Federal Reserve Bank of Chicago

Announcement

2001 Conference on Bank Structure and Competition

First Quarter 2001

Economic perspectives

Economic perspectives

Economic Perspectives is published by the Research Department of the Federal Reserve Bank of Chicago. The views expressed are the authors' and do not necessarily reflect the views of the Federal Reserve Bank of Chicago or the Federal Reserve System.

Single-copy subscriptions are available free of charge. Please send requests for single- and multiple-copy subscriptions, back issues, and address changes to the Public Information Center, Federal Reserve Bank of Chicago, P.O. Box 834, Chicago, Illinois 60690-0834, telephone 312-322-5111 or fax 312-322-5515.

Economic Perspectives and other Bank publications are available on the World Wide Web at http://www.frbchi.org.

ISSN 0164-0682

Contents

First Quarter 2001, Volume XXV, Issue 1

The crisis of 1998 and the role of the central bank

David A. Marshall

Introduction and summary

A key mission of the U.S. Federal Reserve System is to safeguard the economy against systemic financial crises. This concern with financial crises stems from a long-held belief that they are associated with declines in economic activity. In the U.S., there is clear evidence that financial panics and recessions are somehow related (Mishkin, 1991). In the case of the Great Depression, Bernanke (1983) argues that the disruption in financial intermediation transformed a severe downturn into a "protracted depression." More recently, the Asian financial crisis in 1997 was followed by sharp declines in economic activity. (Indonesia, Korea, Thailand, and Malaysia all experienced two-quarter declines in gross domestic product [GDP] of over 12 percent.) This historical record has led to a pervasive belief that systemic crises in the financial sector have consequences that are far more than sectoral. Rather, they appear to affect the entire economy, perhaps through the unique role played by financial intermediation.

The most recent financial crisis in the U.S. occurred in late summer and fall of 1998. On August 17, the Russian government devalued the rouble, defaulted on its rouble-denominated debt, and imposed a moratorium on payments to foreign creditors of Russian financial institutions. Following these actions, asset values fell precipitously in all Group of Seven (G-7) countries, and there is evidence of widespread withdrawal of liquidity from financial markets. Particularly dramatic was the near collapse and eleventh-hour recapitalization in late September of Long-Term Capital Management (LTCM), a large hedge fund.

From a U.S. perspective, these events might be described as an "incipient" crisis, because there is little evidence of damage to western economies. Arguably, this is because of the decisive action by the Federal Reserve in cutting the target federal funds rate in three successive 25 basis-point moves. The second of these moves, on October 15, was particularly noteworthy, since it occurred between regularly scheduled meetings of the Federal Open Market Committee (FOMC). Intermeeting rate cuts of this type are rare; the October 15 action was the first such action since April 1994. In the next section, I provide evidence that the end of this incipient crisis coincided almost exactly with the October 15 rate cut. In particular, credit spreads abruptly narrowed on October 16 (one day after the Federal Reserve move), and stock markets in all G-7 countries started to recover a week to ten days prior to the October 15 move. (That stock markets anticipated the rate cut is no surprise. For at least a week prior to the move the financial press reported rumors of a possible intermeeting rate cut.)

The way this incipient crisis ended is somewhat puzzling. The crisis had a clear trigger: the Russian default and devaluation in mid-August. Western financial institutions were directly affected by the default if they held Russian liabilities. Furthermore, the Russian default may have signaled higher default risk for sovereign debt from other emerging or transition economies. So it is not surprising that uncertainty grew about institutions' solvency (with attendant increases in asset price volatility and credit spreads). What is puzzling is the way the crisis appears to have abated with the Fed's second rate cut. Why would the problems (both direct and informational) associated with the Russian default be reduced by a mere 50 basis-point cumulative cut in the overnight interest

David A. Marshall is a senior economist and economic advisor at the Federal Reserve Bank of Chicago. *The author would like to thank Jim Moser, Eric Swanson, François Velde, and Ruilin Zhou for extensive discussions and Darrin Halcomb and Genny Pham-Kantor for research assistance.*

rate? If the crisis was associated with higher default risk of emerging-economy debt, why would the Fed's rate cut have dramatically reduced this default risk? Similarly, if the crisis was associated with an increased informational asymmetry among financial institutions, why would the 50 basis-point cut in the federal funds rate have reduced this asymmetry?

In this article, I argue that the crisis can be characterized as an episode of potential *coordination failure*, triggered by—but ultimately distinct from—the events in Russia. I propose a simple model of financial crises as coordination failure. The model qualitatively matches the following features typically associated with financial crises:

1. Abrupt shifts between a state of adequate liquidity provision and a state of aggregate illiquidity (the latter being a case where institutions with liquidity refuse to lend to those needing liquidity);
2. A "flight to quality," whereby institutions with funds to invest preferentially choose a low-risk, low-return asset;
3. Fear among lenders that credit quality among potential borrowers has deteriorated;
4. Real costs in economic output;
5. Sudden declines in asset values; and
6. A role for the central bank's open market operations in containing the crisis.

In particular, the model provides one potential explanation for why the Federal Reserve action on October 15, 1998, eliminated the danger of a full-blown crisis. This model contributes to the growing literature developing formal models of financial crises and financial fragility. Notable examples of these include Chang and Velasco (1998), Louganoff and Schreft (1998), DenHaan, Ramey, and Watson (1999), and Chari and Kehoe (1998, 2000).

Coordination failure can emerge in any economy where the profitability of a given agent's investment depends on the decisions of the other agents in the economy. In the model of this article, the possibility of coordination failure arises from the essential function of financial markets: to match potential users of capital (borrowers) with potential providers of capital (lenders) in an environment of asymmetric information. Borrowers and lenders match via a search procedure. In this model, multiple equilibria are possible. A *high-coordination equilibrium* can occur in which all lenders and all borrowers enter the match, maximizing the expected output of the economy. However, there are times when a *low-coordination equilibrium* is possible in which all good quality (that is, highly creditworthy) borrowers refrain from entering the match. Knowing that only poor quality borrowers seek loans, potential lenders refuse to lend. I identify this low-coordination equilibrium with a financial crisis.

In the model, the low-coordination equilibrium cannot exist if the risk-free real interest rate is sufficiently low. This suggests a potential role for the central bank. If monetary policy can affect real interest rates, the central bank can extinguish the low-coordination equilibrium if it reduces the real interest rate sufficiently via an aggressive monetary expansion. That the Federal Reserve has the power to do so is suggested by the events of 1991–93. As I discuss in a later section, there is evidence that banks cut back on lending activity in the early 1990s. Following a shift to a more expansionary monetary policy in mid-1991, in which the real federal funds rate fell from 2.5 percent to 0.5 percent, lending activity moved back to normal levels.

Unlike the Federal Reserve action of 1991–93, the monetary expansion in fall 1998 was too small, and the consequent effect on real interest rates too marginal, to have a substantial direct effect on lender incentives. Rather, I interpret the intermeeting rate cut of October 15 as a signal that the Federal Reserve's *policy rule* had changed. Before the intermeeting move, market participants were uncertain whether the Federal Reserve would compromise its focus on price stability (and the associated tight money policy) even in the face of severe financial market strains. I argue that the intermeeting move was interpreted by market participants as signaling a shift to a *state-contingent* policy: focus on price stability *unless* a financial crisis becomes imminent; temporarily abandon that focus if the threat of financial crisis becomes severe. In this article, I formally model such a policy, and I show that, in principle, such a policy can extinguish the low-coordination equilibrium. Furthermore, if this policy is credible, it never has to be implemented: The policy itself removes the possibility of coordination failure. That is, monetary expansion is an "off-equilibrium path" that enforces the high-coordination equilibrium.

Below, I review the facts of the crisis of fall 1998, highlighting key features that I will seek to replicate in the theoretical model. Then, I describe the basic coordination failure model. Finally, I show how the central bank can avert coordination failure by implementing an appropriate and credible state-contingent monetary policy.

Brief review of the events of fall 1998

Here, I review the crisis and provide evidence for the following assertions:

1. The crisis was associated with large declines in equity values, increased volatility in financial markets, widening credit spreads, and an increased demand for U.S. Treasury securities.
2. During the crisis, there was a reduction in available liquidity, as institutions with loanable funds reduced the volume of funds available to the market.
3. The crisis rapidly abated following the Federal Reserve's intermeeting cut in the federal funds rate on October 15, 1998.

Financial markets showed evidence of potential problems starting around mid-July 1998. However, the onset of the crisis is usually associated with the Russian devaluation and default in August 1998. This denouement was in large part forced by declining hard currency inflows over the preceding several months as oil prices fell. On August 17, Russia defaulted on its rouble-denominated public debt. At that time, this stock of debt represented $61 billion, 17 percent of Russian GDP. At the same time, Russia declared a 90-day moratorium on all foreign obligations of Russian financial institutions. Finally, the rouble exchange rate zone was substantially widened, amounting to a de facto devaluation of 25 percent. The exchange rate zone was completely abandoned ten days later.[1] As I discuss below, western financial markets reacted negatively to these developments. In response, the FOMC cut the federal funds rate by 25 basis points at its next regular meeting on September 29. This move by the Federal Reserve did not calm the financial markets.[2] On October 15, in an unusual intermeeting move, the FOMC made an additional 25 basis-point rate cut. Observers point to this intermeeting move as marking the end of the crisis.

The data in figure 1 characterize more fully the impact of these events on western financial markets. As shown in figure 1, panel A, the U.S. S&P 500 index peaks in mid-July 1998, with a small local peak in mid-August 1998 (vertical dashed line) following the Russian default. Thereafter, there is a sharp decline in stock values, amounting to more than 18 percent over the three months from peak to trough. The S&P 500 index bottomed out on October 8, one week before the FOMC's intermeeting rate cut on October 15 (vertical solid line). The biggest close-to-close rise of this period was from October 14 to October 15. It is no surprise that the market trough occurred one week before the Fed intermeeting action, since there was speculation prior to the Fed's action that an intermeeting rate cut was likely.[3] The behavior of the federal funds futures market supports this interpretation. Through October 7, futures prices implied an expected federal funds rate through the end of October of 5.22 percent to 5.24 percent, implying little probability of a rate cut. On October 8, this expected federal funds rate dropped to 5.18 percent, which is consistent with a 50 percent probability of a quarter-point rate cut around mid-October.[4]

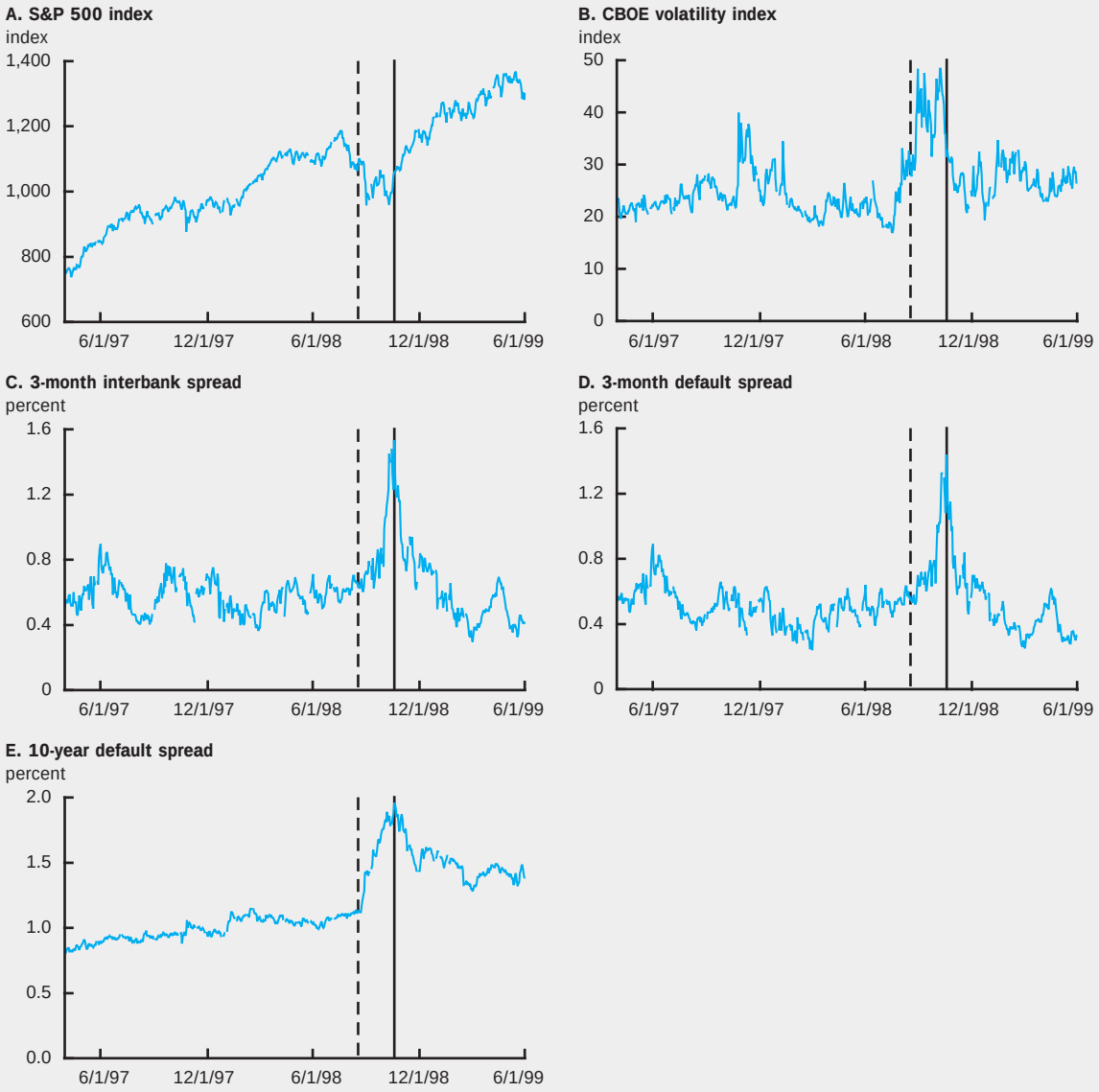
The behavior of stock indexes for the other six countries in the G-7 is roughly comparable to that of the U.S. indexes. In all cases, the market peaks in mid-July, falls steeply, and starts to turn up about one week before the October 15 rate cut. The total market declines over this three-month period were quite pronounced, ranging from 18 percent in Japan to over 28 percent in Canada, Italy, and France.[5]

Figure 1, panel B displays the value of the Chicago Board Options Exchange (CBOE) volatility index (a measure computed by the CBOE from implied volatility on a number of option contracts).[6] These data show that uncertainty (and associated risk) in financial markets rose steeply in mid-August 1998. The date of the first pronounced jump was actually August 27, when the closing value of this index rose to 39.16 (compared with the previous day's close of 30.66). This date corresponds to the Russian government's announcement that it was abandoning its trading band for the rouble. In trading during August 27, the rouble fell 40 percent against the deutschemark. In addition, on that date Deutsche Bank lost its AAA rating from Standard and Poor's when it revealed that it had unsecured Russian credit risk amounting to almost $750 million. For the next seven weeks the volatility index stayed at a level that was unprecedented, except for the period around the 1987 stock market crash. The index remained at or near 40 until October 15 (the date of the intermeeting rate cut),[7] when it fell to 35.95 (compared with the previous day's close of 41.31). Within two trading days the index had fallen to around 30, remaining between 20 and 30 through the end of 1999.

Figure 1, panels C, D, and E display three U.S. credit spreads: the interbank spread (three-month interbank yield minus three-month T-bill yield), the short-term credit spread (three-month commercial paper yield minus three-month T-bill yield), and the long-term credit spread (ten-year AAA corporate bond yield minus ten-year T-bond yield). These credit spreads confirm the inference from figure 1, panel B that there was an abrupt increase in perceived credit risk from mid-August to mid-October. In particular, they show a pronounced spike starting in late September 1998 (around the time of the LTCM rescue) and continuing until October 16, one day after the FOMC's intermeeting rate cut. The peak in the long-term credit

FIGURE 1

U.S. financial market indicators, March 1997–June 1999



Notes: The vertical dashed lines indicate August 17, 1997, the date of the Russian default. The vertical solid lines indicate October 15, 1997, the date of the Federal Reserve's intermeeting cut in the federal funds rate. Panels C, D, and E display three U.S. credit spreads: the interbank spread (panel C) is the three-month interbank yield minus three-month Treasury bill yield. The three-month default spread (panel D) is the three-month commercial paper yield minus three-month Treasury bill yield. The ten-year default spread is the ten-year AAA corporate bond yield minus ten-year Treasury bond yield.

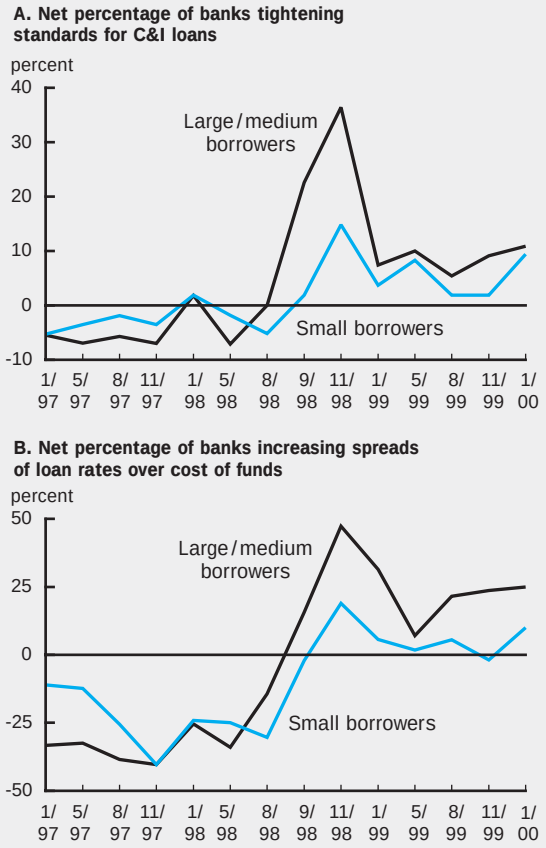
Sources: Board of Governors of the Federal Reserve System (panels A, C, D, E); Chicago Board Options Exchange (panel B).

spread during this period is the highest in the 1990s, and the peak in the other two credit spreads is only exceeded during this decade by that observed during the 1990–91 recession.[8]

There is evidence that the increase in perceived credit risk was associated with a substantial drying-up of liquidity. That is, institutions with loanable funds became more reluctant to extend unsecured loans. The Federal Reserve Board of Governors' Senior Loan Officers Survey in September[9] revealed a marked increase over the August survey in the number of banks tightening loan standards and raising loan rates. (See figure 2, panels A and B.) A principal reason reported by banks for these actions was "a

FIGURE 2

Measures of credit market tightness



Notes: This figure plots data from the Federal Reserve Board's Senior Loan Officers Surveys from January 1997 through January 2000. Responses to the January, March, August, and November surveys report bank credit policies over the preceding three months. The September 1998 survey reports bank credit policies over the preceding month only.

Source: Board of Governors of the Federal Reserve System, Senior Loan Officers Survey.

reduced tolerance for risk." Interestingly, there was a substantial increase in the number of banks reporting *decreased* loan demand. The respondents generally attributed this reduced demand to reductions in both merger and acquisition activity and fixed investment. This suggests that the reduction in loan activity was due to both a reduced willingness of lenders to bear default risk *and* a reduced interest of borrowers in expanding economic activity. While reports of tightened loan standards continue through the November survey, the reduction in loan demand appears to have reversed by November.

The Bank for International Settlements (BIS, 1999) surveyed a number of market participants about the events of fall 1998. The survey results confirm the perception that risk levels were elevated and liquidity provision diminished in the period from mid-August through mid-October 1998. They point to an "unprecedented" widening in bid/ask spreads and even to "one-sided markets," where sellers of risky securities could not find a buyer at any price. On numerous occasions, market makers in government securities simply withdrew from trading and refrained from posting quotes.[10] The BIS interviewees report a flight to the most liquid, "on-the-run" (that is, most recently issued) Treasury securities. For example, by early October, the yield spread between 28-year and 30-year Treasury bonds had widened to 29 basis points from just 7 basis points in mid-August (although the 28-year issues are just as free from default risk as the 30-year on-the-run bonds). Salomon Smith Barney reported that this spread was the widest it had ever recorded.[11] Continued ability to trade Treasury securities in any desired quantity was assured only for the on-the-run issues. The flight to quality even devolved, "for a brutal but short-lived period"[12] to a flight to cash. A number of participants reported reductions in credit lines to other financial institutions. This drying-up of liquidity exacerbated price volatility and increased credit risk associated with institutions that relied on market funding. Interestingly, the infusion of funds to LTCM, facilitated by the Federal Reserve Bank of New York in late September 1998,[13] seemed to *exacerbate* the liquidity crisis. Participants in the BIS survey interpreted the Federal Reserve's role as a signal that the Federal Reserve believed that the crisis was far worse than previously thought. Finally, the BIS interviewees perceived the October 15 rate cut as the turning point of the crisis. In its summary of interviews with market participants, the BIS states, "The second monetary easing by the Federal Reserve (15 October) signaled the beginning of the abatement of financial strains. At that time, traders clearly understood the commitment of the Federal Reserve to fix the problems."

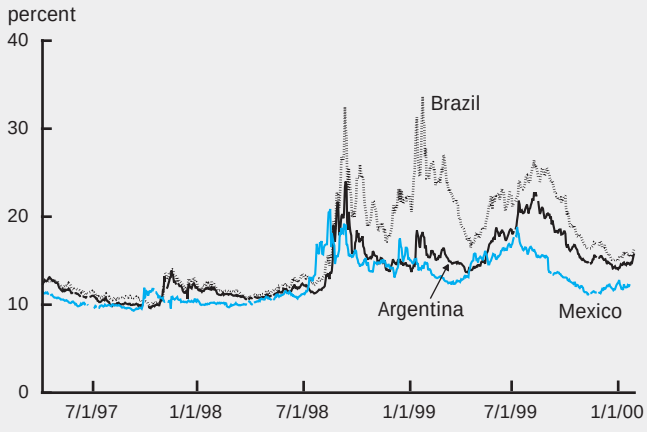
To summarize, the period from mid-August to mid-October 1998 was characterized by rapid declines in stock values, rapid increases in uncertainty, and a reluctance of institutions with loanable funds to provide loans. The crisis appears to have abated in U.S. financial markets with the Fed's intermeeting rate cut of October 15. In particular, the stock market recovery, the narrowing of credit spreads in fixed income markets, and the decline in the CBOE volatility index all commenced around October 15. Other more qualitative measures of the crisis, such as the Board of Governor's Senior Loan Officers Survey, the BIS interviews with market participants, and reports in the financial press, are also consistent with this interpretation.

What generated the crisis?

It is perhaps no surprise that the Russian default and devaluation triggered turmoil in western financial markets. There was a good deal of uncertainty about the direct exposure of western financial institutions to the Russian default. Furthermore, western investors may have interpreted the Russian default as evidence against the creditworthiness of other emerging economies. Investors were particularly concerned about Argentina, Brazil, and Mexico, which are far more important than Russia for U.S. trade.[14] (In fact, Brazil devalued its currency in mid-January 1999.) Figure 3, which plots Brady bond yields,[15] shows how the Russian default triggered an increase in perceived credit risk for these three Latin American countries that eclipsed the increase following the 1997 Asian crisis. In all three countries, the yields more than doubled following the Russian default in mid-August 1998. (The yield spike for Brazil was particularly pronounced.) However, this explanation for the crisis does not fully account for the way it ended. It is hard to imagine that the exposure of western institutions to emerging and transitional economies or the informational asymmetry about these exposures would have been reduced by a 50 basis-point reduction in the federal funds rate. Similarly, the creditworthiness of borrowers in these economies would not have been affected substantially by the Federal Reserve's action. Thus, while the Russian default clearly triggered the financial crisis, the crisis appears to have taken on a self-fulfilling aspect over and above the damage attributable to the actions of the Russian government.

FIGURE 3

Brady bond yields for Argentina, Brazil, and Mexico
(April 1997–January 2000)



Note: This figure plots yields from dollar-denominated sovereign debt ("Brady bonds") issued by Argentina, Brazil, and Mexico.
Source: Bloomberg.

Other financial crises have also involved sudden shifts between crisis and non-crisis states without a commensurate change in fundamentals. The Asian crisis of 1997 provides an example, although in that case the sudden shift occurred at the beginning of the crisis. The Asian crisis was completely unforeseen by financial markets. In particular, in none of the Asian crisis countries do interest rates or forward exchange rates move prior to the speculative attacks leading to the initial Thai devaluation.[16] Furthermore, the Asian crisis was not triggered by any shock to fundamentals commensurate with the magnitude of the subsequent debacle. While there were clear problems with market fundamentals in these countries (in particular, the poor state of their banking sectors), these problems were well known months or even years prior to the crisis.[17] It appears that any theory of systemic financial crisis must incorporate the possibility of sudden, untriggered shifts between crisis and non-crisis states.

Modeling financial crisis as coordination failure[18]

As described earlier, the financial crisis of fall 1998 had a number of characteristics that have been associated with crises more generally. There was a sudden shift between crisis and non-crisis states without a commensurate change in fundamentals. The crisis state was characterized by a sharp reduction in liquidity provision with a corresponding flight to quality. The crisis was associated with a decline in asset values, as reflected in stock market indexes.[19] In addition, the crisis of 1998 shows clear evidence of an increase in perceived default risk. Finally, the end of the 1998 crisis was associated with an unusual action by the central bank (a change in the target federal funds rate between regularly scheduled FOMC meetings).

In this section, I propose a simple model of financial crisis that, in principle, can accommodate these patterns. My approach focuses on the possibility of coordination failure. In coordination models, an investor benefits if he chooses the same strategy as other investors. Thus, investors will tend to "coordinate" on a particular strategy. A multiplicity of equilibria can emerge, each associated with a different pattern of coordination. Suboptimal equilibria are then associated with coordination failure: the failure to coordinate on the socially optimal choices. In a familiar example, known as *external increasing*

returns to scale, the productivity of a particular firm's capital investment is high only if there is a high level of aggregate economic activity. Therefore, a firm may only want to choose a high level of investment if enough *other* firms also choose a high level of investment (thereby assuring a high level of aggregate activity). If other firms choose low investment, aggregate activity will be low, and an individual firm's investment productivity may be too low to justify a high investment level. In this example, there are two equilibria: one where all firms "coordinate" on high investment, the other where all firms have low investment.[20]

In the model I present here, the possibility of coordination failure arises from the essential nature of financial relationships—the need to match potential borrowers with potential lenders in an environment of asymmetric information. In particular, lenders must search for borrowers and vice versa. As the total number of borrowers and lenders rises, this search process becomes more productive. That is, the rate at which borrowers and lenders match goes up. In other words, the matching process exhibits a *thick markets externality*: Everyone benefits as the number of participants in the market rises.[21]

This thick markets externality gives rise to the possibility of a coordination failure equilibrium: If lenders believe that there are few high-quality borrowers searching for loans, and simultaneously the high-quality borrowers believe that there are few lenders willing to extend credit, an equilibrium can emerge where both lenders and borrowers forsake the loan market in favor of alternative investments. In effect, all parties have "coordinated" on nonparticipation, so the optimal strategy for any individual agent is not to participate.

Basic structure of the model

The basic model is completely static. There are two types of risk-neutral agents: borrowers (N_{borr} in number), who are endowed with a project but no liquidity; and lenders (N_{lend} in number), who are endowed with one unit of liquidity but no project. A borrower can operate his project in two mutually exclusive ways: *autarkically*, without any liquidity inflow from outside; or *with investment*, which requires borrowing one unit of liquidity from a lender. A borrower must decide at the beginning of the period whether to operate the project autarkically or whether to seek a loan. In other words, once the borrower has decided to seek a loan, the possibility of autarkic production is precluded.

Borrowers are randomly assigned one of two types of projects, *bad* (assigned with probability p^b), and *good* (assigned with probability $(1 - p^b)$). The quality of the project is private information to the borrower. Good projects pay $R^{autarky}$ with certainty if operated autarkically; they pay R with certainty if operated with investment, *provided* the borrower has found a lender willing to lend. Bad projects pay 0 if operated autarkically; if operated with investment, bad projects pay R with probability θ and $R^{salvage}$ with probability $(1 - \theta)$, again provided borrower has found a lender willing to lend. Informally, a bad borrower defaults on his loan with probability $(1 - \theta)$; $R^{salvage}$ represents the salvage value of the project that is available to satisfy the lender's claim. Finally, if a borrower seeks a loan but fails to match with a lender, he receives zero.

An interpretation[22] of these two types of borrowers is that bad borrowers are in severe financial distress. If they do not get an immediate liquidity infusion, they will be forced into bankruptcy. Even if they do receive liquidity, financial distress may impair their productivity with probability $(1 - \theta)$. In contrast, good borrowers can stay in operation without liquidity, albeit at a lower output. However, there is an up-front cost to structuring the project to utilize liquidity. My assumption that a good borrower who tries to obtain a loan and fails receives zero is equivalent to a specification where the up-front cost equals $R^{autarky}$ and the output with liquidity (before the up front cost is paid) equals $R + R^{autarky}$.

Lenders have one unit of liquidity, which they can use in two mutually exclusive ways. First, they can invest it at a gross risk-free rate R^f. Second, they can attempt to find a borrower to whom to lend. If a borrower and a lender match, the loan contract takes the following exogenous specification:[23] If R is produced, the lender receives R_{lend} (where R_{lend} is an exogenous parameter satisfying $R_{lend} \geq R^{salvage}$) and the borrower receives $R_{borr} \equiv R - R_{lend}$; if $R^{salvage}$ is produced, the borrower is in default, so the lender receives the full salvage value $R^{salvage}$ and the borrower receives nothing. Finally, if a lender does not find a borrower, she simply ends up with her unit of liquidity.[24] To summarize, the payoffs are as follows:

1) Payoff to good borrower =

R_{borr}	if borrower matches with lender
0	if borrower attempts to match with lender and fails
$R^{autarky}$	if borrower operates project autarkically.

2) Payoff to bad borrower =

R_{borr}	if borrower matches with lender and project produces R

0 if borrower attempts to match with lender and fails *or* if borrower matches with lender and project produces $R^{salvage}$ *or* if borrower operates project autarkically.

3) Payoff to lender =

R_{lend} if lender matches with borrower and project produces R

$R^{salvage}$ if lender matches with borrower and project produces $R^{salvage}$

1 if lender attempts to match with a borrower and fails

R^f if lender uses the risk-free investment and does not attempt to match with a borrower.

The matching procedure

According to equations 1, 2, and 3, the expected payoff to an agent who attempts to match depends on the probability of consummating the match. Suppose that there are a total of B borrowers seeking loans and L lenders seeking to match with borrowers. I denote the probability that a given borrower matches with a lender by $prob_{borr}(B,L)$. Similarly, the probability that a given lender matches with a borrower is denoted $prob_{lend}(B,L)$. (In equilibrium, the expected number of matches equals $B \times prob_{borr}(B,L) = L \times prob_{lend}(B,L)$.)

If either B or L equals zero, both $prob_{borr}$ and $prob_{lend} = 0$. (That is, if there are no borrowers or lenders, there can be no matches.) It is also natural to assume, in the language of Mortensen and Pissarides (1998), that borrowers and lenders are *complements*. That is, it is easier for a borrower to find a match if there are more lenders, and vice versa. (Formally, $\frac{\partial prob_{borr}}{\partial L} > 0$ and $\frac{\partial prob_{lend}}{\partial B} > 0.$) In addition, following Mortensen and Pissarides (1998), I assume that there is a *congestion effect*. An increase in the number of borrowers decreases the probability that a given borrower will match, and vice versa. (Formally, $\frac{\partial prob_{borr}}{\partial B} < 0$ and $\frac{\partial prob_{lend}}{\partial L} < 0.$)

Finally, I assume that the expected number of matches displays *increasing returns to scale*. This is equivalent to the condition that as the number of borrowers and lenders increases equiproportionally, both $prob_{lend}$ and $prob_{borr}$ increase. This is a natural assumption to make for many types of matching problems. Consider the problem of finding a taxi cab in a medium-sized city. If there were only one rider looking for a cab and one cab looking for a fare (as might be the case at 2:00 am), the probability of a match would be very low. If there were 10,000 riders and 10,000 cabs, the probability that a given rider would find a cab would be much higher. (This intuition is formalized in the model developed in appendix A.)[25]

Increasing returns is implied by a number of search models that have been proposed in the literature. In appendix A, I discuss a number of these and I develop one model in detail. Diamond (1982) and others note that increasing returns in the matching technology can give rise to multiple search equilibria. I exploit this feature below.[26]

High-coordination and low-coordination equilibria

The matching technology implies that the decision of borrowers whether to enter the match affects the probability that a given lender will match and, therefore, affects the expected payoff to the lender from entering the match. Similarly, the decisions of lenders affect the expected payoff of the borrower. This implies the possibility of coordination failure between borrowers and lenders and, thus, the possibility of multiple equilibria. I define a *high-coordination equilibrium* as one in which all lenders enter the match and all borrowers enter the match. Of course, a lender will enter the match if, and only if, her expected payoff from entering the match equals or exceeds R^f. Using the payoffs given in equation 3, in a conjectured high-coordination equilibrium this condition can be written as

4) $prob_{lend}(N_{borr}, N_{lend})((p^b\theta + 1 - p^b)R_{lend} + p^b(1-\theta)R^{salvage}) + (1 - prob_{lend}(N_{borr}, N_{lend})) \geq R^f.$

Similarly, a good borrower will enter the match, if and only if, his expected payoff from entering the match equals or exceeds $R^{autarky}$. In a conjectured high-coordination equilibrium, this condition is

5) $prob_{borr}(N_{borr}, N_{lend})R_{borr} \geq R^{autarky}).$

Note from equation 2 that the bad borrowers always enter the match, since the payoff from entering the match dominates the autarkic payoff to the bad borrower of zero. Therefore, equations 4 and 5 are sufficient for the existence of a high-coordination equilibrium.

