bank debt and unsecured and subordinated long-term bonds. Similarly, comparing a sample of low-risk and high-risk public firms, Rauh and Sufi found the same pattern; low-risk firms secure funds mainly using equity and unsecured debt, while high-risk firms borrow through a mixture of short-term, secured bank loans and subordinated long-term debt.[23]

Deposits and Lines of Credit Are Complements. In their article, Anil Kashyap and coauthors argue that because banks are funded by deposits, they have a cost advantage in providing lines of credit. As long as depositors' and firms' demands for funds are not perfectly correlated, that is, as long as borrowing firms don't always borrow under their line of credit at the same time that depositors withdraw their funds, banks can meet all commitments while holding a relatively small amount of (low-yield) cash balances. Furthermore, if firms' and depositors' demands for liquidity are negatively correlated — meaning that firms borrow at times when savers are holding more of their savings in deposit accounts — the cost complementarity is even stronger.

Evan Gatev and coauthors provide empirical evidence for this strong type of cost complementarity. They focus on periods in which stress in money markets restricts many firms' access to the commercial paper market. At times of stress, firms borrow on their lines of credit. In stressful times, funds on deposit with banks also increase. Funds flow into the banking system, probably because firms and households view banks as safe places to put their savings in a financial storm.

Lines of Credit Are Part of a Firm's Optimal Financing Mix. Recently, a number of economists have modeled firms' financing decisions as if they were part of an *optimal long-term contract*. The approach in this research is to figure out what the best long-term financing contract would look like — including the pattern of loan payments and the conditions under which the firm is placed in default — and then to ask whether some mix of securities could reproduce the terms of this contract. Interestingly, in Peter DeMarzo and Michael Fishman's model, the terms of the optimal long-term contract can be mimicked by a financing mix of equity, long-term debt, and a line of credit.

Broadly, the optimal long-term contract is designed to solve two types of problems. First, borrowing is rife with conflicting incentives: Borrowing firms have incentives to take too much risk, to cover up problems until it is too late, or to consume excessive perks. Uncontrolled, these conflicts would increase default risk and raise borrowing costs (or even make financing infeasible altogether), so financial contracts are designed to control incentive problems. Second, firms operate in an intrinsically risky business environment. Even when a firm's managers are making cautious and thrifty decisions, the firm's cash flows are variable and uncertain. An optimal long-term contract must impose discipline on managers without closing down the firm every time it suffers a setback.

I have argued that bank loans' mixture of tight control and renegotiation is one solution to these problems, but it is not the only one. The mix of securities the firm uses to finance operations is another solution. In DeMarzo and Fishman's model, the borrowing firm has an incentive to use cash flows to consume perks, and the lender can't directly observe the firm's cash flows or how the firm is using its cash flows. The authors show that the optimal financing mix is a combination of equity, long-term debt, and a line of credit, a combination that looks a lot like the mix of contracts used by many real world firms. The long-term debt forces the firm to make some payments to the lender, but because principal is paid back later, the firm has more flexibility to pay workers, suppliers, etc. The line of credit provides even more flexibility in the event of temporary setbacks; the firm can draw down the line of credit to cover long-term debt payments and to meet operating costs even when the business environment is tough. In addition to the long-term debt payments, discipline is imposed on the firm in two ways. No payments can be made to the firm's stockholders unless the firm stays current on all debt payments. Furthermore, if the firm can't make payments on its credit line, the lender imposes default.

While depository institutions have a cost advantage in providing lines of credit, they have no such advantage in holding long-term debt in their portfolio. Furthermore, we have already seen

While depository institutions have a cost advantage in providing lines of credit, they have no such advantage in holding long-term debt in their portfolio.

[23] I am simplifying their results slightly. The higher risk firms also include some senior unsecured debt and convertible bonds in their debt structures. This slightly complicates but doesn't contradict my interpretation of their evidence.

that there are benefits from having separate investors hold the firm's short- and long-term debt. Consistent with both theory and real world practice, firms with access to public debt markets borrow through a mixture of bank loans — here, loans borrowed under a line of credit — and public bonds.

CONCLUSION

Banks and markets interact in a number of ways. Firms and households view banks and markets as substitute ways to borrow funds and to hold their savings. Many of the distinctive features of banking services are based on cross-subsidies among the bank's customers, but these are only feasible if banks retain market power over borrowers and depositors. Thus, increasing competition in financial markets — driven primarily by deregulation in the last quarter of the 20th century — tends to undermine the profitability of banks and to increase the share of activities carried out through financial markets. But there are limits to how far the banking sector can shrink because banks and markets are also complementary. Many firms, not just those too small to access bond markets, lower borrowing costs using a mix of financial contracts, including bank loans. In particular, banks retain a comparative advantage in providing lines of credit because they provide deposits.

There is a wealth of evidence that the mix of bank loans and bonds has real effects at the firm level — for example, a heavier reliance on bank loans increases the recovery rates for firms that enter bankruptcy — but the evidence that the mix of banks and securities markets matters at the macroeconomic level is much weaker. Ross Levine's comprehensive review of the evidence concludes that while financial development has a significant role in promoting economic growth, there is not much evidence that the relative scale of the activities carried out through banks or through markets has a large effect on a country's economic growth.

That said, the recent financial crisis and the ensuing deep recession are likely to force economists to revisit and rethink the evidence about economic performance in the last few decades, a period that witnessed the rapid growth of financial markets, especially the growth of securitized assets. Some analysts view the heavy shift toward securitized markets in the U.S. as a major cause of the crisis.[24] At the same time, economies dominated by banks, for example, Spain and Ireland, also experienced a lending boom and an attendant bust. It will take some time for us to absorb the lessons of the financial crisis and to determine whether it provides any lessons about the mix of banks and markets going forward.

[24] See Ronel Elul's *Business Review* article for a review of the evidence about securitization and mortgage default.

REFERENCES

Berglof, Eric, and Ernst Ludwig von Thadden. "Short-Term Versus Long-Term Interests: Capital Structure with Multiple Investors," *Quarterly Journal of Economics*, 109 (1994), pp. 1055-84.

Berlin, Mitchell, and Loretta Mester. "Debt Covenants and Renegotiation," *Journal of Financial Intermediation*, 2 (1992), pp. 95-133.

Boot, Arnoud, and Anjan Thakor. "Financial System Architecture," *Review of Financial Studies*, 10 (Fall 1997), pp. 693-733.

Brav, Alon, Wei Jiang, Randall Thomas, and Frank Partnoy. "Hedge Fund Activism, Corporate Governance, and Firm Performance," *Journal of Finance*, 63, 2008, pp. 1729-75.

Carey, Mark, and Michael Gordy. "The Bank as Grim Reaper: Debt Composition and Bankruptcy Thresholds," Board of Governors of the Federal Reserve System Working Paper (May 4, 2009).

Chava, Sudheer, and Michael Roberts. "How Does Financing Impact Investment? The Role of Debt Covenants," *Journal of Finance*, 63 (2008), pp. 2095-2121.

Degryse, Hans, Moshe Kim, and Steven Ongena. *Microeconometrics of Banking: Methods, Applications and Results*. Oxford University Press, 2009.

DeMarzo, Peter, and Michael Fishman. "Optimal Long-Term Financial Contracting," *Review of Financial Studies*, 20 (2007), pp. 2079-2128.

Diamond, Douglas. "Liquidity, Banks and Markets," *Journal of Political Economy*, 105 (1997), pp. 928-56.

Diamond, Douglas, and Philip Dybvig. "Bank Runs, Deposit Insurance, and Liquidity," *Journal of Political Economy*, 91 (1983), pp. 401-19.

Dichev, Ilia, and Douglas Skinner. "Large-Sample Evidence on the Debt Covenant Hypothesis," *Journal of Accounting Research* (September 2002), pp. 1091-1123.

Edmans, Alex, Itay Goldstein, and Wei Jiang. "The Real Effects of Financial Markets: The Impact of Prices on Takeovers," *Journal of Finance*, forthcoming.

Elul, Ronel. "What Have We Learned About Mortgage Default?," Federal Reserve Bank of Philadelphia *Business Review* (Fourth Quarter 2010), pp. 12-19.

Fink, Jason, Kristin Fink, Gustavo Grullon, and James Weston. "What Drove the Increase in Idiosyncratic Volatility During the Internet Boom?," *Journal of Financial and Quantitative Analysis,* 45 (2010), pp. 1252-78.

Gatev, Evan, Til Schuermann, and Philip Strahan. "Managing Bank Liquidity Risk: How Deposit-Loan Synergies Vary with Market Conditions," *Review of Financial Studies*, 22 (2009), pp. 995-1020.

Glode, Vincent, Richard Green, and Richard Lowery. "Financial Expertise as an Arms Race," *Journal of Finance*, forthcoming.

Gorton, Gary, and Andrew Metrick. "Securitization," in George Constantinides, Milton Harris, and Rene Stulz, eds., *Handbook of the Economics of Finance,* forthcoming.

Kashyap, Anil, Raghuram Rajan, and Jeremy Stein. "Banks as Liquidity Providers: An Explanation for the Coexistence of Lending and Deposit Taking," *Journal of Finance*, 58 (2002), pp. 33-73.

Leitner, Yaron. "Stock Prices and Business Investment," Federal Reserve Bank of Philadelphia *Business Review* (Fourth Quarter 2007), pp. 12-18.

Levine, Ross. "Finance and Growth: Theory and Evidence," in Philippe Aghion and Steven Durlauf, eds., *Handbook of Economic Growth.* The Netherlands: Elsevier Science, 2005.

Mester, Loretta, Leonard Nakamura, and Marie Renault. "Transactions Accounts and Loan Monitoring," *Review of Financial Studies*, 20 (2007), pp. 529-56.

Park, Cheol. "Monitoring and the Structure of Debt Contracts," *Journal of Finance,* 55 (2000), pp. 2157-95.