I define a *low-coordination equilibrium* as one where no lenders enter the match and only bad borrowers enter the match. The payoff to lenders in the low-coordination equilibrium is R^f. The payoff to good borrowers is $R^{autarky}$, and the payoff to bad borrowers is zero. If there are no lenders in the match, there is clearly no incentive for good borrowers to defect from the equilibrium strategy of autarky (since

$prob_{borr}(1,0) = 0$). However, there may be an alternative strategy for a lender that could break the low-coordination equilibrium. Let the total number of bad borrowers be denoted N_{bad}. (The expected value of N_{bad} is simply $p^b N_{borr}$.) If, starting in a low-coordination equilibrium, a lender decides to defect from the equilibrium strategy by entering the match, her probability of matching with a borrower is $prob_{lend}(N_{bad}, 1)$ (since only bad borrowers are in the match in a low equilibrium). Her expected payoff conditional on a successful match is $\theta R_{lend} + (1 - \theta) R^{salvage}$. Therefore, a low-coordination equilibrium can only be sustained if the expected payoff to this alternative strategy is less than the payoff to a lender in the low-coordination equilibrium:

$$6) \quad prob_{lend}(N_{bad}, 1)(\theta R_{lend} + (1 - \theta)R^{salvage}) + (1 - prob_{lend}(N_{bad}, 1)) \leq R^f.$$

The left-hand sides of equations 4 and 6 give the value to a lender of entering the match in the high-coordination and low-coordination equilibria, respectively. Similarly, the left-hand side of equation 5 gives the value to a good borrower of entering the match in the high-coordination equilibrium. For a particular base line parameterization,[27] figure 4 displays how these values are affected by changes in the model parameters. Specifically, the left-hand column of plots in figure 4 shows how the left-hand sides of equations 4 (black lines) and 6 (colored lines) change as a particular model parameter is varied; the right-hand column does the same for the left-hand side of equation 5 (colored lines). The five parameters that are varied in figure 4 are: number of lenders, as a fraction of total population (first row of plots); R^{lend}, as a fraction of R (second row); $R^{salvage}$, as a fraction of R (third row); θ (fourth row); and p^b (fifth row).

The behavior of these values is intuitive. The value to lenders of entering the match for both equilibria is strictly decreasing in the ratio of lenders to total population (reflecting the greater competition from other lenders); the corresponding value to good borrowers is strictly increasing in this ratio (reflecting the higher probability of matching with a lender). Not surprisingly, increasing R_{lend}/R, the fraction of output received by lenders, increases the value of the match to lenders, but decreases that value to borrowers. For both equilibria, the value of the match is increasing for lenders in $R^{salvage}$ and θ (the probability that a bad project produces R). Neither of these parameters affects the value of the match for good borrowers. Finally, an increase in p^b, the probability of bad projects, reduces the value of the match for lenders in the high-coordination equilibrium, but increases the value of the match for lenders in the low-coordination equilibrium. In the high-coordination equilibrium, increasing p^b simply increases the probability of borrower default. In the low-coordination equilibrium, however, an increase in p^b increases the number of borrowers seeking loans. (Recall that only bad borrowers seek loans in the low-coordination equilibrium.) This increases the probability of a match for a lender contemplating deviating from the equilibrium strategy.

Suppose the borrower condition for a high-coordination equilibrium (equation 5) holds. That is, suppose $R^{autarky}$ lies below the colored line in any of the plots in the right-hand column of figure 4. Then the existence of the high- or low-coordination equilibrium depends on the level of the risk-free rate R^f relative to the solid and colored lines in the plots in the left-hand column. If R^f is above both lines, then neither equilibrium exists for these parameter values. If R^f is below the solid line but above the colored line, then both low-coordination and high-coordination equilibria exist. If R^f is below both the solid line and the colored line, then a high-coordination equilibrium exists but no low-coordination equilibrium exists. Thus, if equation 5 holds, the high-coordination equilibrium can be enforced by setting R^f sufficiently low.

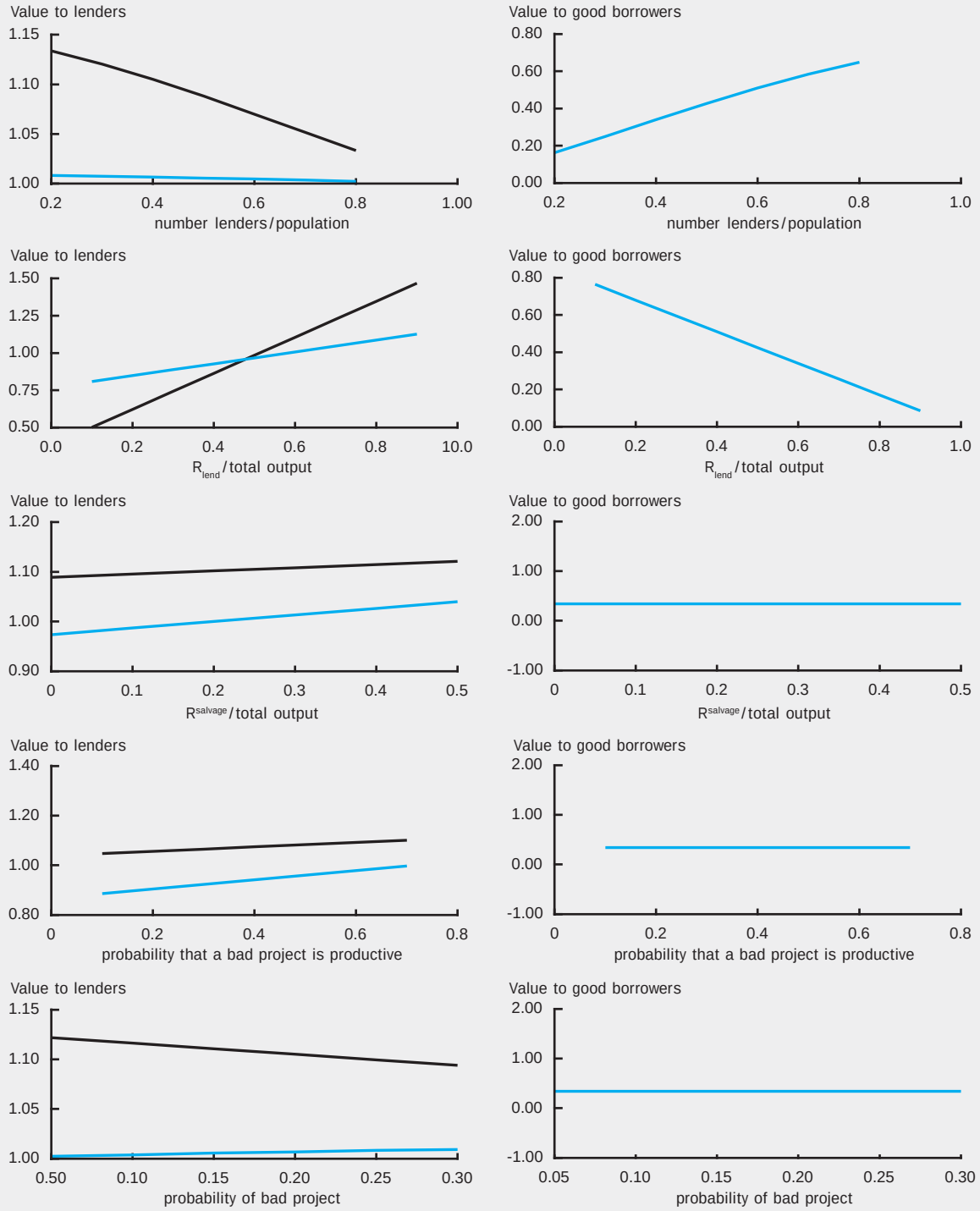
Finally, there may also be additional "mixed" equilibria where a fraction of lenders and/or good borrowers enter the match, while the remaining agents choose the alternative strategies (investing risk-free for lenders, operating the project autarkically for the borrowers.) I discuss the conditions for these mixed equilibria in box 1. The possibility of mixed equilibria complicates the analysis of this model. For the purposes of this article, I assume that these mixed equilibria are never observed. For the remainder of this section, I focus only on the low- and high-coordination equilibria.

Interpreting the model as a theory of financial crises

I associate the low-coordination equilibrium in the model with a financial crisis. This equilibrium captures many characteristics associated with financial crises. In this simple model, asset values and output can both be measured by the expected payoff to a borrower's project; both are clearly lower in the low-coordination equilibrium than in the high-coordination equilibrium.[28] There is a clear flight to quality in the low-coordination equilibrium, coupled with a drying-up of liquidity: Lenders invest in the risk-free asset instead of making loans, so the aggregate quantity of liquidity provided falls to zero. There is a perception of declining credit quality: If we were to ask a lender

FIGURE 4

Effect of changes in model parameters on the value of entering the match



Notes: For a particular set of baseline parameters, this figure illustrates how the value of entering the match implied by the model changes as five parameters of the model are varied. The left-hand column of figures plots the value of a lender entering the match in the high-coordination equilibrium (left-hand side of equation 4, represented by the black lines) and the low-coordination equilibrium (left-hand side of equation 6, represented by the colored lines) changes as the following five parameters change: number of lenders, as fraction of total population (first subplot); payoff to the lender R_{lend}, as a fraction of total output (second subplot); salvage value of a bad borrower's project $R^{salvage}$, as a fraction of total output (third subplot); probability that a bad project is productive θ, (fourth subplot); and the probability that a given project is bad, p^b (fifth subplot). The right-hand column of figures plots the value of a good borrower entering the match in the high-coordination equilibrium (left-hand side of equation 5), as the same five parameters are varied. The baseline parameters are as follows: N_{lend} = 20; N_{borr} = 30; R = 2; R_{lend} = 1.2 (so R_{borr} = 0.8); $R^{salvage}$ = 0.5; p^b = 0.2; θ = 0.75. Parameter N_{bad} is set equal to its expected value of 6. I use the model of $prob_{lend}$ and $prob_{borr}$ described in appendix A, equations 19 and 20, with parameter M = 10.

why she refrained from making loans, she would answer that the risk of default was too high (since all borrowers actually entering the match in the low-coordination equilibrium are bad borrowers). This is the sort of response given by lending institutions in the BIS interviews and the Board of Governors'

BOX 1

Mixed equilibria

In a mixed equilibrium, some lenders and/or good borrowers enter the match, while the remaining agents choose the alternative strategies (investing risk-free for lenders, operating the project autarkically for the borrowers). If there were a continuum of agents, these mixed equilibria would require agents to be indifferent between entering the match and using the alternative strategies. If one takes seriously the constraint that the number of agents of each type be an integer, then the conditions for a mixed equilibrium must take into account the effect on the matching probabilities were an agent to deviate from the equilibrium.

To write down the conditions for a mixed equilibrium, it is convenient to define functions V_{lend} and V_{borr} that measure the value of entering the match for lenders and borrowers, respectively. Let B_{good} denote the number of good borrowers entering the match, and let L denote the number of lenders entering the match. Recall that all bad borrowers enter the match (since the value of autarky for bad borrowers is zero). Therefore, the fraction of bad borrowers in the match is $\frac{N_{bad}}{N_{bad}+B_{good}}$, and (analogously with the left-hand side of equation 4) the value to a lender of entering the match is

$$\text{B1)}\quad V_{lend}\left(B_{good},L\right)\equiv prob_{lend}(N_{bad}+B_{good},L)\times\left(\left(\frac{\theta N_{bad}}{N_{bad}+B_{good}}+\frac{B_{good}}{N_{bad}+B_{good}}\right)R_{lend}+\frac{(1-\theta)N_{bad}}{N_{bad}+B_{good}}R^{salvage}\right)+(1-prob_{lend}(N_{bad}+B_{good,}L)).$$

The value of entering the match for a good borrower is given by the analogue to the left-hand side of equation 5:

$$\text{B2)}\quad V_{borr}(B_{good},L)\equiv prob_{borr}(N_{bad}+B_{good},L)R_{borr}.$$

If there were a continuum of agents, so the defection of a single agent from the equilibrium strategy would not affect the matching probabilities, then a mixed equilibrium would be a pair $\{B_{good}, L\}$ satisfying

$$\text{B3)}\quad 0<B_{good}<N_{borr}-N_{bad}$$
$$0<L<N_{lend}$$

for which

$$\text{B4)}\quad V_{lend}(B_{good},L)=R^{f}$$
$$V_{borr}(B_{good},L)=R^{autarky}.$$

If (as I assume throughout this article) there are an integer number of agents of each type, then a conjectured defection from the equilibrium strategy changes $prob_{lend}$ or $prob_{borr}$ and, therefore, changes V_{lend} or V_{borr}. To take this explicitly into consideration, I must modify equation B4. Assume that

$$\text{B5)}\quad V_{lend}\text{ is decreasing in }L.$$

Since $prob_{lend}(B,L)$ is strictly decreasing in L, a sufficient condition for assumption B5 is

$$\text{B6)}\quad \left(\frac{\theta N_{bad}}{N_{bad}+B_{good}}+\frac{B_{good}}{N_{bad+}B_{good}}\right)R_{lend}+\frac{(1-\theta)N_{bad}}{N_{bad}+B_{good}}R^{salvage}>1.$$

In other words, the expected payoff to a lender from a successful match exceeds the payoff from entering the match but failing to match. Note that V_{borr} is decreasing in B_{good}, because $prob_{borr}(B,L)$ is strictly decreasing in B. Under assumption B5, a mixed equilibrium is a pair $\{B_{good}, L\}$ satisfying equation B3 and

$$\text{B7)}\quad V_{lend}(B_{good},L+1)\le R^{f}\le V_{lend}(B_{good},L)$$
$$V_{borr}(B_{good}+1,L)\le R^{autarky}\le V_{borr}(B_{good}+1,L).$$

The logic behind equation B7 is straightforward. If the first set of inequalities in equation B7 holds, then a lender in the match has no incentive to switch to the risk-free asset (since the value of being in the match exceeds R^{f}), and a lender investing risk-free has no incentive to switch to entering the match (since, by entering the match, the total number of lenders in the match will equal $L + 1$, and the value to being a lender in the match when the total number of lenders equals $L + 1$ is dominated by the risk-free rate). A similar logic holds for borrowers if the second set of inequalities in equation B7 holds.

Senior Loan Officers Survey, discussed earlier. It is also consistent with widening credit spreads. Furthermore, there is a reduction in demand for liquidity on the part of borrowers, a pattern that was also reported in the September Senior Loan Officers Survey.

This model is also consistent with sudden switches between normal and crisis states without any change in underlying fundamentals (as represented by the model's parameters). While I do not model dynamics explicitly, a multiple-equilibrium model of this type can be incorporated into a dynamic model in which switches between coordination states are driven solely by changing expectations. If enough lenders in the economy become pessimistic about the aggregate number of borrowers entering the match (or vice versa), then a low-coordination equilibrium will emerge, validating their pessimism ex post. Thus, all that would be needed to model the abrupt switches between crisis and non-crisis states would be to model switching between optimism and pessimism in the economy.[29]

Financial crises and the role of the central bank

Perhaps the most interesting feature of the model presented here is that it suggests a role for the central bank in dealing with financial crises. We can see from equation 6 that a liquidity crisis (that is, a low-coordination equilibrium) is only possible if the real risk-free rate is sufficiently high. If the central bank can affect the real risk-free rate through open market operations, it can extinguish the possibility of a liquidity crisis by reducing the risk-free rate until the left-hand side of equation 6 exceeds the right-hand side. Intuitively, if the risk-free rate is so low that a lender expects a higher return by seeking to match with a borrower *even if all borrowers are believed to be of bad quality*, then the low-coordination equilibrium cannot be sustained.

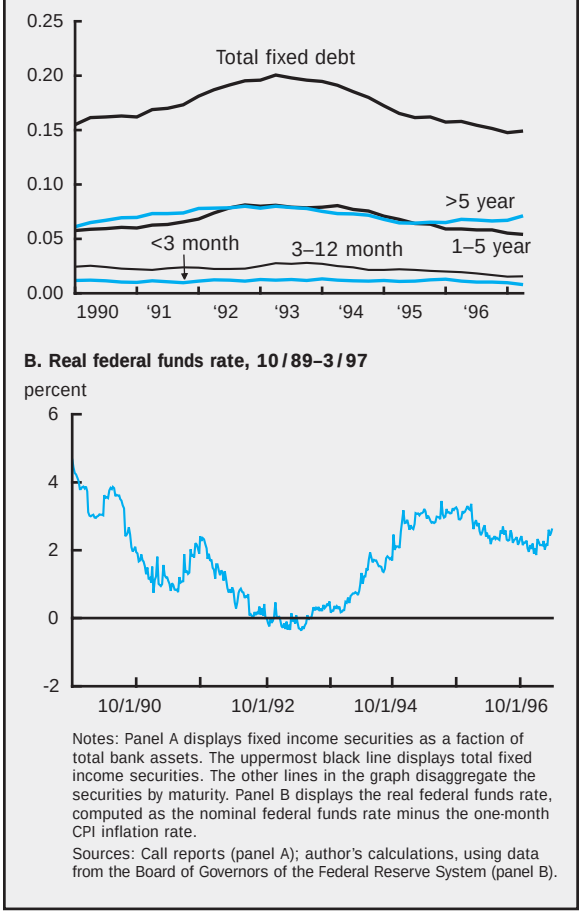
Example of central bank action: The events of 1991–93

One interpretation of monetary policy in the early 1990s is that the Federal Reserve used open market operations in the manner suggested in the preceding paragraph. The recovery from the 1990–91 recession appeared to be impeded by a so-called credit crunch. Responding perhaps to the introduction of risk-based capital requirements, banks reduced their volume of loan provision, investing instead in Treasury securities and other low-risk assets. One can see this process in figure 5, panel A, which displays fixed income securities as a fraction of total banking assets. Note that fixed income securities as a percentage of total assets rose from just over 15 percent at the beginning of 1990 to over 20 percent at the beginning of 1993. While this is far less dramatic than the complete coordination failure in the low-coordination equilibrium of the model, this process can be interpreted as a slow shift away from full coordination.

In mid-1991, the FOMC started reducing the federal funds rate in an effort to encourage more lending. This policy shift is evident in figure 5, panel B, which displays the real federal funds rate (defined here as the difference between the nominal federal funds rate and the ex post monthly CPI inflation rate) from 1989 through 1997. Note that the real funds rate declines to an extremely low level (between 0.5 percent and 0 percent) between late 1992 and February 1994. As in the simple model, the effect is to reduce the return on alternative assets, making even the relatively low risk-adjusted return on loans seem reasonably attractive.[30] As shown in figure 5, panel A, banks did shift

FIGURE 5

Bank portfolios and monetary policy, 1989–97.



Notes: Panel A displays fixed income securities as a faction of total bank assets. The uppermost black line displays total fixed income securities. The other lines in the graph disaggregate the securities by maturity. Panel B displays the real federal funds rate, computed as the nominal federal funds rate minus the one-month CPI inflation rate.

Sources: Call reports (panel A); author's calculations, using data from the Board of Governors of the Federal Reserve System (panel B).

away from non-loan assets to loans following the implementation of this policy. It is possible that the increased loan growth was due to some change in the economic or regulatory environment other than the extremely low real interest rates. However, these patterns in the data are certainly consistent with the intuition that lenders are more willing to lend when the return to alternative investments is low, and that the central bank can influence this alternative return.

Particularly interesting is what happened after the FOMC reversed course starting in February 1994 and allowed the real interest rate to return to its level of early 1990. If banks' willingness to lend depended only on the return on alternative assets, the banks presumably would then have cut back on their loan provision. In fact, figure 5, panel A shows that banks continued to increase their lending. This suggests that there may have been an element of coordination failure in the credit crunch. Once the economy had securely moved to a high-coordination state, it could remain there even after the FOMC raised real interest rates to a higher level.

Policy alternatives implied by the model

According to the model, a central bank can prevent financial crises by keeping interest rates extremely low all the time. However, this strategy conflicts with the central focus of monetary policy: to establish a reputation as a force for price stability. Even if the Fed could act against the possibility of a low-coordination equilibrium by decisively reducing Treasury yields, such an action would be costly, not only in its direct effect on future inflation, but also in eroding the credibility of the Fed's commitment to containing inflationary pressures.

In principle, a central bank could reconcile these two competing imperatives by establishing a credible *state-dependent* policy—enforce a low interest rate *only* when there is clear evidence that a financial crisis is imminent. The advantage of such a policy is that the low interest rate is rarely implemented, yet the possibility that it might be implemented moves the economy to a preferred equilibrium. In fact, when I incorporate such a state-dependent policy into the simple model developed earlier, the low interest rate is *never* implemented. In the language of economic theory, it is an *off-equilibrium path* that enforces the high-coordination equilibrium.

As an example of such a state-dependent policy, consider the aftermath of the October 1987 stock market crash. There is considerable anecdotal evidence that many banks were reluctant to provide the liquidity needed to settle trades made during the day of the crash. This withdrawal of liquidity may have represented a low-coordination equilibrium: If a given bank is likely to be repaid only if the aggregate provision of liquidity is high, it may be individually rational for each bank to withhold liquidity. In response, the Federal Reserve announced a state-contingent policy: "The Federal Reserve ... affirmed today its readiness to serve as a source of liquidity to support the economic and financial system."[31] The operative word is "readiness." With the Fed standing ready to ensure adequate liquidity in the market, it became rational for individual banks to provide liquidity to their clients. In the event, no significant liquidity disruptions were observed,[32] yet the Fed itself did not actively provide the liquidity—discount window borrowing by member banks did not increase significantly, and the increase in non-borrowed reserves was small.

In September 1998, investors were uncertain whether a state-dependent policy was in place.[33] One can interpret the intermeeting rate cut on October 15, 1998, as a credible signal that the Fed had shifted to this sort of state-dependent policy. It seems more plausible to interpret the effect of this rate cut to its role as a signal than to any direct effects. Certainly, the interest rate cuts in fall of 1998 were much smaller than those in 1990–91, clearly not sufficient to substantially change investors' incentives directly. There is evidence that financial markets perceived the intermeeting rate cut as signaling a policy change. According to the *Wall Street Journal,* "the economic indicators that the Fed usually tracks—the unemployment rate, the pace of orders for factory goods and retail sales, among other things—don't explain the Fed's sudden action, since many of those indicators have suggested that the economy is relatively healthy. Rather, officials at the Fed ... have been focused on unusual signs of stress in the financial markets."[34] In the words of analysts at Morgan Stanley Dean Witter, the Fed's "unexpected easing" signaled a "new aggressiveness."[35] Market participants' perception of a change in Fed emphasis is consistent with the minutes of the FOMC's deliberations. In the minutes of the September 29 meeting,[36] the financial market turmoil is noted, but it is seen primarily as one factor among many affecting inflationary pressures through aggregate demand. In particular, "The members did not believe that the tightness in credit markets and strong demand for safety and liquidity were likely to lead to a 'credit crunch.' ..." A 50 basis-point cut is explicitly ruled out because "the risk of rising inflation ... was still present, especially in light of the persistence to date of very tight labor markets and relatively robust economic growth." In contrast, the minutes of the FOMC teleconference preceding the intermeeting rate cut of October 15 do

not mention inflationary pressures at all. Rather, an additional rate cut is motivated "to help settle volatile financial markets and cushion the effects of more restrictive financial conditions on the ongoing expansion." Following the October 15 action, members of the FOMC describe the move as a (temporary) shift of focus from price stability to financial stability. For example, the *Wall Street Journal* cited Federal Reserve Bank of St. Louis President William Poole as saying that "[the recent market instability] and the circumstances surrounding it are so unusual in the context of U.S. history that policy makers must concentrate on dealing with this situation for the time being."[37] According to the same article, Governor Roger Ferguson "indicated that the Fed would be willing to cut rates aggressively at any hint of a recession."

In the following section, I incorporate such a state-dependent policy into the simple model outlined earlier. I argue that such a policy, if credible, may have been sufficient to eliminate the possibility of coordination failure without requiring the central bank to actually implement any substantial interest rate reductions. To formalize this idea, I extend the model so that the central bank acts in real time. I then demonstrate that the low-coordination equilibrium can be eliminated if agents believe that the monetary authority will act in the future to eliminate coordination failure should coordination failure become likely.

Modeling a state-dependent central bank policy

To model a central bank interest rate policy that actively responds to a developing liquidity crisis, I need to modify the simple model to make precise the notion of "incipient crisis." I do so in the following (admittedly highly stylized) way. Suppose that there is a preliminary period before the matching of borrowers and lenders. At the beginning of this preliminary period, each lender is assigned at random a mood of *pessimism* or *optimism*. A pessimist believes that the low-coordination equilibrium will prevail *provided* a low-coordination equilibrium could exist (that is, provided equation 6 could hold). An optimist believes that the high-coordination equilibrium will prevail, again provided this equilibrium could exist (that is, equations 4 and 5 could hold). In addition, in the beginning of the preliminary period the N_{lend} lenders are assigned an index $i = 1, ..., N_{lend}$. They then must declare (irrevocably) in order of their index assignment whether they will enter the match or invest in the risk-free technology. After all N_{lend} lenders have declared, the match is held, projects are operated, and payoffs are made, as in the simple model.

Let Q_i denote the number of lenders who have declared that they are in the match through the ith lender, so $Q_{N_{lend}}$ denotes the total number of lenders committed to entering the match at the end of the preliminary period. As I show in proposition 1 below, if $Q_{N_{lend}}$ is sufficiently high (that is, if $Q_{N_{lend}}$ is greater than or equal to a particular threshold N^*), a low-coordination equilibrium cannot exist. A pessimist with index i will assume that the low-coordination equilibrium will prevail *unless* Q_{i-1} (weakly) exceeds this threshold.

Now, I develop this idea more fully. For simplicity, I consider a case where the central bank can choose one of two interest rates: R^{high} and R^{low}, where R^{high} is consistent with both equilibria. That is, when $R^f = R^{high}$, equations 4, 5, and 6 all hold. (I discuss conditions on R^{low} below.) If $R^f = R^{high}$, the equilibrium that emerges depends on the beliefs of the lenders about $Q_{N_{lend}}$. In particular, there exists an N^* such that, if it is believed that $Q_{N_{lend}} \geq N^*$, it is optimal for *all* good borrowers to enter the match. The smallest such value of N^* is given by

7) $N^* = \min N$ s.t. $prob_{borr}(N_{borr}, N) R_{borr} \geq R^{autarky}$.

The existence of $N^* \leq N_{lend}$ follows from the assumption that a high-coordination equilibrium exists (that is, equation 5 holds.)

To proceed, I must make an additional assumption. Let V_{lend} denote the value to a lender of entering the match. V_{lend} depends on both the total number of lenders in the match and the number of good borrowers in the match. An explicit expression for V_{lend} is given in equation B1 in box 1. I assume that

8) V_{lend} is strictly decreasing in the total number of lenders in the match.

A sufficient condition for assumption 8 to hold is given in equation B6 in box 1.

Proposition 1

Suppose equations 4 and 8 hold. If all lenders believe that $Q_{N_{lend}}$ will be at least as big as N^*, then all lenders enter the match, so $Q_{N_{lend}} = N_{lend} \geq N^*$. This implies, first, that their beliefs are ratified ex post, and, second, that the high-coordination equilibrium prevails. (The proofs of all propositions are in appendix B.)

Proposition 1 tells us that the central bank can ensure that the high-coordination equilibrium will prevail if it can ensure that at least N^* lenders commit to entering the match. As in the previous model, it can do so by setting R^f sufficiently low. To formalize this possibility, let us assume that

9) $(\theta R_{lend} + (1 - \theta) R^{salvage}) > 1$

and let R^{low} satisfy

10) $prob_{lend}(N_{bad}, N^*)(\theta R_{lend} + (1-\theta) R^{salvage}) + (1 - prob_{lend}(N_{bad}, N^*)) \geq R^{low}$.

Equation 9 means that the expected payoff to a lender conditional on matching with a bad borrower exceeds the payoff from failing to match. It ensures that the left-hand side of equation 10 is increasing in $prob_{lend}$.

Proposition 2

If the central bank sets $R^f = R^{low}$ and equations 9 and 10 hold, the high-coordination equilibrium is enforced.

Proposition 2 tells us that the central bank can eliminate the low-coordination equilibrium by permanently setting the risk-free rate sufficiently low (in particular, low enough so equation 10 holds). In reality, however, this low interest rate policy is a very costly way to deal with the possibility of financial crisis. As I discussed above, the excessively expansionary monetary policy needed to keep interest rates at R^{low} may directly conflict with the central bank's primary mission of price stability. If so, a better central bank rule is to set $R^f = R^{high}$, but commit to switching to R^{low} if there is evidence of an incipient crisis. Informally, the central bank can measure the tone of the market by looking at the ratio Q_i/i. This ratio gives the fraction of the first i lenders who will enter the match, so this ratio measures the "skittishness" of the market. If the central bank observes a low value of Q_i/i (presumably because the random assignment of the first i indexes fell disproportionately on pessimists), it may be concerned that a financial crisis is brewing.

In the formalism of this model, let "incipient crisis" be defined as any point i^* in the declaration sequence such that

11) $Q_{i^*} + (N_{lend} - i^*) = N^*$.

In words, if such an i^* is reached, then *all* of the remaining lenders must declare themselves in the match to ensure that there are N^* lenders seeking to match with borrowers. Since the goal of the central bank is to ensure that at least N^* lenders enter the match, this is the "last chance" for the central bank to do so.

Central bank rule

The proposed central bank rule is as follows:

- Set $R^f = R^{high}$ as long as no $\{i^*, Q_{i^*}\}$ satisfying equation 11 is reached.
- The first time $\{i^*, Q_{i^*}\}$ satisfying equation 11 is reached, set $R^f = R^{low}$ from that point on.

Proposition 3

If this rule is credible, the only equilibrium is high-coordination with $Q_i = i$, $\forall i$, and $R^f = R^{high}$.

According to proposition 3, the second branch of the rule is an off-equilibrium path that is never observed in equilibrium. Thus, the best of all possible worlds is obtained: Liquidity crises are ruled out without compromising the goal of price stability. Proposition 3 specifies that the rule must be credible. I do not attempt to formalize how "credibility" is to be established. Authors such as Christiano, Chari, and Eichenbaum (1998) and Christiano and Gust (2000) stress the importance of the central bank establishing a credible commitment to price stability if expectations-driven inflationary episodes are to be avoided. Proposition 3 suggests that a credible commitment to financial stability may serve an analogous role in avoiding financial crises.

Discussion

Is this what happened in October 1998?

One interpretation of the FOMC's interest rate cut on October 15, 1998, is that it was an intentional signal that Federal Reserve had shifted from an unequivocal focus on price stability to a policy of "price stability unless there is a pressing need to deter a financial crisis." In the formalism of the model, the former policy sets $R^f = R^{high}$ always, while the latter policy is given by the policy rule described above.

There are clearly other possible explanations for the ending of the fall 1998 crisis. One such explanation is that the reduced interest rates increased the collateral value of firms' fixed income portfolios, thereby increasing their borrowing capacity. But the effect of a 50 basis-point interest rate cut on the value of debt holdings is small, especially for the short-term securities generally used as collateral. In any event, the turmoil in fall of 1998 was associated with a "flight to quality," which raised the value of the Treasury securities that typically collateralize liquidity loans. Another explanation is that the open market operations used to implement these interest rate cuts increased the total supply of reserves in the system, increasing the amount of liquidity available to be borrowed. Again, this explanation seems wanting. In contrast to the period from 1991 to 1993, when there was an extended and pronounced increase in the volume of reserves in circulation, the amount of reserves in fall 1998 was relatively unchanged.

Perhaps a more straightforward explanation for the abrupt reversal of the 1998 crisis is that financial

intermediaries believed that the Federal Reserve had implicitly agreed to provide all financial institutions with a guarantee. In particular, the role of the Federal Reserve Bank of New York in the recapitalization of LTCM may have been interpreted as a commitment to provide similar services to other intermediaries with similar problems. I do not believe that the facts support this explanation. Following the announcement of the LTCM rescue plan, *the crisis actually deepened*. Measures of credit spreads and market volatility deteriorated during the two weeks between the LTCM rescue and the Federal Reserve's intermeeting action on October 15. Furthermore, market participants reported that the Federal Reserve's role in the rescue served to exacerbate market fears, not ameliorate them. Thus, the data seem to contradict the hypothesis that the LTCM rescue was interpreted as an extension of the safety net.

Finally, the October 15 rate cut may have signaled a changed policy stance regarding International Monetary Fund (IMF) funding rather than monetary policy. The Russian fiscal crisis virtually assured that a good deal of IMF resources would flow to Russia. Without an increase in funding levels, the IMF's resources to deal with other countries' problems (most importantly, Latin America) would have been substantially reduced. The increase in Brady bond yields, documented in figure 3, may have reflected concerns that less IMF funding would be available to deal with future Latin American problems following the Russian crisis. During 1998 the U.S. Congress was considering an increase in America's IMF funding quota. However, there was considerable congressional opposition to increased funding. Perhaps the October 15 rate cut was interpreted as a signal that the Federal Reserve would work with greater intensity to secure increased IMF funding.

This interpretation is certainly possible. However, it relies on a less direct mechanism than the monetary policy interpretation I put forth in this article. It places a good deal of weight on the Federal Reserve's influence with Congress. Furthermore, the Federal Reserve was already on record supporting the proposal to increase IMF funding (see Chairman Greenspan's testimony to Congress on May 21, 1998), so the October 15 rate cut would have represented at best a strengthening of this position, not a reversal of a previously held position. Finally, the data are not entirely consistent with this explanation. As shown in figure 3, the peak in Latin American Brady bond yields during 1998 happened in mid-August (Mexico) or mid-September (Brazil and Argentina), not in mid-October when the presumed signal occurred.

Costs of a state-contingent policy for financial crises

In the theoretical model described here, the state-contingent policy rule is costless to implement, since the low interest rate is never actually imposed in equilibrium. Of course, the real world is not so simple. In reality, there would doubtless be crises that could not be extinguished by the belief that the central bank's rule specifies a particular off-equilibrium path. As a practical matter, this sort of policy rule would require aggressive monetary expansion from time to time. Actions of this type have costs. Each time such a monetary expansion is implemented, the central bank compromises its primary objective of price stability. Any time it injects liquidity into financial markets in an effort to counter potential liquidity it faces the difficult task of negotiating a "soft landing"—removing the liquidity after the crisis has abated without triggering a recession. Furthermore, if the state-contingent policy rule weakens the commitment to price stability, the resulting instability might even increase the possibility of financial crises. Finally, if private market participants believe that the central bank will always act to successfully counter financial turmoil, they may engage in less vigilant risk management than they would otherwise. This so-called moral hazard problem may actually increase the chances of an incipient crisis. Policymakers must take all of these issues into consideration before adopting a state-contingent rule as a practical policy doctrine.

Conclusion

In this article, I propose a precise characterization of financial crisis. I argue that coordination problems arise generically in financial markets. I associate financial crisis with a condition of coordination failure, in which low levels of financial intermediation become self-justifying. I also argue that the central bank, through its ability to affect real interest rates, may be able to extinguish the low-coordination trap. This argument supports a role for the central bank in countering systemic financial disruptions.

Having said this, there may be circumstances in which the central bank's power to affect real rates is insufficient to stave off a crisis. In particular, if potential lenders are sufficiently pessimistic about returns from lending, crisis aversion may require a real interest rate below that achievable by open market operations. In addition, the use of open market operations to counter financial crises is not without cost. Open market operations can only have a temporary effect on real rates. Prolonged use of this tool to reduce real interest rates would run directly counter to the central bank's primary goal of price stability. In principle,

the central bank is better off establishing a credible contingent policy, whereby a liquidity injection is only made when there is evidence that a crisis is forthcoming. In the simple model presented here, a credible policy of this type never has to be implemented in equilibrium. In reality, of course, life is not so simple. There would doubtless be cases where the central bank would have to implement an expansionary monetary policy to counter an incipient crisis. Thus, this article's policy implications have benefits and costs that must be carefully weighed by policymakers when considering practical policy formulation. Nonetheless, this article does provide a formal justification for the central bank as an essential institution in dealing with financial crises.

APPENDIX A

Increasing returns to scale in matching

Increasing returns to scale in matching can be derived from a number of more primitive search models. For example, Mortensen and Pissarides (1998) describe an environment in which each lender has a list of telephone numbers that includes the numbers of potential borrowers, and each borrower has a similar list that includes the numbers of potential lenders. If borrowers and lenders choose numbers at random, the probability of a match displays increasing returns. Kultti (1998) describes a somewhat more elaborate model.[1] In his approach, lenders are posted at fixed locations, and borrowers randomly choose a location. If there is a lender at the location and there are no other borrowers, a match is made with certainty. If there is a lender and more than one borrower at the location, a borrower is chosen at random to match with the lender. Finally, if there is no lender at the location, no match is made. One can think of these locations as bank branches, where some branches have exhausted their loan capacity. Kultti (1998) shows that if the number of locations does not change as the number of borrowers and lenders increases, the matching probabilities display increasing returns.[2]

Now, I consider in greater detail a micro model of matching, similar to Kultti's (1998), that implies increasing returns. Suppose there are M locations. To successfully match, a lender and a borrower must go to the same location. They cannot communicate before traveling to a location, so the event of a lender and a borrower being in the same location is purely random. Ex ante, all locations look the same to both lenders and borrowers. I assume that lenders and borrowers make their location decision at random, independently of the other lenders and borrowers. Therefore, the probability that a given borrower or a given lender arrives at any particular location is $1/M$.[3]

An interpretation of this set-up is that the "locations" are banks or other intermediaries. Lenders are agents with excess liquidity. To match with borrowers, the lenders must go through an intermediary. Lenders choose the intermediary at random. Similarly, borrowers visit intermediaries at random to apply for a loan. The loan application process is sufficiently time-intensive that a borrower can only apply at one intermediary.

Suppose there are l lenders and b borrowers at a given location. If $l = 0$ or $b = 0$, no matches take place at that location. If $l = b$, all the lenders and borrowers at that location match with probability one. If $l > b$, the b borrowers are allocated randomly among the lenders, so the probability of a given lender obtaining a match is b/l, and all borrowers obtain a match with probability one. Similarly if $b > l$, the probability of a given borrower obtaining a match is l/b, and all lenders obtain a match with probability one. To summarize,

$$Prob\{lender\ matching \mid b,l\} \min\left(\frac{b}{l},1\right);$$

$$Prob\{borrower\ matching \mid b,l\} min\left(\frac{l}{b},1\right).$$

I now compute the unconditional probability that a lender will match. Let B denote the number of borrowers seeking loans, and let L denote the number of lenders seeking to match with borrowers. The probability that a given lender will be at a particular location is $1/M$. For $n = 0, 1, ..., L$, the probability that n lenders will arrive at a particular location is denoted by $p_l(n|L)$, as follows:

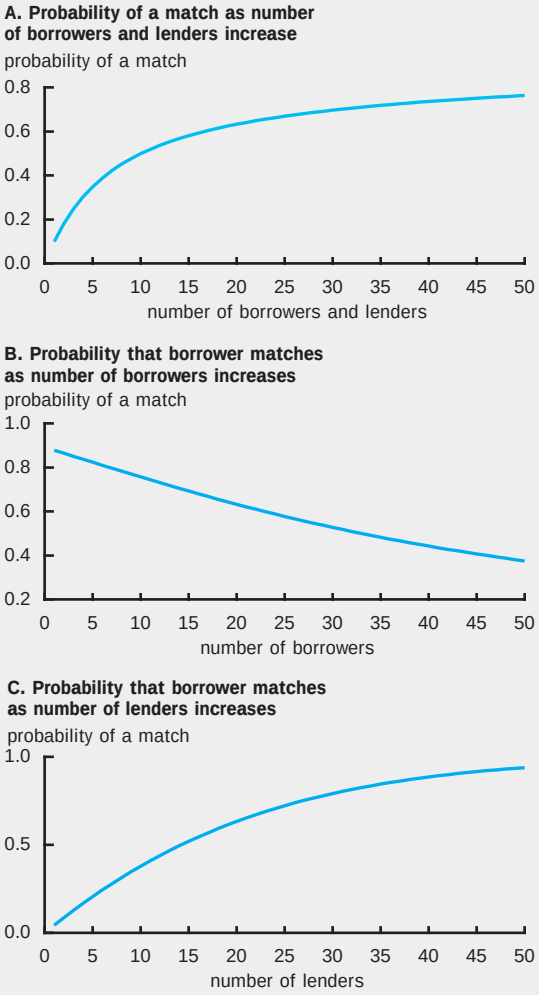
$$p_l(n|L) = \frac{L!}{n!(L-n)!}\left(\frac{1}{M}\right)^n\left(1-\frac{1}{M}\right)^{L-n}.$$

Similarly, for $n = 0, 1, ..., B$, the probability that n borrowers will arrive at a particular location is denoted by $p_b(n|B)$, as follows:

$$p_b(n|B) = \frac{B!}{n!(B-n)!}\left(\frac{1}{M}\right)^n\left(1-\frac{1}{M}\right)^{B-n}.$$

FIGURE A1

Effect of matching probabilities of increase in numbers of borrowers or lenders



Notes: The top panel gives the probability that a lender or borrower matches as both the number of lenders and the number of borrowers increase at the same rate. The middle panel gives the probability that a borrower matches as the number of borrowers increases (holding lenders fixed). The bottom panel gives the probability that a borrower matches as the number of lenders increases (holding borrowers fixed).

To determine $prob_{borr}(B,L)$, the probability that a given borrower matches, one must sum overall possible values of l and b:

A1) $$prob_{borr}(B,L) = \sum_{l=0}^{L}\sum_{b=0}^{B-1} \min\left(\frac{l}{b+1},1\right) p_b(b|B-1)\,p_l(l|L).$$

In the second summation, I sum only to B–1 because we are concerned with the number of *other* borrowers that show up at the same location as the given borrower. There are only B–1 other borrowers. The expression for $prob_{lend}(B, L)$, the probability that a given lender matches with a borrower, is analogous:

A2) $$prob_{lend}(B,L) = \sum_{b=0}^{B}\sum_{l=0}^{L-1} \min\left(\frac{b}{l+1},1\right) p_b(b|B)\,p_l(L-1).$$

The increasing returns property is illustrated in the first panel of figure A1. As both L and B rise, the probability of a match (given by equation 19 or 20) increases. As illustrated in the last two panels of figure A1, equation A1 implies that the probability of a given borrower matching is increasing in the number of lenders, L, (holding B constant) and decreasing in the number of borrowers B (holding L constant). Similarly, the probability of a given lender matching is increasing in B and decreasing in L.

[1]See also Hall (1999).

[2]More generally, this result holds if the number of locations increases at a slower rate than the number of borrowers and lenders.

[3]This explicitly rules out equilibria of the form, "All borrowers and all lenders choose to go to the kth location with probability one."

APPENDIX B

Proofs of propositions

Proposition 1

Let functions V_{borr} and V_{lend} be defined as in equations B1 and B2 of box 1. $Prob_{borr}(B, L)$ is increasing in L and decreasing in B. Therefore, the condition

$$Q_{N_{lend}} \geq N^*$$

and equation 7 imply that $V_{borr}(B_{good}, Q_{N_{lend}}) \geq R^{autarky}$ for all $B_{good} \leq N_{borr} - N_{bad}$. This implies that all good borrowers enter the match. However, if all good borrowers enter the match, then equations 4 and 8 imply that the value of entering the match to a lender exceeds R^f, regardless of the decisions of the other lenders.[1] Therefore, all lenders enter the match. This ends the proof.

Proposition 2

If $R^f = R^{low}$, equations 9 and 10 together imply that if $Q_{i-1} < N^*$, it is optimal for the ith lender to enter the match, *even if no good borrowers enter the match*. This follows because equation 9 ensures that, for arbitrary $\tilde{N}$,

$$prob_{lend}(N_{bad}, \tilde{N})(\theta R_{lend} + (1-\theta) R^{salvage}) + (1 - prob_{lend}(N_{bad}, \tilde{N}))$$

is decreasing in $\tilde{N}$. (Recall that $prob_{lend}(N_{bad}, \tilde{N})$ is decreasing in $\tilde{N}$: As more lenders compete with a given lender, the probability that a lender matches goes down.) This in turn means that at least N^* lenders will enter the match, regardless of what any of the other lenders or good borrowers do, so $Q_{N_{lend}} \geq N^*$. According to proposition 1, this is sufficient for the high coordination equilibrium to prevail. This completes the proof.

Proposition 3

To prove proposition 3, note that the monetary policy rule ensures that there will always be at least N^* lenders committing to enter the match. By proposition 1, this is sufficient to ensure that the high-coordination equilibrium prevails. This completes the proof.

---

[1]This assertion uses the fact that assumption 8 implies that $V_{lend}(B,L)$ is decreasing in L. Therefore, if $V_{lend}(N_{borr} - N_{bad}, N_{lend}) \geq R^f$, then $V_{lend}(N_{borr} - N_{bad}, L) \geq R^f, \forall L \leq N_{lend}$.

NOTES

[1]An account of the events surrounding the Russian default can be found in Perotti (2000).

[2]Immediately following the announcement of the rate cut, the Dow Jones Industrial Average dropped. Press reports indicate that many investors expected a bigger rate cut. The U.S. Chamber of Commerce described the move as "underwhelming in its modesty" (Schlesinger and Wessel, 1998b), and investors in bond futures "were treating a half-point cut by year's end as a certainty, and a half-point cut tomorrow as a reasonable possibility" (Schlesinger, 1998). The federal funds futures market supports this assertion. From September 25 through 28, the price of the October federal funds futures contract implied an expected fed funds rate of around 5.16 percent for the month of October (down from 5.5 percent before September 29). This implies that investors put substantial probability on a cut of 50 basis points or more at the September FOMC meeting.

[3]"[M]arkets had been rife with rumors about the possibility of an intermeeting move for the past week or so. …" in Greenlaw (1999).

[4]The fed funds futures market gives a forecast of the 30-day average federal funds rate over the month of October. The funds rate was already at 5.25 percent. A rate cut in mid-October to 5.00 percent would move the 30-day *average* of October rates to 5.125 percent. If investors only assigned a 50 percent probability to such a rate cut, the expected average rate would be approximately 5.18 percent.

[5]The magnitudes of the declines were as follows: UK, 23 percent; Germany, 19 percent; Japan, 18 percent; Canada, 28 percent; Italy, 32 percent; and France, 29 percent.

[6]I would like to thank Eileen Smith of the Chicago Board Options Exchange for providing me with these data.

[7]The actual peak in the volatility index came on October 8, 1998, the date of the trough in the S&P 500 index.

[8]The behavior of default spreads for other G-7 countries gives a less clear picture of the start and end of the crisis. For Canada, France, Italy, and the UK, these spreads move roughly in line with the U.S. data (although the interbank spread for France is so volatile that it is difficult to identify peaks and troughs with any degree of certainty). The German interbank spread peaks in November 1998, several weeks after the peak in U.S. data. Finally, the behavior of default spreads in Japanese data is rather different from the other G-7 countries. In particular, the Japanese interbank spread shows little evidence of a liquidity crisis until a pronounced spike in mid-November 1998, well after the crisis abated in the U.S.

[9]The Board of Governors' Senior Loan Officers Survey can be found on the Internet at www.federalreserve.gov/boarddocs/SnLoanSurvey/. The surveys in August and November 1998 followed the Board's usual procedure of asking respondents about credit conditions over the preceding three months. Therefore, the November survey reflected most of the crisis period. In contrast, the September 1998 survey was a special survey that only asked about credit conditions over the previous month.

[10]This characterization of markets during the crisis is confirmed by other sources. For example, Morgan Stanley Dean Witter reported a sharply reduced volume of activity across the corporate borrowing spectrum. (Roach, 1998.) Similarly, a strategist at Merrill Lynch asserted that "there were literally occasions when you could not get a bid of any kind for debt that was a reasonable risk" (The Economist Newspaper Limited, 1998, p. 75).

[11]See Schlesinger and Wessel (1998a).

[12]BIS (1999, p. 42).

[13]The role of the Federal Reserve Bank of New York in the recapitalization of LTCM was limited to providing meeting facilities for the involved parties. The Federal Reserve provided no funds in the LTCM workout.

[14]According to the U.S. Department of Commerce, Bureau of the Census (1998), the U.S. exports of goods and services to Argentina, Mexico, and Brazil in 1997 exceeded $93 billion (13.5 percent of total U.S. exports), while U.S. imports from these countries totaled almost $98 billion (11.2 percent of U.S. total). In contrast, U.S. exports to Russia totaled $3.4 billion (0.49 percent of the U.S. total) with imports from Russia totaling $4.3 billion (0.50 percent of the U.S. total).

[15]Brady bonds are U.S. dollar-denominated obligations of various developing countries, mainly in Latin America.

[16]For Indonesia and Malaysia, swap yields for maturities up to two years and forward exchange rates for up to one year closely track the spot exchange rates for these currencies. That is, these forward-looking markets did not anticipate the currency devaluations. Data on these particular markets are not available for Korea and the Philippines, but yields on government bonds (five-year maturity for Korea, one-year maturity for the Philippines) display the same patterns as the Indonesian and Malaysian swap yields: they did not budge until these countries devalued their currency. The only country where interest rates (as measured by swap yields) and forward exchange rates moved before the devaluation was Thailand. For that country, both rates started to increase six weeks before the devaluation of the baht. However, this coincided with the initial speculative attacks on the baht and the Thai government's initial attempts to defend its currency peg. See Halcomb and Marshall (2001).

[17]See Burnside, Eichenbaum, and Rebelo (1998).

[18]In developing this model, I benefited from extensive discussions with François Velde.

[19]These second and third characteristics were also associated with the Asian crisis of 1997. The reduction in liquidity provision took the form of a reversal of short-term capital flows from western countries. Needless to say, stock markets fell precipitously in all five Asian crisis countries.

[20]In this example, which is discussed in Cooper and John (1988), there can be more than two equilibria.

[21]The classic demonstration of how a search model can give rise to a thick markets externality is Diamond (1982).

[22]I'd like to thank Eric French for suggesting this interpretation.

[23]It would be preferable to have the contract emerge as the equilibrium outcome of a bargaining game. This approach is not straightforward to implement. The Nash (1953) axiomatic approach to solving the two-person bargaining problem does not generalize to a game with asymmetric information. An alternative would be to specify a noncooperative bargaining game. For example, one could assume that either the borrower or lender is randomly given the right to make a single take-it-or-leave-it offer. (Mortensen and Pissarides, 1998, note that this game under full information implies the same solution as a particular version of the Nash bargaining problem.) It is beyond the scope of this article to explore the range of noncooperative game theoretic approaches to this problem. As a result, I adopt the simple expedient of an exogenous contract.

[24]My assumption that a lender who fails to match gets a zero net return captures the idea that there is some opportunity cost to committing to provide loans, rather than investing exclusively in the risk-free investment. However, the analysis would not be changed substantially if a lender who fails to match received a small positive return.

[25]Lagos (2000) develops a model of passenger–cab matching that implies constant returns to scale. The difference between this model and the model I develop in appendix A is that Lagos (2000) assumes a continuum of passengers and cabs, whereas my model assumes that both passengers and cabs are discrete and finite in number.

[26]More recently, DenHaan, Ramey, and Watson (1999) and Burdette, Imai, and Wright (2000) propose search models that give rise to nontrivial multiple equilibria even with constant returns to scale.

[27]The baseline parameters used in figure 4 are: $N_{lend} = 20$; $N_{borr} = 30$; $R = 2$; $R_{lend} = 1.2$ (so $R_{borr} = 0.8$); $R^{salvage} = 0.5$; $p^b = 0.2$; $\theta = 0.75$. Parameter N_{bad} is set equal to its expected value of 6. I use the model of $prob_{lend}$ and $prob_{borr}$ described in appendix A, equations A2 and A1, with parameter $M = 10$.

[28]Of course, this static model cannot address the question of why existing assets decline in value. If existing productive assets utilize a continued flow of credit to maintain high profitability, a reluctance of lenders to provide credit would presumably reduce asset values. However, it would require a dynamic extension of this model to analyze this effect formally.

[29]In the literature on financial fragility, this is typically done by assuming an exogenous "sunspot process," whose realization determines the state of optimism in the economy. (See, for example, Chang and Velasco 1998, Christiano and Harrison 1996, and Burnside, Eichenbaum, and Rebelo 2000.) I do not explicitly implement this approach for rendering the model dynamic, but to do so would be straightforward.

[30]While figure 5, panel B displays the real overnight interest rate, the relevant rate for bank incentives is the expected rate of return from holding longer-term fixed income securities. Forecasting expected real holding period returns is a process fraught with difficulty, and I do not attempt to do so here. However, simple term structure models (such as the expectations hypothesis) imply that expected real holding period returns move with short-term real interest rates. Figure 5, panel B suggests that the expected returns relevant to bank decisions fell substantially as a result of the Federal Reserve's expansionary policy from 1991 to 1993.

[31]Quoted in Murray (1987).

[32]For an account of these events, see the Securities Exchange Commission report excerpted in Kamphuis et al. (1989).

[33]For example, the Morgan Stanley Dean Witter *Global Economic Forum* of September 25, 1998, noted that a large rate cut at the FOMC meeting four days later "… may reflect a potentially profound transformation in the Fed's basic philosophy." The forum argued "… that the biggest risk of all is that the world's policymakers may not be up to the task at hand." Similarly, Jacob Schlesinger wrote in the *Wall Street Journal* of September 28, 1998, that "an easing of monetary policy would mark a swift and amazing turnaround in the central bank's fundamental economic outlook. For well over a year, the Fed's greatest concern has been inflation, not recession."

[34]Schlesinger and Wessel (1998a), p. A3.

[35]Morgan Stanley Dean Witter (1999).

[36]Minutes of the FOMC can be obtained from the Federal Reserve website at www.federalreserve.gov/fomc/minutes/.

[37]Schlesinger and Ip (1998).

REFERENCES

Bank for International Settlements, 1999, "A review of financial market events in Autumn 1998," Committee on the Global Financial System, Basel, Switzerland, October.

Bernanke, Ben S., 1983, "Nonmonetary effects of the financial crisis in the propagation of the Great Depression," *American Economic Review*, Vol. 73, No. 3, pp. 257–276.

Burdette, Kenneth, Ryoichi Imai, and Randall Wright, 2000, "Unstable relationships," University of Pennsylvania, manuscript.

Burnside, Craig, Martin Eichenbaum, and Sergio Rebelo, 2000, "On the fundamentals of self fulfilling currency attacks: The role of government guarantees to foreign bank creditors," Northwestern University, working paper.

_____, 1998, "Prospective deficits and the Asian Currency Crisis," Federal Reserve Bank of Chicago, working paper, No. WP-98-5.

Chang, Roberto, and Andres Velasco, 1998, "The Asian liquidity crisis," Federal Reserve Bank of Atlanta, working paper, No. 98-11.

Chari, V. V., Lawrence J. Christiano, and Martin Eichenbaum, 1998, "Expectation traps and discretion," *Journal of Economic Theory*, Vol. 81, No. 2, pp. 462–492.

Chari, V. V., and Patrick J. Kehoe, 2000, "Financial crises as herds," Federal Reserve Bank of Minneapolis, working paper, No. 600.

_____, 1998, "Hot money," Federal Reserve Bank of Minneapolis, staff report.

Christiano, Lawrence J., and Christopher Gust, 2000, "The expectations trap hypothesis," *Economic Perspectives*, Federal Reserve Bank of Chicago, Second Quarter, pp. 21–39.

Christiano, Lawrence J., and Sharon G. Harrison, 1996, "Chaos, sunspots, and automatic stabilizers," Federal Reserve Bank of Chicago, working paper, No. WP-96-16.

Cooper, Russell, and Andrew John, 1988, "Coordinating coordination failures in Keynesian models," *Quarterly Journal of Economics*, Vol. 103, No. 3, pp. 441–463.

DenHaan, Wouter J., Garey Ramey, and Joel Watson, 1999, "Liquidity flows and fragility of business enterprises," National Bureau of Economic Research, working paper, No. 7057.

Diamond, Peter A., 1982, "Aggregate demand management in search equilibrium," *Journal of Political Economy*, Vol. 90, No. 5, pp. 881–894.

Economist Newspaper Limited, The, 1998, "Alan Greenspan's miracle cure," *The Economist*, October 24, p. 75.

Evans, Charles L., and David A. Marshall, 1998, "Monetary policy and the term structure of nominal interest rates: evidence and theory," *Carnegie-Rochester Conference Series on Public Policy*, Vol. 49, pp. 53–112.

Greenlaw, David, 1999, "U.S.: A shocker!," *Global Economic Forum*, Morgan Stanley Dean Witter, October 16, available on the Internet at www.msdw.com/GEFdata.

Halcomb, Darrin, and David Marshall, 2001, "A retrospective on the Asian Crisis of 1997: Was it foreseen?," *Chicago Fed Letter*, Federal Reserve Bank of Chicago, No. 161, January.

Hall, Robert, 1999, "Reorganization," National Bureau of Economic Research, working paper, No. 7181.

Kamphuis, Jr., Robert W., Roger C. Kormendi, and J. W. Henry Watson, (eds.), 1989, "The October 1987 market break," in *Black Monday and the Future of Financial Markets*, Homewood, IL: Irwin.

Kultti, Klaus, 1998, "Scale returns of a random matching model," *Economics Letters*, Vol. 58, pp. 277–280.

Lagos, Ricardo, 2000, "An alternative approach to search frictions," *Journal of Political Economy*, Vol. 108, No. 51, pp. 851–873.

Louganoff, Roger, and Stacey Schreft, 1998, "Financial fragility with rational and irrational exuberance," Federal Reserve Bank of Kansas City, manuscript.

Mishkin, Frederic, 1991, "Asymmetric information and financial crises: A historical perspective," in *Financial Markets and Financial Crises*, National Bureau of Economic Research, project report, London and Chicago: University of Chicago Press, pp. 69–108.

Morgan Stanley Dean Witter, 1999, *Global Economic Forum*, October 19, available on the Internet at www.msdw.com/GEFdata.

Mortensen, Dale, and Christopher Pissarides, 1998, "New developments in models of search in the labor market," Northwestern University, manuscript.

Murray, Alan, 1987, "After the Crash," *Wall Street Journal,* October 21.

Nash, John, 1953, "Two-person cooperative games," *Econometrica*, Vol. 21, pp. 155–162.

Perotti, Enrico, 2000, "Banking regulation under extreme legal underdevelopment: Lessons from the Russian meltdown," University of Amsterdam, manuscript.

Roach, Stephen, 1998, *Global Economic Forum*, Morgan Stanley Dean Witter, New York, October 16, available on the Internet at www.msdw.com/gef/.

Schlesinger, Jacob M., 1998, "Whys, wherefores of a possible rate cut," *Wall Street Journal*, September 28.

Schlesinger, Jacob M., and Greg Ip, 1998, "Fed officials are upbeat on economy, signal further rate cuts if necessary," *Wall Street Journal*, October 23.

Schlesinger, Jacob M., and David Wessel, 1998a, "Fed cuts interest rates a quarter point in a surprise move to shore up markets," *Wall Street Journal*, October 16.

_____, 1998b, "Fed cuts short-term rates by 0.25 point," *Wall Street Journal*, September 30, p. A3.

Sprague, O. M. W., 1910, *History of Crises under the National Banking System*, U.S. Senate, 61st Congress, Government Printing Office, Washington, DC, No. 538.

U.S. Department of Commerce, Bureau of the Census, 1998, "U.S. international trade in goods and services—Annual revision for 1997," Washington, DC, No. FT 900, available on the Internet at www.census.gov/foreign-trade/Press-Release/97_press_releases/Final_Revisions_1997/.

Market discipline and subordinated debt: A review of some salient issues

Robert R. Bliss

Introduction and summary

Academics and regulatory economists have long been concerned that mispriced deposit insurance undermines monitoring of banks by investors and increases incentives for bank risk-taking. Government supervision provides a partial substitute for the private corporate governance services provided by a firm's shareholders and creditors. As financial firms have become more complex, however, government supervisors have found it more difficult to monitor them in a timely manner. This is particularly true of large, complex banking organizations. Accordingly, many analysts—both inside and outside the regulatory agencies—have suggested that supervisors should rely on "market discipline" to supplement their traditional supervisory methods (Meyer, 1999). The Basel Committee on Banking Supervision's consultative paper on capital adequacy (Basel Committee, 1999) asserts that "[m]arket discipline imposes strong incentives on banks to conduct their business in a safe, sound, and efficient manner," and designates market discipline as one of the three pillars on which future financial regulation should be based. (The other two pillars are minimum capital standards and supervisory review of capital adequacy.)

The Basel Committee's (1999) consultative paper puts forth few concrete proposals for achieving a greater role for market discipline. The paper's proposals were concerned primarily with requiring greater transparency—certainly a *sine qua non* of effective market discipline. However, a more concrete potential market discipline mechanism, in the form of subordinated debt requirements, has long been discussed in academic and some regulatory circles.[1] Such proposals are currently gaining regulatory prominence, particularly in the U.S. A Federal Reserve task force recently investigated whether requiring large banking firms to issue subordinated debt on a regular basis would enhance supervision. The resulting study, Kwast et al. (1999), includes a summary of 11 different previous proposals (table 1 in their study). In their response to the Basel Committee (1999), the U.S Shadow Financial Regulatory Committee came out strongly in favor of mandated subordinated debt as a mechanism for realizing enhanced market discipline of banks (see Kaufman et al., 2000). Evanoff and Wall (2000) provide another summary of subordinated debt arguments and address some potential criticisms. The wide-ranging Gramm–Leach–Bliley Act of 1999 mandates that the Treasury and Federal Reserve conduct a study of the advisability and optimal design of a mandated subordinated debt requirement and report back to Congress in 2001. In the interim, the 50 largest nationally insured banks, if nationally chartered, are required to have at least one issue of debt outstanding rated A or better.

Despite its increasing popularity as a potential means for controlling bank risk-taking, the term "market discipline" is commonly used with two importantly different meanings. Bliss and Flannery (2000) argue that the concept of (effective) market discipline incorporates two distinct components: investors' ability to *evaluate* a firm's true condition; and the *responsiveness* of firm managers to the investor feedback impounded in security prices or, alternatively,

Robert R. Bliss is a senior economist and economic advisor at the Federal Reserve Bank of Chicago. *The author thanks seminar participants at the Federal Reserve Bank of Chicago, the Bank of England, the Sveriges Riksbank, the Federal Reserve System Committee on Financial Structure and Regulation, Daniel Covitz, Simonetta Ianotti, George Kaufman, Helen Koshy, Don Morgan, Anthony Saunders, Kevin Stiroh, and an anonymous referee for comments and discussion. He offers especial thanks to Mark Flannery for long and constructive discussions on the issues discussed herein. Donald Morgan and Kevin Stiroh kindly provided the data used in this article.*

regulatory feedback triggered by changes in security prices. Although the banking literature often fails to distinguish clearly between these components, their implications for regulatory reform differ substantially. Following Bliss and Flannery, this article defines two distinct aspects of market discipline: market monitoring and market influence.

- *Market monitoring* refers to the hypothesis that investors accurately understand changes in a firm's condition and incorporate those assessments promptly into the firm's security prices. Monitoring generates the market signals to which managers are thought to respond.
- *Market influence* is the process by which a security price change engenders firm (manager) responses to counteract adverse changes in firm condition.[2]

Most studies of market discipline in the banking literature are in fact studies of market monitoring. For market discipline to be effective, it is necessary that both components of market discipline obtain.[3] The purpose of market discipline in the context of bank regulation is, after all, to control or effect changes in bank behavior.

The intuition underlying efforts to enhance market discipline of banking firms through mandated subordinated debt issuance is both simple and intuitively appealing. Risky-bond investors can enhance their wealth and welfare by evaluating default risks carefully and demanding adequate compensation for the risks they assume. Shareholders and their agents the firm managers then evaluate the full range of costs and benefits associated with any decision, conscious that increasing asset risk is likely to raise the cost of debt. Ideally, the resulting firm decisions will be socially appropriate.

This article examines this intuition from a number of perspectives. The large literature on corporate governance—the problems that lead to a need for discipline and the mechanisms that markets have evolved to discipline firms—points to many ways that markets in unregulated industries can discipline firms. The central focus of this literature (both theoretical and empirical) is on the manager as the decision-making agent in the firm and on the manager's incentives. Disciplinary mechanisms include boards of directors, hostile takeovers, actions by large stockholders, performance-based compensation, and the managerial labor market. The little evidence on how these factors operate in banking suggests that regulation is accompanied by a weakening of normal market disciplinary mechanisms. In this literature, to the extent that investors play a direct role at all, it is the equity holders who are presumed to have the strongest ability to influence managers.

A thread running through the subordinated debt literature is that equity holders, while they may be able to influence managers, indeed they are frequently assumed to be the managers, nonetheless have incentives that are opposed to the interests of bondholders and regulators in ensuring that firms avoid undue risk-taking. This presumption leads to downplaying the role of equity holders in potentially enhanced market discipline of banks and of equity prices as signals of bank problems. The appendix discusses the sources for this argument and argues that a more realistic view of equity holders' interests suggests that ignoring this source of discipline and market information might be unwise.

Subordinated debt proposals rely in part on evidence that bond yield spreads reflect bank risk. This has been confirmed in numerous studies. In this article, I raise two important questions. The first question is "What else do bond yield spreads reflect?" As is occasionally noted, bond yield spreads are noisy measures of risk. The noise however is not random, rather it reflects numerous other factors that systematically affect bond yields—stale quotes, liquidity, embedded options, supply–demand factors, and Treasury bond market factors. Given the presence of substantial noise in bond yields, the second question is "Can we therefore use yields to reliably infer risk?" I show that when risk is measured by bond rating (a statistically significant predictor of yield spreads), yields spreads are poor predictors of risk. The underlying source of this failure, the presence of non-risk-related factors in bond yields, is likely to produce the same effects when "risk" is measured in other ways.

My goal is to point out previously underemphasized issues, raise questions, and examine assumptions in order to better inform the ongoing discussion, rather than to offer firm conclusions regarding how best to make market discipline an adjunct to regulatory discipline in banking. My analysis suggests that consideration should be given to examining existing regulatory barriers to the normal corporate governance mechanisms; that an undue emphasis on any single market signal may be suboptimal; and that the assumed superiority of yield spreads as measures of insolvency risk (vis-à-vis equity returns and accounting-based models) needs to be assessed empirically.

First, it is helpful to note how subordinated debt proposals differ in structure and rationale. This will provide a context for examining in detail the literature on corporate governance and the informativeness of bond yields.

Overview of subordinated debt proposals

The numerous current subordinated debt proposals vary in their underlying goals and objectives. These can, however, be broken down into three underlying rationales:

1) Subordinated-debt holders may *directly influence* banks to prevent them from taking on too much risk.
2) Yields on subordinated debt may provide *additional* useful *information* to regulators to assist in supervision.
3) Prompt corrective action (PCA) or automatic bank portfolio changes can be tied to yields on subordinated debt, preventing unwise regulatory forbearance and providing, in effect, a *regulatory fail-safe mechanism*.

Each rationale relies on slightly different definitions of the problem subordinated debt proposals are intended to solve, and each is predicated on different, though not mutually exclusive, assumptions and analysis of the extant evidence. The direct influence rationale seeks to supplement existing regulatory influence with market influence.[4] The desire to do so is in part motivated by resource limitations of supervisors—markets are thought to provide continuous monitoring and influence, while supervisors examine banks infrequently—and presumed informational advantages of market participants. The evidence that is usually cited to support this rationale includes the observation that risky banks pay higher yield spreads and issue less uninsured debt. Another piece of supporting evidence is the observation that derivatives market participants frequently decline to trade with low-rated counterparties (Greenspan, 2000). The direct influence rationale relies on the operation of market discipline across all banks—preventing sound banks from becoming marginal in the first place and causing marginal or unsound banks to become less so.

The additional information rationale is predicated on the assumption that market participants are able to analyze public, and perhaps private, information in ways not available to regulators—perhaps due to a better understanding of the bank's operating environment or a better understanding of the valuation of complex positions. The additional information rationale seeks to harness this market information, reflected in subordinated debt prices, to supplement the information obtained through examinations to improve the efficacy of regulatory influence. Extensive evidence supports the hypothesis that markets can effectively identify a firm's true financial condition, at least on average and on a contemporaneous basis.[5]

The regulatory fail-safe mechanism rationale seeks to address perceived problems in supervisory incentives, which result in costly forbearance. Eisenbeis and Horvitz (1994) analyze the theoretical arguments for and against forbearance and survey the empirical literature on the efficacy of past instances of forbearance. Forbearance may be optimal when there exist market frictions (for example, bankruptcy costs) and information asymmetries, and supervisors have greater ability to assess the viability of a particular bank than the bank's customers, shareholders, or the market in general. However, Eisenbeis and Horvitz (1994) conclude that "[w]hile some forbearance decisions have worked out, recent research has suggested that … agencies' ability to predict is limited at best." A key example of forbearance and its adverse consequences is the failure of regulators to intervene in a timely manner during the savings and loan crises of the early and late 1980s.[6] The explicit policy of forbearance applied to the savings and loan industry by the Federal Home Loan Bank Board in 1981–82 was followed by congressionally mandated forbearance in the form of the Competitive Equity in Banking Act of 1987.[7] Kane (1990), Eisenbeis and Horvitz (1994), Kaufman (1995), and others have argued that this forbearance created perverse incentives for uneconomic risk-taking that substantially increased the eventual cost to the taxpayer. Forbearance also occurs for individual banks—Continental Illinois being the most famous example—both due to too-big-to-fail concerns and because a bank's failure may be perceived as a supervisory failure, creating incentives to avoid regulatory recognition of problems in the hope they will resolve themselves.