Parlour, Christine, and Guillaume Plantin. "Loan Sales and Relationship Banking," *Journal of Finance,* 63 (2008), pp. 1291-1314.

Rauh, Joshua, and Amir Sufi. "Capital Structure and Debt Structure," *Review of Financial Studies*, 12 (2010), pp. 4242-80.

Roberts, Michael. "The Role of Dynamic Renegotiation and Asymmetric Information in Financial Contracting," University of Pennsylvania Working Paper (December 29, 2010).

Roberts, Michael, and Amir Sufi. "Renegotiation of Financial Contracts: Evidence from Private Credit Agreements," *Journal of Financial Economics*, 93 (2009), pp. 159-84.

Schenone, Carola. "Lending Relationships and Informational Rents: Do Banks Exploit their Information Advantages?," *Review of Financial Studies*, 23 (2010), pp. 1149-99.

Song, Fenghua, and Anjan Thakor. "Financial System Architecture and the Coevolution of Banks and Markets," *Economic Journal*, 120 (2010), pp. 1021-55.

Sufi, Amir. "Bank Lines of Credit in Corporate Finance: An Empirical Analysis," *Review of Financial Studies*, 22 (2009), pp. 1057-88.

Contingent Capital*

BY YARON LEITNER

Government bailouts during the recent financial crisis were controversial because of the burden on taxpayers and because even if taxpayers eventually get their money back, such bailouts can undermine banks' incentives not to take excessive risk in the future. New regulatory reforms aim to avoid such crises in the future. One proposal is to require banks to hold "contingent capital." In this article, Yaron Leitner explains what contingent capital is and discusses some of the arguments in favor of it. He also discusses potential implementation problems and looks at some of the alternatives.

The recent financial crisis has illustrated the problems that can be caused by a failure of a large financial institution and the government's reluctance to let it fail. These government interventions, or bailouts, have been controversial because of the burden they impose on taxpayers and because even if taxpayers eventually get their money back, such bailouts can undermine banks' incentives not to take excessive risk in the future. New regulatory reforms aim to avoid such crises in the future. One proposal is to require banks to hold "contingent capital."

Indeed, the Dodd-Frank Wall Street Reform and Consumer Protection Act, passed by Congress on July 21, 2010, allows the Federal Reserve to require large banks and other financial firms supervised by the Fed to "maintain a minimum amount of contingent capital that is convertible to equity in times of financial stress." However, this can be mandated only after a study by the Financial Stability Oversight Council to be completed by June 2012.[1] Regulators in several other countries have also shown interest in adding contingent capital to their supervisory toolkit to improve crisis management.[2]

Yaron Leitner is a senior economist in the Research Department of the Philadelphia Fed. This article is available free of charge at www.philadelphiafed.org/research-and-data/publications.

WHAT IS CONTINGENT CAPITAL?

Before explaining what contingent capital is, it is useful to say what we mean by capital. *Bank capital* is the value of the bank's assets minus the value of its liabilities (its debt). Alternatively, this is what the bank's shareholders own, or their equity.[3] Examples of banks' assets are loans that banks make to households and firms and financial securities that banks hold, such as government bonds. Examples of banks' liabilities are the amounts of money that banks obtain by borrowing or by taking deposits from households and firms. Essentially, banks earn interest on their assets and pay interest on their debt.[4, 5]

[1] The Financial Stability Oversight Council was established by the Dodd-Frank Act to identify threats to the financial stability of the United States, promote market discipline, and respond to emerging risks to the stability of the U.S. financial system. The act contains details about the council's organizational design (members, meetings), duties, and authority.

*The views expressed here are those of the author and do not necessarily represent the views of the Federal Reserve Bank of Philadelphia or the Federal Reserve System.

[2] See, for example, the recent regulatory proposal (July 20, 2011) by the European Commission. More details are available at http://ec.europa.eu/internal_market/bank/regcapital/index_en.htm.

[3] Regulators may define capital a bit differently to account for the fact that, in practice, banks hold complex securities other than simple debt and equity (see *Bank Capital Regulation* in the *Business Review* article by Mitchell Berlin). But, to simplify, we will use the simple definition in the text.

[4] Throughout the article, we use the word banks, but the article also applies to other financial firms that might pose systemic concerns.

[5] People sometimes confuse capital requirements and liquidity requirements. The terms "capital," "capital requirements," and "capital structure" refer to the way the bank is funded and, in particular, to the mix between debt and equity. In contrast, the term "liquidity requirements" refers to the type of assets and the asset mix the bank holds. For example, if the bank has a lot of cash and Treasury securities, it is considered

Contingent capital refers to debt that automatically converts into equity in times of financial stress if certain prespecified triggers are hit. For example, in November 2009, Lloyds Bank issued a bond that converts into common stock if the bank's tier 1 capital ratio falls below 5 percent. Tier 1 capital is a measure of a bank's capital used by regulators.[6] The tier 1 *capital ratio* is the value of the bank's tier 1 capital divided by the risk-weighted value of the bank's assets.[7] It is a measure of the bank's financial health and its ability to absorb losses. When the bank's capital ratio is high, a significant loss can be absorbed by the bank's shareholders and does not trigger bankruptcy. In contrast, when the ratio is low, losses may trigger bankruptcy, since the bank may not be able to pay off its debt or make the required interest payments.

Figure 1 illustrates how contingent capital works. The numbers are for illustration purposes only and are chosen so that the algebra is simple. Suppose that initially (Panel A) the value of the bank's assets is $10 billion and its capital ratio (equity divided by assets) is 30 percent. The bank also has $2 billion of contingent capital. Suppose that the trigger for conversion is a 15 percent capital ratio; that is, conversion is automatic whenever the capital ratio falls below 15 percent. If this trigger is hit, $1 of the face value of contingent capital converts to $1 of common stock (i.e., the conversion ratio is that $1 of contingent capital converts to $1 of equity). Now suppose the bank suffers a loss and the value of its assets drops to $8 billion. The loss is absorbed by the banks' equity holders, and the bank's capital ratio falls to 12.5 percent, which is below the trigger (Panel B). Since the trigger is hit, the $2 billion of contingent capital converts to $2 billion of equity. The bank's capital ratio then rises to 37.5 percent (Panel C) and has returned to a safe level.

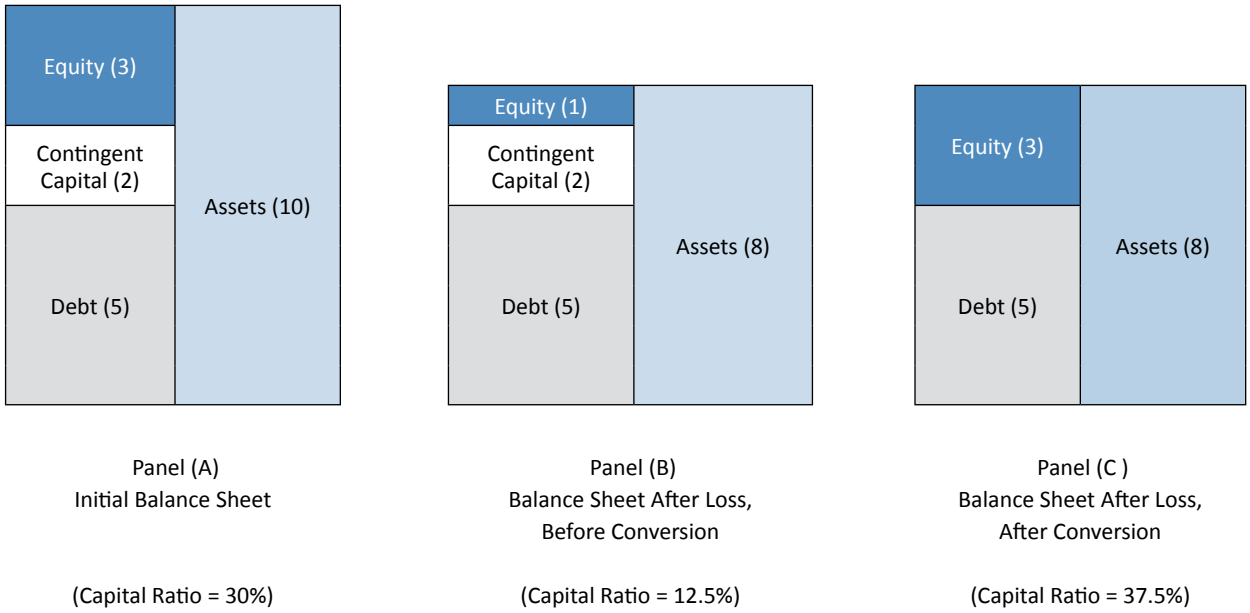
very liquid. Otherwise, the bank is considered less liquid because even though the bank may own a lot of assets, it may not be able to sell them at short notice, or it may obtain less than the fair value in a sale.

[6] Tier 1 capital consists mainly of the bank's common stock and retained earnings, but it may also include more complex securities, such as preferred equity, which is a special type of equity that is senior to common stock but subordinate to bonds.

[7] For a definition of risk-weighted assets, see *Bank Capital Regulation* in the *Business Review* article by Mitchell Berlin.

FIGURE 1

How Contingent Capital Works



Panel A shows the balance sheet of a bank that has contingent capital in its capital structure. (Numbers in brackets represent billions of dollars.) Panel B shows the balance sheet of the same bank after it suffers a loss. As you can see, the bank's capital ratio (equity over total assets) falls drastically, and so the trigger for conversion occurs. Panel C shows the balance sheet of the bank after conversion occurs. Now the bank's capital ratio is back to a "safe" level.

ARGUMENTS IN FAVOR OF CONTINGENT CAPITAL

When a large bank fails, its failure can spread to other banks in a domino effect, which economists call contagion. Regulators may then be forced to bail out the bank, using taxpayers' money, because of the potential damage a single bank's failure can do to the banking system and to the whole economy. Even if taxpayers eventually get their money back, bailouts have a social cost because they may induce banks to take excessive risks, i.e., risks that benefit the bank's shareholders but are harmful to society. If the risky investment succeeds, the bank's shareholders gain a lot; if the risky investment fails, the shareholders are protected by their limited liability.[8]

Contingent Capital May Reduce the Need for Bailouts. The idea behind requiring banks to hold contingent capital, or more generally capital, is that a bank that suddenly loses money can absorb losses and does not need to be bailed out. First, since the debt is converted to equity, the bank is relieved from paying interest on its debt. Second, since the bank obtains more equity, it is easier for the bank to absorb additional losses in the future.

In one view, the main role of contingent capital is to prevent failures of large banks to begin with. Under this view, the trigger for conversion should be hit at a relatively early stage, when there is still a chance to save the bank by recapitalizing it (i.e., increasing its level of capital). Conversion would then be a relatively frequent event and would not be limited to financial crises.[9]

In another view, the purpose of contingent capital is not to prevent single bank failures but instead to create procedures to deal with the failure of large banks in situations in which many banks experience problems at the same time. More generally, the idea is to have an out-of-court resolution mechanism so that if large banks get into financial problems, the regulator does not need to rely on ad hoc measures or lengthy and costly bankruptcy procedures. In this case, the trigger should apply at a later stage and conversion would occur only during a full-blown financial crisis.[10]

A Distressed Bank May Not Take Appropriate Measures on Its Own. When a bank suffers a loss, the value of its assets drops, and this reduces the value of its equity (assuming the value of its liabilities remains unchanged). Hence, the bank has a smaller capital cushion to absorb additional losses in the future, and its chances of going bankrupt increase.

To reduce the likelihood of bankruptcy, the bank can recapitalize by issuing more equity. However, the bank may be reluctant to do so because of a problem that economists call *debt overhang*, one variant of stockholders' incentive to take excessive risks. A debt overhang problem refers to a situation in which a bank has a lot of outstanding debt and there is a significant likelihood of default. Since the money raised by issuing equity must first go to satisfy existing debt obligations (debt holders get first priority in payments), and since new shareholders must at least break even on their investment or else they would not provide the bank with any capital, issuing new equity is essentially a transfer from existing shareholders to existing debt holders. In particular, issuing equity increases the likelihood that existing debt holders will be repaid but, at the same time, dilutes the shares of existing shareholders. Moreover, if some of the bank's debt is insured, issuing new equity is not only a transfer from equity holders to debt holders, but it is also a transfer from equity holders to the deposit insurer. Hence, if a bank is managed in the interests of existing stockholders, it will not issue equity unless it is forced to do so.[11]

The bank can also recapitalize by selling assets. But again, the bank may be reluctant to do so because of the debt overhang problem. Moreover, selling assets can also impose problems on other banks and on the whole economy. If other banks, which are the potential buyers of the assets, also face financial problems, they may be reluctant to buy the assets. Alternatively, they may agree to buy, but only at "fire sale" prices, which are well below the price they would normally pay. Such a significant drop in prices further amplifies problems because it reduces the value of assets that other banks own and, thus, the banks' capital ratios.

When a large bank fails, its failure can spread to other banks in a domino effect, which economists call contagion.

[8] I talk about contagion in my 2002 *Business Review* article and in my paper. I also discuss private-sector bailouts, in which banks help each other without using taxpayers' money and the regulator acts only as a coordinator.

[9] See, for example, Mark Flannery's proposal.

[10] See, for example, the Squam Lake Group's proposal.

[11] The problem of debt overhang was first discussed in the seminal paper by Stewart Myers. More generally, it refers to a situation in which a firm with a lot of debt forgoes profitable investment opportunities.

Instead of selling existing assets, the bank can simply stop acquiring new assets, but this means that the bank will lend less to households and firms. The regulator may then be forced to step in to avoid the potential damage to the economy.

Contingent capital may reduce the need for bailouts because when the bank gets into trouble and its capital ratio drops, its debt converts to equity automatically, and so its capital ratio increases back to what the regulator perceives to be a safe level.

One lingering question is whether contingent capital has an advantage (to banks or to society as a whole) over simply requiring banks to hold more capital. There are views on both sides.

Contingent Capital May Be "Cheaper" Than Capital. Some economists argue that contingent capital is cheaper than standard capital. Underlying this argument is the tradeoff between debt and equity and the notion that contingent capital captures the benefits of debt while avoiding most of its problems. In particular, they argue that contingent capital captures the tax benefits and disciplinary role of debt while avoiding the problems of debt when the bank is in financial distress and may not be able to pay off its debt.[12]

Under existing tax law, debt has an advantage over equity (to the issuing bank) because the bank can deduct interest payments, but it cannot deduct dividend payments. In addition, in some economic models, debt has a disciplinary role. To make sure they get their money back, debt holders monitor the bank so that the bank's managers don't waste money or take excessive risks. Moreover, if the debt is short term, debt holders may choose not to renew it after poor performance. The threat of insolvency if short-term debt holders refuse to roll over their claims imposes discipline on bank managers.[13]

However, as we saw earlier, too much debt can create problems both to the issuing bank (e.g., debt overhang) and to society as a whole (e.g., contagion and costly bailouts). Assuming that investors are willing to hold it, contingent capital can help avoid these problems because the automatic conversion helps to recapitalize the bank before the problems spill over to the rest of the economy.

Contingent capital might also help banks avoid the costs of lengthy bankruptcy procedures. Once a bank is in bankruptcy, it may take a long time for creditors to get paid (either partially or fully). The automatic conversion of contingent capital helps to avoid this problem.

Later in this article we discuss alternative views as to whether contingent capital is indeed less costly than equity, but before that, we discuss some issues with implementation.

Under existing tax law, debt has an advantage over equity (to the issuing bank) because the bank can deduct interest payments, but it cannot deduct dividend payments.

HOW SHOULD CONTINGENT CAPITAL BE DESIGNED?

Suppose we believe that contingent capital is beneficial. How should we design it? In particular, what event should trigger conversion? What should the conversion ratio be? That is, how many shares of stocks should $1 of face value of debt convert to?

We start with the event or events that should trigger conversion. Table 1 provides examples of three different proposals. We explain the features in these proposals below.

TABLE

Examples of Specific Proposals for Contingent Capital

Proposal	Trigger for Conversion	Dual Trigger?	Book or Market Values?
Mark Flannery	1. Bank's stock price falls below some threshold.	No	Market
Squam Lake Group	1. The regulator declares a systemic crisis. 2. The bank's tier 1 capital ratio falls below some threshold.	Yes	Book
Robert McDonald	1. A broad financial stock's index falls below some threshold. 2. Bank's stock price falls below some threshold.	Yes	Market

[12] Note, however, that it is unclear whether *any* form of contingent capital will qualify as debt for tax purposes. See the discussion of this issue in the paper by Robert McDonald.

[13] See, for example, the paper by Douglas Diamond and Raghuram Rajan.

The Trigger for Conversion. Some economists have suggested that conversion should depend only on the bank's own condition; that is, conversion should occur whenever the issuing bank has serious financial problems. Other economists have suggested using a dual trigger, meaning that conversion should occur only if both the issuing bank and the whole financial system are in trouble. Clearly, under a dual trigger, conversion occurs less often, and any individual bank with financial problems is less likely to be recapitalized. One advantage of this is that the disciplinary role of debt is enhanced, since the threat of bankruptcy is stronger. One disadvantage is that the failure of a single large bank can have negative consequences for the whole economy.

Now suppose we decide on a dual trigger. How should we determine whether the financial system is in trouble? Should we rely on the regulator to declare a systemic crisis, or should we use a more objective criterion, such as a broad financial stock index? Each option has some pros and cons. One problem with regulatory discretion is that market participants may be uncertain as to how the regulator will interpret the data. A second problem is that regulators may be concerned about maintaining confidence in the financial system and, hence, may be reluctant to declare a financial crisis until it is too late. An objective rule may avoid these problems. However, it is impossible to come up with a rule that is always accurate, and blindly following some decision rule may be misleading. The more likely outcome is that regulators would choose not to follow the rule when it looks like it is mistakenly calling for conversion. Thus, they are likely to use discretion in practice. Nonetheless, specifying some rule for intervention may help to the extent that regulators may have difficulty pre-committing to declaring a crisis. At the minimum, regulators will have to explain to the public why they are acting contrary to the rule.

Another issue is whether the triggers for conversion should be based on book values (meaning accounting numbers) or market values. An example of a trigger based on book values is regulatory tier 1 capital, which is derived from the bank's financial statements. Examples of triggers based on market values are the bank's stock price or its credit default swap spread (see *Examples of Market-Based Triggers*).

One advantage of using market values is that they are more forward looking and rapidly adapt to changes in the bank's financial condition. In contrast, financial statements are updated only periodically and reflect the bank's financial condition with a lag. In addition, accounting numbers can be distorted by the bank. For example, if a bank is subject to mark-to-market accounting, meaning that assets are valued based on the recent market price of identical or similar assets, the bank may stop trading in these assets so that losses are not reflected on its balance sheet.[14]

However, relying on market prices may also create problems because market prices could trigger conversion based on factors other than the bank's true financial condition. One such problem is that market prices create opportunities for manipulation. Second, market prices may create the possibility of multiple self-fulfilling equilibria.

Manipulation. An investor who holds contingent capital may attempt to drive down the stock price, so that conversion will occur, and then drive the price back up to make a profit.[15] For example, suppose an investor has $1000 of face value of debt that converts into 20 shares of stock when the

Examples of Market-Based Triggers

While most proposals that use market values as triggers rely on stock prices, it is also possible to use the prices of credit default swaps.[a,b] Credit default swaps are a form of insurance against default on the bank's debt, and so their prices reflect whether the bank is in financial trouble. The advantage of using prices of credit default swaps over stock prices is that prices of credit default swaps capture only the likelihood of default, whereas stock prices capture both the expected profits of the bank when it doesn't default as well as the likelihood of default. One disadvantage is that credit default swaps may also reflect the likelihood of government bailouts, and at another extreme, they can also reflect the likelihood that the firm that provides insurance may itself default.