By explicitly tying mandated regulatory intervention to subordinated debt signals, proponents of this approach seek to prevent unwise forbearance, in effect making supervisors the agents of the market.[8] The regulatory fail-safe mechanism rationale and, to a somewhat lesser degree, the additional information rationale focus attention on marginal, or potentially problem, banks rather than equally across all banks. The information debt yields may provide is clearly most important for banks that are most in need of supervisory intervention.

Agency costs and market discipline

Discussions of market discipline in general, and subordinated debt proposals in particular, frequently assume, without qualification, that market discipline would obtain if only regulatory distortions such as too-big-to-fail were eliminated and transparency increased.[9] However, the highly idealized view of the

relevant economic agents implicit in these discussions masks critical issues that impact the functioning of market discipline mechanisms (both monitoring and influence).

One major theoretical benefit of choosing subordinated debt to be the market instrument supervisors use to supplement their own examination efforts is that junior debt has a payoff structure closely resembling the payoff facing regulators.[10] Proponents argue that increasing bond market discipline is consistent with regulatory objectives of minimizing risks to the deposit insurance fund and ultimately the taxpayers. Regulators, the deposit insurance fund, and subordinated debt investors all risk losses when bank condition deteriorates. By contrast, none of these groups shares meaningfully in the potential upside rewards of risk-taking that accrue to bank equity holders. Banks are assumed to have incentives to increase the risk of underlying assets to maximize the value of deposit insurance and to increase the value of the residual (equity) claim on the bank's profits. Uninsured creditors, including bondholders, are reasonably thought to prefer less risk in general. While intuitively appealing, this "moral hazard" view of the bank regulation problem is not the only possible perspective. The arguments underlying this analysis importantly make no distinction between banks, bank managers, and bank equity holders, thus ignoring an important element of the corporate governance problem. Furthermore, both equity holders and bondholders are harmed when the banks they invest in make poor investment decisions. (The appendix assesses the evidence concerning these two potential regulatory problems.)

The idealized worldview underlying subordinated debt proposals ignores what are called principal–agent problems, or simply agency problems. These problems and their underlying cause, separation of ownership and control, were noted as far back as Smith (1796 [1776], Vol. III, Book 5, p.124).

> The directors of such [joint-stock] companies, however, being the managers rather of other people's money than their own, it cannot well be expected, that they should watch over it with the same anxious vigilance with which partners in a private copartnery frequently watch over their own. … Negligence and profusion, therefore, must always prevail, more or less, in the management of the affairs of such a company.

A formal theory of agency was first developed by Jensen and Meckling (1976) and applied to the modern corporation by Fama (1980) and Fama and Jensen (1983). Agency costs are created by the separation of ownership, or provision of capital, and control (management) in an environment characterized by information asymmetries, costly monitoring, and incomplete contracting. These unavoidable costs arise because investors cannot reliably ensure that managers will act in the investors' interest and not the managers'—that is, equity (and bond) holders cannot perfectly discipline managers. There is an extensive empirical literature analyzing the determinants, extent, and magnitude of agency costs in the economy.[11] The success of corporate capitalism clearly demonstrates that the benefits of separation of investment and management far outweigh the agency costs that this separation gives rise to.[12] However, it would be incorrect to conclude that agency costs are negligible and that, therefore, market discipline is not significantly affected by these costs.

Agency costs have several important implications for the market discipline/subordinated debt discussion:

1) A firm is not a single, rational economic agent, but rather a legal fiction. Modern corporate finance views a firm as a "nexus of contracts" among managers, equity holders, bondholders, workers, and customers, and these are the entities that respond (in their own interests) to economic incentives. Anthropomorphizing firms obfuscates these important issues.[13]
2) The principal–agent conflict exists between managers and equity-holders/bondholders. Therefore, managers and equity holders cannot be viewed as a single economic agent whose incentives are opposed to those of the bondholder/regulator.
3) The modern capitalist economy incorporates, and has evolved mechanisms to deal with, substantial impediments to outside investors' ability to make managers act in their best interests.

Agency costs can be mitigated by various mechanisms: delegated monitors (boards of directors, regulatory supervision), reducing information costs (required disclosures of relevant information), and reducing managers' incentives to abuse their position (fiduciary, fraud, and insider trading laws; threat of a takeover; and performance incentives such as managerial stock options). However, agency costs cannot be entirely eliminated, thus market discipline of managers is inevitably imperfect.

Equity holders and bondholders can surely influence managers *in extremis*. For example, when money market participants refused to roll over Penn Central's commercial paper in 1971, management was forced to take action. They filed for Chapter 11 bankruptcy protection from the firm's creditors. Equity holders can also vote out management, and poor firm performance increases the likelihood of managerial turnover.

Sufficiently disgruntled equity holders may create an environment that facilitates a hostile takeover.

Direct and reasonably certain discipline of managers is possible only in certain circumstances. The market for corporate control (takeovers) and direct control by large external equity holders who have effective-control blocks are forms of market discipline.[14] Major equity holders can themselves effect changes in board composition or form, at relatively low cost, coalitions to do so; inducing the board to change management. Major investors or other firms can mount hostile takeovers, which if successful will result in a change in management. These situations, while they represent the ultimate sanction against management (other than prosecution for malfeasance), are rare events and affect only the top managers of a few very large firms. Informal or ad hoc restrictions of derivatives dealing to highly rated counterparties are another form of market discipline, by preventing managers from engaging in certain forms of excessive risk-taking if they wish to participate in those markets. Empirical evidence confirms the existence, though not the invariable effectiveness, of all these disciplinary forces.

Labor market discipline is another form of market discipline. For most senior managers, the hope of more lucrative jobs at other firms induces them to work to establish their reputations as value-enhancing agents acting in the equity holders' interests.[15] Examining the post-resolution placement of bank managers following a number of Texas bank failures in the 1980s, Cannella, Fraser, and Lee (1995) find results consistent with the managerial labor market discriminating between managers who were likely to have been responsible for their bank's problems and those who were not. Managers likely to have been responsible for bank failures tended not to be subsequently employed in the industry, while those arguably not responsible were frequently employed by other banks. Farrell and Whidbee (2000) find a similar result for outside directors. In this case, outside directors who were aligned with forcibly removed chief executive officers (CEOs), owned little equity, and made poor choices in replacing the CEO, on average, subsequently lost their positions. Directors who were not aligned with the fallen leader (and/or had large equity stakes) not only kept their current directorships, but also were appointed to additional directorships at other firms.

Policy proposals for using market discipline to enhance banking supervision usually envisage something more commonplace, constructive, and benign than precipitating bankruptcy or replacing management through takeovers. Yet we have virtually no empirical evidence, outside the managerial labor market literature, concerning equity holder and bondholder market influence in non-extreme situations.[16]

The hypothesized form of market discipline underlying subordinated debt proposals is through the prices investors demand in return for providing capital to a firm. The secondary market prices of outstanding securities provide an indication of the rates of return investors will demand when the firm next comes to market. For a firm that can fund investments through internally generated cash flows, such market signals may have little direct effect should the manager choose to ignore them. For firms that have to raise new capital in the market, negative market signals in the form of depressed security prices will eventually translate into an increased cost of funds if price declines for one security class are not offset by price rises in another.[17] This increased cost of funds reduces the return on existing projects the firm invests in and may discourage the firm from investing in marginal projects. However, even if this effect is material, does this prevent managers from taking risks?

There is, of course, no reason to assume that bank investment opportunity set expected returns are identical. Indeed, financial theory teaches us that expected returns and risk should be positively related. If a riskier investment portfolio is associated with a high enough increase in expected return to compensate for the increase in cost of funds, the manager will rationally choose to take the riskier position, bondholders will be compensated for their increased default risk, and equity holders will be better off. Thus, there is no unambiguous disincentive to taking on risky projects per se, as evidence presented in a later section demonstrates. There remain, of course, disincentives to taking on projects, be they very risky or relatively safe, that do not have the required risk-adjusted expected return.[18]

Only a few papers look at banking and corporate control (agency cost) issues, and most of the economics of regulation literature (which considers among other things the incentives of regulators) is concerned with market power regulation (for example, utilities) or safety regulation (for example, airlines), rather than firm financial safety regulation. To draw implications from our review of the corporate literature for the problem of bank regulation, we can only make tentative extrapolations of the existing theory and consider the few existing empirical studies.

The theoretical effects of bank regulation, supervision and deposit insurance on agency costs are potentially ambiguous. Examiners are in an excellent position to act as delegated monitors. They have

unparalleled access to information, and they can compel remedial action. Only the board of directors is in as strong a position to monitor and discipline management. Early empirical investigations of this hypothesis found little evidence that the supervisors' theoretical comparative advantage translated into measurable benefits. However, DeYoung, Flannery, Lang, and Sorescu (2001), using an improved research methodology and a unique data set, find strong evidence that exams do reveal information that is not known to the market. Thus, examiners functioning as effective delegated monitors may serve to reduce agency costs.

On the other hand, much of the information examiners develop is confidential, and other aspects of bank regulation may have a negative effect. Deposit insurance obviously eliminates most, if not all, incentives for insured creditors to monitor. However, this may not be material. Insured depositors are unlikely to produce much useful information in any case—they typically have small stakes, reducing incentives to engage in costly monitoring, and are unsophisticated in valuation and risk assessment. Explicit too-big-to-fail policies in the 1980s undermined the incentives of uninsured creditors as well. This effect may have continued even after the passage of the Federal Deposit Insurance Corporation Improvement Act (FDICIA) in 1991, while the credibility of regulators in foreswearing forbearance remained untested. Even if one can argue that regulators will now let individual banks fail, imposing costs on uninsured creditors, one can also argue that diversified holders of uninsured claims might still rely on regulators' unwillingness to allow a large number of banks to fail. These factors would tend to increase free-riding and, therefore, undermine market discipline.[19]

One of the few studies that directly examines the agency cost consequences of bank regulations is Prowse (1997). Prowse examines the frequency of friendly mergers, hostile takeovers, management turnover initiated by the board of directors, and intervention by regulators in U.S. bank holding companies (BHCs) from 1987 to 1992, and compares this with data on the frequency of the first three of these corporate control events in nonfinancial firms. Prowse concludes that

> …while market-based mechanisms of corporate control in BHCs appear to operate in the same [broad] fashion as manufacturing firms they may be weakened because hostile takeovers are precluded by regulation and bank boards of directors are not as aggressive in removing poorly performing managers. These weaknesses leave intervention by regulators as the primary force in disciplining management. (Prowse, 1997, p. 525)

This evidence suggests that, whatever the informational benefits of examination, one effect of regulation on banks is to reduce the effectiveness of other corporate governance mechanisms. However, we cannot say anything on the basis of this sparse evidence as to whether regulatory discipline has been effective in *replacing* the market's usual disciplinary mechanisms.

The evidence of bond yields as measures of bank risk

The "additional information" and "regulatory fail-safe mechanism" rationales for subordinated debt proposals are predicated on the informativeness of subordinated debt yields. To be useful for their intended purpose, the yields on the bonds of problem, or potentially problem, banks must provide early and accurate warning of latent problems in sufficient time for supervisors to step in or for PCA triggers to take effect and avert the danger. Flannery (1998) and Kwast et al. (1999) both provide extensive reviews of the evidence on the accuracy and timeliness of the information in various bank and BHC debt yields.

The evidence of bond yield informativeness, needed to support subordinated debt proposals, has several components: the responsiveness of yields to bank risk; the timeliness of this response; the incremental informativeness of bond yield changes as to changes in bank risk; and the relevant sample of banks to be considered when examining these issues.

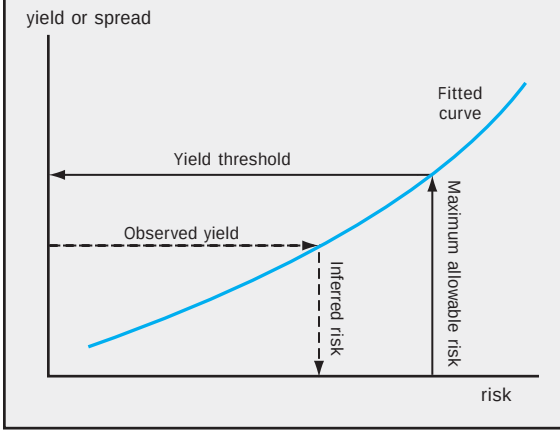
Cross-sectional studies

The informativeness of subordinated debt yields has generally been measured by regressing yields or yield spreads against various accounting measures of risk. A few studies have also used examiners' ratings as measures of risk. Authors of these studies implicitly assume that if cross-sectional variations in yields or yield spreads reflect issuer risk, then yields or yield spreads can be used as an indicator of issuer risk. Figure 1 illustrates their logic. An upward sloping regression line is seen as evidence that yields respond to risk.[20] If the slope is significantly positive, it is assumed that issuer risk can be inferred by observing the yield or yield spread and then, in effect, translating this into the corresponding issuer risk using the regression line. Proposals that contemplate setting a yield-spread threshold would first determine a maximum acceptable level of risk and then the yield-spread/bank-risk regression line would provide the corresponding yield or yield-spread threshold.

Early studies, using data through the mid-1980s, found little relation between measures of bank risk and subordinated debt yields. This is generally considered

FIGURE 1

Bond yields as a signal of bank quality



to be due to the residual effects of an explicit too-big-to-fail policy in the early 1980s and uncertainty as to the credibility of PCA subsequently mandated under FDICIA. Later studies, including Flannery and Sorescu (1996) and Jagtiani et al. (2000) find a statistically significant relation between bond yields and accounting measures of risk.

These studies are supportive of the hypothesis the bond yields *in general* respond to changes in bank or BHC risk. However, cross-sectional studies by their nature cannot illuminate two key issues: relevance and timeliness of the signals. The banks included in these studies are generally all banks of a certain size with subordinated debt outstanding, together with a few minor qualifications. The additional qualifications are unlikely to be misleading, but the "all banks" sample is problematical. The behavior of bond yields for marginal banks, the ones of regulatory interest, cannot be inferred from cross-sectional studies of all banks. The timeliness of the bond yield signals also cannot be inferred from cross-sectional studies.

None of the cross-sectional studies finds anywhere near perfect agreement between the bond yields and the accounting (and/or examiner rating) measures of risk. For instance, the Jagtiani et al. (2000) paper finds that the accounting measures of risk or rating agency debt ratings or examiner ratings, together with various control variables, explain between 60 percent and 65 percent of the variation in bond yields. It is impossible to say whether the remaining 35 percent to 40 percent is uninformative random variation in bond yields, or informative variation picking up information not available from accounting data, rating agency, or examiner ratings. The unexplained variation between alternative risk measures provides a rough starting point for thinking about the signal precision of predictors of insolvency.

Furthermore, cross-sectional studies use publicly available accounting information to measure bank risk. It is thus only possible to assess whether bond yields provide *redundant* information. It may well be the case that bond yields provide additional information, as may other sources of public information (for example, news stories), but that cannot be concluded from these studies. Thus, by their nature, cross-sectional studies cannot resolve the issue of whether bond yields provide *additional* information not already available from other sources.

What is in a credit spread?

Credit spreads are defined as the differences in yield to maturity between risky bonds and equivalent risk-free bonds. "Equivalent bonds" are found using one of several methods. The simplest is to use a similar-maturity Treasury bond. A risk-free term structure, estimated using Treasury bonds, can be used to value a fictitious maturity- and coupon-matched risk-free bond. Some mandated subordinated debt proposals advocate looking at spreads over a rating-based index of corporate or bank bonds. It is also possible to account for options embedded in the corporate bond by using option-adjusted spreads, though the methodology for doing so is not standardized, and different researchers may estimate substantially different option-adjusted spreads. In any event, option-adjusted spreads are rarely used. However, default-risk, maturity, coupon, and embedded options are not the only differences among groups of bonds. As a result, any difference that has not been factored into computing the "equivalent risk-free bond" will show up in the credit spread. Many of these factors are not credit-related.

Studies of the determinants of corporate bond spreads find that non-default-risk-related factors are also determinants of average corporate bond spreads. These other priced factors include liquidity (Cornell, 1992), the level of the Treasury term structure (Duffee, 1998), the level and slope of the term structure (Minton, 1997, studying swap rates) and the supply of alternative investment (Sloane, 1963, and Jaffee, 1975). Time and cross-sectional variation in information asymmetries (arising from bank opacity) and perceptions of agency costs are likely to add non-default-risk-related components to individual bond yield spreads (for example, Crabbe and Turner, 1995). Other studies find that short- and long-maturity credit spreads differ significantly and frequently move in opposite directions (VanHorne, 1979, and Fama, 1986). Duffee (1999) concludes that a single-factor model (capturing a single

generalized "default risk") cannot explain credit spreads. Duffee (1998) argues that changes in interest rate volatility change the value of embedded call options. This can affect measured credit spreads as most corporate bonds are callable and benchmark Treasury bonds are generally not. The opposite situation may complicate the interpretation of mandated subordinated debt yields—benchmark bond indices are based on callable bonds, but most subordinated debt proposals contemplate non-callable subordinated debt issues.

This is not to say that credit spreads are not influenced by default risk. They undoubtedly are. However, the potential presence of a host of non-credit-related priced factors in credit spreads means that changes or cross-sectional variation in credit spreads cannot reliably be interpreted as solely arising from changes or differences in credit risk.

Studies of timeliness

A very few studies have tackled the timeliness issue directly. Berger et al. (2000), using Granger causality tests, find that examiner ratings changes lead both stock returns and bond ratings changes, and that both stock returns and bond ratings changes lead examiner ratings changes. This apparently anomalous result reflects both stock returns and bond ratings changes containing information not contemporaneously available in examiner ratings and vice versa. Berger et al. do not examine the relative informativeness of stock returns versus bond ratings changes. Bond ratings changes are not bond yields and other evidence shows that bond yields tend to lead bond ratings changes. One cannot infer that use of bond yields would have strengthened the lead relation over examiner rating changes, though it is not unlikely.

The Berger et al. (2000) study, like the cross-sectional studies discussed above, compares one risk proxy against another risk proxy. It provides no direct evidence of timeliness vis-à-vis actual bank problems, albeit confidential examiner ratings are likely to carry a great deal of information. The Berger et al. study does, however, address the problem of average versus problem banks. They focus on examiner rating changes from grade 3 (weak but sound) to 4 (marginal), the most critical break in the rating scale. The lowest rating of 5 is reserved for banks in the process of "resolution," that is, banks beyond help.

Randall (1989) examined the particular circumstances of 40 large BHCs that developed problems between 1980 and mid-1987. Randall first looked for early signs of bank solvency problems in accounting measures—variables such as nonperforming loans and return on assets—and news stories to identify when, in hindsight, the problem first became apparent in public information.[21] He then looked at the timeliness of stock returns, bond ratings changes, and examiner ratings changes. He finds that none of these exhibit any marked ability to anticipate events, even though signs of the problems were already in the accounting data. Randall's study is subject to a number of criticisms, including the possible effects of too-big-to-fail forbearance and whether examiner ratings changes lag examiner awareness of, and informal actions to resolve, problems. Nonetheless, this remains the only study of which I am aware that tackles the crucial question of whether bond markets (or rating agencies in this case) sufficiently anticipate problems to enable timely intervention or merely react when the problem becomes patently obvious and it is too late to avoid costly resolution.[22] Absent any convincing conflicting empirical evidence on this issue, one cannot dismiss the Randall result out of hand.

In summary, while the evidence of correlation between bond yields and risk measures used heretofore to support subordinated debt proposals is consistent with their being indicators of bank risk, there are important caveats to this evidence. The current empirical evidence lacks specific information on marginally performing banks and the timeliness of bond yield changes. To this we must add the considerable evidence that so-called credit spreads respond to a number of factors in addition to changes in default risk. This last issue becomes especially critical when we consider using subordinated debt yields as a regulatory fail-safe mechanism.

The interpretation of yield spreads

Even if bond yield spreads only embedded credit-risk-related factors, the interpretation of yield spreads for supervisory purposes would be anything but straightforward.

- Depositor preference causes deposit insurers (supervisors) and bondholders to care about different parts of the distribution of bank asset values. When comparing two theoretically correctly priced bonds, the higher coupon need not be associated with the higher default risk. When these are subordinated bank bonds, the bank with the higher coupon bond need not have the higher probability of incurring losses to the deposit insurer, nor the higher expected losses to the deposit insurer.
- Regulators may also have different attitudes toward risk than bondholders. Systemic risk concerns may make regulators concerned about default per se, while bondholders are clearly concerned about recovery in the event of default. Jagtiani et al.

(2000) argue that examiners focus on default probability, while bondholders are concerned with (their own) expected losses, incorporating recovery in the event of default as well as default probability.

- Subordinated bondholders of firms with little or no economic equity begin to behave like equity holders (Black and Cox, 1976). They will prefer the bank take on risky projects as they are almost certain to suffer losses otherwise, while deposit insurers protected by the remaining buffer provided by subordinated debt would prefer less risk to lock in the protection.
- Covitz, Hancock, and Kwast (2000) point out that yield spreads are based on promised yields. These overstate actual returns bondholders expect to receive after factoring in both the probabilities of default and likely recoveries in the event of default.[23]

Consequences of bond yield imperfections

Much of the subordinated debt discussion to date has implicitly assumed that bond yields are nearly perfect indicators of insolvency risk,[24] or, alternatively, that the unspecified costs of classification errors are less than the (not quantified) expected gains resulting from imposing the subordinated debt requirement.[25] However, evidence presented in the previous section shows that bond yields are less than perfect indicators of insolvency risk.

Explicit in some and implicit in most subordinated debt proposals is the idea that banks must maintain subordinated debt yields at an acceptable level. While the "or else" sanctions needed to give force to the regulation are infrequently discussed, the idea of "an acceptable level" necessarily implies the choice of a discrete trigger point for doing something. This is intentional, as the objective of many mandated subordinated debt proponents is to reduce supervisory discretion. The argument for a rule is that policymakers often made mistakes, because they were overconfident about their assessments or because of the temptation to favor short-term goals over long-term objectives. Conversely, the argument for discretion is that a mechanical rule would ignore pertinent, useful information about the economy's course.

These same issues are pertinent in the current discussion of how best to incorporate subordinated debt yields into the bank supervisory process. The requirement that supervisors should take PCA steps whenever a bank's debenture yield rises above some threshold is a fixed rule. Unless one believes that bond yields *perfectly* reflect bank default risk, however, such a rule will potentially penalize some truly solvent banks, as well as potentially permitting some dangerously undercapitalized banks to remain open. The "additional information" alternative would be for supervisors to evaluate the implications of debenture yield spreads in conjunction with other sources of information. This approach preserves some element of supervisory discretion. Nor does it constitute a policy change for those BHCs with subordinated debt already outstanding. Analysts who feel that supervisors are overly inclined to delay action will find this level of discretion too permissive.

It is beyond the scope of this article to investigate the costs of unnecessarily taking action against a bank erroneously thought to be in danger of insolvency. These depend crucially on the details of how mandated subordinated debt requirements are implemented and the costs of failing to take action against a bank that is in a more precarious position than is thought.[26] However, we do know that small amounts of signal noise can produce high frequencies of misclassification when attempting to identify the most risky banks,[27] and concrete evidence using actual yields is provided in the section analyzing yield spreads.

"Single-signal solutions" versus "all available information"

Single-signal subordinated debt trigger proposals use the yields or yield spreads on subordinated debt to override information from other sources. Even moderate proposals to (somehow) use subordinated debt yields as input to supervision stress that subordinated debt yields are a preferred or superior measure of bank condition. Reasons cited for doing so include the public nature of the signal, unlike examination results, and the simplicity of interpreting the information to various parties, unlike a sophisticated default-risk model. These are valid points, but using a single signal also has an unavoidable cost—it is likely to result in increased numbers of regulatory mistakes.

Supervisors have a plethora of sources of information available to them. These include confidential examination information (infrequently updated), publicly available accounting information (updated quarterly), various activity information reported electronically to supervisors on a frequent basis, equity returns and market value of equity, insured deposit rates and issuance amounts, uninsured CD rates, and subordinated debt yields. All of these indicators of bank quality will include some noise.

Subordinated debt yields are not the only candidates for a simple, single signal of bank quality. Equity is an obvious alternative. FDICIA originally required that supervisors prescribe a minimum level of the ratio of market value to book value of equity for publicly

traded shares "to the extent feasible." For instance, the Shadow Financial Regulatory Committee (1992), in discussing this provision, noted that reductions in the ratio of market value to book value of equity could provide a signal to supervisors of declining bank quality. Using equity prices is not, however, without controversy. Equity prices are noisy as tests of equity pricing models invariably find. Kaufman (1992) argues that BHC equity prices embed the effects of mispriced deposit insurance, which complicates their interpretation. A common criticism is that the incentives of equity-holder/managers are not aligned with those of supervisor/deposit insurers, and that, therefore, equity prices are not likely to be informative.[28] This idea is formalized in the Merton (1977) model discussed in the appendix. On the other hand, Levonian (2000) shows, using the same model, that equity prices contain the same theoretical information as debt yields, and that the theoretical interpretation of debt yields is no less subtle than the interpretation of equity prices. Several nonbank failure prediction models use traded equity prices as an input variable (see Altman and Saunders, 1998, for a survey of this literature).[29] Furthermore, equity markets are generally more liquid than debt markets. Equity is, of course, issued at the bank holding company level and this may complicate interpretation of price information. But then so are most currently issued subordinated bonds, which are considered to be sufficiently informative to make the empirical case for subordinated debt proposals.

Nonetheless, in practice no single signal is likely to be as informative as the simultaneous examination of several sources of information. It is a simple fact of statistics that when we observe a number of more or less independent noisy measures of some unobservable factor, it is usually best to combine the various measures rather than to pick one.[30] If the measurement errors in the various signals are not perfectly correlated they will tend to cancel out, resulting in the combined signal having a lower level of noise. This is always true if signals are combined in the correct way: more noisy signals should be given less weight and more accurate signals more weight. Only if one of the signals is perfectly accurate or all of the signals are perfectly correlated does it make statistical sense to disregard the other signals entirely.

In practice one would build, calibrate, and test a model that combines all available information to assess what combination of signals is optimal. Such a study would also provide a measure of the costs of using only one signal: one could compare the optimal-signal-mix results with the results obtained using only the proposed single signal.

Pettway and Sinkey (1980) used abnormal equity returns to predict problem banks, an accounting model to do the same thing, and then examined the results of combining the two forecasts. Each of the two models results in a classification of banks into "good" and "bad." The Pettway and Sinkey results demonstrate that, for their sample, both accounting information and equity returns contained useful information of subsequent bank problems and that the combination produced even better forecasts. The latter results demonstrate the benefits of combining noisy sources of information, and the former results demonstrate the potential usefulness of equity returns in failure prediction.

Given the evidence concerning the potential informativeness of alternative measures of bank insolvency risk, the sole use of subordinated debt yields contemplated in regulatory fail-safe mechanism proposals is almost certainly trading simplicity for accuracy. To date this tradeoff has not been examined. The potential benefits of using a single, simple, public measure of bank risk are perhaps not quantifiable; however, the costs, in terms of predictive accuracy, of disregarding other information are quantifiable.

Analysis of yield spreads and ratings of newly issued bonds

Kwast et al. (1999) and other authors stress the informational superiority of newly issued bonds. Kwast et al. assert that information revelation by companies increases at the time of new security issuance. Thus, merely requiring regular security issuance, it is hypothesized, may increase transparency. The study considers yields of newly issued bonds to be "market prices," because money is changing hands in the process.[31] Secondary market prices for all but the most liquid bonds are apt to be "indicative prices," set by individual dealers for purposes of marking positions to market rather than firm bids and offers.

Morgan and Stiroh (2000) have compiled a database of financial and nonfinancial bond issues, their ratings by both Moody's and Standard & Poor's (S&P), and their yields at time of issue. This database permits us to examine the transparency of information available at time of bond issue by examining agency-rating agreement.[32] Ratings are frequently criticized as potential measures of risk because they are frequently stale. However, the Morgan–Stiroh database permits us to study simultaneously determined market yields and agency assessments of default risk.

The rating scales used by Moody's and S&P are comparable in their definitions, and with slight adjustments for labeling of risk classes, are used

interchangeably. One measure of transparency is the extent to which rating agencies agree on the default risk of bonds. This can only be reliably done when ratings are simultaneous determined, as at time of issue of a new security. Table 1 presents the Moody's and S&P ratings for financial and nonfinancial firms from 1993 through 1998.[33]

Earlier, I argued that bonds do not necessarily restrict risk-taking: So long as bondholders are adequately compensated, they will be willing to lend to risky companies. Using S&P ratings, fully 46 percent of the new nonfinancial bond issues fail the Gramm–Leach–Bliley test of "top three ratings categories," while 16 percent of newly issued nonfinancial bonds are below investment grade.[34] The figures for Moody's ratings are comparable at 44 percent and 14 percent, respectively. Risky-bond issuance by financial institutions is considerably less frequent. Using S&P ratings, 21 percent of new financial bonds failed the Gramm–Leach–Bliley threshold, while only 2 percent were not

TABLE 1

A comparison of Moody's and S&P ratings for newly issued bonds (1993–98)

A. Nonfinancial institutions

	Moody's																
S&P	**Aaa**	**Aa1**	**Aa2**	**Aa3**	**A1**	**A2**	**A3**	**Baa1**	**Baa2**	**Baa3**	**Ba1**	**Ba2**	**Ba3**	**B1**	**B2**	**B3**	
AAA	**175**	7	–	–	–	–	–	–	–	–	–	–	–	–	–	–	182
AA+	9	**7**	31	2	–	–	–	–	–	–	–	–	–	–	–	–	49
AA	2	11	**34**	46	3	–	–	–	–	–	–	–	–	–	–	–	96
AA–	5	5	48	**101**	29	2	2	–	–	–	–	–	–	–	–	–	192
A+	–	–	18	86	**185**	83	6	–	–	–	–	–	–	–	–	–	378
A	–	1	5	5	128	**358**	40	–	10	–	–	–	–	–	–	–	547
A–	–	–	–	7	13	131	**196**	87	1	–	–	–	–	–	–	–	435
BBB+	–	–	–	–	1	26	62	**170**	78	8	8	–	–	–	–	–	353
BBB	–	–	–	–	1	5	28	84	**214**	39	2	–	–	–	–	–	373
BBB–	–	–	–	–	–	1	–	27	96	**178**	3	2	–	–	–	–	307
BB+	–	–	–	–	–	–	–	1	13	37	**18**	12	4	1	–	–	86
BB	–	–	–	–	–	–	–	–	2	14	25	**17**	11	3	3	–	75
BB–	–	–	–	–	–	–	–	–	1	–	6	11	**45**	22	4	1	90
B+	–	–	–	–	–	–	–	–	–	–	1	6	22	**46**	37	3	115
B	–	–	–	–	–	–	–	–	–	–	–	1	4	26	**82**	30	143
B–	–	–	–	–	–	–	–	–	–	–	–	1	1	12	37	**58**	109
	191	31	136	247	360	606	334	369	414	277	63	50	87	110	163	92	3,530

B. Banks and savings and loans

	Moody's																
S&P	**Aaa**	**Aa1**	**Aa2**	**Aa3**	**A1**	**A2**	**A3**	**Baa1**	**Baa2**	**Baa3**	**Ba1**	**Ba2**	**Ba3**	**B1**	**B2**	**B3**	
AAA	**2**	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	2
AA+	4	–	2	–	–	–	–	–	–	–	–	–	–	–	–	–	6
AA	–	23	**19**	1	–	–	–	–	–	–	–	–	–	–	–	–	43
AA–	–	1	8	**34**	14	8	–	–	–	–	–	–	–	–	–	–	65
A+	–	–	2	13	**42**	42	19	–	–	–	–	–	–	–	–	–	118
A	–	–	–	3	4	**70**	35	3	–	–	–	–	–	–	–	–	115
A–	–	–	–	–	–	20	**63**	23	1	–	–	–	–	–	–	–	107
BBB+	–	–	–	–	–	2	16	**10**	11	1	–	–	–	–	–	–	40
BBB	–	–	–	3	–	–	4	13	**5**	10	–	–	–	–	–	–	35
BBB–	–	–	–	–	–	–	–	2	7	**19**	1	–	–	–	–	–	29
BB+	–	–	–	–	–	–	–	–	–	–	**1**	–	–	–	–	–	1
BB	–	–	–	–	–	–	–	–	1	1	–	–	–	–	–	1	3
BB–	–	–	–	–	–	–	–	1	–	–	–	1	–	–	–	–	2
B+	–	–	–	–	–	–	–	–	–	–	–	–	2	**3**	–	–	5
B	–	–	–	–	–	–	–	–	–	–	–	–	1	1	**1**	–	2
B–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	1	–	1
	6	24	31	54	60	142	137	52	25	31	2	1	2	4	2	1	574

Notes: Ratings agreements indicated in boldface. Moody's ratings totals provided in right-hand column; S&P ratings totals provided in last row. Table is broken into quadrants based on Gramm–Leach–Bliley reference to "top three broad ratings classes."

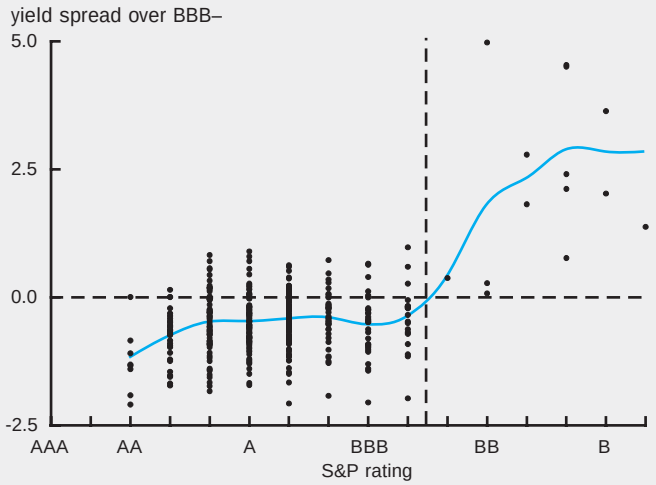
Source: Morgan and Stiroh (2000); provided by authors.

investment grade. Moody's numbers are again comparable at 20 percent and 2 percent, respectively.[35]

There are clear differences in the ratings characteristics between financial and nonfinancial institutions. This may reflect in part the incorporation of regulatory supervision into the rating agencies' assessments of default risk; or it may reflect the access that banks have to insured funds, which may make them choose to issue low-rated bonds less frequently. Nonetheless, the bond market does appear to support issuance of debt by risky firms. Nonfinancial firms frequently find that issuing marginal or sub-investment-grade debt, even given the generally higher yields required to do so, is a viable corporate finance decision. This latter statistic suggests that if banks were forced simply to issue subordinate bonds without restrictions on yields or ratings, the bond markets would accommodate their doing so at a price, and the price extracted might not in general be prohibitive.