[a] See, for example, the article by Oliver Hart and Luigi Zingales.

[b] A credit default swap is a contract that is written between the seller of the swap and the buyer of the swap in reference to some credit event, such as a default by Bank ABC on a specific bond (long-term debt) it issued. The buyer of the swap pays a premium to the seller of the swap, just like the buyer of car insurance pays a premium to the company that sells insurance. In return, the seller of the swap promises to make a payment to the buyer of the swap if the credit event occurs.

[14] See, for example, my paper with Philip Bond and the paper by Konstantin Milbradt.

[15] One way to drive down the price is by short selling the stock. Short selling is explained in Ronel Elul's *Business Review* article.

stock price falls below $50. Suppose the stock price is currently $51 (the true value of the firm) and the investor can drive the price down to $49. Then conversion occurs, and instead of owning $1000 of debt, the investor now owns 20 shares of stock. When the price returns to $51, the investor has a position worth $1020, which is $20 more than what he originally had; that is, he gained 2 percent.

One way to minimize this problem is to set a low conversion ratio so that the value of equity that the holder of contingent capital obtains after conversion is lower than the debt's face value. For example, if the original $1000 of debt converts into 19 shares of stock, rather than 20 shares, then after conversion occurs, the price would need to go up to $52.63 (=1000/19) for the investor to make a profit. Hence, manipulation becomes harder.[16]

Triggers that are based on the average stock price over some given period rather than just one day may also make it more difficult to manipulate conversion.

Multiple Equilibria. Another possible problem when conversion is based on market prices has to do with what economists call "multiple equilibria."[17] We usually think of stock prices as aggregating investors' views about the firm's true value, given the information they have. However, stock prices may also reflect investors' expectations as to what other investors will do, which may affect whether the debt will convert to equity. This may lead to situations in which conversion depends on self-fulfilling expectations rather than on whether the bank is truly facing financial problems. It may also lead to situations in which the market "breaks down," as it is impossible to come up with a price that is consistent with investors' expectations. The problem in both cases is that the information role of the stock price in aggregating investors' views about the bank's true value is reduced.

For example, consider a situation in which the bank is not in financial trouble, so the stock price should remain high and conversion should not occur. Suppose that the conversion ratio is high, so after conversion the shares of existing equity holders are diluted and the value of each share drops. There are two self-fulfilling outcomes: In the first outcome, stock holders expect that conversion will occur (say, tomorrow) and their shares will be diluted; hence they attempt to sell their stock today before conversion occurs. But because of these sales, the stock price falls today, and this by itself induces conversion.[18] Alternatively, what happens if investors do not expect conversion to occur? In this case, the stock price remains high, and conversion indeed does not occur.

Now suppose instead that the conversion ratio is low so that stockholders' existing shares are not diluted and the price of each share rises after conversion. If investors expect that conversion will occur tomorrow, the price will increase today in anticipation, but then conversion will not occur. In contrast, if investors do not expect conversion to occur, the price remains low and conversion occurs as a result. Hence, in both cases, the stock price is inconsistent with investors' expectations. One interpretation of this is that the market breaks down, e.g., investors may not trade because they cannot determine what the price should be, which economists refer to as "no equilibrium." More broadly, the stock price may be arbitrary and may not reflect investors' views about the bank's true value.

Note that in the first case, when the conversion ratio is high, conversion induces a transfer of wealth from existing equity holders to investors in contingent capital, whereas in the second case, when the ratio is low, conversion induces a transfer in the other direction. To avoid the problems above, Suresh Sundaresan and Zhenyu Wang have argued that the conversion ratio must be such that there should never be any wealth transfers between equity holders and investors in contingent capital at the time of conversion. This means that the conversion ratio should depend on the market price of the convertible debt at the time of conversion. However, as for now, it is hard to tell whether the market for contingent capital, if introduced, will be liquid enough so that market prices will be readily available. In addition, the condition above (no wealth transfer) is inconsistent with the condition of a low conversion ratio, thereby removing one of the tools to prevent manipulation.

One may ask why we do not run into the problems above (multiple equilibria or no equilibrium) with the standard, widely used convertible debt. The reason is that the holder of convertible debt can *choose* whether to convert and so will convert only if conversion makes him better off. If conversion induces a wealth transfer from existing equity holders to holders of convertible debt, the holders of convertible debt will choose to convert their debt to equity. If the wealth transfer is in the other direction, they will choose not to convert. Hence, in both cases, we obtain a unique outcome. In contrast, holders of contingent capital cannot choose whether to convert, since the conversion occurs automatically whenever the stock price

[16] The numerical examples above are taken from Robert McDonald's paper.

[17] In a seminal paper, Douglas Diamond and Philip Dybvig have illustrated this problem in the context of bank runs.

[18] Such expectations may also cause a "death spiral," in which the price drops drastically. This can happen when the price after conversion is very low.

hits the trigger. As we saw above, this may lead to multiple outcomes or, alternatively, to a market breakdown.

ALTERNATIVES TO CONTINGENT CAPITAL

The technical issues above have raised concerns about whether contingent capital will satisfy its intended role. In addition, some economists have provided compelling arguments as to why we should not take the notion that "equity is costly" for granted.[19] Under this view, instead of requiring banks to issue contingent capital, the regulator should simply raise capital requirements, i.e., require banks to have more equity in their capital structure. Another suggestion is to increase the liability of equity holders in the event of loss.

Why Bank Equity Should Not Be Viewed as "Costly." The well-known Modigliani-Miller theorem says that in a frictionless world, it does not matter to the firm or to society whether the firm finances itself with debt or with equity. Although investors require a higher return to hold equity than the return they require to hold debt, equity should not be viewed as "costly." The higher return in equity simply compensates investors for the extra risk they take.

The importance of the Modigliani-Miller theorem is that it helps us identify the frictions under which capital structure does matter. When designing capital regulations, we should take these frictions into consideration, remembering that some of them are inherent, while others are created by policy. For example, conflict of interest between investors and firm managers is an inherent friction, while the tax advantage of debt is created by policy. We should also be aware that the "cost" to the bank and the "cost" to society are not necessarily the same.

For example, tax deductions make it cheaper for a bank to issue contingent capital (assuming that contingent capital qualifies for the same tax benefits as debt). However, instead of contingent capital, we can require banks to hold more equity, and we can neutralize the impact of increased equity requirements on the bank's tax liabilities by replacing any tax benefit lost due to the reduction in leverage with alternative deductions or tax credits.

More broadly, it is not clear that the tax advantage of debt over equity should exist in the first place. After all, this tax advantage is costly to society because it induces banks to take too much debt (relative to equity) and therefore too much risk, which can create contagion, costly bailouts, and other spillovers to the whole economy. The issue is then whether we should let the tax code, which may be suboptimal to begin with, drive new capital regulations. An even broader perspective might be to redesign the tax system, along with redesigning capital regulations.

Economists have also questioned the disciplinary role of debt. First, it relies on the idea that what disciplines managers is early liquidation or the fact that short-term debt may not be renewed if the bank performs badly. However, these forms of discipline increase the fragility of the banking system as a whole and may cause other problems to society, such as credit freezes.[20] Using long-term debt may also create problems, such as the debt overhang problem discussed earlier. In particular, when the bank is in financial distress, the managers may pass over some profitable investment opportunities because taking them hurts existing equity holders. Second, it is unclear whether debt is the only (or best) way to discipline the bank's managers. For example, the bank can commit to pay stock holders a prespecified level of dividends, which, if not maintained, would trigger a shareholder vote to replace the bank's managers. Third, since bank deposits are insured by the government through the Federal Deposit Insurance Fund, banks may still be induced to take excessive risk, since the cost is ultimately borne by taxpayers. The potential for bailouts also distorts banks' incentives.[21]

> Some economists have provided compelling arguments as to why we should not take the notion that "equity is costly" for granted.

Increasing Equity Holders' Liability. Some economists have proposed increasing the liability of equity holders on the grounds that bank failures and financial instability have large social costs. For example, we can require that equity holders add money to the firm whenever there is an immediate risk that the firm will default on its debt. If equity holders don't add money, their equity will be wiped out and they will lose control.[22] In particular, Oliver Hart and Luigi Zingales propose that equity holders should be

[19] In particular, see the paper by Anat Admati, Peter DeMarzo, Martin Hellwig, and Paul Pfleiderer.

[20] See, for example, the paper by Viral Acharya, Douglas Gale, and Tanju Yorulmazer.

[21] A potential solution, suggested by Viral Acharya, Lasse Pedersen, Thomas Philippon, and Matthew Richardson, is to tax each bank during good times based on its expected loss conditional on the occurrence of a systemic crisis.

[22] This is analogous to margin accounts and margin calls that are in place to guarantee payments on future obligations.

forced to add money whenever the price of a credit default swap moves above some prespecified threshold, meaning that the price of insurance against default on the bank's debt is too high.[23] Anat Admati and Paul Pfleiderer also propose to increase the liability of equity holders, but in their proposal, equity holders must set aside some cushion of safe assets upfront to back up their guarantees. One way to think of these equity injections is as a more general form of contingent capital.

Viral Acharya, Hamid Mehran, and Anjan Thakor take the idea above even further by proposing that if a bank gets into financial trouble, a part of the safe asset buffer will belong to the regulator, rather than to the bank's creditors. Their idea is to provide both debt holders and equity holders with incentives to monitor. Since the regulator takes over part of the bank's equity cushion when the bank is troubled, debt is risky from the perspective of creditors and they are induced to monitor the bank managers. (If debt is completely safe, debt holders will have no incentive to monitor.) At the same time, since equity holders have a lot of equity at stake (if the bank fails, they lose the buffer of safe assets), they will make sure that bank managers are not tempted to take excessive risk.

CONCLUSION

Contingent capital may be a step toward reducing failures of large banks, but it is unclear whether it can fulfill its intended role. In particular, problems such as price manipulation, multiplicity of equilibria, and incentive issues must be addressed.

Some alternatives to contingent capital, such as increasing capital requirements and increasing shareholders' liability, also exist. Contingent capital must be evaluated against these as well as other alternatives.

[23] They also propose that equity capital protect only against "systemically relevant" obligations (such as bank deposits, short-term debt, interbank borrowing, and derivative contracts) but not against long-term debt obligations. Hence, long-term debt will sometimes need to absorb losses. This provides an extra protection, and it also provides an underlying asset on which credit default swaps can be traded

REFERENCES

Acharya, Viral V., Douglas Gale, and Tanju Yorulmazer. "Rollover Risk and Market Freezes," *Journal of Finance* 66:4 (August 2011), p. 1177-1209.

Acharya, Viral V., Hamid Mehran, and Anjan Thakor. "Caught Between Scylla and Charybdis? Regulating Bank Leverage When There Is Rent Seeking and Risk Shifting," Federal Reserve Bank of New York Staff Report 469 (revised November 2010).

Acharya Viral V., Lasse H. Pedersen, Thomas Philippon and Matthew Richardson. "A Tax on Systemic Risk," New York University Working Paper (2010).

Admati, Anat R., and Paul Pfleiderer. "Increased-Liability Equity: A Proposal to Improve Capital Regulation of Large Financial Institutions," Stanford Graduate School of Business Working Paper (June 2010).

Admati, Anat R., Peter M. DeMarzo, Martin F. Hellwig, and Paul Pfleiderer. "Fallacies, Irrelevant Facts, and Myths in the Discussion of Capital Regulation: Why Bank Equity Is *Not* Expensive," Stanford Graduate School of Business Research Paper 2065(R) (October 2010).

Berlin, Mitchell. "Can We Explain Banks' Capital Structures?," Federal Reserve Bank of Philadelphia *Business Review* (Second Quarter 2011).

Bond, Philip, and Yaron Leitner. "Market Run-Ups, Market Freezes, Inventories, and Leverage," Federal Reserve Bank of Philadelphia Working Paper 12-8 (February 2012).

Diamond, W. Douglas, and Philip H. Dybvig. "Bank Runs, Deposit Insurance and Liquidity," *Journal of Political Economy* 9 (1983), pp. 401-19.

Diamond, W. Douglas, and Raghuram G. Rajan. "A Theory of Bank Capital," *Journal of Finance* 55 (2000), pp. 2431-65.

Elul, Ronel. "Regulating Short-Sales," Federal Reserve Bank of Philadelphia *Business Review* (Second Quarter 2009).

Flannery, Mark. "Stabilizing Large Financial Institutions with Contingent Capital Certificates," Working Paper, University of Florida (October 2009).

Hart, Oliver, and Luigi Zingales. "A New Capital Regulation for Large Financial Institutions," *American Law and Economics Review*, 13:2 (Fall 2011), pp. 453-90.

Leitner, Yaron. "Financial Networks: Contagion, Commitment, and Private-Sector Bailout," *Journal of Finance*, 60 (2005), pp. 2925-53.

Leitner, Yaron. "A Lifeline for the Weakest Link? Financial Contagion and Network Design," Federal Reserve Bank of Philadelphia *Business Review* (Fourth Quarter 2002).

McDonald, Robert L. "Contingent Capital with a Dual Price Trigger," Northwestern University Working Paper (February 2010).

Milbradt, Konstantin. "Level 3 Assets: Booking Profits and Concealing Losses," *Review of Financial Studies*, 25 (2012), pp. 55-95.

Myers, Stewart. "Determinants of Corporate Borrowing," *Journal of Financial Economics*, 5 (1977), pp. 147-75.

Squam Lake Group. *The Squam Lake Report: Fixing the Financial System.* Princeton: Princeton University Press, 2010.

Sundaresan, Suresh, and Zhenyu Wang. "Design of Contingent Capital with a Stock Price Trigger for Mandatory Conversion," Federal Reserve Bank of New York Staff Report 448 (May 2010).

Time-Consistency and Credible Monetary Policy After the Crisis*

BY JAMES M. NASON AND CHARLES I. PLOSSER

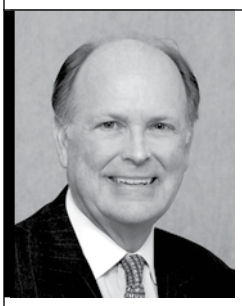
The economic crisis and its aftermath have posed significant challenges to policymakers. To help meet those challenges, the Federal Reserve deployed several innovative policy tools to help relieve the stress in financial markets during the crisis. These tools have created their own significant challenges for the conduct of monetary policy in the post-crisis era. The wider range of policy options now available to policymakers makes it more difficult to credibly commit to a particular policy course, and this discretion poses a problem. This is because monetary policy is subject to a time-inconsistency problem. The new monetary policy tools introduced during the crisis can make such time-inconsistency problems worse by reinforcing the incentives for financial institutions or other sectors of the economy to take on excessive risk. In this article, Jim Nason and Charles Plosser discuss why it is important for central banks to consider ways in which they can limit discretion and use these new tools in a systematic way.

The economic crisis and its aftermath have posed significant challenges to policymakers around the world. To help meet those challenges, the Federal Reserve developed and deployed some innovative policy tools, including liquidity programs, to help stressed markets and large-scale purchases of mortgage-backed securities and longer maturity Treasury securities, which altered the size and composition of the Fed's balance sheet. These tools have created their own significant challenges for the conduct of monetary policy in the post-crisis era. We do not yet have well-developed theories about how such tools can be optimally deployed, but we can draw on earlier economic research to reach some conclusions. In particular, we know that the wider range of policy options now available to policymakers and the lack of fully articulated models make it more difficult for policymakers to credibly commit to a particular policy course and that this discretion poses a problem. Research since the late 1970s, including important contributions by Henry Simons, Guillermo Calvo, Finn Kydland, and Edward Prescott, indicates, perhaps paradoxically, that when policymakers take a more systematic approach to policy and use less discretion, their policies yield better outcomes.

This is because monetary policy is subject to what economists call a time-inconsistency problem — what might seem like the best policy when first announced may not be viewed as best when the time comes to act. But if policymakers yield to temptation and renege on that announced policy, this can lead to worse outcomes than if they were able to stick with the original plan. This is a well-known aspect of many forms of policymaking and one reason monetary policymak-

Jim Nason is a vice president and economist in the Research Department of the Philadelphia Fed and the head of the Macroeconomics section. This article is available free of charge at www.philadelphiafed.org/research-and-data/publications.

*The views expressed here are those of the authors and do not necessarily represent the views of the Federal Reserve Bank of Philadelphia or the Federal Reserve System.



Charles Plosser is the president and chief executive officer of the Federal Reserve Bank of Philadelphia. This article is available free of charge at www.philadelphiafed.org/research-and-data/publications.

ers often talk about the importance of commitment, credibility of previous policy promises, and reputation, and they look for ways to limit the use of discretion in their policymaking.

But the new monetary policy tools introduced during the crisis can make such time-inconsistency problems worse by reinforcing the incentives for financial institutions or other sectors of the economy to take on excessive risk — so-called moral hazard. These firms know that in the midst of severe problems in financial markets, policymakers might find it extremely difficult to refrain from acting as lender of last resort to rescue them, even if doing so would lead to better risk-taking incentives in the future. Such a policy of refraining from bailouts would not be credible or time-consistent. But knowing this, firms have less incentive to refrain from excessive risk-taking in the first place, making the likelihood that policymakers might face a situation in which a firm will need to be rescued even greater. Thus, with new tools now at policymakers' disposal, it is important that they understand the interplay between time inconsistency and moral hazard. As we'll discuss below, we think it is important for central banks to consider ways in which they can limit discretion and use these tools in a systematic, or rule-like, way.

A TIME-CONSISTENT BEDTIME STORY

Although "time-inconsistency" is an economic concept, it affects almost everyone at some time or another. Consider Jane and her parents, who have given Jane a bedtime rule. On a school night, Jane must turn off the TV, laptop, and smartphone, as well as unplug the electric guitar and amplifier at 9 p.m. and go to bed. The parents' goal of setting the rule is for Jane to be well rested for school. They also anticipate that without a rule, it will be very costly negotiating bedtime every school night with Jane, whose goal is to watch TV, IM her friends, and play guitar (all at the same time!).

How the rule works in practice, however, is different than intended. One school night at 9 p.m., Jane complains loudly and persistently to her parents that there is a TV show that she must watch. Her complaints impose high enough costs on her parents that they relent and give Jane more time before going to bed.

What do you think will happen on future school nights? Jane reasons that because her parents have deviated from their bedtime rule once, they are likely to do so again. In other words, the bedtime rule has lost its credibility. The loss of credibility gives Jane the incentive to test her parents' willingness to stand by their bedtime rule on future school nights. She doubts that her parents are prepared to bear the costs necessary to enforce the bedtime rule in the future. Jane's parents have lost their reputation for following their bedtime rule.

This bedtime story shares much in common with the canonical monetary policy "game" facing central banks.[1] The key elements are: (a) the central bank engages in discretionary policy because it is next to impossible to credibly commit today to follow rules in the future; (b) the central bank and the private sector can have different goals; (c) the monetary policy "game" is repeated over and over; and (d) the central bank and the private sector each believe the other will act in its own best interest, given previous actions and outcomes. When the central bank engages in discretionary monetary policy, say, by cutting interest rates in an attempt to boost employment in the short to medium run, the private sector has incentives to challenge the credibility of the central bank's commitment to price stability over the medium to longer run, as Jane did with her parents' bedtime rule. Over time, the public will learn by experience whether the central bank's commitment is sustained, and the central bank will develop a reputation for either making credible commitments or not.