FIGURE 2

Yield spreads of newly issued bank bonds



Note: Yield spreads are measured relative to contemporaneous newly issued nonbank BBB bond yields.
Source: Morgan and Stiroh (2000), provided by authors.

Exact ratings-agreement occurs 53 percent of the time for nonfinancial issues and 47 percent of the time for financial firms. However, there are slight variations in risk assessment. Ratings differences of two or more sub-categories, for instance from Baa1 to Baa3, occurred 8 percent of the time for nonfinancials and 9 percent of the time for financial bond issues. It is not obvious on the basis of these numbers that financial institutions are more or less transparent than nonfinancial institutions; however, rating agencies differ in their assessments by more than a trivial amount sufficiently often to raise the question whether transparency is adequate at time of bond issuance. Nonetheless, ratings appear to be generally consistent across rating agencies.

However, yields are not very consistent within ratings. To examine the collective impact of the sources of priced non-default-risk factors on bond yields, I examine the yield spreads plotted against S&P ratings for financial firms in figure 2.[36] To eliminate the most severe term-structure effects, I only consider issues with five or more years to maturity. To compute spreads, a time-varying BBB benchmark rate was computed by averaging nonfinancial institution yields-at-issue for newly issued bonds each date. These average yield observations were then interpolated between nonfinancial bond issue dates to create a BBB benchmark rate for each financial bond issue date. The yield-spread-at-issue for each financial institution bond was then computed by subtracting the yield-at-issue from the BBB benchmark rate for that day. This methodology approximates the Calomiris (1999) suggestion of requiring subordinate bonds to have yields comparable to other BBB-rated bonds. Focusing on yields over BBB rather than yields over Treasury bonds eliminates some of the common non-default-risk factors that influence all risky-bond spreads (for example, the level and slope of the Treasury term structure). Ideally one would wish to eliminate industry effects as well by using a financial BBB benchmark rather than a nonfinancial one. Unfortunately, there are insufficient financial institution issues to construct a time-varying "financials only" index, without using illiquid seasoned bond yields (which may not be market prices) and stale ratings.

Ratings-at-issue are not the only measure of risk against which yield spreads may be plotted, but they are certainly plausible. Research already discussed shows that ratings provide some incremental information over examiner assessments. Agency-issued ratings are generally used as measures of risk by portfolio managers, are a major component of proposed revisions to bank capital standards (Basel Committee, 1999), and are enshrined in law (for example, Gramm–Leach–Bliley).[37]

The continuous line in figure 2 is interpolated from the average within-rating-class yield-spreads. This is the summary information produced by a regression of yields against rating, for example, in Morgan and Stiroh (2000) figure 2 and table 3. This picture has

the advantage of not imposing linearity on the yield-spread/risk relation as is usually done (though not in Morgan and Stiroh). This permits us to see the non-linear nature of the relation between yield spread and risk. While yield spreads are overall increasing in risk, they appear to be more or less flat for a large range of ratings (A+ to BBB–). The relation is strongly upward sloping for sub-investment-grade bonds and for bonds rated AA to A+. In addition, regression results do not reveal the variation of yields about the estimated averages.[38] This variation has important implications for the use of yield spreads to infer risk. The yield spreads for similarly rated bonds with ratings between AA and BBB vary by 50–150 basis points above and below their average values. This variation is far greater than the variation in across-ratings-class mean yield spreads. While one would expect that bonds rated higher than BBB would have negative yield spreads (relative to BBB– yields), approximately 25 percent of bonds rated A+ through BBB+ have positive yield spreads. Variation seems to increase for sub-investment-grade bonds—BB-rated bonds have yield spreads ranging from 0 to almost 500 basis points—though there are too few observations for reliable inference.

Critical to the use of bond yields as triggers for PCA is the belief that bond yield spreads can be used to infer bank risk. Using the Calomiris threshold of spread over BBB as an example, figure 2 shows that 100 percent of the bonds with yields below the BBB benchmark are indeed rated BBB or better. Of the bonds with positive yield spreads, only 20 percent are actually rated worse than BBB. This means that in four out of five cases where the "BBB yield or better" requirement would have triggered regulatory actions of some sort, these actions would have been taken against banks with acceptable risk levels. Whether this is an acceptable rate for imposing unnecessary regulatory costs on banks depends on the actions that threshold violations trigger, the resources available to regulators to act (for example, staff to immediately examine threshold-violating banks), and the availability of alternatives for reducing errors in identifying potentially problem banks.

While rating is not the only measure of risk, I believe that it is unlikely that other risk measures will materially improve the picture, though this remains to be tested. Arguably much of the variation of within-ratings yield spreads comes from the non-default-risk factors discussed earlier. It is conceptually possible to construct an "adjusted yield spread" to back out these other factors. However, two problems arise: feasibility and complexity. While models exist for pricing (and backing out) some factors such as embedded options, these models are not particularly reliable. Models simply do not currently exist for pricing other factors such as liquidity. Building a reasonably effective model to account for a substantial portion of the major non-default-risk factors is highly speculative at this time. Even if such a model were possible, the use of the model would destroy some of the main attractions of bond yields: that they are intuitive, simple, and directly observable. Once the process of building complex models to extract the desired information is begun, the question naturally arises as to whether to include other inputs.

Summary and recommendations

The growing tide of interest in regulatory circles, both in the U.S. and abroad, in bringing market discipline to bear on bank regulation joins an old academic belief that markets provide information and mechanisms for solving corporate control and incentive problems. Strangely though, the regulatory discussion has been devoid of the issues raised in the academic literature on these subjects. To date, the regulatory discussion has begun with the proposition that "markets discipline" and proceeded to discuss practical issues of how best to use that discipline in bank regulation. The result has been two lines of discussion: transparency and mandated subordinated debt issuance.

The corporate governance literature says a great deal that is informative about why market discipline is needed (in non-banking contexts) and what mechanisms markets have evolved to deal with the central agency problems. Market discipline is a complex issue with causes arising in agency problems—the separation of provision of capital (risk-bearing) and decision-making (management). Mechanisms that unregulated markets have evolved to deal with these problems include: the systems of laws and penalties that govern the economic environments, separation of monitoring and decision-making, delegated monitors, managerial incentives either through the labor market or through (usually equity-based) incentive contracts, the market for corporate control, and the voting power of equity holders. Notably absent in the corporate governance literature is any discussion of debt as a disciplinary mechanism.

The corporate governance literature suggests that improvements in direct market influence of managerial decision-making may best be achieved through addressing the relation between regulation and the effectiveness of the usual mechanisms of market discipline. Evidence cited here suggests that these mechanisms may be relatively weak in the banking sector.

Improvements in transparency are almost surely beneficial[39] but are nonetheless proving extremely difficult to implement. This article only notes that adequate transparency is the *sine qua non* of any efforts to improve market discipline.

Mandating subordinated debt issuance is the primary specific proposal under wide discussion for altering the supervisory framework to bring market discipline more forcefully to bear on banks, either directly through the bondholders influencing bank risk-taking or indirectly through the information provided through debt yields. The latter route can be broken into relatively vague "additional information" arguments or relatively explicit "regulatory fail-safe mechanism" arguments. The additional information rationale has its roots in the idea that banks are increasingly complex and markets are better able than examiners to analyze them. The regulatory fail-safe mechanism rationale has its roots in the fear that supervisory incentives are not aligned with public goals and that supervisors are apt to forebear closing problem banks even when they are aware of the problems. This was all too painfully evident in the savings and loan crisis (although the contribution of politics should not be overlooked). FDICIA was intended to cure the worst of these supervisory incentive problems, but many remain unconvinced.

My review of the corporate governance literature suggests that we should not put undue reliance on mandated subordinated-debt holders directly influencing bank management. While the additional information and regulatory fail-safe mechanism rationales originate in very different views of the regulatory problem to be solved, and few explicit and detailed regulatory fail-safe mechanism proposals exist, the regulatory drafting process may be drawn toward some form of regulatory fail-safe rule if a mandated subordinated-debt proposal is implemented. This could arise for two reasons: mandated issuance alone may not be sufficient to alter the behavior of managers to avoid undertaking excessive risks; and the necessary regulatory "or else" provision requires clarity regarding when the or-else will be invoked.

Both additional information and regulatory fail-safe mechanism rationales rely on the informativeness of yields. I postulate that the relevant criteria for judging the value of any closely monitored signal of bank quality should be: that it be timely (to intervene before doing so becomes costly); accurate, particularly for potentially insolvent banks (to reduce the frequency and costs of unnecessary interventions); and, if regulatory fiat is required to produce a signal, that signal should be superior to alternative sources of information already at hand.

The existing evidence on subordinated debt yield informativeness does not allow us to reliably ascertain whether these conditions hold. Cross-sectional studies only tell us that *in general* subordinated debt yields respond to contemporaneous measures of bank risk. Because most of the studies cover all banks with subordinated debt outstanding, we cannot determine if the results apply equally well to marginal banks (the more numerous sound banks tend to overwhelm the empirical results). Because these studies use only contemporaneous measures of risk, they cannot establish the relative efficacy of available risk measures. This is not to say that subordinated debt yields are more or less informative than alternatives, only that, excepting the absence of bond yields in bankruptcy prediction models, we have no empirical information or even hint of their relative efficacy; notwithstanding oft-made theoretical arguments that bond yields *must* be informative (and equity prices not).

The key criticism in this article is regarding reliance on any single signal as a (sufficient, though not necessary) basis for regulatory intervention. Numerous alternative sources of information are available. Both empirical evidence and theoretical analysis suggest that these alternatives are not to be dismissed out of hand. The large body of evidence that factors other than default risk are embedded in bond yields suggests that while bond yields respond to default risk, they also respond to many other factors. Hence, inferring default risk from bond yields is apt to be imprecise. While some authors note that yields are noisy, the analysis presented here of actual yields-at-issue of bank subordinated debt provides a quantitative indication of the potential effects of trying to use yields to measure risk. The problem appears to be substantial.

Basic statistics and the analysis in this article argue that where signals are imprecise, it is better to combine available information rather than use any one signal. This argues for optimally combining (existing) subordinated debt yields with equity prices and accounting information into a unified problem prediction framework. Whether in such a context subordinated debt yields contribute enough new information to justify mandating their issuance is yet to be determined and depends also on the costs (direct and indirect) of imposing such a requirement.

Central to the discussion of "why subordinated debt" and "why not equity" is the argument that a moral hazard problem leads equity-holder/managers to invariably attempt to expropriate bondholders and increase the value of their call-option-like claim by increasing asset risk. Concomitantly bondholders,

like deposit-insurer/regulators, have a shared interest in limiting risk-taking. This framework ignores equity holder versus manager agency issues and important differences between bondholders' and regulatory payoffs. This overly simple model can be extended in a number of ways with marked changes in the implications. As the appendix discusses, the unconditional equity-holder appetite for risk conclusion can be mitigated or reversed by considering more realistic multiperiod frameworks in which bank charters are valuable, managers can dynamically adjust asset risk, insolvency may occur continuously or at random times, and the investment opportunity set can vary in expected return as well as risk.[40]

Poor (apparently irrational) investments are as problematic as excessively risky projects (with positive risk-adjusted returns). Evidence suggests that poor investments are likely to be the major explanation for banks getting into trouble, and agency problems readily explain how they may come about (albeit incompetence per se is not an agency problem). The a priori argument that equity prices cannot be informative has been countered using the same model employed (informally) to draw that conclusion.

Recommendations

My first major recommendation is that the central question of optimal bank insolvency prediction be addressed. Failure prediction and econometric methods have advanced since Pettway and Sinkey (1980), and it is desirable to include all available information in an optimal bank failure (or "problem short of failure") forecasting model. The efficacy of an optimal, full-information model can then be compared with single-signal models using subordinated debt yields, equity returns, and other alternatives. Such a study would provide crucial information on the best ways to incorporate available market information into the supervisory process and the potential costs of relying solely on subordinated debt yields as a PCA trigger. The data and econometric techniques are readily available.

My second recommendation is that the discussions of enhancing market discipline take cognizance of the evidence from other literatures and explore a range of approaches. The demonstrated importance of managerial labor markets and the market for corporate control in disciplining managers suggests a need to evaluate the impact of regulation and supervision on the functioning of these mechanisms in the banking sector.

My final recommendation is that if a mandated subordinated debt proposal is implemented, it lean heavily toward solely providing information to supervisors and providing supervisors with incentives to consider that information. For example, requiring that whenever a large bank's debenture credit spread exceeds some level, the head of the examining agency must either

1) Promptly initiate PCA, or
2) Report to a national oversight authority such as the Federal Financial Institutions Examination Council why he or she believes that no current supervisory action is appropriate. The oversight authority would have the authority to direct PCA be taken if it was unconvinced.

Such a "rule" would raise the supervisor's cost of forbearance, while maintaining the flexibility to incorporate other sources of information. Thus, highlighting the invoking of, and reasons for, forbearance would address a major source of concern—the temptation for the immediately responsible supervisors to bury a problem.

The temptation to impose yield-spread thresholds and costly related regulatory consequences on banks should be avoided. Either the thresholds will be set so high as to fail to provide timely triggers or, alternatively, will be almost certain to produce many unnecessary interventions. The more meaningful the automatically triggered regulatory interventions are, the more costly the unnecessary interventions will be.

Conclusion

The concerns about the limitations and distortions attendant to the current regulatory environment are well placed, and markets undoubtedly can provide important influences on the safe and sound operation of banks and BHCs. However, the limitations of market discipline are as important as the successes. Ardent advocacy of one possible mechanism for achieving market discipline should not blind us to unresolved issues, countervailing evidence, or the full range of possibilities. While no regulatory or market mechanism can be perfect, we should consider all available alternatives and not simply ask if one proposal is "good enough." I do not advocate pursuit of an unattainable ideal—"making the best the enemy of the good." Rather I advocate not closing our minds to strengths and weaknesses of all approaches—not "making the good the enemy of the better." If these issues are addressed head on, rather than ignored or brushed aside, the resulting policy decision will have a better chance of achieving regulatory objectives and avoiding unintended consequences.

I do not pretend to have shown that mandated subordinated debt is materially worse than alternative approaches to incorporating market discipline into

bank and BHC supervision, or doing nothing for that matter, only that the evidence underlying subordinated debt proposals is incomplete and open to question and to note alternatives. Some readers will feel that the costs of mandated subordinated debt, while not quantified, are unlikely to be high and the likely benefits, while not quantified, are sufficient to proceed immediately. These readers may also feel that there is an urgency to regulatory reform that precludes waiting until open questions have been investigated further. Others may feel that the imposing of regulations should be done grudgingly and only with the best possible (albeit imperfect) information or that an exclusive focus on subordinated debt may preempt or preclude the investigation of alternative, possibly more efficacious, approaches. These are questions of judgment and belief.

Whether the questions raised in this article are sufficient to justify delaying implementation of a subordinated debt proposal depends on one's priors. I leave the conclusion to the reader.

APPENDIX

Appendix: Is moral hazard the major problem?

> There have been more financial institutions ... that have disappeared by getting their strategy wrong than by being overexposed to risk. (Reed, 1999)

Firms can lose money in two ways: bad investments and bad luck. Bad luck can occur with good (positive risk-adjusted expected return) investments. However, bad outcomes may be more likely to occur with bad investments. The discussion of market discipline and subordinated debt proposals has focused almost exclusively on bank portfolio risk. The quality (expected returns) of bank investments has been, for the most part, ignored. This focus on risk is derived from the widespread believe that banks suffer from particularly severe moral hazard problems due to *de facto* (though perhaps past) too-big-to-fail policies and mispriced deposit insurance.[1] This section describes the conceptual framework that underpins most discussions of moral hazard in the subordinated debt literature. The thrust of this analysis is that equity holders have incentives to increase asset risk and thus expropriate bondholders. I then show that more complex and realistic models of incentives lead to less clear-cut conclusions and that in some circumstances equity holders and bondholders may have similar interests in ensuring the long-term viability of the firm. I end by reviewing the few empirical articles that have examined whether bank failures are due to "bad luck" (together with excessive risk-taking) or "bad investments."

The bank moral hazard argument is simple and intuitively appealing. The essence of moral hazard is the "heads I win; tails you lose" situation created when gains accrue to decision-makers while losses are borne by other agents. The definition of moral hazard used here arises out of equity-holder/manager versus bondholder conflicts and is a form of agency problem as the bondholders provide capital but do not control asset risk. The equity-holder appetite for risk implications of this form of moral hazard derives from the observation first made in Black and Scholes (1972) that equity can be viewed as a call option on the value of the assets of the firm. A similar framework was used to study deposit insurance moral hazard in Merton (1977).[2] Under this model, once bondholders have committed their capital at a fixed rate, equity-holders/managers (who are assumed to be identical) can increase the value of their option by increasing the volatility of the value of the firm by taking on riskier projects, thereby expropriating bondholders. The increase in the value of equity for a given increase in asset volatility is proportional to the amount of debt (including insured deposits if any) that the firm has.

This argument is however overly simplistic. The conclusion that equity holders unambiguously prefer more risk depends on the assumptions of the Black–Scholes–Merton framework: that all transactions are concluded when the bonds are (or are not) paid off, that equity is a European-style call option, and that returns to the underlying assets of the firm evolve continuously with normally distributed innovations. Changes in any of these assumptions can weaken or reverse the unqualified "equity holders like risk" conclusion.

It is possible to show that even in the single-period European call option context it is not necessarily the case that equity holders prefer risk if the "log-normally distributed returns" assumption made in most equity option pricing models is dropped. Whether the assumption that all asset prices in fact do follow a log-normally distributed stochastic process is realistic is an empirical question, seldom examined.

Another critical assumption of the Merton model is that the amount and cost of debt are locked in before

equity-holder/managers decide asset quality. Levonian (2000) has shown that the moral hazard incentive is reduced if subordinated debt is repriced after asset volatility is changed. Within the context of the Levonian model, if repricing is frequent, as for example with commercial paper, then the benefits to equity holders of increasing asset volatility are proportional only to the amount of debt that is not repriced (for example, deposits) and not to the total amount of debt. If frequently repriced debt is substituted for equity, leaving deposits unchanged, the moral hazard incentives are actually increased as the change in equity value for a given change in asset volatility remains unchanged, but now accrues to a smaller equity base. If repriced subordinated debt replaces risk-insensitive deposits, the moral hazard incentives are reduced by reducing the amount of debt that is not repriced when asset risk changes. Replacing the same amount of deposits with equity instead of repriced debt produces an even larger diminution in moral hazard incentives by spreading the reduced increase in equity value for a given increase in asset volatility over a larger equity base.

If firms are declared bankrupt as soon as the value of equity declines to zero then the equity holders hold a down-and-out barrier option, not the simple European call option envisioned in Black and Scholes (1972) and Merton (1977). Unlike holders of simple European options who always prefer more risk to less, holders of a down-and-out barrier call option prefer more risk to less only up to a point, after which increased risk reduces the value of their option. The option-value-maximizing level of risk declines as the barrier is approached (the firm approaches insolvency).

The moral hazard bondholder-expropriation argument is a single-period argument, while financing is in fact usually a multi-period problem. Expropriating bondholders in one period will raise the costs of borrowing in subsequent periods as bondholders reassess upwards the moral hazard agency costs they face vis-à-vis equity-holder/managers, thus reducing expected future profits. Merton (1978) modeled bank equity as a quasi-down-and-out perpetual option. The equity holders derive ongoing benefits from risk-insensitive deposit insurance; however, they face random audits (the down-and-out barrier option is equivalent to continuous audits). If during an audit the bank is found to be insolvent, the bank is liquidated and equity holders lose future rents from the mispriced deposit insurance. Merton found that as the barrier is approached (the firm approaches insolvency), the equity-value-maximizing level of asset risk declines. However, if the bank becomes insolvent, equity holders' incentives are to maximize risk in the hopes of re-achieving solvency before the next audit.

Access to debt financing is also made more valuable by the tax code that effectively favors debt financing over equity financing. Leland (1998) shows that tax advantages of debt outweigh the moral hazard agency costs attendant to issuing debt. He argues that equity-holders/managers therefore have incentives to control firm risk to reduce moral hazard/agency costs of debt.

Ritchken et al. (1993) found that permitting equity-holder/managers to dynamically adjust their portfolios prior to expiration of the equity call option also resulted in an optimal level of portfolio risk beyond which equity holders would not wish to go. Their model also included a "charter value," the present value of future profits, that would be lost to the equity holders if the firm becomes insolvent.

Geske and Shastri (1981) examine American options with uncertain (suspendable) discrete dividends. They show that in this case the relation between option value and underlying asset risk is not always monotonically increasing.

Deposit insurance confers benefits on equity-holders/managers—the value of the deposit insurance subsidy and value of the bank charter. Attempts by managers to exploit moral hazard may ignore the potential costs of doing so and may be short-sighted. Only banks already in deep trouble (low charter value) are likely to see "betting the bank" as a viable option.

The moral hazard analysis focuses solely on the incentives of equity-holder/managers to increase risk, implicitly assuming that the portfolio choices do not differ in their expected returns. This is unlikely to be the case. Where equity holders are not the managers, and agency problems are significant, equity holders and bondholders will have the same interests as regulators in higher expected asset returns, holding risk constant. Absent a model of the investment opportunity set faced by banks, it is impossible to theoretically determine the relative magnitude of these conflicting factors. A few papers have, however, examined the issue empirically.

Gorton and Rosen (1995) used a simple theoretical model, tested using bank call report data from 1984 to 1990, to examine the causes of declining bank profitability and increasing bank risk during the 1980s. They conclude that:[3]

> … managerial entrenchment played a more important role than did the moral hazard associated with deposit insurance in explaining the recent behavior of the banking industry. (Gorton and Rosen, 1995, p. 1377)

This is in contrast to the widely held belief that deposit-insurance moral hazard underlay the problems.

Randall (1989) concluded that all but four of the 40 BHC problems he examined were related to the poor credit quality of the loans the bank was making rather than its taking excessive risks that failed to pay off. Both equity holders and bondholders would be equally concerned about poor loan quality. Studies such as Pettway and Sinkey (1980) and Berger et al. (2000) that find equity returns contain information useful in predicting subsequent bank problems are consistent with this result.

It is thus arguable, on theoretical grounds, that agency problems in the form of poor managerial decision-making are a potentially important alternative cause of banks getting into trouble, that moral hazard is not the only problem, and that equity-holder incentives as not clear cut as is frequently assumed. The scant empirical evidence we have on the relative importance of "bad luck" versus "bad investments" suggests that the moral hazard risk-taking problem is usually only operative once banks have become materially impaired. Unfortunately, common examples of poor managerial decision-making including fraud, self-dealing, lack of internal controls, inadequate/incompetent credit screening, and overpaying for acquisitions are not rare. In these cases equity holders have as strong incentives to monitor and influence managers to avoid excessive risk-taking and poor investment choices.

[1]The models most frequently used in the discussion of moral hazard derive from an option pricing framework wherein it is assumed that all investments have the same (risk neutral) expected return. These models do not admit the possibility of "bad investments."

[2]Kaufman (1992) has criticized these applications of option pricing models to deposit insurance, noting that the option holder (the equity holder) does not control the timing of the option "exercise." Rather the effective option writer—the FDIC—controls the timing and manner of bank closures. Kaufman also questions the static nature of asset risk inherent in the Black–Scholes and Merton models (*vide infra* the discussion of Ritchken et al., 1993).

[3]Gorton and Rosen do not deny the plausibility of moral hazard arguments for banks with low levels of capital; however, they suggest that it is unclear that moral hazard arguments can explain how banks got into low-capital positions in the first place.

NOTES

[1]Subordinated debt proposals would require banks to issue bonds that are subordinate to all other claims, excepting only equity and preferred stock. These bonds would not be covered by guarantees explicit or conjectural.

[2]This taxonomy differs from the more usual one of "direct market discipline" and "indirect market discipline" (see, for example, Kwast et al., 1999; and Covitz, Hancock, and Kwast, 2000). The usual terminology essentially defines market discipline in terms of incentives, while the definition proposed here requires that incentives translate into desired managerial actions.

[3]Regulatory discipline triggered by market signals relies on market monitoring and not market influence and hence is not market discipline per se.

[4]No subordinated debt proposal suggests actually *replacing* regulatory supervision with market mechanisms. Other proposals to do away with both regulatory oversight and the deposit insurance safety net have been long discussed, though they have made little headway.

[5]See the recent survey by Flannery (1998) and earlier papers by Gilbert (1990) and Berger (1991).

[6]See Kane (1990), Eisenbeis and Horvitz (1994), and Kaufman (1995) for detailed discussions.

[7]In both cases accounting standards were redefined in violation of generally accepted accounting practices (GAAP) to allow institutions that were plainly insolvent to report positive net worth and remain open.

[8]It is important to note that no mandatory subordinated debt advocate is suggesting that examiners cannot or should not act on alternative information sources when bond yields fail to signal a problem.

[9]Subordinated debt proposals are aimed in part at overcoming the too-big-to-fail distortions by designating a class of securities that will absolutely, positively not be bailed out in the event of problems (caveats to this rule somewhat undermine its intent). However, subordinated debt proposals generally do not discuss the need for resolving transparency issues as a necessary precondition for the proposals to work.

[10]This is not strictly true, *vide infra* the discussion on interpreting bond yield spreads.

[11]Shleifer and Vishny (1997) and Short (1994) provide useful introductions to this literature.

[12]It was not always thus. Adam Smith continues the previously quoted discussion saying "They [joint-stock companies] have, accordingly, very seldom succeeded without an exclusive privilege [legal monopoly]; and frequently have not succeeded with one." One might hypothesize from this contrast between then and now that well-disciplined markets are not a natural state of economic nature, spoilt only by the intrusiveness of modern regulation.

[13]To talk of incorporeal firms acting or having incentives can be a convenient shorthand for managers' actions (in the name of the firm) or managerial incentives. Where these simplifying assumptions are innocuous, it matters little that they may be counterfactual and may be necessary to clarify a particular unrelated question. For instance, capital budgeting theory is concerned in part with what managers should do if they are acting in shareholders' interests. The assumption that managers are acting in the shareholders' interests leads to the desired prescriptive conclusions and is harmless in this context. However, if we design a policy predicated on the assumption that managers' and equity holders' incentives are aligned, when in fact they may not be, that policy is apt to be less efficacious than one that acknowledges the full complexity of interactions and incentives of the parties involved.

[14]Managers, on the other hand, can sometimes protect (immunize) themselves against involuntary replacement through golden parachutes and antitakeover amendments. Fama (1980) notes that adversarial resolution of manager/investor conflicts is very expensive. While golden parachutes apparently reward outgoing managers for failure, they may constitute the least costly means of removing managers who are willing to use the firm's (investors') own resources to contest their removal.

[15]This labor market discipline is an important source of managerial discipline; indeed Fama (1980) argues that "[t]he viability of the large corporation with diffuse security ownership is better explained in terms of a model where the primary disciplining device comes through managerial labor markets, both within and outside the firm, ..."

[16]Bliss and Flannery (2000) and Calomiris and Powell (2000) are two exceptions. Bliss and Flannery find little clear evidence of equity or bond market influence on U.S. bank holding companies, while Calomiris and Powell purport to find evidence consistent with bondholder influence in Argentina under very different circumstances.

[17]See the appendix discussing moral hazard and theoretical reasons why equity and debt prices may or may not react differently to changes in total asset risk. Other financial theories suggest that equity prices respond to changes in systematic risk, while bond prices respond to changes in default risk, which is related to total asset risk.

[18]Myers and Majluf (1984) argue that managers will underinvest even in profitable risky projects that require outside financing.

[19]Park (2000) develops a formal model in which senior-debt holders monitor firms for moral hazard problems and junior-debt holders free-ride. This is consistent with observed debt priority, ownership, and maturity structure in nonbanks. The reasons that junior-debt holders do not monitor (benefits accrue first to senior claimants) carry over to subordinated bank debt. It is less clear that Park's arguments for why senior-debt holders do monitor (gain full benefit of their monitoring efforts) would carry over to bank supervisors, who are agents rather than principals with their own funds at risk. On the other hand, Park argues that senior-debt holders will tend to have lower monitoring costs, an observation that carries over to supervisors who are *paid* to monitor.

[20]Most studies use multiple measures of risk and impose linearity on the risk–yield (–spread) relation and then test for significantly positive coefficients. A few studies, such as the first part of Morgan and Stiroh (2000), use methodologies that do not impose linearity or monotonicity on the risk–yield relation.

[21]Since most of these measures lag when problems develop—poor lending procedures take time to show up as nonperforming loans—these measures provide a conservative estimate of when problems actually began.

[22]Pettway and Sinkey (1980), *vide infra*, find evidence that equity returns do anticipate problems. They do not test bond returns.

[23]This important point has been generally ignored in studies of the informativeness of yield spreads.

[24]A notable exception is the Kwast et al. (1999) report, which notes several of the non-credit-risk priced factors discussed above, and advises that subordinated debt yields be interpreted with caution.

[25]Equivalently the subordinated debt literature has assumed that the cost–benefit tradeoffs under a mandated subordinated debt proposal would be superior to the cost–benefit tradeoffs under the current regulatory approach. This certainly is a key question, on which we have little concrete evidence.

[26]The "costs of subordinated debt requirements and related supervisory actions" question has, to my knowledge, not been investigated in detail in any of the subordinated debt proposals. Kwast et al. (1999) and Kaufman et al. (2000) address the costs of misclassification issue only in passing and present no quantitative estimates.

[27]This can be shown with simple simulations (available from the author on request).

[28]Black and Cox (1976) show that for banks approaching or actually in insolvency, subordinated debt holders, like equity holders, ceteris paribus prefer more risk to less.

[29]Curiously, I could find no studies that employed bond yields as inputs to failure prediction models for either banks or nonfinancial corporations.

[30]Levonian (2000) also makes this point.

[31]It is worth noting that a large literature on initial public offerings of equities suggests that equities are offered at prices substantially removed from their post-offering equilibrium levels. (My thanks to S. Ianotti for this observation.) Little evidence exists as to whether similar factors are or are not present in fixed-income primary markets, though liquidity-related factors such as on-the-run premia in Treasury bonds and seasoned versus newly issued pricing differentials in corporate bonds are known to exist.

[32]This analysis can say nothing about the *relative* transparency discussed in Kwast et al. (1999).

[33]For comparison with the Gramm–Leach–Bliley Act, I broadly define A-rated bonds to include A– (A3). Other authors suggest "investment grade" as the threshold. For this purpose, I include BBB– (Baa3) in the range of qualifying ratings.

[34]These numbers probably understate the frequency of low- and sub-investment-grade bond issuance. The Morgan–Stiroh database contains only rated bonds. Unrated bonds are frequently issued and these tend to be from more risky companies, though this is not invariably the case.

[35]Where ratings agencies disagree, it may be presumed that the higher rating will suffice for meeting the Gramm–Leach–Bliley requirements (16 percent of newly issued financial institution bonds failed on both Moody's and S&P ratings). Furthermore, the act requires that only one adequately rated bond be outstanding, which may be a senior or seasoned issue.

[36]Using Moody's ratings does not alter the results materially.

[37]Alternatives to ratings are apt to be complex. Studies discussed earlier of the risk-sensitivity of bond yields use numerous accounting variables to collectively proxy for risk. It is conceptually possible to construct an econometric model that produces a single measure of risk. Though how to calibrate such a model without having another measure of risk as a "true value" is somewhat problematical. Examiner ratings are not publicly available and are apt to be stale.

[38]Standard errors of the fitted bond yield spreads are rarely reported and would apply in any case to the aggregate across all ratings.

[39]There are some concerns that forcing information revelation, for example, risk exposures, will cause banks to modify their models, thus degrading the quality of the information they produce.

[40]See, for example, Merton (1978) and Ritchken et al. (1993).

REFERENCES

Altman, Edward I., And Anthony Saunders, 1998, "Credit risk measurement: Developments over the last 20 years," *Journal of Banking and Finance*, Vol. 21, No. 11&12, December, pp. 1721–1742.

Basel Committee on Banking Supervision, 1999, "A new capital adequacy framework," consultative paper, Basel, Switzerland, June.

Berger, Allen N., 1991, "Market discipline in banking," *Proceedings of the Conference on Bank Structure and Competition*, Federal Reserve Bank of Chicago, pp. 419–439.

Berger, Allen N., Sally M. Davies, and Mark J. Flannery, 2000, "Comparing market and regulatory assessments of bank performance: Who knows what when?," *Journal of Money, Credit, and Banking*, Vol. 24, No. 7, August, Part 2, pp. 641–667.

Black, Fisher, and John C. Cox, 1976, "Valuing corporate securities: Some effects of bond indenture provisions," *Journal of Finance*, Vol. 31, No. 2, May, pp. 351–367.

Black, Fisher, and Myron Scholes, 1972, "The valuation of option contracts and a test of market efficiency," *Journal of Finance*, Vol. 27, No. 2, May, pp. 399–417.

Bliss, Robert R., and Mark J. Flannery, 2000, "Market discipline in the governance of U.S. bank holding companies: Monitoring vs. influencing," Federal Reserve Bank of Chicago, working paper, No. WP-2000-03.

Calomiris, Charles W., 1999, "Building an incentive compatible safety net," *Journal of Banking and Finance*, Vol. 23, No. 10, October, pp. 1499–1519.

Calomiris, Charles W., and Andrew Powell, 2000, "Can emerging market bank regulation establish credible discipline: The case of Argentina 1992–1999," Columbia University, working paper.

Cannella, Albert A., Donald R. Fraser, and D. Scott Lee, 1995, "Firm failure and managerial labor markets: Evidence from Texas banking," *Journal of Financial Economics*, Vol. 38, No. 2, June, pp. 185–210.

Cornell, Bradford, 1992, "Liquidity and the pricing of low-grade bonds," *Financial Analysts Journal*, January-February, pp. 63–67.

Crabbe, Leland E. and Christopher M. Turner, 1995, "Does the liquidity of a debt issue increase with its size? Evidence from the corporate bond and medium-term note markets," *Journal of Finance*, Vol. 50, No. 5, December, pp. 1719–1734.

Covitz, Daniel M., Diana Hancock, Myron L. Kwast, 2000, "Mandatory subordinated debt: Would banks face more market discipline?," Board of Governors of the Federal Reserve System, working paper, June.

DeYoung, Robert, Mark J. Flannery, William W. Lang, and Sorin M. Sorescu, 2001, "The informational content of bank exam ratings and subordinated debt prices," *Journal of Money, Credit, and Banking*, forthcoming.

DeYoung, Robert, Kenneth Spong, and Richard J. Sullivan, 2001, "Who is minding the store? Motivating and monitoring hired managers at small, closely held firms' commercial banks," *Journal of Banking and Finance*, forthcoming.