The new monetary policy tools introduced during the crisis can make such time-inconsistency problems worse by reinforcing the incentives for financial institutions or other sectors of the economy to take on excessive risk — so-called moral hazard.

TIME-INCONSISTENCY AND COMMITMENT

While time-inconsistency problems will arise when a central bank changes its goals for inflation and output growth over time,[2] this is not the only, or even main way, it arises in monetary policymaking. Similar to Jane's parents' experience with their

[1] Our interest is in the class of monetary policy "games" that focus on the economic costs the private sector can impose on a central bank engaged in discretionary monetary policy. Economic costs are generated by, for example, disagreements over the (specific) goals of monetary policy. There are monetary policy "games" that include political costs. Political costs can arise because pre-election promises and post-election outcomes are not time-consistent.

[2] The seminal reference is the paper by Robert Strotz.

bedtime rule, time-inconsistency is driven by incentives that encourage a central bank to deviate today from a previously announced policy.[3] Even though there are well-known benefits to having policymakers commit to engage in systematic or rule-like behavior over time and limit their use of discretion, full commitment by a central bank is not possible.[4] That is, there is no credible way for a policymaker, before the fact, to promise always and everywhere not to take discretionary actions and follow rules announced earlier. However, while full commitment is not feasible for real world central banks, economics can provide alternative ways to limit discretion and tie the hands of policymakers, thereby yielding better economic outcomes. One such method is reputation; another is central bank independence.

Reputation. The current U.S. monetary policy regime has been associated with great diversity on the FOMC led by Chairmen as different as William McChesney Martin, Arthur Burns, Paul Volcker, Alan Greenspan, and Ben Bernanke.[5] Under Chairman Martin, the FOMC established its reputation with a long period of low inflation, which is often credited with the sustained real growth experienced during his tenure. Responsibility for the disinflation of the 1980s is often attributed to the FOMC because it rebuilt its anti-inflation reputation under Chairman Volcker after having lost it in the 1970s.[6]

These examples suggest that time-consistency problems can be solved by reputation when it fills in gaps created by a lack of commitment.[7] A central bank can use a reputation for being intolerant of high inflation to commit to low inflation.[8] Although reputation can surmount time-consistency problems, a study by Kenneth Rogoff shows that reputation alone cannot determine the inflation rate for an economy.[9] Instead, private-sector market participants must agree about the inflation rate on which to focus, given that many inflation rates are possible. But this creates a problem for central banks. Because the way in which market participants determine which inflation rate they believe will prevail is beyond the control of policymakers, a central bank depending only on its reputation of being tough on inflation will eventually lose control of inflation in the long run.

There is at least one partial solution to this public-private coordination problem. The solution is to have the government and the central bank strike a contract that creates incentives for the central bank to adopt low-inflation policies. In a 1995 study, Carl Walsh provides the first example in which a contract between the government and its central bank is a means to coordinate market participants' beliefs about policies that yield low inflation.[10] Since the 1980s, several countries have altered the design of their central banks by offering monetary policymakers a contract with inducements to achieve, say, a low inflation rate averaged over three to five years. While such contracts can help, experience with these contracts teaches us that central bank reputations are fragile unless supported by actual achievements.

Central Bank Independence. Designing a central bank so that its monetary policy decisions are inde-

While full commitment is not feasible for real world central banks, economics can provide alternative ways to limit discretion and tie the hands of policymakers, thereby yielding better economic outcomes.

[3] These incentives are part of an economic environment in which market participants and policymakers act rationally.

[4] See the 2008 and 2010(a) speeches by Charles Plosser and the paper by John Taylor. Although rule-like behavior generally yields better outcomes than discretion, it may not be desirable to entirely rule out discretion by policymakers in all cases, especially in a democracy. In the U.S. democratic system, the make-up of the legislative, executive, and judicial branches changes over time by design. Future Congresses, Presidents, and federal judges can repudiate current law constrained only by the U.S. Constitution and its legal interpretation.

[5] Martin served as Federal Reserve Chairman from 1951 to 1970; Burns served 1970-1978; Volcker from 1979-1987; Greenspan from 1987-2006; and Bernanke from 2006 to the present.

[6] See the 2007 speech by Charles Plosser.

[7] For a nontechnical exposition of how reputation can support a central bank in attaining low inflation, see the article by Herb Taylor.

[8] The 1985 paper by Kenneth Rogoff provides an example of the importance of the central bank's reputation for supporting a low inflation policy that has been interpreted as explaining the Volcker deflation of the early 1980s. For a nontechnical discussion of these issues, see the article by Herb Taylor.

[9] See the 1989 study by Rogoff.

[10] Walsh studies a wage contract between the government and the central banker. The contract specifies an inflation goal and a wage for the central banker. The more negative actual inflation is net of the inflation goal, the higher is the central banker's wage. Walsh argues that his interpretation of the contract between the central bank and the government can be generalized. The contract can include rewards and punishments based on things other than wages to encourage the central bank to achieve the outcomes desired by the government.

pendent of the political process can also help moderate time-consistency problems.[11] In the U.S., several Congresses and Presidents have delegated monetary policy to the Federal Reserve and Federal Open Market Committee (FOMC). This sort of independence seems to be necessary for a central bank to wield its reputation as a pledge to keep inflation low. Nonetheless, central bank independence is no guarantee that similar pledges will always lead to low inflation as the Fed and the country learned during the Great Inflation of the 1970s.

TIME-INCONSISTENCY AND MONETARY POLICY AFTER THE CRISIS

The Fed responded to the crisis with some innovative policy tools, including paying interest on the excess reserves that banks hold in their accounts at the Fed, setting up special lending facilities, and engaging in large-scale purchases of assets that increased the size of the Fed's balance sheet. These tools have expanded the scope of discretion available to the Fed. But as the time-inconsistency literature indicates, regardless of how necessary the exercise of discretion during the crisis may have been, it presents the Fed with the possible loss of credibility and independence.

Interest on Excess Reserves: Fed Independence and the Balance Sheet. In October 2008, Congress and the President gave authority to the Board of Governors to pay *interest on excess reserves* (IOER). This authority means the Fed compensates private banks on reserves they hold at the Fed that are in excess of the reserves required by Fed regulation. The return on the reserves in excess of the required reserves is the IOER. The Board of Governors has discretion to set the IOER, which it does in consultation with the FOMC.

The economist and Nobel laureate Milton Friedman argued that a central bank should pay interest on reserves to improve financial market efficiency.[12] When a central bank pays IOER, the difference between the rate of return a bank earns on reserves it holds at the Fed and what it would earn by investing in short-term assets (e.g., in the money markets) is reduced, if not eliminated. Thus, banks no longer have a reason to look for ways to avoid an implicit tax on their reserve holdings, thereby increasing efficiency in the intermediation process. However, Friedman also noted that inherent in IOER is a trade-off. Greater financial market efficiency resulting from paying IOER may align the interests of the fiscal authority too closely with those of the central bank. Friedman's argument is that when taxes fund IOER payments to private banks, the balance sheet of the central bank becomes entwined with the fiscal authority.[13]

> Designing a central bank so that its monetary policy decisions are independent of the political process can also help moderate time-consistency problems.

This raises questions about the central bank's independence.

The trade-off becomes starker when we consider the impact the IOER operating mechanism can have on the Fed's balance sheet. Two different kinds of IOER operating mechanisms have been proposed for the Fed once monetary policy returns to more normal operating conditions.[14] Under the *corridor system*, the Fed's policy rate would lie within a range bounded below by the IOER rate and above by the Fed's *discount rate*. The discount rate is the rate the Fed charges on short-term loans it makes to banks facing temporary liquidity needs that come to borrow at the Fed's discount window. The corridor system is consistent with the monetary policy procedures the Fed employed for 25 years or more prior to the financial crisis. Since the Fed did not pay interest on excess reserves during this period, the IOER rate was implicitly zero, and the Fed's policy rate, the fed funds rate, lay below the discount rate.

Whether the IOER rate is zero or not, the corridor system only requires that the IOER rate be less than the policy rate for private banks to want to minimize the opportunity cost of holding excess reserves. This cost reflects the best alternative use of these funds for banks (i.e., the return banks can earn by investing in other assets). An implication is that banks' demand for excess reserves will fall as the policy rate rises and the opportunity cost increases. This demand does not go all the way down to zero because the rate charged at the discount window, which is greater than the policy rate, is an incentive for a bank to hold at least some excess reserves as insurance against unexpected liquidity needs. Converse-

[11] See the speech by Charles Plosser (2010b).

[12] See the study by Milton Friedman and the paper by Thomas Sargent.

[13] When IOER are financed by taxes, the budget constraints of the central bank and the rest of the government are explicitly tied together. Thomas Sargent discusses IOER and central bank independence in the context of Friedman's monetary policy proposals.

[14] The study by Marvin Goodfriend and the one by Todd Keister, Antoine Martin, and James McAndrews analyze the IOER operating mechanisms in full.

ly, a fall in the policy rate encourages banks to hold more reserves, since the opportunity cost of holding those reserves has fallen. When the IOER rate is non-zero and less than the policy rate, the opportunity cost of holding excess reserves limits the amount that banks are willing to hold. Reserves are a liability on the Fed's balance sheet, which are offset by the assets the Fed holds, mainly marketable securities. By limiting the amount of reserves, the corridor system puts a constraint on discretionary use of the Fed's balance sheet by placing an upper bound on its size — it gives the balance sheet a "small footprint."

The other IOER operating mechanism is called the floor system. Under the floor system, the central bank's policy rate is set equal to the IOER rate. This equality implies that private banks face no opportunity cost when holding excess reserves. Under this system, the demand for excess reserves does not respond to the IOER policy rate, and the central bank always supplies the amount of reserves that meets the demand for reserves of private banks.