Duffee, Gregory R., 1999, "Estimating the price of default risk," *Review of Financial Studies*, Vol. 12, No. 1, Spring, pp. 197–226.

_____, 1998, "The relation between Treasury yields and corporate bond yield spreads," *Journal of Finance*, Vol. 53, No. 6, December, pp. 2225–2242.

Eisenbeis, Robert A., and Paul M. Horvitz, 1994, "The role of forbearance and its costs in handling troubled and failed institutions," in *Reforming Financial Institutions and Markets in the United States,* George G. Kaufman (ed.), Boston: Kluwer Academic Publishers, pp. 49–68.

Evanoff, Douglas D., and Larry D. Wall, 2000, "Subordinated debt as bank capital: A proposal for regulatory reform," *Economic Perspectives*, Federal Reserve Bank of Chicago, Vol. 24, No. 2, Second Quarter, pp. 40–53.

Fama, Eugene F., 1986, "Term premiums and default premiums in money markets," *Journal of Financial Economics*, Vol. 17, No. 1, September, pp. 175–196.

_____, 1980, "Agency problems and the theory of the firm," *Journal of Political Economy*, Vol. 88, No. 2, April, pp. 288–307.

Fama, Eugene F., and Michael C. Jensen, 1983, "Separation of ownership and control," *Journal of Law and Economics,* Vol. 26, June, pp. 301–325.

Farrell, Kathleen A., and David A. Whidbee, 2000, "The consequences of forced CEO succession for outside directors," *Journal of Business*, Vol. 73, No. 4, October, pp. 597–527.

Flannery, Mark J., 1998, "Using market information in prudential bank supervision: A review of the U.S. empirical evidence," *Journal of Money, Credit, and Banking*, Vol. 30, No. 3, August, Part I, pp. 273–305.

Flannery, Mark J., and Sorin M. Sorescu, 1996, "Evidence of bank market discipline in subordinated debenture yields" *Journal of Finance*, Vol. 51, No. 4, September, pp. 1347–1377.

Gilbert, R. Alton, 1990, "Market discipline of bank risk: Theory and evidence," *Review*, Federal Reserve Bank of St. Louis, Vol. 72, No. 1, January-February, pp. 3–18.

Gorton, Gary, and Richard Rosen, 1995, "Corporate control, portfolio choice and the decline of banking," *Journal of Finance*, Vol. 50, No. 5, pp. 1377–1420.

Greenspan, Alan, 2000, "Banking evolution," remarks presented at the 36th annual Conference on Bank Structure and Competition, Federal Reserve Bank of Chicago, May 4.

Jaffee, D. M., 1975, "Cyclical variations in the risk structure of interest rates," *Journal of Monetary Economics*, Vol. 3, No. 1, July, pp. 309–325.

Jagtiani, Julapa, George Kaufman, and Catharine Lemieux, 2000, "Do markets discipline banks and bank holding companies? Evidence from debt pricing," Federal Reserve Bank of Chicago, working paper.

Jensen, Michael C., and William H. Meckling, 1976, "Theory of the firm: Managerial behavior, agency costs, and ownership structure," *Journal of Financial Economics*, Vol. 3, No. 4, October, pp. 305–360.

Kane, Edward J., 1990, "Defective regulatory incentives and the Bush initiative" in *Restructuring the American Financial System,* George G. Kaufman (ed.), Boston: Kluwer Academic Publishers, pp. 117–127.

Kaufman, George G., 1995, "The banking debacle of the 1980s: An overview and lessons," *The Financier*, Vol. 2, No. 2, May, pp. 9–26.

_____, 1992, "Some shortcomings of analyzing deposit insurance as a put option: A communication," *Journal of Financial Engineering*, Vol. 1, No. 3, December, pp. 376–383.

Kaufman, George G., Robert E. Litan, Richard C. Aspinwall, George J. Benston, Charles W. Calomiris, Franklin R. Edwards, Scott E. Harrington, Richard J. Herring, Paul M. Horvitz, Roberta Romano, Hal S. Scott, Kenneth E. Scott, and Peter J. Wallison, 2000, *Reforming Bank Capital Standards: A Proposal by the U.S. Shadow Regulatory Committee*, Washington, DC: American Enterprise Institute for Public Policy Research, Studies on Financial Market Deregulation.

Kwast, Myron L., Daniel M. Covitz, Diana Hancock, James V. Houpt, David P. Adkins, Norah Barger, Barbara Bouchard, John F. Connolly, Thomas F. Brady, William B. English, Douglas D. Evanoff, and Larry D. Wall, 1999, "Using subordinated debt as an instrument of market discipline," report of a study group on subordinated notes and debentures, Board of Governors of the Federal Reserve System, staff study, No. 172, December.

Leland, Hayne E., 1998, "Agency costs, risk management, and capital structure," *Journal of Finance*, Vol. 53, No. 4, August, pp. 1213–1244.

Levonian, Mark, 2000, "Subordinated debt and the quality of market discipline in banking," Federal Reserve Bank of San Francisco, working paper, September.

Merton, Robert C., 1978, "On the cost of deposit insurance when there are surveillance costs," *Journal of Business*, Vol. 51, July, pp. 439–452.

_____, 1977, "An analytic derivation of the cost of deposit insurance and loan guarantees: An application of modern option pricing theory," *Journal of Banking and Finance*, Vol. 1, June, pp. 3–11.

Meyer, Laurence H., 1999, "Market discipline as a complement to bank supervision and regulation," remarks before the Conference on Reforming Bank Capital Standards, Ernst & Young and AEI-Brookings Joint Center on Regulatory Studies, Council on Foreign Relations, New York, June 14.

Minton, B., 1997, "An empirical examination of basic valuation models for plain vanilla U.S. interest rate swaps," *Journal of Financial Economics*, Vol. 44, No. 2, May, pp. 251–277.

Morgan, Donald P., and Kevin J. Stiroh, 2000, "Bond market discipline of banks: Is the market tough enough?," Federal Reserve Bank of New York, staff report, No. 95.

Myers, Stewart C., and Nicholas S. Majluf, 1984, "Corporate financing and investment decisions when firms have information that investors do not have," *Journal of Financial Economics*, Vol. 13, No. 2, pp. 184–221.

Nickell, Stephen, Daphne Nicolitsas, and Neil Dryden, 1997, "What makes firms perform well?," *European Economic Review*, Vol. 41, No. 3-5, April, pp. 781–796.

Park, Cheol, 2000, "Monitoring and structure of debt contracts," *Journal of Finance*, Vol. 55, No. 5, October, pp. 2157–2195.

Pettway, Richard H., and Joseph F. Sinkey, Jr., 1980, "Establishing on-site bank examination priorities: An early-warning system using accounting and market information," *Journal of Finance*, Vol. 35, No. 1, March, pp. 137–150.

Prowse, Stephen, 1997, "Corporate control in commercial banks," *Journal of Financial Research*, Vol. 20, No. 4, Winter, pp. 509–527.

Randall, Richard E., 1989, "Can markets evaluate asset quality exposure in banks?," *New England Economic Review*, July/August, pp. 3–24.

Reed, John, 1999, address to the Office of the Comptroller of the Currency conference on Measuring Financial Risk in the Twenty-first Century, Washington, DC, October 14.

Ritchken, Peter, James B. Thomson, Ramon P. DeGennaro, and Anlong Li, 1993, "On flexibility, capital structure and investment decisions for the insured bank," *Journal of Banking and Finance*, Vol. 17, No. 6, December, pp. 1133–1146.

Shadow Financial Regulatory Committee, 1992, "Statement No. 89: Standards for Safety and Soundness (Implementation of Section 132 of FDICIA)," *Journal of Financial Services Research,* Vol. 10, No. 4, Supplement, December, p. S-44.

Shleifer, Andrei, and Robert W. Vishny, 1997, "A survey of agency costs," *Journal of Finance*, Vol. 52, No. 2, June, pp. 737–783.

Short, Helen, 1994, "Ownership, control, financial structure and the performance of firms," *Journal of Economic Surveys*, Vol. 8, No. 3, pp. 203–249.

Sloane, P. E., 1963, "Determinants of bond yield differentials," *Yale Economic Essays*, Vol. 3, No. 1, pp. 3–55.

Smith, Adam, 1796[1776], *An Inquiry into the Nature and Causes of the Wealth of Nations*, eighth edition, London: Strahan, Cadell, and Davies, the Strand, now in the public domain.

Smith, Clifford W., Jr., and Jerald B. Warner, 1979, "On financial contracting: An analysis of bond covenants," *Journal of Financial Economics*, Vol. 7, No. 2, June, pp. 117–162.

VanHorne, James C., 1979, "Behavior of default risk premiums for corporate bonds and commercial paper," *Journal of Business Research*, pp. 301–313.

Evidence of the North–South business cycle

Michael A. Kouparitsas

Introduction and summary

This article examines the economic fluctuations of two regional economies: the developed, industrial goods exporting countries of the world (which I call "North") and the developing, non-fuel commodity exporting countries (or "South"). My principal objective is to document the salient features of their business cycles from 1970 to 1995. I frame the discussion around two questions. The first question is whether these very different regions share similar business cycle characteristics. The second question is whether there is an international business cycle, in other words, whether cyclical fluctuations in one region are positively correlated with cyclical fluctuations in the other.

This study is a natural extension of the modern approach to measuring business cycles initiated by Burns and Mitchell's (1946) study of U.S. industrial data. Just as in this earlier work, the goal of this article is to summarize properties of the data without imposing much theoretical structure. The hope is that the resulting regularities of North–South data can then serve as a guide for future theoretical and empirical studies of international business cycles.

While the study of short-run interaction between the North and South is relatively new, the study of long-run interaction between the North and South is as old as international trade theory, with roots dating back to Ricardo's theoretical analysis of the British Corn Laws.[1] The conventional view of the North and South that emerged from this long-run analysis separates the world economy into the developed North, which is a net exporter of manufactures, and the developing South, which is a net exporter of primary non-fuel commodities (that is, agricultural and mining products). This view incorporates two criteria for classifying countries as North and South, namely their level of development and the composition of their export and import bundles, but these may conflict. For example, Australia is an industrial country that principally exports primary commodities in exchange for manufactured imports. Another challenge posed by this definition is that countries may move from one category to the other over time. There are many economies, including newly industrialized Asian countries such as Taiwan and Korea, that have changed from predominantly agricultural to highly industrialized. For the purposes of this article, I adopt a more general definition of the North by expanding it to include countries that were classified by the International Monetary Fund (IMF) as developed countries over the data sample used in this article, which runs from 1970 to 1995. This definition includes primary commodity exporting industrial countries, such as Australia, while it excludes the newly industrialized countries of Asia.[2] The South, in contrast, is limited to countries that were net exporters of primary non-fuel commodities and classified as developing by the IMF from 1970 to 1995, a definition that matches the traditional view of the South. A country is considered to be a net exporter of primary non-fuels if its average export share exceeded its average import share of primary non-fuels over the data sample.

Using these definitions, I have assembled data for 22 northern countries and 46 southern countries, which include 25 years of annual real and nominal sectoral output, consumption, investment, export, and import data.[3] I combine these individual country data to form regional aggregates for the North and South; this greatly simplifies the discussion. It is also necessary because the focus of this article is on the behavior of regional rather than individual country business cycles.

I begin my discussion of the aggregate regional data by comparing the industrial and trade characteristics of the North and South. In contrast to stylized models typically used to study long-run growth linkages

Michael A. Kouparitsas is an economist at the Federal Reserve Bank of Chicago. *The author would like to thank seminar participants at the Federal Reserve Bank for useful comments on an earlier draft.*

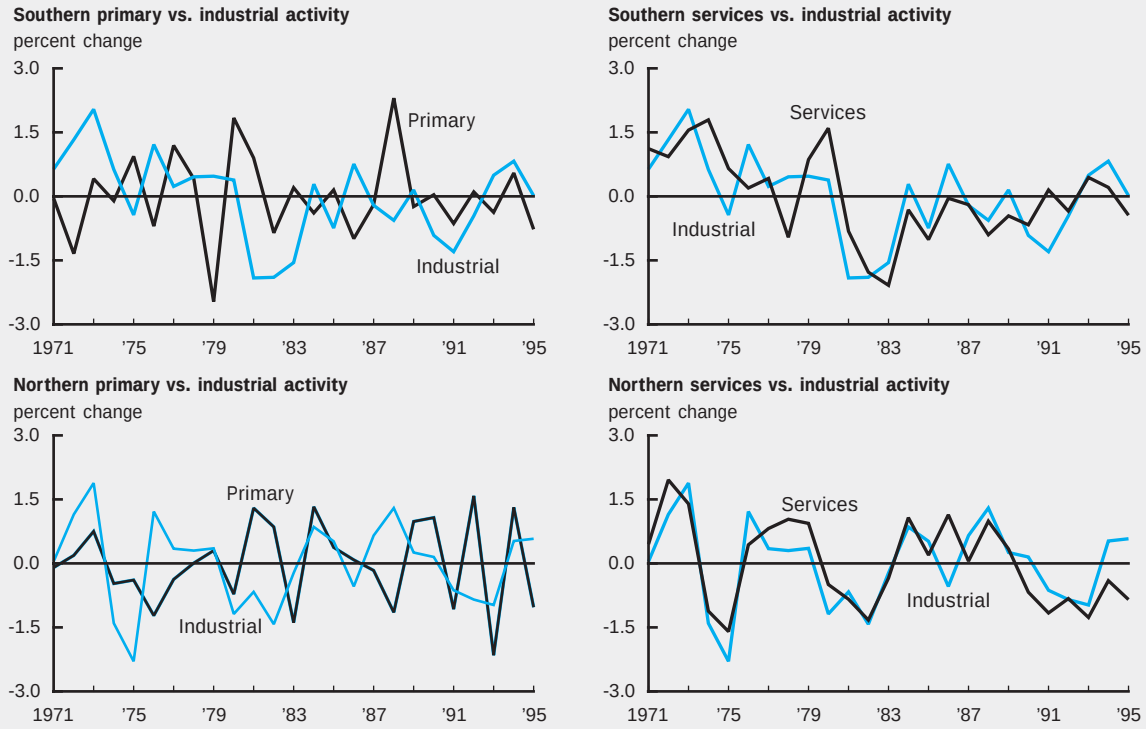
between the North and South, I find that the South is not completely specialized in the production and exportation of primary commodities. In fact, I find that the South devotes a significant share of its production and export trade to manufactures and services. On the other hand, I find that the North fits closely with the stylized view of North–South trade models by devoting a small share of its production and trade to primary commodities and a significant share of production and trade to manufactures and services.

I break up the description of the cyclical features of the data by focusing on four areas: intraregional activity; interregional activity; the behavior of relative prices and quantities; and the behavior of trade flows and quantities. The intraregional data reveal many similarities in the cyclical fluctuations of the North and South. For example, one of the highlights of the regional dataset used in this article is its relatively rich sectoral output data covering primary, industrial, and service sector activity. These data allow me to explore whether there are regional business cycles in the North and South. Building on Burns and Mitchell's (1946) definition of a national business cycle, a regional business cycle comes about if upswings or downturns in one regionwide production sector are matched by upswings or downturns in other regionwide sectors. Lucas's (1977) summary of Burns and Mitchell's work points out that cyclical fluctuations in the growth rates of U.S. non-primary production sectors are closely related to each other, while cyclical fluctuations in the growth rates of U.S. primary activity have a low coherence with the rest of the U.S. economy. This suggests that the U.S.'s national business cycle is limited to non-primary activities. I find that this result extends to the broader northern and southern economies, which indicates there are regional business cycles in non-primary activities. More specifically, I find that both regional economies display a high positive correlation between fluctuations in growth rates of industrial and service sector activity, while registering a low correlation (in many cases a negative one) between fluctuations in the growth rates of primary activity and non-primary activity (see figure 1).

The next feature of regional economies that I focus on is the persistence of their economic fluctuations.

FIGURE 1

Intraregional business cycles



Notes: For graphical purposes, the growth rate series have been transformed so that they have a zero mean and unit standard deviation. Standard deviations of the underlying growth rate series are listed in table 9.

Sources: Author's calculations based on data from World Bank, *World Development Indicators* (1998) and OECD, *International Sectoral Database* (1998).

Policymakers are particularly interested in the persistence of economic fluctuations, since more persistent fluctuations imply greater economic hardship during downturns and longer expansions. I find that fluctuations in the growth rate of aggregate output tend to be more persistent in the South. One possible interpretation of this result in that policy (monetary and fiscal) responds with a much longer lag in the South.

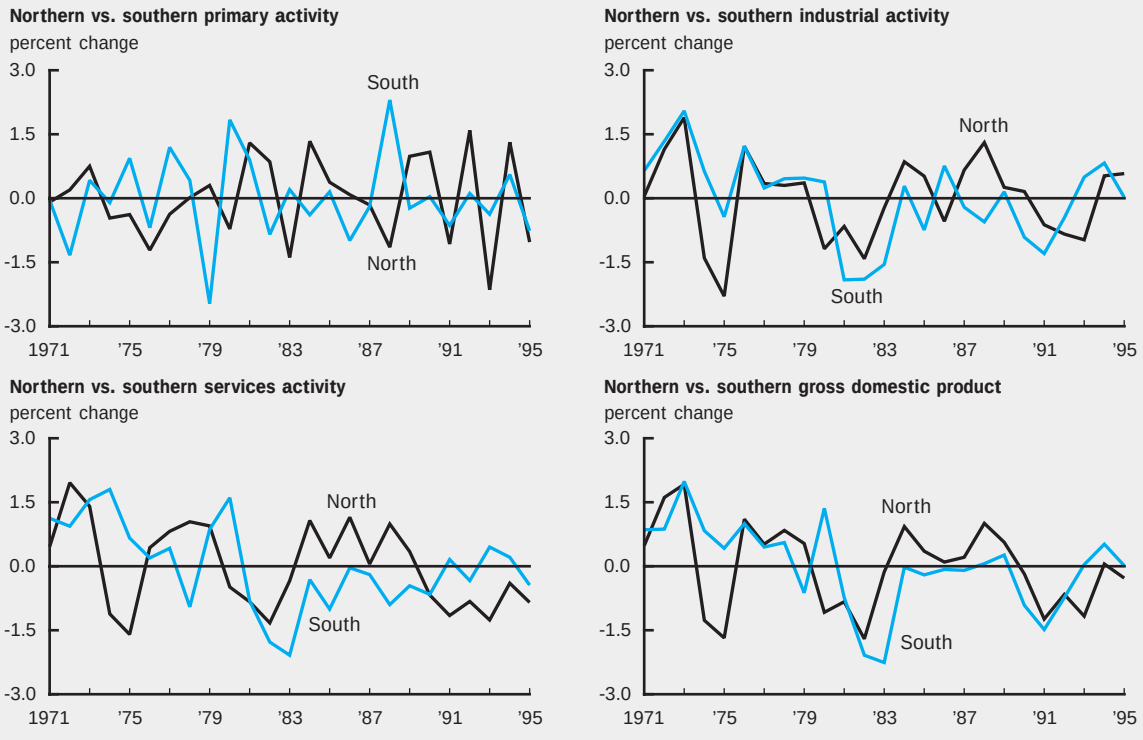
Many studies have documented the positive correlation of the growth rates of gross domestic product (GDP) of the major industrialized countries of North America and Europe.[4] Extending Burns and Mitchell's definition of a business cycle along the international dimension, these data suggest that there is an international business cycle. First, the data show that when one major industrial country is in an upswing or downturn, the other major industrial countries tend to be in an upswing or downturn. This could be referred to as the North–North business cycle. In addition, I find that the international business cycle extends to the South by showing that there is a strong positive correlation of the growth rates of northern and southern GDP (see figure 2). Using the sectoral data, I show that the North–South business cycle is supported by a strong positive correlation of the growth rates of northern and southern industrial activity and a somewhat weaker correlation of the growth rates of northern and southern services.

There are two basic explanations for why these regional economies move together. One view is that business cycles simply reflect the fact that regions are influenced by the same source of innovations and that they respond to these innovations in the same way. The other is that business cycles are transmitted from one region to another via interregional trade. I lay the seeds for future empirical and theoretical analysis of this issue by examining the cyclical relationship between interregional relative prices, interregional trade flows, and regional expenditure and production.

The data suggest that the North–South business cycle is the byproduct of a strong North–South business cycle transmission mechanism. There are two important transmission channels. First, the relative price, trade, and production data suggest that fluctuations in the South's terms of trade (ratio of southern export to import prices) are caused by fluctuations in the North's demand for southern exports. Second, the

FIGURE 2

Interregional business cycles



Notes and sources: See figure 1.

relative price, trade, and expenditure data suggest that fluctuations in southern consumption and investment are largely driven by fluctuations in the South's terms of trade. Overall, these data suggest that fluctuations that originate in the North are transmitted to the South through interregional trade.

TABLE 1

Industry value-added as share of total value-added

	North			South		
	1970	**1980**	**1995**	**1970**	**1980**	**1995**
Primary	5	4	2	25	19	15
Industrial	44	42	35	29	35	33
Services	51	54	63	45	46	52
Total	**100**	**100**	**100**	**100**	**100**	**100**

Source: Author's calculations based on data from World Bank, *World Development Indicators* (1998) and OECD, *International Sectoral Database* (1998).

Data and methodology

The developing country time series data are from the World Bank's *World Development Indicators* database (WDI) (1998). This database contains real and nominal annual data on value-added by sector, consumption, investment, government spending, exports, and imports for 46 southern countries as defined earlier, covering the period from 1970 to 1995. (Other countries in the WDI fall outside the definition of a southern country adopted here.) Time-series data for the northern countries come from the Organization for Economic Cooperation and Development's (OECD) *International Sectoral Database* (ISD) (1998). The ISD contains real and nominal annual data on value-added by sector, consumption, investment, government spending, exports, and imports for 22 northern countries, covering the period from 1970 to 1995.

Disaggregated merchandise trade data are from the World Bank world tables (1991). This database contains annual merchandise trade flows for 100 countries from 1969 to 1988 (the data were not collected after 1988).[5] Data on interregional trade flows come from various issues of the United Nations' *International Trade Statistics Yearbook*, special table B.

Following in the tradition of modern business cycle analysis, I describe the cyclical characteristics of the North and South through various second-moment statistics (standard deviations and correlation coefficients) of a stationary component of the time-series data. Following the literature, I focus on the so-called business cycle components of the data, which are captured in annual data by simple growth rates. All statistics reported in this article refer to regional aggregates for the North or South. The regional aggregates are the sum of the individual North or South real and nominal country data, converted to a common currency, U.S. dollars.

Stylized features on North–South industry, expenditure, and trade

Table 1 describes the structure of industry in the North and South from 1970 to 1995. The stylized view of the South is of a small economy that specializes in the production and export of primary goods, while the stylized view of the North is of a large economy that specializes in the production and export of industrial goods. Table 1 reveals that while the South devoted considerably more activity than the North to primary production from 1970 to 1995, it was not specialized in the production of primary goods over this period. In fact, the South allocated roughly 25 percent of its activity to primary production in 1970, which steadily declined to 15 percent by 1995. The North, on the other hand, devoted 5 percent or less of its activity to primary production from 1970 to 1995. As table 1 shows, the share of activity devoted to industrial production was roughly similar in the North and South in 1995, which stands in stark contrast to the stylized view of the North and South. This reflects the fact that the share of activity allocated to industrial production has been steadily declining in the North and generally rising in the South since 1970. The final row of table 1 reveals that a growing share of activity in the North and South has been devoted to services since 1970. The South shifted from a service share of 45 percent in 1970 to 52 percent in 1995, while the North shifted from a service share of 51 percent in 1970 to 63 percent in 1995.

Table 2 divides activity by expenditure category. With the exception of trade, the North's expenditure shares were largely constant from 1970 to 1995. The South, in contrast, has experienced a steady decline in its consumption share, which has largely been offset by a steady increase in its investment share. Table 2 also reveals that the North and South experienced a significant increase in the share of activity they devoted to trade from 1970 to 1995. In 1970, the share of activity devoted to trade was roughly similar in the North and South at around 13 percent for exports and 13 percent for imports in both regions. By 1995, trade activity had grown considerably to 19 percent for exports and 19 percent for imports in the North and 23 percent for exports and 25 percent for imports in the South.

TABLE 2

Expenditure as share of total value-added

	North			South		
	1970	**1980**	**1995**	**1970**	**1980**	**1995**
Consumption	60	60	63	72	67	65
Investment	22	23	21	20	24	25
Government	17	18	16	10	11	12
Exports	13	20	19	11	17	23
Imports	13	20	19	13	19	25
Total	**100**	**100**	**100**	**100**	**100**	**100**

Note: Government indicates government spending.
Source: Author's calculations based on data from World Bank, *World Development Indicators* (1998) and OECD, *International Sectoral Database* (1998).

Table 3 takes a closer look at trade by breaking up the export and import bundles into three categories, primary non-fuels, primary fuels, and industrial goods. A typical stylized model of North–South trade assumes that the North is specialized in the export of industrial goods, while the South is specialized in the export of primary goods. The upper panel of table 3 shows that this is consistent with recent data. For example, with 75 percent of its exports devoted to industrial goods, the North was roughly specialized in the export of industrial goods from 1969 to 1988. Similarly, the other half of the upper panel reveals that with 72 percent of exports devoted to non-fuel and fuel commodities, the South was roughly specialized in the export of primary goods from 1969 to 1988. The middle panel of table 3 reveals that the North and South have similar import baskets. The largest share is devoted to industrial goods, 58 percent in the North and 66 percent in the South, while primary non-fuels and fuels are in the range of 16 percent to 24 percent of the regions' import baskets. The similarity of the North and South import baskets reflects the fact that a significant share of the North's trade is intraregional. We can see this in table 4, which describes in detail intraregional and interregional trade flows between the North and South from 1970 to 1995. The upper panel describes the flow of exports by destination, while the lower panel describes the flow of imports by the source. For example, reading across the first row of the upper panel we see that 85 percent of all exports that originate in the North are actually shipped to other countries in the North. Continuing across this row we see that 15 percent of all exports from the North are destined for the South. The next row indicates that 81 percent of all exports from the South are shipped to the North, while 19 percent of exports from the South are destined for countries within the South. The columns of the lower panel indicate the source of imports to the North and South. The column on the right indicates that 87 percent of the North's imports come from other countries in the North, while 13 percent of its imports come from the South. The middle column reveals roughly similar sources for southern imports, 83 percent and 17 percent, respectively, from the North and South.

Another way to think about table 4 is that it tells us that the regional gross trade data of the South

TABLE 3

Composition of merchandise trade

	North	South
Export shares		
Primary non-fuels	19	67
Primary fuels	5	6
Industrial	75	28
Total	**100**	**100**
Import shares		
Primary non-fuels	24	17
Primary fuels	19	16
Industrial	58	66
Total	**100**	**100**
Net export shares		
Primary non-fuels	–4	49
Primary fuels	–13	–11
Industrial	18	–39

Source: Author's calculations based on data from World Bank, "World tables" (1991).

TABLE 4

Intra- and inter-regional trade flows

A. Trade by destination region
(distribution of exports)

From / to	**North**	**South**	**World**
North	85	15	**100**
South	81	19	**100**
World	80	20	**100**

B. Trade by source region
(distribution of imports)

From / to	**North**	**South**	**World**
North	87	83	86
South	13	17	14
World	**100**	**100**	**100**

Source: Author's calculations based on data from United Nations, *International Trade Statistics Yearbook* (various issues).

largely reflects trade between the North and South, while the regional gross trade data of the North largely reflects trade between northern countries or intra-North trade. In my subsequent discussion, I measure the trade flows between the North and South as the regional gross exports/imports of the South and the terms of trade between the North and South as the ratio of regional gross export prices to import prices of the South. This means that in the following discussion the exports and imports of the North are constrained to be the same as gross regional imports and exports of the South.

The lower panel data in table 4 also provide information about the relative size of the regional economies. In my discussion of table 2 I noted that the North and South have roughly similar gross import and export shares; therefore, the relative size of the regional economies is directly proportional to their share of world trade.[6] Combining the results from tables 2 and 4, the North accounts for roughly 86 percent of the world's output, while the South provides the remaining 14 percent. In other words, the North is roughly six times the size of the South, so the stylized view that the South is a small open economy fits with the data.

Is there a regional business cycle in the North and South?

One of the highlights of my regional dataset is its relatively rich sectoral output data covering primary, industrial, and service sector activity. These data allow me to explore whether there are regional business cycles in the North and South. My definition of a regional business cycle is a logical extension of Burns and Mitchell's (1946) definition of a national business cycle. Burns and Mitchell argue that a nation has a business cycle if upswings or downturns in one national production sector are matched by upswings or downturns in other nationwide production sectors. With this in mind, I argue that a regional business cycle comes about if upswings or downturns in one regionwide production sector are matched by upswings or downturns in other regionwide sectors.

Table 5 describes the contemporaneous correlation coefficients of regionwide sectoral output growth rates from 1970 to 1995. The upper panel lists the correlation coefficients for the North and the lower panel for the South. Reading across the first row of the upper panel, we see that the correlation of northern GDP and primary output growth is 0.19, while the correlation of northern GDP and industrial and service sector output growth is considerably higher at 0.96 and 0.91, respectively (a coefficient of 1.0 would imply a perfect correlation). The remaining rows of the upper panel reveal the precise relationship between the sectors. The correlation of primary and service sector output growth is 0.15, while the correlation of industrial and service sector output growth rates is 0.77.[7] This suggests that there is a regional business cycle in non-primary activity in the North.

TABLE 5

Intraregional cyclical output correlations

	GDP	Primary	Industrial	Services
North				
GDP	1.00	0.19	0.96	0.91
Primary		1.00	0.10	0.15
Industrial			1.00	0.77
Services				1.00
South				
GDP	1.00	0.19	0.89	0.87
Primary		1.00	–0.16	–0.04
Industrial			1.00	0.71
Services				1.00

Source: Author's calculations based on data from World Bank, *World Development Indicators* (1998) and OECD, *International Sectoral Database* (1998).

The correlation statistics in the lower panel of table 5 display the same pattern as the upper panel. Just as in the northern case, the southern data reveal a low coherence between fluctuations of primary and non-primary activity and a high coherence between fluctuations of non-primary sectors, which suggests that there is also a regional business cycle in non-primary activity in the South.

Table 6 extends the analysis of regional business cycles to include expenditure aggregates. The upper panel of table 6 describes the correlation of aggregate and sectoral output growth rates (discussed above) and the growth rates of consumption, investment, government spending, exports, and imports in the North. The lower panel describes the same set of correlation statistics for the South. Reading across the first row of the upper panel, we find that the growth rates of consumption, investment, government spending, exports, and imports are all positively correlated with the growth rate of GDP in the North. Turning to the sectoral outputs, we see that the growth rates of industrial and service sector activity are also strongly correlated with the growth rates of the expenditure variables. Primary activity fluctuations, in contrast, display much lower correlations with the expenditure variables.

Moving onto the southern correlation statistics in the lower panel of table 6, there are a number of similarities to, and some significant differences from, the northern case. Just as in the North, the South shows

TABLE 6

Intraregional cyclical output and expenditure correlations

	Consumption	Investment	Government	Exports	Imports
North					
GDP	0.91	0.94	0.61	0.12	0.49
Primary	0.10	0.18	0.14	–0.04	–0.12
Industrial	0.77	0.92	0.64	0.22	0.64
Services	0.87	0.86	0.54	0.02	0.35
South					
GDP	0.72	0.75	0.39	0.19	0.48
Primary	0.18	0.34	–0.11	–0.19	0.16
Industrial	0.70	0.54	0.28	0.38	0.42
Services	0.69	0.65	0.21	0.02	0.51

Note: Government indicates government spending.
Source: Author's calculations based on data from World Bank, *World Development Indicators* (1998) and OECD, *International Sectoral Database* (1998).

a higher correlation between fluctuations in non-primary activity and fluctuations in the expenditure variables. However, the correlation coefficients tend to be smaller than their northern counterparts. The main difference between the regions is the South's considerably lower correlation of output and government spending growth rates. For example, the correlation of the growth rates of GDP and government spending in the North is 0.61, while it is only 0.39 in the South.

Table 7 completes the matrix of intraregional statistics by reporting on the cyclical relationship between the five regional expenditure variables. The North variables in the upper panel reveal a high level of coherence between fluctuations in northern consumption, investment, government spending, exports, and imports. This is not surprising given the number of multicountry business cycles studies, such as Backus and Kehoe (1992), that have shown that expenditure components tend to be highly correlated in individual northern countries. I also find, as did these earlier national studies, that exports display the lowest correlation statistics.

The South statistics in the lower panel of table 7 echo many of the findings of table 6: there appears to

TABLE 7

Intraregional cyclical expenditure correlations

	GDP	Consumption	Investment	Government	Exports	Imports
North						
GDP	1.00	0.91	0.94	0.61	0.12	0.49
Consumption		1.00	0.82	0.36	–0.13	0.34
Investment			1.00	0.50	0.20	0.59
Government				1.00	0.11	0.37
Exports					1.00	0.38
Imports						1.00
South						
GDP	1.00	0.72	0.75	0.39	0.19	0.48
Consumption		1.00	0.51	–0.10	0.23	0.61
Investment			1.00	0.12	0.04	0.79
Government				1.00	–0.29	–0.26
Exports					1.00	0.38
Imports						1.00

Note: Government indicates government spending.
Source: Author's calculations based on data from World Bank, *World Development Indicators* (1998) and OECD, *International Sectoral Database* (1998).

be high coherence between fluctuations of consumption, investment, and imports, albeit less than in the North, and relatively low coherence between fluctuations of exports and the other expenditure variables. I also find, as in the North, that fluctuations of government spending in the South are poorly correlated with fluctuations of the other expenditure variables.

How persistent are regional business cycles?

The next feature of regional economies that I focus on is the persistence of their economic fluctuations. Policymakers are particularly interested in this issue, since more persistent fluctuations imply greater economic hardship during downturns and longer expansions. Persistence is typically measured by the first-order autocorrelation of a variable. Table 8 reports autocorrelation statistics for the North and South in the first and second columns, respectively.

The first row of table 8 reveals that fluctuations in the growth rate of aggregate output tend to be more persistent in the South. In fact, the correlation between fluctuations in current and lagged southern activity is 0.51, which is considerably higher than the North's correlation of 0.33. The next three rows of table 8 reveal a similar pattern for the persistence of sectoral output fluctuations in the North and South. Fluctuations in the growth rate of primary sector output tend to die out very quickly, with an expansion one year followed by a downturn the next year. On the other hand, fluctuations in the growth rate of the service sector and, to a lesser extent, the industrial sector tend to die out more slowly over time, with the lowest correlation standing at 0.17 for current and lagged industrial activity in the North.

TABLE 8

Persistence

	North	South
Output		
GDP	0.33	0.51
Primary	–0.51	–0.35
Industrial	0.17	0.23
Services	0.46	0.31
Expenditure		
Consumption	0.41	0.10
Investment	0.39	0.37
Government	–0.37	0.13
Exports	0.38	0.22
Imports	0.22	0.38
Relative prices		
Terms of trade	–0.05	–0.05

Note: Government indicates government spending.
Source: Author's calculations based on data from World Bank, *World Development Indicators* (1998) and OECD, *International Sectoral Database* (1998).

How volatile are regional business cycles?

The final regional dimension along which I compare the North and South is the volatility of their production and expenditure activity. I do this by reporting percentage standard deviations of fluctuations in the regional growth rates of output and expenditure in the first two columns of table 9. The third column reports on the volatility of total North–South activity. I postpone my discussion of these world statistics until the next section. To simplify the comparison of North and South, I have divided the sector activity and disaggregated expenditure standard deviations by the standard deviation of regional and world GDP to form relative standard deviations. Values greater than one indicate the variable is more volatile than GDP, while values less than one indicate the variable is less volatile than GDP.

The first row of table 9 tells us that, with a standard deviation of 1.91 percent, aggregate northern activity is roughly 50 percent more volatile than aggregate southern activity, which has a standard deviation of 1.27. This seems counterintuitive given the high volatility of individual developing countries and the relatively low volatility of the major industrial countries. The outcome reflects the fact that fluctuations tend to be more highly correlated across developed countries. This comes about because the northern economies are more homogenous in production, in

TABLE 9

Cyclical volatility

	North	South	World
GDP	1.91	1.27	1.75
Relative volatility			
Output			
Primary	1.96	1.63	1.27
Industrial	1.53	1.82	1.57
Services	0.75	0.99	0.75
Expenditure			
Consumption	0.86	0.97	
Investment	2.37	4.18	
Government	1.65	2.22	
Exports	3.98	2.38	
Imports	2.10	4.50	
Relative prices			
Terms of trade	2.68	3.04	
Real commodity price			2.72

Note: Government indicates government spending.
Source: Author's calculations based on data from World Bank, *World Development Indicators* (1998) and OECD, *International Sectoral Database* (1998).

the sense that they devote similar shares to primary, industrial, and service activities. The southern economies tend to be more heterogeneous in production, with a greater deal of variation in the share of activity they allocate to primary and non-primary production.

Moving onto the relative volatility statistics for the North in the first column of table 9, we see that primary activity in the North is roughly twice as volatile as aggregate output. Industrial activity is also more volatile than aggregate output, but less volatile than primary activity. Services, in contrast, are less volatile than aggregate output, with a relative volatility statistic of 0.75. The second column of table 9 reveals a similar relationship for the South: service activity is less volatile than aggregate activity, while primary and industrial activity are both more volatile than aggregate activity. The only difference between the North and South is that the ranking of the relative volatility of primary and industrial activity is reversed.

Results on the relative volatility of expenditure variables at the national level are well known, from individual and multicountry studies of northern economies (see, for example, Backus and Kehoe, 1992). These studies show that consumption is less volatile than aggregate output, while investment, government spending, exports, and imports tend to be more volatile than aggregate output. The first two columns of table 9 show that these results extend to the regional economies of the North and South. With the exception of the trade flows, the ordering of the relative volatilities of the expenditure variables is identical for the two regions. In contrast, the data reveal that exports are more volatile than imports in the North, while exports are less volatile than imports in the South. In other words, exports from the North to the South are more volatile than exports from the South to the North.

Is there a North–South business cycle?

Many studies have documented the positive correlation of output fluctuations of the major industrialized countries. Extending Burns and Mitchell's (1946) definition of a business cycle along the international dimension, these data suggest that there is an international business cycle. In particular, the data for industrial countries show that when one major industrial country is in an upswing or downturn, the other major industrial countries tend to be in an upswing or downturn. This could be referred to as the North–North business cycle. Here, I examine whether the international business cycle has a North–South component.

Table 10 describes in detail the lead, lag, and contemporaneous correlation statistics of fluctuations of sectoral and aggregate economic activity in the North and South. The rows of the upper panel describe the cyclical relationship between a particular North output variable and the four South output variables at the same point in time. The columns describe the cyclical relationship between a particular South output variable and the four North output variables at the same point in time. For example, the first cell of the first column of the upper panel indicates that the contemporaneous correlation of the growth rates of northern and southern GDP is 0.45. This finding is similar to Backus, Kehoe, and Kydland's (1995) estimate of the correlation of fluctuations of output of the major developed regions, such as Europe and the U.S. This suggests that there is as much coherence between the North and South as there is between regions of the North, which implies that the international business cycle does indeed extend to the South.

The other elements of the first column of the upper panel reveal that the North–South business cycle is the byproduct of a strong correlation of the growth rates of aggregate southern activity and northern non-primary activity. Tracing across the

TABLE 10

Interregional cyclical output correlations

	South at time t			
North at time t	GDP	Primary	Industrial	Services
GDP	0.45	–0.09	0.59	0.08
Primary	–0.04	–0.08	–0.06	–0.11
Industrial	0.33	–0.08	0.45	–0.02
Services	0.43	–0.13	0.58	0.14
	South at time t+1			
	GDP	Primary	Industrial	Services
GDP	0.44	–0.03	0.40	0.45
Primary	–0.32	–0.26	–0.27	–0.12
Industrial	0.31	0.06	0.26	0.32
Services	0.46	0.01	0.37	0.41
	South at time t–1			
	GDP	Primary	Industrial	Services
GDP	0.10	–0.01	0.16	–0.13
Primary	0.02	0.19	–0.05	–0.02
Industrial	0.06	–0.03	0.15	–0.17
Services	0.09	–0.04	0.12	–0.17

Source: Author's calculations based on data from World Bank, *World Development Indicators* (1998) and OECD, *International Sectoral Database* (1998).

rows for northern non-primary activity, we see that these statistics are supported by a strong correlation of the growth rates of northern non-primary and southern industrial activity, with correlation coefficients of 0.45 for industrial activity and 0.58 for service sector activity. The other highlights of the panel are the low coherence of growth rates of northern and southern primary and service sector activity.

These observations are echoed in the world volatility statistics presented in table 9. The low coherence between northern and southern primary activity is revealed by the relatively low volatility of world primary output fluctuations, while the high coherence between northern and southern industrial activity is evident in the relatively high volatility of world industrial output fluctuations.

The middle panel of table 10 examines whether the North leads the South. Just as in the upper panel, northern output variables are listed in the rows and southern output variables are listed in the columns, but now the southern variables refer to fluctuations one year ahead. For example, the first cell of the first column reveals that the growth rate of northern GDP is positively correlated with the growth rate of southern GDP in the following year. This suggests that northern fluctuations lead southern fluctuations by at least one year. The lower right-hand block of the panel reveals that this aggregate relationship is supported by a strong correlation of the current growth rate of northern and future growth rate of southern non-primary activity. The lower panel of table 10, in contrast, points to a weak relationship between lagging southern activity and contemporaneous northern activity, implying that southern activity does not lead northern activity.

Next, I look at the relationship between fluctuations in northern and southern expenditure components, reported in table 11. The format of table 11 is the same as table 10: North variables form the rows and South variables form the columns. The upper panel reports contemporaneous correlation statistics, the middle panel looks at the relationship between current northern and future southern activity, and the lower panel reports the relationship between current northern and lagged southern activity.

Much has been written in the North–North business cycle literature about the fact that the international contemporaneous correlation of fluctuations of consumption expenditure is lower than the international contemporaneous correlation of fluctuations of GDP (see, for example, Backus, Kehoe, and Kydland, 1995, table 11.2). This statistic has garnered a lot of attention because it suggests that there is very little risk-sharing across industrial countries. Although it is less well known, it is also true that the international correlation of fluctuations of investment expenditure

TABLE 11

Interregional cyclical expenditure correlations

	Southern variable at time t					
Northern variable at t	GDP	Consumption	Investment	Government	Exports	Imports
GDP	0.45	0.31	0.14	0.13	0.49	0.12
Consumption	0.40	0.11	0.02	0.36	0.34	–0.13
Investment	0.31	0.21	0.11	0.03	0.59	0.20
Government	0.17	0.34	–0.05	–0.17	0.37	0.11
	Southern variable at time t+1					
	GDP	Consumption	Investment	Government	Exports	Imports
GDP	0.44	0.41	0.51	0.10	–0.08	0.43
Consumption	0.57	0.60	0.46	0.09	0.11	0.44
Investment	0.31	0.31	0.46	0.00	0.07	0.48
Government	0.09	–0.02	0.30	0.08	–0.36	0.13
	Southern variable at time t–1					
	GDP	Consumption	Investment	Government	Exports	Imports
GDP	0.10	–0.04	–0.38	0.31	0.25	–0.44
Consumption	0.12	–0.05	–0.28	0.36	0.05	–0.41
Investment	–0.04	–0.08	–0.45	0.17	0.38	–0.37
Government	–0.05	–0.32	–0.37	0.31	–0.03	–0.59

Note: Government indicates government spending
Source: Author's calculations based on data from World Bank, *World Development Indicators* (1998) and OECD, *International Sectoral Database* (1998).

is lower than the international correlation of fluctuations of GDP in contemporaneous North–North data. Table 11 reveals that these results also carry over to the contemporaneous North–South data. However, the correlation coefficients are considerably lower than those typically found in North–North studies. One interpretation of this finding is that there is even lower risk-sharing along the North–South dimension.

The middle panel of table 11 reinforces our earlier finding of a strong lead–lag relationship for North and South. Focusing on the first three rows and columns, we see that the growth rates of current northern consumption and investment are highly correlated with future growth rates of southern consumption and investment. In contrast to the contemporaneous statistics, the lagging interregional correlation coefficients of consumption and investment fluctuations exceed the lagging interregional correlation coefficient of GDP.

Are fluctuations transmitted from North to South?

There are two basic explanations for why regional economies move together. One view is that the regional business cycle simply reflects the fact that the regions are influenced by the same sources of innovation and that they respond to these innovations in the same way. The other is that fluctuations that originate in one region are transmitted to another through trade and/or production linkages. This section lays the seeds for future empirical and theoretical analysis of this issue by examining the cyclical relationship between interregional relative prices, interregional trade flows, and regional expenditure and production.

Support for the notion that business cycles are transmitted from the North to South via trade links appears in tables 6, 7, 12, and 13.[8] Table 12 describes in detail the contemporaneous relationship between fluctuations in northern aggregate and sectoral output and components of southern expenditure. This table reveals a strong direct relationship between northern and southern activity. In particular, the first row of the table indicates that the strong correlation between fluctuations in northern and southern aggregate activity is due in large part to a strong correlation between fluctuations in northern activity and southern exports. Moving down the exports column, the sectoral data reveal that the correlation between northern GDP and southern exports is actually the byproduct of an even stronger correlation of 0.64 between northern industrial production and southern exports (see also figure 3).

Before discussing the implications of these results, I examine the relationship between activity and relative prices, as shown in table 13. I study two relative prices: the real commodity price, which is the ratio of primary non-fuel prices to industrial goods prices, and the South's terms of trade, which is the ratio of export prices to import prices in the South. An improvement in the South's terms of trade means that the price of its exports has risen relative to the price of its imports.

The lower panel of table 13 reveals that fluctuations in the real commodity price are highly positively correlated with fluctuations in industrial activity and weakly negatively correlated with fluctuations in primary activity (see also figure 4). The upper and middle panels reveal a similar relationship between fluctuations in regional activity and the real commodity price. Non-fuel commodities are largely consumed as intermediate inputs in the production of northern industrial goods, which has led researchers to argue that fluctuations in the real commodity price are caused by fluctuations in the North's demand for non-fuel commodities (see, for example, Borensztein and Reinhart, 1994).

Linking these results on trade flows and relative prices, it follows that a typical expansion in the North is associated with increased demand for the South's exports, which causes a significant rise in the relative price of the South's exports.

The flip side to this relationship is the relationship between southern imports and southern activity.

TABLE 12

Interregional cyclical output and expenditure correlations

	Southern variable at time t					
Northern variable at t	GDP	Consumption	Investment	Government	Exports	Imports
GDP	0.45	0.31	0.14	0.13	0.49	0.12
Primary	–0.04	–0.29	0.14	0.05	–0.12	–0.04
Industrial	0.33	0.24	0.13	–0.07	0.64	0.22
Services	0.43	0.32	0.06	0.21	0.35	0.02

Note: Government indicates government spending.
Source: Author's calculations based on data from World Bank, *World Development Indicators* (1998) and OECD, *International Sectoral Database* (1998).

TABLE 13

Relative prices and activity

Correlation with	Terms of trade	Real commodity price
North		
Output		
GDP	0.54	0.69
Primary	–0.05	0.04
Industrial	0.58	0.74
Services	0.40	0.53
Expenditure		
Consumption	0.23	0.62
Investment	0.44	0.63
Government	0.54	0.40
Exports	0.50	0.29
Imports	0.44	0.55
South		
Output		
GDP	0.54	0.36
Primary	0.06	–0.18
Industrial	0.51	0.54
Services	0.43	0.20
Expenditure		
Consumption	0.56	0.24
Investment	0.37	0.27
Government	–0.36	–0.12
Exports	0.44	0.55
Imports	0.50	0.29
World		
Output		
GDP		0.71
Primary		–0.02
Industrial		0.75
Services		0.54

Note: Government indicates government spending.
Source: Author's calculations based on data from World Bank, *World Development Indicators* (1998) and OECD, *International Sectoral Database* (1998).

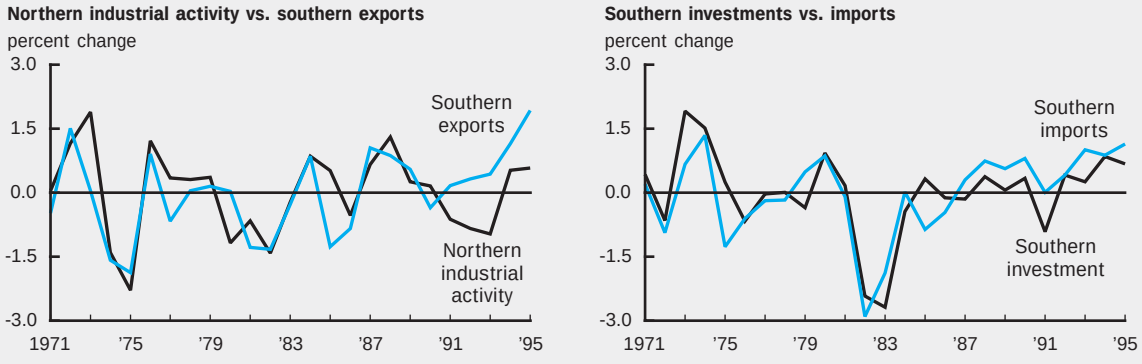
The correlation statistics for fluctuations in southern imports and expenditure are reported in the lower panel of table 7. This table reveals a strong correlation of 0.48 between fluctuations in southern GDP and imports. Reading down the imports column, we see that this correlation is supported by strong correlation coefficients for both consumption and investment. The investment correlation is particularly high at 0.79 (see also figure 3). Table 6 reveals a slightly weaker correlation between fluctuations in southern imports and non-primary production.

Looking again at the relative price and activity statistics in table 13, we find that the correlation statistics (in the middle of the table) reveal a strong positive correlation between fluctuations in the South's terms of trade and non-primary sector output. Fluctuations in southern consumption and investment expenditure are also positively correlated with fluctuations in the South's terms of trade (see also figure 5).

Bringing together these results on southern trade flows and relative prices, it follows that a typical improvement in the terms of trade of the South is associated with greater consumption and investment. The import and output data suggest that this increased demand for goods is satisfied by greater imports of industrial goods from the North and increased non-primary production in the South. One interpretation of these statistics is that an improvement in the South's terms of trade raises the South's real income and demand for both southern and northern final goods. The volatility statistics in table 9 suggest that terms of trade fluctuations may have a large effect on southern activity. In fact, the results in table 9 suggest that a typical terms of trade shock leads to a change in the South's real income that exceeds the typical increase in the growth rate of southern GDP.

FIGURE 3

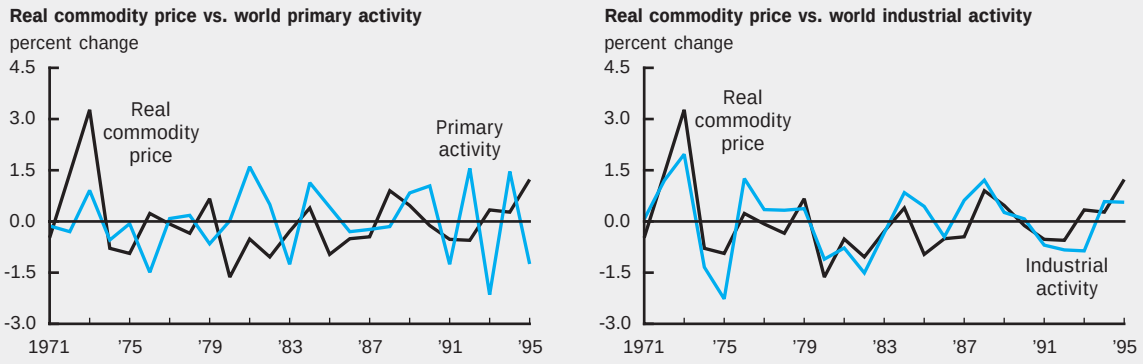
Trade flows vs. southern expenditure and northern activity



Notes and sources: See figure 1.

FIGURE 4

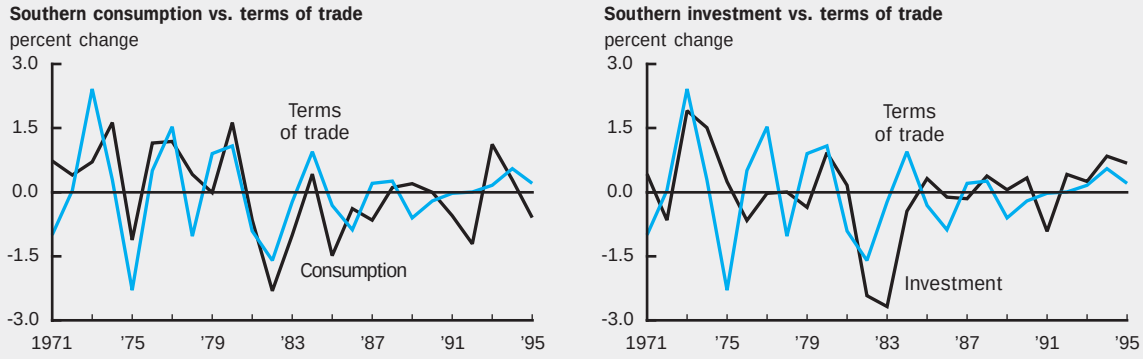
Real commodity price vs. real activity



Notes and sources: See figure 1.

FIGURE 5

Southern terms of trade vs. real expenditure



Notes and sources: See figure 1.

On the whole, the relative price, trade flow, and activity data suggest that economic fluctuations that originate in the North are transmitted to the South through fluctuations in the South's terms of trade. To recap, the data suggest that a typical expansion in the North leads to increased demand for the South's exports, which causes a significant increase in the terms of trade of the South through higher real commodity prices.[9] And, as discussed above, the data suggest that the South's typical response to an improvement in its terms of trade is higher consumption and investment, which is satisfied by greater imports from the North and increased production in the South. All told, the evidence suggests that a typical expansion in the North causes an expansion in the South.

Conclusion

The answer to the question of whether there is a North–South business cycle boils down to the question of whether cyclical fluctuations in one region are positively correlated with cyclical fluctuations in the other. With this in mind, I develop and document in this article a new dataset that describes in detail the short-run characteristics of the North and South regions. I find that fluctuations in the North's aggregate output are indeed strongly positively correlated with fluctuations in the South's aggregate output, which provides evidence in favor of the North–South business cycle. Using these data, I also argue that the North–South business cycle comes about because fluctuations that originate in North are transmitted to the South through fluctuations in their terms of trade. This analysis is the first step on the path to developing a better understanding of the way in which the North and South interact in the short-run. It will hopefully serve as a useful guide for future theoretical and empirical studies of international business cycles and North–South business cycles, in particular.[10]

NOTES

[1]See the extensive literature review by Currie, Muscatelli, and Vines (1988) for details.

[2]It is important to note that the results of the article would not change if the North only included industrial countries that were net exporters of manufactured goods from 1970 to 1995.

[3]The North comprises Australia, Austria, Belgium, Canada, Denmark, Finland, France, Germany, Greece, Iceland, Ireland, Italy, Japan, Netherlands, New Zealand, Norway, Portugal, Spain, Sweden, Switzerland, UK, and U.S. The South comprises Argentina, Bangladesh, Botswana, Brazil, Burkina Faso, Cameroon, Central African Republic, Chile, Colombia, Congo, Cost Rica, Côte d'Ivoire, Dominican Republic, Ecuador, El Salvador, Gambia, Ghana, Guatemala, Guinea-Bissau, Honduras, India, Indonesia, Jamaica, Kenya, Lesotho, Madagascar, Malawi, Malaysia, Mali, Mauritania, Mauritius, Mexico, Morocco, Nicaragua, Pakistan, Papua New Guinea, Paraguay, Philippines, Rwanda, Senegal, South Africa, Sri Lanka, Thailand, Tunisia, Turkey, Zambia, and Zimbabwe.

[4]See, for example, Backus, Kehoe, and Kydland (1995) and references therein.

[5]Baxter and Kouparitsas (2000) discuss these data in much greater detail.

[6]The North has a positive export balance with the South, which explains why the worldwide distribution of regional imports and exports is not identical.

[7]These findings add to the well-known result that there is high coherence between fluctuations of non-primary activities and low coherence between primary and nonprimary activity in individual industrial countries; see the discussion in Lucas (1977) for details.

[8]Kouparitsas (2001) provides a more formal time-series analysis of these data. He shows, using vector autoregression (VAR) techniques, that a significant share of the variation in southern activity is due to shocks that originate in the North. He also finds that the transmission mechanism identified here is supported by the VAR analysis.

[9]Baxter and Kouparitsas (2000) show that the North–South terms of trade is a broader relative price than the real commodity price. In fact, their work suggests that fluctuations in the growth rate of the real commodity price account for roughly 60 percent of the variation in the growth rate of the North–South terms of trade.

[10]Kouparitsas (1996) has already made substantial progress on the development of a quantitative model on the North–South business cycles. Kouparitsas' multisector interregional trade model is able to replicate many of the stylized facts presented in this article, including the observed pattern of interregional output correlation coefficients and the strong coherence between northern industrial activity and real commodity prices.

REFERENCES

Backus, D. K., and P. J. Kehoe, 1992, "International evidence on the historical properties of business cycles," *American Economic Review*, Vol. 82, pp. 864–888.

Backus, D. K., P. J. Kehoe, and F. E. Kydland, 1995, "International business cycles: Theory and evidence," in *Frontiers of business cycle research*, T. F. Cooley (ed.), Princeton, NJ: Princeton University Press.

Baxter, M, and M. A. Kouparitsas, 2000, "What can account for fluctuations in the terms of trade?," Federal Reserve Bank of Chicago, working paper, No. 00-25.

Borensztein, E., and C. M. Reinhart, 1994, "The macroeconomic determinants of commodity prices," International Monetary Fund, staff papers, No. 41, pp. 236–261.

Burns, A., and W. Mitchell, 1946, *Measuring Business Cycles*, New York: National Bureau of Economic Research.

Currie, D. A., A. Muscatelli, and D. Vines, 1988, "Introduction," in *Macroeconomic Interactions between North and South*, D. A. Currie and D. Vines (eds.), (Cambridge, UK: Cambridge University Press.

Kouparitsas, M. A., 2001, "To what extent are developing country business cycles caused by developed countries? A VAR analysis of international business cycles," Federal Reserve Bank of Chicago, working paper.

_____, 1996, "North–South business cycles," Federal Reserve Bank of Chicago, working paper, No. 96-9.

Lucas, R. E., 1977, "Understanding business cycles," in *Stabilization of the Domestic and International Economy*, K. Brunner and A. H. Meltzer (eds.), *Carnegie-Rochester Series on Public Policy*, Vol. 5, Amsterdam: North Holland.

Organization for Economic Cooperation and Development, 1998, *International Sectoral Data Base*, Paris.

United Nations, various years, *International Trade Statistics Yearbook: Trade by Country*, Vol. 1, New York.

World Bank, 1998, *World Development Indicators*, Washington, DC.

_____, 1991, "World tables and stars retrieval system version 2.00," computer disc, Washington, DC.

The financial performance of pure play Internet banks

Robert DeYoung

Introduction and summary

The number of banks and thrifts that offer financial services over the Internet is increasing rapidly in the U.S. By using "transactional websites," customers can check account balances, transfer funds, pay (and perhaps receive) bills, apply for loans, and perform a variety of other financial transactions without leaving their home or place of business. Approximately 1,100 U.S. banks and thrifts operated transactional websites at year-end 1999—an elevenfold increase over year-end 1997—and projections by bank regulators suggest that nearly half of U.S. banks will offer transactional websites by late 2001 or early 2002 (Furst, Lang, and Nolle, 2000).

Most banks and thrifts that operate over the Internet use a *click and mortar* business strategy, maintaining traditional networks of brick and mortar branches along with their transactional websites. Only a small number of banks and thrifts have completely abandoned physical branches in favor of a *pure play Internet* business strategy, relying exclusively on transactional websites to deliver banking services. As of mid-year 2000, less than two dozen of these virtual banks and thrifts were operating in the U.S., and their market penetration rates were in the low single digits. Various surveys report that Internet-only banks have captured less than 5 percent of the U.S. online banking market, and less than 1 percent of all Internet banking customers consider an Internet-only bank or thrift to be their primary bank.[1]

In theory, the pure play Internet model offers advantages for both banks and their customers. The central financial advantage stems from the savings associated with not having to operate branches. If being branchless substantially reduces physical overhead expenses, and if these savings are not offset by reductions in revenues or increases in other expense items, then, all else equal, Internet-only banks will earn high profits. Customers benefit not only from increased convenience, but also because these banks (again, in theory) can use some of their overhead cost savings to pay higher interest rates. The ability to pay above-market interest rates, combined with access to a much wider base of potential depositors, arguably allows these banks to grow faster than traditional banks.

In practice, however, the degree to which pure play Internet banks can actually deliver these benefits is not yet clear. The pure play business model, the banks that deploy it, and the technology on which it relies are still relatively young, so learning effects have not yet been exhausted. Furthermore, most of the existing evidence on Internet bank performance is anecdotal, and the few systematic studies of Internet bank performance do not distinguish between the pure play model and the click and mortar model.

This article represents a first attempt to analyze systematically the financial performance of pure play Internet banks. Unlike previous studies of Internet banks that include any branching or branchless bank that operates a transactional website, this article focuses on a small sample of six branchless banks and thrifts that distribute financial services exclusively through their websites. The pure play banks and thrifts in this sample are all newly chartered institutions, so I evaluate their financial performance relative to a benchmark sample of newly chartered banks and

Robert DeYoung is a senior economist and economic advisor at the Federal Reserve Bank of Chicago. *The author thanks Cynthia Bonnette for her expert advice on Internet banking; Susan Yuska and Darrin Halcomb for assistance with the data; Hesna Genay, Helen Koshy, and David Marshall for commenting on an earlier version of the manuscript; and seminar participants in the Research Department, the Supervision and Regulation Department, and the Emerging Payments Group at the Chicago Fed for their helpful suggestions.*

thrifts that have branches. I compare the two samples across 17 different measures of financial performance, using multiple regression analysis to control for differences in age, local economic environment, and regulatory conditions.

For this set of relatively young banks, my tests indicate that the average pure play Internet bank is significantly less profitable than the average branching bank. A number of factors contribute to this poor financial performance, including high labor expenses, low noninterest income, and difficulty attracting core deposits. My results also bring to light two fallacies about the standard Internet banking model, at least as implemented by the institutions in this sample: Overall overhead expenses are not necessarily lower, and overall deposit interest rates are not necessarily higher, compared with branching banks. However, consistent with the standard Internet banking model, my results indicate that Internet banks tend to grow faster than traditional branching banks. In sum, the early financial performance of these pure play Internet banks is reminiscent of the early financial performance of many nonfinancial dot-com companies: fast growth but low (or no) profits.

These results are intriguing because they imply that pure play Internet banking may not be a financially viable business model. However, the data presented here—which come from a small number of relatively young banks and thrifts using a largely untested business model—are not sufficient by themselves to support such a strong conclusion. As the pure play institutions analyzed here become more financially mature, as additional banks and thrifts adopt a pure play Internet approach, and as all of these institutions learn from each other's experiences, the financial performance of this business model may well improve. This article is an early attempt to analyze the financial performance of pure play Internet banks, and future studies using larger data sets and different analytic approaches may come to different conclusions. The results of this article should be interpreted with these caveats in mind.

The Internet and bank distribution channels

As the number of banks with fully transactional Internet sites increases—from zero only a few years ago to well over 1,000 today—the overall mix of bank distribution channels is also changing. As the number of commercial banks in the U.S. declined from roughly 12,000 to 8,500 during the 1990s, the number of branch locations increased from about 64,000 to 74,000, and the number of ATMs (automated teller machines) soared from around 80,000 to well over 200,000.[2] At the same time, the definitions of branch and ATM are changing. Some banks are converting their ATMs into "kiosks" that combine a telephone, an ATM, and an Internet terminal.[3] Increasingly, limited service branches are located in supermarkets or other retail establishments, and some of these "mini-branches" feature Internet kiosks in place of, or along with, teller windows.

This movement toward a less-centralized distribution system affects both customer convenience and banking costs. Convenience may increase because customers do not have to travel as far to perform basic banking transactions, and banks could potentially have lower overhead expenses as the number of full service branches declines. For example, it has been estimated that branch banking costs about $1.07 per transaction, telephone banking costs about $0.55 per transaction, ATM banking costs about $0.27 per transaction, and Internet banking costs about $0.01 per transaction.[4]

These distribution channels are not perfect substitutes. Checking an account balance, transferring funds, paying bills, and applying for credit cards do not require personal contact or a large physical space, and hence are well suited for delivery over the Internet channel. But setting up a new account, applying for a business loan, retirement planning, closing a mortgage, and other complex transactions often require a secure physical space and/or person-to-person communication. Furthermore, getting cash is impossible over the Internet and requires either branches or ATMs. Because some banking transactions are more conducive to some channels than to others, and because some customers prefer certain delivery channels, most (but not all) banks deploy a combination of delivery channels.

Most large and mid-sized banks treat different distribution channels as complements, and augment their physical branch locations with ATMs, call centers, and transactional Internet websites. The *click and mortar* banking strategy mentioned above is a good example of this approach. Although maintaining a network of branch offices requires substantial overhead expenditures, this strategy provides both convenient high-tech distribution and low-tech branch-based service options, and allows banks to sell a full range of banking services to a wide range of customers. Sometimes click and mortar banks use a *trade name* strategy, in which they create a separate brand identity for their Internet channels. This is simply a marketing distinction—trade name "banks" do not have separate banking charters and do not report separate financial statements—and this strategy is successful

only if the separate brand identity generates enough additional revenue to offset the additional marketing expenditures. (Perhaps the best known example of this strategy is Wingspan Bank, which is operated by First USA, an affiliate of Bank One.)

Other banks treat different distribution channels as substitutes, and serve their customers predominantly through a single channel. The *pure play Internet* banking strategy mentioned above is a good example of this approach. Because some products are difficult to deliver, and some customers are difficult to serve, over a single delivery channel, this approach is most likely to be effective as part of a niche strategy. For example, a recent study found that 70 percent of online banking customers said they would consider opening a new account at a bank with physical locations, but only 40 percent would consider doing so at an Internet-only bank.[5] Along these same lines, the traditional *brick and mortar* banking strategy may be profitable for community banks that specialize in products or customers that require person-to-person service. But as customers become more familiar with the Internet, there may be less room in the market for banks that completely exclude the Internet channel.

A financial model of pure play Internet banks

The central financial characteristic of the pure play Internet banking model is reduced overhead spending. By eliminating its physical branch locations, the pure play bank can substantially reduce expenses on rent (or mortgage payments), on upkeep and maintenance, and, most importantly, on the labor needed to run branch locations. Banks can use these savings to increase the per-unit profit on their existing business. Or banks can use the savings to increase their market share, attracting customers by paying higher interest rates on deposits or charging lower interest rates on loans. Although this will reduce the bank's interest margin, increasing the bank's size could create beneficial scale effects by spreading administrative costs over a greater volume of business or allowing the bank to market fee-based services (like investment or insurance products) to a greater number of captive customers.

The simple financial statements displayed in table 1 illustrate the potential financial advantages of the pure play Internet strategy. The balance sheet shown in panel A leaves out many items normally

TABLE 1

Potential advantages of pure play Internet model
($millions, unless stated otherwise)

A. Balance sheet

Cash	40	Deposits	450
Securities	140	Other liabilities	5
Loans	310		
Plant and other	10	Equity	45
Total assets	500	Liabilities and equity	500

B. Income statements

	Brick and mortar bank	Internet bank 1	Internet bank 2
Assumptions			
Rate on loans, securities (%)	7.50	7.50	7.50
Rate on deposits (%)	3.33	3.33	4.00
Noninterest income	7	7	7
Noninterest expense	15	12	12
Interest revenue	33.75	33.75	33.75
Interest expense	15.00	15.00	18.00
Net interest income	18.75	18.75	15.75
Noninterest income	7.00	7.00	7.00
Noninterest expense	15.00	12.00	12.00
Before tax profit	10.75	13.75	10.75
Tax (40%)	4.30	5.50	4.30
Net income	6.45	8.25	6.45
Return on assets (%)	1.29	1.65	1.29
Return on equity (%)	14.33	18.33	14.33

found on bank balance sheets, but it offers a reasonable representation of the composition of assets, liabilities, and equity at the typical U.S. commercial bank with $500 million in assets in 2000. The income statements are derived using the numbers on the balance sheet plus four additional numbers for the typical $500 million bank: the average interest rate paid on deposits, the average interest rate (including loan origination fees) received on loans and securities investments, total noninterest revenues, and total noninterest expenses.

Three different versions of the income statement are presented in panel B. The first column presents the income statement for a hypothetical brick and mortar bank that pays on average an interest rate of 3.33 percent on its deposit liabilities, and earns an average interest rate of 7.50 percent on its investments in loans and securities. Given these rates, the brick and mortar bank earns an interest margin of 4.17 percent and has an interest margin-to-assets ratio of about 3.75 percent. The bottom line is that the brick and mortar bank earns a 1.29 percent return on assets and a 14.33 percent return on book equity.