Essentially, the Fed has been operating under the floor system for the past two years since the depths of the crisis. Notice that the independence of the IOER policy rate from the supply of excess reserves in the floor system gives the Fed two policy tools: the IOER rate and the size (and composition) of the Fed's balance sheet.[15] Under the floor system, the Fed has been able to saturate banks with excess reserves to satisfy liquidity needs in financial markets without having to change the IOER policy rate, that is, without having to alter the stance of monetary policy.

This benefit that the floor system can bestow during a financial crisis needs to be weighed against the potential costs of such a system. A central bank using the floor system faces the potential of a very large balance sheet, that is, one with a "big footprint." In fact, reserves are potentially in unlimited supply at the IOER policy rate, so the floor system calls into question the credibility of commitments to limit the size of the Fed's balance sheet. As long as the Fed's balance sheet is seen as a policy tool with little or no costs, there likely will be those who want to employ it to solve problems even if the expected benefits are small or to achieve goals outside the realm of monetary policy, such as supporting particular industrial sectors of the economy. Such policy actions risk not only the Fed's credibility but its independence as well.

The Fed Balance Sheet: Section 13(3). During the financial crisis, for the first time since the 1930s, the Board of Governors invoked section 13(3) of the Federal Reserve Act to offer liquidity to particular financial market participants. Prior to being amended by passage of the Dodd-Frank Wall Street Reform and Consumer Protection Act in July 2010, section 13(3) granted the Board the authority to discount securities of "corporations, partnerships, and individuals" when it deemed there were "unusual and exigent circumstances." The Board used this authority to set up several term lending facilities and create entities that discounted private securities.[16] The Dodd-Frank Act put some limits on the Board's discretionary authority under 13(3). In particular, the Board must now act in concert with the Treasury to broadly supply liquidity to the financial system rather than to assist a financial firm that is in trouble. Nonetheless, the Board retains substantial discretion to employ section 13(3) because the Dodd-Frank Act imposes few other restrictions on the uses to which the Fed can put its balance sheet.

The Fed Balance Sheet: Large Scale Asset Purchases. Large scale asset purchases refer to policies in which the Fed buys long-term non-Treasury and Treasury securities. Purchases of mortgage-backed securities (MBS) and government agency (i.e., Freddie Mac, Fannie Mae, and Federal Home Loan Banks) debt helped to support housing finance. Buying MBS and agency debt broke a self-imposed Fed rule of a "Treasuries-only balance sheet" that dated at least to the Treasury-Fed Accord of March 1951.[17] However, these purchases were motivated by a desire to shore up distressed financial markets and were justified by the Fed's responsibilities as the U.S.'s lender of last

A central bank using the floor system faces the potential of a very large balance sheet, that is, one with a "big footprint."

[15] Under the floor system, the demand for reserves is not determined by the price — the IOER policy rate — in the sense that the supply of reserves is any amount that is consistent with the IOER policy rate; see the article by Goodfriend and the one by Keister, Martin, and McAndrews.

[16] Examples are the Asset-Backed Commercial Paper Facility, Money Market Fund Liquidity Facility, the Term Asset-Backed Securities Loan Facility, Term Asset Lending Facility, and Maiden Lane, Maiden Lane II, and Maiden Lane III. For more details, see the 2010(b) speech by Charles Plosser.

[17] From 1971 to 2003, the Fed's balance sheet held agency debt. These holdings of non-Treasury securities were tiny compared with the stock of Treasury securities on the balance sheet. The Fed's balance sheet consisted only of Treasury securities from January 2004 to September 2008.

resort.[18] The FOMC has also scaled up its purchases of longer-dated Treasury securities for the Fed's balance sheet — the so-called quantitative easing program. Having already reduced its policy rate, the fed funds rate, to essentially zero, the Fed began purchasing longer maturity Treasuries with the goal of lowering long-term interest rates. The large scale asset purchases programs enlarged the Fed's balance sheet as well as altered its composition and maturity structure.[19]

A concern of holding MBS on the Fed's balance sheet is that moral hazard becomes incorporated into the time-consistency problem.[20] Having seen the Fed purchase non-Treasury assets, market participants may come to expect that the Fed will adopt a policy to purchase other assets with credit risk greater than Treasuries or even than MBS.[21] This might induce these participants to take on excessive risk. The expense of such policies could fall on taxpayers.[22]

Without constraints on the composition of the Fed's balance sheet, discretion may also encourage time-inconsistent policies independent of the amount of credit risk or interest-rate risk the Fed might take on its balance sheet.[23] Instead, MBS holdings can prompt expectations that the Fed's balance sheet is a tool that could be put to uses outside the realm of monetary policy. When market participants come to embrace these expectations, the Fed's credibility and reputation suffer, and this makes it more costly for the Fed to achieve its monetary policy mandates.

Without constraints on the composition of the Fed's balance sheet, discretion may also encourage time-inconsistent policies independent of the amount of credit risk or interest-rate risk the Fed might take on its balance sheet.

Proposals to Sustain Fed Independence and Constrain the Balance Sheet. Without systematic policy or rule-like behavior to constrain discretion, are there other actions that could raise the hurdle for the Fed to deviate from widely agreed-to policy rules? Let's discuss a few possibilities.

A good initial step is the restrictions imposed by the Dodd-Frank Act on the Board's use of section 13(3). Another step could be for the Board to announce that, in the future, it would limit the use of the discretionary powers of section 13(3), say, to a 120-day window during which the Fed would seek public support from the Treasury and congressional leadership to continue the emergency lending in a crisis for an additional 60-day period. Assuming the extra 60-day period is granted, the Fed would have six months to manage the crisis during which time Congress and the President could enact legislation aimed at resolving the crisis.

By adopting this proposal, the Board of Governors would impose constraints on a future Board. This is a theme that runs through the academic literature on time-inconsistency: designing constraints to minimize the discretion available to future policymakers. It is always possible for a future Board to decide to deviate from this constraint, but such an action could entail its own costs in the form of stronger congressional prohibitions on Fed discretion. For example, the Fed could find itself restricted to lending under section 13(3) only when there is a request from the Treasury and Congress.

Another possibility for limiting Fed discretion is to have the Treasury and Congress become increasingly responsible for taking discretionary actions about lending during the "unusual and exigent circumstances" of a financial crisis, which seems reasonable given that this type of lending is part of fiscal policy. For example, the Treasury and Fed could negotiate and commit to an accord under which the Treasury could agree that during a

[18] David Small and James Clouse discuss the legal restrictions on monetary policy and the Fed's balance sheet prior to IOER and the Dodd-Frank Act. James Clouse, Dale Henderson, Athanasios Orphanides, David Small, and Peter Tinsley extend this analysis to environments in which the zero bound on the federal funds rate binds.

[19] Prior to the crisis, the Fed aimed its Treasuries-only balance sheet at replicating approximately the maturity structure of outstanding Treasury securities.

[20] See the 2009 speech by Charles Plosser.

[21] Current policy for the Fed's MBS and agency debt holdings is contained in the FOMC statement of August 10, 2010 in which the FOMC announced that the Fed's balance sheet will not be allowed to shrink and that it would reinvest principal payments from agency debt and agency mortgage-backed securities in longer-term Treasury securities.

[22] For example, the U.S. Treasury has committed to absorbing the Maiden Lane facilities on the Fed's balance sheet, according to the Treasury-Fed joint statement, "The Role of the Federal Reserve in Preserving Financial and Monetary Stability," of March 23, 2009. However, this commitment has yet to be fulfilled and, at the moment, appears to have fallen by the wayside.

[23] Holding MBS on the Fed's balance sheet most likely does not generate much credit risk given that market participants believe that at least some of these securities have implicit U.S. government guarantees. The potential for interest-rate risk from holding MBS is greater because the maturity duration of the Fed's balance sheet is longer.

financial crisis it would exchange its own securities for non-Treasury securities purchased and held by the Fed, say, after 120 days.[24] With such an accord, fiscal policy remains outside the province of the Fed, but policy has the flexibility to respond to a crisis in the short run. Once again, this would give Congress and the Treasury time to prepare a legislative response to a crisis. Committing to a corridor system, with a positive IOER rate, would dovetail with these proposals.

CONCLUSION

Congress delegated authority for monetary policy to the Federal Reserve System beginning with its founding in 1913. Inherent in monetary policy are time-inconsistency problems that are not eliminated by making the central bank independent. Time-consistent policy will remain a problem for central banks because current and future policymakers will not conduct policy in a systematic manner without credible commitments to explicit rules. Ample theory and empirical evidence exist to support the view that limiting discretionary behavior yields better economic outcomes over the long run.

The Fed reacted to the recent financial crisis by employing its balance sheet in innovative ways. Much credit should be given to the Fed for these actions. However, the Fed may find it increasingly difficult to reduce the size of its balance sheet and return it to a Treasuries-only balance sheet without a commitment to explicit rules to do so. In the absence of such rules, the Fed's balance sheet remains a discretionary tool carrying the risk of being used for activities unrelated to the Fed's monetary policy mandate. Engaging in these policies would present the Fed with a loss of credibility and independence.

The Treasury-Fed Accord of March 1951 helped William McChesney Martin and his colleagues on the FOMC to establish a tradition of Fed independence and an admirable record of monetary policy. The accord helped release the Fed from an obligation to support the price of U.S. government debt, which it had done since World War II. This history indicates that, at this moment, there is a need for a new Treasury-Fed accord. A new accord should contribute to maintaining a credibly independent Fed by correctly aligning incentives between it, the Treasury, and Congress. The proposals suggested here are not the final words on monetary policy reform, but such reforms are of profound importance for the future of the Federal Reserve System and the U.S. economy in the post-crisis world.

[24] See the 2010(b) speech by Plosser.

REFERENCES

Calvo, Guillermo A. "On the Time Consistency of Optimal Policy in a Monetary Economy," Econometrica, 46:6 (1978), pp. 1411–28.

Clouse, James, Dale Henderson, Athanasios Orphanides, David H. Small, and Peter A. Tinsley. "Monetary Policy When the Nominal Short-Term Interest Rate Is Zero," B.E. Journal of Macroeconomics, 3:1 (2003), Article 12.

Friedman, Milton. A Program for Monetary Stability. New York: Fordham University Press, 1960.

Goodfriend, Marvin. "Interest on Reserves and Monetary Policy," Federal Reserve Bank of New York Economic Policy Review, 8:1 (2002), pp. 13-29.

Keister, Todd, Antoine Martin, and James McAndrews. "Divorcing Money from Monetary Policy," Federal Reserve Bank of New York Economic Policy Review, 14:2 (2008), pp. 41-56.

Kydland, Finn E., and Edward C. Prescott. "Rules Rather Than Discretion: The Inconsistency of Optimal Plans," Journal of Political Economy, 85:1 (1977), pp. 473-91.

Plosser, Charles. "Output Gaps and Robust Policy," speech to the 2010 European Banking and Financial Forum at the Czech National Bank, March 23, 2010a.

Plosser, Charles. "The Federal Reserve System: Balancing Independence and Accountability," speech to the World Affairs Council of Philadelphia, February 17, 2010b.

Plosser, Charles. "Ensuring Sound Monetary Policy in the Aftermath of Crisis," speech for the U.S. Monetary Policy Forum, Chicago Booth School of Business, February 27, 2009.

Plosser, Charles. "The Benefits of Systematic Monetary Policy," speech to the National Association for Business Economics, Washington Economic Policy Conference, March 3, 2008.

Plosser, Charles. "Credibility and Commitment," speech to the New York Association for Business Economics, March 6, 2007.

Rogoff, Kenneth. "The Optimal Degree of Commitment to an Intermediate Monetary Target," Quarterly Journal of Economics, 100:4 (1985), pp. 1169-89.

Rogoff, Kenneth. "Reputation, Coordination, and Monetary Policy," in Robert J. Barro, ed., Modern Business Cycle Theory. Cambridge, MA: Harvard University Press, 1989, pp. 236-64.

Sargent, Thomas J. "Where to Draw Lines: Stability Versus Efficiency," New York University Department of Economics, unpublished paper (2010).

Simons, Henry C. "Rules versus Authorities in Monetary Policy," Journal of Political Economy, 44:1 (1936), pp. 1-30.

Small, David H., and James Clouse. "The Limits the Federal Reserve Act Places on the Monetary Policy Actions of the Federal Reserve," Annual Review of Banking Law, 19 (2000), pp. 553-79.

Strotz, Robert. "Myopia and Inconsistency in Dynamic Utility Maximization," Review of Economic Studies, 23:3 (1956), pp. 165-80.

Taylor, Herb. "Time Inconsistency: A Potential Problem for Policymakers," Federal Reserve Bank of Philadelphia Business Review (March/April 1985), pp. 3-12.

Taylor, John B. "A Historical Analysis of Monetary Policy Rules," in John B. Taylor, ed., Monetary Policy Rules. Chicago, IL: University of Chicago Press, 1999, pp. 319-340.

Walsh, Carl. "Optimal Contracts for Central Bankers," American Economic Review, 85:1 (1995), pp. 150–67.

RESEARCH RAP

Abstracts of research papers produced by the economists at the Philadelphia Fed

You can find more Research Rap abstracts on our website at: www.philadelphiafed.org/research-and-data/publications/research-rap/. Or view our working papers at: www.philadelphiafed.org/research-and-data/publications/.

SUPPLY-SIDE POLICIES, LOW AGGREGATE DEMAND, AND THE ZERO LOWER BOUND

This paper examines how supply-side policies may play a role in fighting a low aggregate demand that traps an economy at the zero lower bound (ZLB) of nominal interest rates. Future increases in productivity or reductions in mark-ups triggered by supply-side policies generate a wealth effect that pulls current consumption and output up. Since the economy is at the ZLB, increases in the interest rates do not undo this wealth effect, as will be the case outside the ZLB. The authors illustrate this mechanism with a simple two-period New Keynesian model. They discuss possible objections to this set of policies and the relation of supply-side policies to more conventional monetary and fiscal policies.

Working Paper 11-47, "Supply-Side Policies and the Zero Lower Bound," Jesús Fernández-Villaverde, University of Pennsylvania, and Visiting Scholar, Federal Reserve Bank of Philadelphia; Pablo Guerrón-Quintana, Federal Reserve Bank of Philadelphia; and Juan F. Rubio-Ramírez, Duke University, and Visiting Scholar, Federal Reserve Bank of Philadelphia

FINDING THE OPTIMAL LABOR MARKET-POLICY MIX OVER THE BUSINESS CYCLE

The authors examine the optimal labor market-policy mix over the business cycle. In a search and matching model with risk-averse workers, endogenous hiring and separation, and unobservable search effort, they first show how to decentralize the constrained-efficient allocation. This can be achieved by a combination of a production tax and three labor-market policy instruments, namely, a vacancy subsidy, a layoff tax, and unemployment benefits. The authors derive analytical expressions for the optimal setting of each of these for the steady state and for the business cycle. Their propositions suggest that hiring subsidies, layoff taxes, and the replacement rate of unemployment insurance should all rise in recessions. They find this confirmed in a calibration targeted to the U.S. economy.

Working Paper 11-48, "Optimal Labor-Market Policy in Recessions," Philip Jung, University of Mannheim, and Keith Kuester, Federal Reserve Bank of Philadelphia

WHY HAS THE U.S. NONFINANCIAL CORPORATE SECTOR BECOME A NET LENDER?

The authors document that the U.S. nonfinancial corporate sector became a net lender in the 2000s, using aggregate and firm-level data. They develop a structural model with investment, debt, and equity. Debt is fiscally advantageous but subject to a no-default borrowing constraint. Equity allows the firm to suspend dividends when the cash flow is negative. Firms accumulate financial assets for precautionary reasons, yet value equity as partial insurance against shocks. The calibrated model replicates the prevalence of net savings in the period 2000-2007 and attributes the rise in corporate savings over the past 40 years to lower dividend taxes.

Working Paper 12-1, "The Macroeconomics of Firms' Savings," Roc Armenter, Federal Reserve Bank of Philadelphia, and Viktoria Hnatkovska, University of British Columbia and the Wharton School

ASSESSING VARIOUS APPROACHES TO MODELING THE SEPARATION MARGIN

This paper assesses how various approaches to modeling the separation margin affect the ability of the Mortensen-Pissarides job matching model to explain key facts about the aggregate labor market. Allowing for realistic time variation in the separation rate, whether exogenous or endogenous, greatly increases the unemployment variability generated by the model. Specifications with exogenous separation rates, whether constant or time-varying, fail to produce realistic volatility and productivity responsiveness of the separation rate and worker flows. Specifications with endogenous separation rates, on the other hand, succeed along these dimensions. In addition, the endogenous separation model with on-the-job search yields a realistic Beveridge curve correlation and performs well in accounting for the productivity responsiveness of market tightness. While adopting the Hagedorn-Manovskii calibration approach improves the behavior of the job finding rate, the volume of job-to-job transitions in the on-the-job search specification becomes essentially zero.

Working Paper 12-2, "Exogenous vs. Endogenous Separation," Shigeru Fujita, Federal Reserve Bank of Philadelphia, and Garey Ramey, University of California, San Diego

ESTIMATING A DSGE MODEL TO EXPLAIN VARIABILITY IN OUTPUT IN SMALL DEVELOPED ECONOMIES

Using an estimated dynamic stochastic general equilibrium model, the author shows that shocks to a common international stochastic trend explain, on average, about 10 percent of the variability of output in several small developed economies. These shocks explain roughly twice as much of the volatility of consumption growth as the volatility of output growth. Country-specific disturbances account for the bulk of the volatility in the data. Substantial heterogeneity in the estimated parameters and stochastic processes translates into a rich array of impulse responses across countries.

Working Paper 12-3, "Common and Idiosyncratic Disturbances in Developed Small Open Economies," Pablo Guerrón-Quintana, Federal Reserve Bank of Philadelphia

BAYESIAN METHODS FOR ESTIMATING NEW KEYNESIAN DSGE MODELS

The authors survey Bayesian methods for estimating dynamic stochastic general equilibrium (DSGE) models in this article. They focus on New Keynesian (NK) DSGE models because of the interest shown in this class of models by economists in academic and policy-making institutions. This interest stems from the ability of this class of DSGE model to transmit real, nominal, and fiscal and monetary policy shocks into endogenous fluctuations at business cycle frequencies. Intuition about these propagation mechanisms is developed by reviewing the structure of a canonical NKDSGE model. Estimation and evaluation of the NKDSGE model rests on being able to detrend its optimality and equilibrium conditions, to construct a linear approximation of the model, to solve for its linear approximate decision rules, and to map from this solution into a state space model to generate Kalman filter projections. The likelihood of the linear approximate NKDSGE model is based on these projections. The projections and likelihood are useful inputs into the Metropolis-Hastings Markov chain Monte Carlo simulator that the authors employ to produce Bayesian estimates of the NKDSGE model. The authors discuss an algorithm that implements this simulator. This algorithm involves choosing priors of the NKDSGE model parameters and fixing initial conditions to start the simulator. The output of the simulator is posterior estimates of two NKDSGE models, which are summarized and compared to results in the existing literature. Given the posterior distributions, the NKDSGE models are evaluated with tools that determine which is most favored by the data. The authors also give a short history of DSGE model estimation as well as point to issues that are at the frontier of this research.

Working Paper 12-4, "Bayesian Estimation of DSGE Models," Pablo Guerrón-Quintana, Federal Reserve Bank of Philadelphia, and James Nason, Federal Reserve Bank of Philadelphia