The second column (Internet bank 1) illustrates how the bank's profitability *might* change if it adopted an Internet distribution strategy, and if such a change in strategy allowed the bank to reduce its overhead expenditures by closing its brick and mortar branches. Note that one of the main assumptions changes—noninterest expenses decline by a hypothetical 20 percent, from $15 million a year to $12 million per year. (Even if a bank closed all its branches and successfully migrated its customers to the Internet, noninterest expenditures would not fully disappear. The bank would still have some physical space requirements, it would have to increase its expenditures on computer equipment, and it would still have labor expense—the biggest expense at banks after interest payments.) Assuming no other offsetting effects, the financial impact of this change would go straight to the bank's bottom line. Return on assets (ROA) would increase to 1.65 percent, and return on equity (ROE) would increase to 18.33 percent.

As discussed above, these increased profits could be simply paid out to the shareholders, or they could be retained and used to grow the bank. The third column (Internet bank 2) assumes that the bank uses the hypothetical overhead savings to attract additional depositors by paying higher rates on deposits. In this example, the bank increases its deposit rate by a hypothetical 20 percent, from 3.33 percent to 4.00 percent. This change reduces the bank's interest margin from 4.17 percent to 3.67 percent, but its return on assets and return on equity remain the same as the brick and mortar bank's. Over time these relatively high deposit rates might attract a greater number of customers to the bank, allowing it to grow faster than its brick and mortar competitors.[6] (Although not shown in table 1, a similar result could be accomplished by reducing the interest rate charged to borrowers from 7.50 percent to 6.834 percent, while leaving the deposit interest rate unchanged.)

Of course, this is a very simple model—in practice, a number of potentially offsetting financial or marketing effects could come into play. On the downside, the Internet bank must be able to generate loans, attract deposits, and sell fee-based services (for example, mutual funds, investment advice, insurance products) of the same amount and quality as the brick-and-mortar bank, despite having fewer physical locations for face-to-face contact with customers. On the upside, switching from physical branches to Internet distribution *may* generate financial and marketing benefits that are not captured in this simple model. Reductions in plant and equipment on the balance sheet could allow more assets to be shifted into revenue-generating loans or securities. The bank could use the Internet to gather deposits and market loans in new geographic locations, potentially increasing its growth rate and allowing for risk-reducing diversification effects. And customers that use the Internet for banking are likely to be more educated, sophisticated, and wealthy, and, therefore, more profitable customers.

Performance of the Internet banking model: Anecdotal evidence

One might persuasively argue that because Internet banking is so new, and because it is such a fundamentally different way to bank, it is too early to gauge the ultimate success of this business model. However, an increasing amount of anecdotal evidence testifies to various weaknesses of Internet banking—weaknesses that will have to be addressed for the pure play banking model to enjoy widespread viability in the future.

Person-to-person service

The U.S. has a relatively recent history of local banking, with tens of thousands of banks, thrifts, and credit unions focusing their efforts on individual cities, towns, and counties. Given this history, many Americans have come to expect in-person service, and very often name recognition, at their bank. Federal Reserve Chairman Alan Greenspan recently said "we should not lose sight of the exceptional value of franchises based on old-fashioned face-to-face interpersonal banking," a clear suggestion that traditional banking,

for at least some customers and/or products, will not wither away any time soon.[7]

Internet banking is the antithesis of high-touch, person-to-person banking. At an Internet bank, customer complaints must be resolved over the telephone or by e-mail, which can be frustrating for an already annoyed client. Customer requests that are simple at a brick-and-mortar bank, such as picking up additional deposit slips, become more complicated at an Internet bank, costing the bank postage and handling and requiring the customer to wait. Potential mortgage borrowers may be willing to shop for loan rates over the Web, but they are often reluctant to apply for these highly complicated financial products without person-to-person contact. A recent survey found that 85 percent of homebuyers use the Web for research but only 10 percent are comfortable getting their mortgage from a Web-only institution; another survey found that Internet banks get the majority of their mortgage originations from third party mortgage brokers.[8] If Internet-only banks have trouble generating mortgages and other types of loans, they have to make up the difference by investing in lower yielding securities (for example, mortgage-backed securities) or purchasing loans on the wholesale market where competition drives down margins.

Deposit pricing

Unable to attract depositors by offering in-person service, Internet banks often attempt to attract depositors that are interest-rate sensitive. A recent survey found that 14 Internet banks (which included both Internet-only and trade name Internet banks) offered an average rate of 6.875 percent on 12-month CDs (certificates of deposit), while 21 traditional banks offered an average rate of 6.29 percent. Another survey found that checking accounts at Internet-only banks generally paid between 3 percent and 6 percent (and were sometimes accompanied by no-fee or low-fee bill-paying services), compared with only about 2 percent at traditional banks.[9]

But these higher deposit rates are often merely short-run teaser rates designed to nab new customers—especially at trade name Internet banks and click and mortar banks where high deposit rates can be subsidized by other parts of the organization—and may not reflect the overall deposit rate structure of the bank.[10] These rates often attract financially savvy "hit-and-run" customers, who search the Web for high deposit rates and do not purchase additional services from the bank. These deposits typically flow out of the bank when interest rates are reduced or when the CD matures and, hence, do not represent long-term, core deposit funding. This is a primary reason that one industry consultant concluded that 70 percent of Internet customers are unprofitable, compared with 50 percent of non-Internet customers.[11] Thus, one of the theoretical financial advantages of the simple Internet banking model—growing the bank based on its ability to profitably pay above-market interest rates on deposits—may not work well in practice.

Getting cash and depositing checks

The most obvious problem for a bank without branches involves cash—how can customers get cash out of their accounts when they need it? Some Internet-only banks, like E*tradebank, maintain their own fleet of ATMs. Although ATMs are, of course, much less expensive than bank branches, they nonetheless represent an unwelcome expense for Internet banks.[12] Some Internet-only banks simply rebate to the depositor $5 or $6 in foreign ATM fees per month (typically enough to cover four to six ATM transactions), while some banks use a combination of the two approaches.

A similar problem arises when customers need to deposit checks into their Internet-only bank accounts. Direct deposit (ACH) works fine for repeating deposits like paychecks, but for non-repeating deposits customers typically must deposit by mail, which can be inconvenient and adds several days to the time a customer must wait before drawing on those funds. Some Internet banks have made alternative arrangements. For example, Wingspan Bank allows customers to make deposits in ATMs that are part of four regional electronic-transfer networks (NYCE, Fifth Third's Jeanie network, Star Systems, and MAC), and NationalInterBank.com allows customers to send their deposits by overnight mail at Mail Boxes Etc. locations. Of course, these arrangements also add to banks' expenses. A related problem involves funding new accounts. A large percentage of new accounts at Internet banks are never funded; depositors complete the online application form but never mail the funds. To combat this problem, NetBank allows new accounts to be funded at the time of application with credit cards or electronic transfers drawn on accounts at other banks.

In the future, smart cards that serve as cash substitutes—easily reloaded at home using a card reader and readily accepted by merchants—may make cash obsolete. When and if this happens, it will remove a major impediment to the pure play Internet model. But predictions of a "cashless" society have been made before and have yet to be fulfilled. No one knows how long it will take for U.S. consumers to willingly abandon cash.

Overhead expenditures

Economists are fond of reminding us that there is no free lunch, and eschewing physical space for cyberspace does not come without costs. A pure play Internet bank requires less physical overhead, but running a high-tech delivery system requires labor that is more highly educated and, therefore, more expensive than, say, window tellers. Unlike trade name Internet banks, pure play Internet banks cannot use the excess systems capacity of their parents for customer support, computer networks, data processing, or loan underwriting—they either must develop these systems from scratch or outsource them. And for Internet-only banks a 24-hour call center is a necessity, not a luxury, because the customers of an Internet bank expect around-the-clock business hours.

Marketing poses a particularly thorny problem. For Internet-only banks, creating a brand identity is at once more difficult (because the bank has to cut through the noise on the Internet) and more crucial (because the bank lacks physical branch locations which would otherwise help establish its presence in the marketplace). Rosen and Howard (2000) report that the average online retailer spends $26 on marketing and advertising per purchase, more than ten times the cost to brick and mortar retailers. Wingspan reportedly spent $19 million on Web advertising during a five-month period in early 2000, compared with $13 million for MBNA and $4.6 million for Fleet Boston Financial Group, both of which are larger than Wingspan but were not relying on a purely Internet distribution channel.[13] Furthermore, the effectiveness of these advertising expenditures is not clear; for example, the CEO (chief executive officer) of Bank One recently called banner ads on the Web to promote an Internet bank website "essentially worthless."[14]

Ellen Seidman, director of the Office of Thrift Supervision (OTS)—an agency that has chartered a number of Internet-only thrift institutions—summarized the overhead situation at Internet banks: "... the savings they have achieved by not having branches have often been offset by the high costs associated with acquiring and retaining customers and with updating and improving their technology infrastructure. The promise of low general and administrative expenses has yet to be proven."[15]

Performance of the Internet banking model: Research studies

Measuring the impact of the Internet on bank financial performance can be difficult, because in most cases the costs and revenues associated with Internet activities are not reported separately from the costs and revenues generated by the rest of the bank. As a result, there is little systematic evidence regarding the financial performance of the Internet banking channel. Most studies simply measure trends in market shares, numbers of accounts, market penetration rates, and similar phenomena using data from surveys of consumers, annual reports of banks, or bank press releases.

Recently, federal regulatory agencies have begun to collect data on Internet banking in a more systematic fashion. The Federal Reserve and the Office of the Comptroller of the Currency (OCC) have used their regularly scheduled safety and soundness examinations as an opportunity to ask banks about their Internet activities. Among other questions, examiners ask if the bank operates a website; whether that website is transactional; which products and services are offered on the website; whether the site is operated by an outside vendor or by the bank; and whether the bank plans to upgrade the website in the future. The resulting databases can be linked to the call report, allowing systematic financial analysis of various Internet banking strategies.

Because these databases are very new, only two studies (to my knowledge, at the time this article was prepared) have thus far used them to examine the financial performance of Internet banks. Both studies broadly define an "Internet bank" as a bank that operates a transactional website. Furst, Lang, and Nolle (2000) use a large database of national banks. They find that the typical Internet bank is more profitable than the typical non-Internet bank and tends to generate greater amounts of noninterest (fee-based) revenue; however, they find that newly chartered banks (less than one year old) that offer Internet banking tend to be less profitable than newly chartered non-Internet banks. Sullivan (2000) uses a database of commercial banks located in the Tenth Federal Reserve District. He finds that Internet banks have substantially higher ROE than non-Internet banks, although this difference is not statistically significant. He further finds that the typical Internet bank generates higher noninterest revenues, relies more on purchased funds financing, has slightly better loan quality, and (contrary to the standard Internet banking model, but consistent with the anecdotal evidence reported above) generates *higher* levels of noninterest expenses.

These studies are important, because they offer the first systematic analysis of whether banks that offer a nontrivial array of services over the Internet are more or less profitable than traditional brick and mortar banks that offer little or no services over the Internet. However, because these studies use such a

broad definition of an Internet bank, they cannot distinguish between the effectiveness of various Internet strategies, such as the pure play, trade name, and click and mortar strategies discussed above. Furthermore, because the databases these researchers have to work with do not identify the amount of business that flows through Internet channel, the banks in these studies may generate as little as 1 percent, or as much as 100 percent, of their business via the Internet. Thus, these studies do not provide (and in their defense, they do not set out to provide) a good test of the model in table 1, because most of the "Internet banks" in these studies are click and mortar banks that employ multiple distribution channels.

In contrast, this article focuses on the financial performance of pure play Internet banks only. The downside of this approach, compared with the earlier studies, is that only a small handful of pure play Internet banks have operated long enough to have established a financial record. But on the upside, this approach allows us to more accurately test the Internet model in table 1, because pure play banks generate 100 percent of their business through the Internet channel.

Identifying pure play Internet banks and thrifts

A financial institution had to meet four conditions to be included in this study as a pure play Internet bank. To start with, the institution had to be previously identified by the Federal Deposit Insurance Corporation (FDIC) as an institution whose primary contact with customers was over the Internet. The FDIC maintains an informal database of Internet activity at commercial banks and thrifts, and at the end of the third quarter 2000, there were 22 "Internet-primary" institutions in this database. Second, the institution had to produce a full range of basic banking services, including taking insured deposits, offering checking accounts, and making loans. Third, the institution had to begin its operations using a new commercial bank charter or new thrift charter. Imposing this condition excludes institutions that began their Internet-only operations using a preexisting charter, and whose assets, liabilities, costs, and revenues unavoidably reflect the preexisting physical branching strategy. Fourth, the institution had to file its first quarterly *Statement of Condition and Income* (call report) before the year 2000. Imposing this condition excludes institutions for which I could observe only one or two full quarters of financial performance.

Only six banks and thrifts met all four of these conditions. The six pure play Internet banks—Ebank, First Internet Bank of Indiana, Gay and Lesbian Bank, Marketplace Bank, NetBank, and Principal Bank—are described in box 1.[16] The other 16 banks and thrifts from the informal FDIC list are also listed in the box, along with a description of how they violated one of the conditions listed above. Although this filtering procedure excludes the majority of banks and thrifts on the initial FDIC list, for my analysis to be meaningful it must focus only on institutions that can clearly be called "pure play Internet" banks and thrifts.[17]

It is important to understand that the tests below reflect the *average* financial performance of these six pure play banks. The results in this study are not meant to imply that any single one of these six institutions performed well or performed poorly during the sample period.

Choosing an appropriate performance benchmark

This careful selection process yields an interesting byproduct: each of the six pure play Internet banks is also a newly chartered, or de novo, bank. This is an important observation, because the financial performance of de novo banks has been shown to differ systematically from the financial performance of established banks (Hunter and Srinivisan, 1990; DeYoung and Hasan, 1998; and DeYoung, 1999). To properly evaluate the financial performance of these pure play Internet banks, it is therefore essential to evaluate their performance relative to newly chartered non-Internet banks, not relative to established non-Internet banks.

Figures 1 and 2 illustrate why this is important. Figure 1 shows how a newly chartered bank's ROA improves over time relative to the average ROA for established banks. Figure 2 shows how a newly chartered bank's capital ratio (equity divided by assets) declines over time relative to the average capital ratio for established banks. Although these figures are highly stylized, they are reasonable representations of results from systematic studies of actual de novo bank performance (for example, DeYoung, 1999). In terms of ROA, the typical de novo bank substantially underperforms the typical established bank during its early years, but as the new bank matures its profitability gradually approaches the level of established banks. As shown in the figure, this maturation process—or learning curve—can take as long as a decade to run its course. A similar, albeit faster, maturation process occurs for the capital ratio, with the typical de novo bank having a substantially larger capital cushion than the typical established bank during its early years.

These learning curve effects must be taken into account to avoid misstating the financial performance of pure play Internet banks. Figure 1 shows

BOX 1

The pure players

It is common knowledge that the number of Internet banks and thrifts is increasing rapidly in the U.S. But it is less well known how few of these banks and thrifts have completely abandoned traditional distribution channels in favor of a pure play strategy of operating exclusively, or nearly exclusively, over the Internet. As of the third quarter of 2000, staff at the FDIC had identified 22 Internet-primary banks and thrifts that do business primarily over the Internet and have no branch locations. "Internet primary" is an informal designation, and is not used formally in any regulatory or supervisory matter.

Starting with the 22 institutions on the FDIC's informal list, I identified six pure play Internet banks and thrifts to include in the regression tests. Each of these six institutions had the following characteristics: 1) it produced a full range of basic banking services, including taking insured deposits, offering checking accounts, and making loans; 2) it began its operations using a new commercial bank charter or new thrift charter; and 3) it filed its first quarterly *Statement of Condition and Income* (call report) before the year 2000.

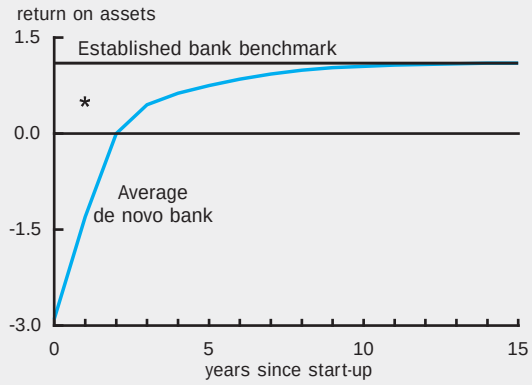
A short description of these six pure play Internet banks and thrifts follows. The accompanying financial information reflects the most current call report data available as of June 30, 2000. Other information was gleaned from informal discussions with other regulators, from the bank websites, and from the National Information Center Database.

- Ebank, www.ebank.com, established in August 1998 as Commerce Bank—a $69 million thrift institution headquartered in Atlanta, Georgia. It adopted its Internet focus in the early months of its life and changed its name to Ebank in 1999. The website has a clear focus on selling credit, transactions, and other services to small businesses. The loan portfolio is a mixture of business and real estate loans.
- First Internet Bank of Indiana, www.firstib.com, established in December 1998—a $188 million state bank headquartered in Indianapolis, Indiana. The loan portfolio is mostly real estate and consumer (non-credit-card) loans.
- Gay and Lesbian Bank, www.glbank.com, established in September 1999—a $41 million thrift institution headquartered in Pensacola, Florida. The bank's mission is to help ensure discrimination-free access to financial products and services, such as mortgage loans for unmarried couples. The loan portfolio is about evenly split among real estate, consumer, and business loans.
- Marketplace Bank, www.marketplacebank.com, established in October 1999—a $465 million national bank headquartered in Maitland, Florida. The bank is owned by Amicus Federal Savings Bank, which itself is owned by Canadian Imperial Bank of Commerce. It operates Internet kiosks (which combine ATMs, telephones, and Internet terminals) in large grocery store chains. The loan portfolio is almost exclusively real estate loans.
- NetBank, www.netbank.com, established in August 1997 as Atlanta Internet Bank—a $1.5 billion thrift institution headquartered in Alpharetta, Georgia. It changed its name to NetBank in 1998. The loan portfolio is almost exclusively real estate loans.
- Principal Bank, www.principal.com, established in February 1998—a $182 million thrift institution headquartered in Des Moines, Iowa. It is affiliated with the Principal Financial Group, which also includes life insurance, financial services, and mortgage banking subsidiaries. The loan portfolio is mostly real estate loans, mixed with some consumer loans.

I excluded the remaining 16 banks and thrifts on the FDIC's informal list for a variety of reasons. A large number of the excluded institutions were chartered after 1999 and had not yet established a financial performance record long enough to be included in the tests. This group includes Bank of Internet, DeepGreen Bank, Echarge Bank, EOS Bank, ING Bank, Lighthouse Bank, TB Bank, and Virtual Bank. Other excluded institutions are "limited purpose banks" that either choose not to offer a full range of banking services or are limited by their charter type from doing so. This includes CompuBank and BMW Bank. Another set of excluded institutions started up by taking over existing bank or thrift charters; in some of these cases a clear date on which these institutions began their pure play Internet strategy could not be identified, and in any event the performance of these institutions could be affected by the residue from their former business strategies. This group includes ClarityBank.com, E*trade Bank, Nexity Bank, and Next Bank. One excluded institution, Security First Network Bank, changed ownership and strategy after receiving its charter. Another excluded institution, Millennium Bank, did not yet have a functioning interactive website at the time this article was prepared.

FIGURE 1

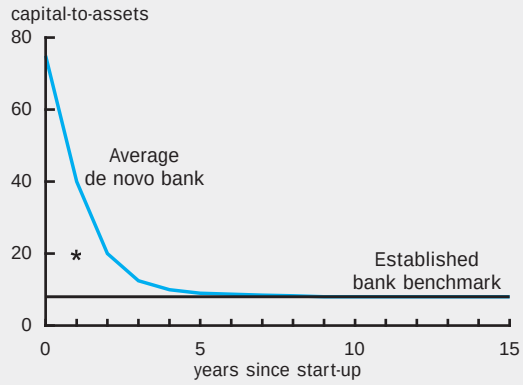
Typical time trend for ROA at a newly chartered bank



* Indicates a hypothetical Internet bank.

FIGURE 2

Typical time trend for capital-to-assets at a newly chartered bank



* Indicates a hypothetical Internet bank.

why. Assume that the asterisk located at about ROA = 0.50 percent represents the ROA of a hypothetical one-year-old pure play Internet bank. (I present some actual ROA data for pure play Internet banks later in this article.) Is this good performance or bad performance? Compared with the ROA of the typical established bank, say around 1.20 percent, this would be a poor performance. But such a comparison would understate the profitability of the pure play Internet bank because it ignores the de novo bank learning curve—that is, it does not separate the financial effects of "newness" from the financial effects of "pure playing." The more appropriate benchmark is the ROA of –1.00 percent generated by the typical one-year-old bank. Compared with this more appropriate benchmark, the hypothetical one-year-old Internet bank would be performing well.

Similarly, assume the asterisk in figure 2 represents the capital ratio of a hypothetical one-year-old pure play Internet bank, about 20 percent. (I present some actual capital ratio data for pure play Internet banks later in this article.) Is this a large capital cushion or a small capital cushion? Compared with the established bank benchmark of 8 percent, this would be a large capital cushion. But such a comparison may overstate the safety and soundness of the pure play Internet bank. The more appropriate benchmark, *ceteris paribus*, is the capital ratio of about 40 percent for the typical one-year-old bank. Compared with this benchmark, the hypothetical one-year old Internet bank would have a relatively small capital cushion.

Similar learning curve patterns exist for other measures of de novo bank financial performance. DeYoung (1999, p. 22) shows that overhead costs, interest revenues, noninterest revenues, and deposit funding at new banks start out worse than established banks but gradually improve over time, while asset growth, equity cushions, and loan quality start out better than established banks but deteriorate over time.

By estimating the following equation for a suitable data set of newly chartered banks, I can test how using a pure play Internet strategy affects the financial performance of banks and thrifts, while at the same time controlling for the effects of "newness" on the financial performance of these firms:

1) $$\textit{financial performance}_{i,t} = \alpha + \beta \times \textit{pure play}_{i} + \gamma \times \ln(\textit{bank age}_{i,t}) + \delta \times \textit{control variables}_{i,t} + \varepsilon_{i,t}.$$

I estimate equation 1 using time-series cross-section data so that each bank or thrift can be observed at different stages of its early development. The subscript i is an index that identifies individual banks, and the subscript t is a time index that represents calendar quarters. On the left-hand-side, *financial performance* takes on the value of a variety of different financial performance ratios (such as return on assets, equity to assets, or asset growth rate) in different regressions. On the right-hand-side, *pure play* is a binary variable equal to one if the bank uses the pure play Internet strategy, and equal to zero otherwise. The variable *bank age* measures the age of a bank or thrift in calendar quarters. This variable controls for the variation in financial performance due to the learning curve effects illustrated in figures 1 and 2, and specifying this variable as a natural log allows the estimated relationship to closely reflect the curve-linear shapes shown in those figures.[18] Similarly, *control variables* are a collection of variables that control for exogenous influences on financial performance, such as local

economic conditions, thrift or national bank status, organizational form, fixed time effects, and quarter dummies.

Because equation 1 holds constant the effects of newness (*bank age*) and other factors (*control variables*), the coefficient β on the *pure play* variable provides a controlled test of whether the financial performance of pure play Internet banks is better, similar, or worse than the financial performance of non-Internet banks. In terms of figures 1 and 2, the sign (positive or negative) of this coefficient indicates whether the asterisks are above or below the de novo bank learning curves. A more detailed description of equation 1 appears in the appendix.

The data set

The time-series cross-section data set includes data for two groups of institutions: the six pure play Internet banks and thrifts described above and 522 "benchmark" banks and thrifts. Like the pure play banks and thrifts, the benchmark banks and thrifts are all located in urban markets, and all are de novo institutions that started their operations in either 1997, 1998, or 1999.[19]

I observe financial information for these institutions each quarter over a 13-quarter window from 1997:Q2 through 2000:Q2. I begin with 1997:Q2 because I exclude data from the start-up quarters—institutions typically operate for less than 90 days during their start-up quarters, and the financial statements reported for these quarters tend to contain low-quality data. The final quarter in the window is 2000:Q2 because this was the most recent data available at the time this article was prepared. The data set is an unbalanced panel. Institutions that started up in 1997:Q1 were observed as many as 13 times, and institutions that started up in 1999:Q4 were observed only twice. Some institutions did not last until the end of the 13-quarter window due either to acquisition or failure. In all, the data set includes 3,263 quarterly observations (38 for the pure play banks and thrifts and 3,225 for the benchmark banks and thrifts). The average pure play observation was 4.45 quarters old, and the average benchmark observation was 4.74 quarters old. Additional details about the data set are included in the appendix.

Regression test results

I estimate 17 different versions of equation 1, applying ordinary least squares (OLS) techniques to the data panel described above. Each of the rows in table 2 reports results from a different regression, and each regression uses a different *financial performance*

TABLE 2

Selected OLS regression results from equation 1

Performance variable	Mean for benchmark banks[a]	Pure play mean vs. benchmark mean (using estimated β)[b]	Change as average bank ages one quarter (using estimated γ)[b]	R^2
Return on assets	–1.36	1.76 lower***	+0.31***	0.3940
Return on equity	–4.61	6.87 lower***	+1.01***	0.2793
Book value of physical assets/assets	3.60	0.85 lower***	–0.11***	0.1817
Expense on physical assets/assets	0.82	0.12 higher	–0.03***	0.2550
Expense on labor/assets	2.52	0.45 higher*	–0.10***	0.3149
Employees per $million assets	0.53	0.007 higher	–0.015***	0.2943
Salary and benefits/employees	$49,005	$7,165 higher*	–$547***	0.2406
Noninterest expense/assets	5.06	1.65 higher***	–0.21***	0.3108
Average deposit interest rate	3.46	0.13 lower	+0.05***	0.2480
Deposits/assets	76.65	12.39 lower***	+1.58***	0.3304
Average loan interest rate	7.66	1.00 lower**	+0.19***	0.4898
Loans/assets	56.53	5.09 lower	+2.04***	0.2420
Nonperforming loans/loans	0.16	0.14 lower***	+0.21***	0.0502
Interest margin/assets	3.63	0.18 higher	+0.05***	0.3505
Noninterest income/assets	0.71	0.76 lower***	+0.03***	0.1020
Annual asset growth rate	94.06	43.97 higher*	–14.59***	0.3275
Equity/assets	19.42	14.77 higher***	–1.56***	0.4916

[a]Results are in annual percent terms, unless stated otherwise.
[b]Results are in percentage points, unless stated otherwise.
Notes: Each row displays results from a separate regression. Data are an unbalanced panel of 3,263 quarterly observations of 528 banks and thrifts between 1997:Q2 and 2000:Q2. The *, **, and *** indicate statistical significance at the 10 percent, 5 percent, and 1 percent levels, respectively, for the estimated coefficients β and γ.

ratio as the dependent variable. The first column reports the financial performance for the average benchmark bank. The second column reports the financial performance of the average pure play bank relative to the average benchmark bank, based on the estimated value of coefficient β in each regression. The third column reports the change in financial performance of the average sample bank as it grows one quarter older (from approximately five quarters old to six quarters old), based on the estimated value of coefficient γ in each regression. All of the regressions were estimated using quarterly data, but the results in table 2 are reported in annualized terms.

Because there is only a small number (38) of pure play observations in the data set, the significance levels, indicated by asterisks in the table, are only suggestive of statistical precision, and should be interpreted with caution.[20] The coefficient estimates themselves, however, are unbiased estimates. Overall, the multiple regression tests performed here are a useful way to evaluate the performance of pure play banks after controlling for a large number of exogenous influences represented on the right-hand-side of equation 1. Full regression results are available upon request from the author.

The coefficient γ on ln(*bank age*) is statistically significant at the 1 percent level in all 17 regressions. This indicates that all aspects of financial performance are in the process of evolving at the typical five-quarter-old bank in this data set. Return on assets, return on equity, rates paid on deposits, total funding from deposits, rates charged on loans, total investment in loans, nonperforming loan ratios, overall interest margins, and the portion of income generated from noninterest activity are all on the increase as these banks mature. On the other hand, the book value of physical assets, spending on physical assets, total labor expenses, workers per million dollars of assets, average compensation, noninterest expenses, asset growth rates, and equity capital cushions are all on the decrease as these banks mature. These results are consistent with the previous research cited above on the evolution of financial performance at newly chartered banks. Furthermore, these results suggest that the primary regression test—that is, the coefficient β on the *pure play* variable—can be interpreted knowing that significant controls are in place to absorb the effect of learning curves on the financial performance of new banks.

The coefficient β is statistically significant in 12 of the 17 regressions. Most importantly, this coefficient is negative and highly significant in the profitability regressions in the first two rows of the table. Thus, after controlling for bank age and other influences on bank performance, pure play Internet banks and thrifts earned lower profits than banks that used more traditional distribution channels. The estimates suggest that ROA and ROE at the average five-quarter-old pure play Internet bank were, respectively, 176 basis points and 687 basis points lower than the ROA and ROE at the typical five-quarter-old benchmark bank. This is consistent with the findings of Furst, Lang, and Nolle (2000) for (non-pure-play) de novo Internet banks. The remaining regressions contain prescriptive evidence for why profitability is relatively lower at the pure play banks.

Do the regressions provide evidence that overhead spending at pure play Internet banks is relatively low, a central tenet of the Internet banking model in table 1? Consistent with the model, the book value of physical assets at pure play Internet banks was significantly lower than at the benchmark banks. However, ongoing expenditures on physical assets were not lower, perhaps reflecting a tradeoff between lower spending on branch locations but higher ongoing spending on technology.[21] Furthermore, spending on labor—a substantial component of which is nonvariable overhead spending—was significantly higher at the pure play banks. These high labor expenses appear to be associated with relatively high salaries and benefits, not excessively large numbers of employees. The average pure play bank paid its employees about $7,000 more per year than the average benchmark bank.

These results suggest a nontraditional overhead structure at the pure play Internet banks, featuring less physical overhead but more highly paid (and presumably more highly skilled) workers. This overhead structure generates substantially higher noninterest expenses. These high noninterest expenses include some unknown amount of spending on marketing and advertising. Unfortunately, it is not possible to test whether the pure play banks incurred higher marketing expenses than the benchmark banks—as suggested by the anecdotal evidence presented above—because marketing expenses are not reported separately in the call reports.

Do the regressions support the other central feature of the model in table 1, that pure play Internet banks pay systematically higher interest rates to depositors or charge systematically lower rates to borrowers? The regressions show no significant difference in average deposit rates between the pure play and benchmark samples. This is consistent with the anecdotal evidence that Internet banks periodically offer deposit products with high interest rates, but

that deposit rates are not systematically higher across all deposit products. In contrast, the regressions show that the average loan interest rate is about 100 basis points lower at the pure play banks than at the benchmark banks. This result must be interpreted with some caution, as it may simply indicate that the pure play banks in this sample made lower risk loans. Indeed, the nonperforming loan ratio is relatively low at the pure play banks, suggesting that this may be the case.

Consistent with the results of Sullivan (2000), the pure play banks in this sample had difficulty attracting deposit funding. For the average pure play bank, the deposits-to-assets ratio is over 12 percentage points lower than at the average benchmark bank. This funding shortfall is offset by higher levels of expensive equity capital: The equity-to-assets ratio at the average pure play bank was nearly 15 percentage points higher than at the average benchmark bank. There is no significant difference in the loans-to-assets ratios between the two sets of banks. Combining the effects of loan levels, deposit levels, loan rates, and deposit rates, there is no significant difference in net interest margins across the two sets of banks.

In addition to their relatively high noninterest expenses, the pure play banks had significantly lower noninterest income ratios than the benchmark banks. This suggests that it is difficult to cross-sell fee-based financial products to loan and deposit customers over a distribution channel that minimizes person-to-person contact. This is consistent with the notion that a large portion of Internet banking customers do not view the Internet bank as their main financial institution.

Finally, do the regressions provide evidence that pure play Internet banks—unencumbered by physical location and able to reach across geographic boundaries via the Internet—grow faster than more traditional banks? The regressions confirm this conventional wisdom. Assets at the pure play banks grew at a substantially faster rate than assets at the benchmark banks, more than 40 percentage points faster per year. This torrid asset growth rate, combined with the deposit funding problems intrinsic to this business model, helps explain why the pure play banks have below average deposit-to-assets ratios. The rapid asset growth rate also helps explain why these banks hold above average capital ratios, which are needed not only to fund fast growth, but also to absorb the large initial losses that these banks generate. For these reasons, federal and state chartering authorities typically require higher levels of start-up capital for de novo Internet banks than for traditional de novo banks.

Conclusion

The results of this article should be interpreted carefully. On one hand, I find statistical evidence of poor financial performance at pure play Internet banks and thrifts. On the other hand, this evidence is based on a small number of newly chartered banks and thrifts that are struggling with two different learning curves: They are passing through the financial maturation process common to all de novo banks, and they are pioneering the use of a new business model. Although my empirical framework carefully controls for the effects of the first of these two learning curves, the newness of the pure play business model precludes me from controlling for the effects of the second of the two learning curves.

Putting these potential limitations aside for the moment, the tests indicate that the average pure play Internet bank in my data set was less profitable than the average branching bank of similar age and circumstance. The tests also imply the existence of two fallacies about the Internet banking model: that this strategy does not necessarily reduce overall overhead expenses, and that banks that use this strategy do not necessarily pay higher overall interest rates on deposits.

I find that pure play Internet banks tend to have relatively low physical overhead, chiefly due to not operating brick and mortar branches. However, I find relatively high levels of other noninterest expenses, chiefly related to labor costs, which more than offset any expense savings from lower physical overhead. Contrary to anecdotal evidence, I find no evidence that pure play Internet banks pay higher than average deposit rates on a systematic basis. My results also suggest that the Internet-only distribution strategy used by these banks makes it difficult to cross-sell fee-based financial products to their loan and deposit customers, depressing revenue growth and contributing to their low profitability. Despite these troubles, I find evidence that pure play banks grow faster than non-Internet banks at similar stages of development. On average, rapid asset growth outstrips these banks' ability to raise deposit funding, requiring large amounts of expensive equity capital funding to fuel their growth.

While these results are based on historical data from 1997 through 1999, they are consistent with more recent reports in the banking press about the difficulties of "going virtual." During the first week of 2001 alone, two Internet-only banks announced measures to boost their lagging profits. First Internet Bank of Indiana announced it was laying off 20 percent

of its already small work force in an effort to cut overhead expenses, and WingspanBank.com announced changes in checking account fees and interest rates in an effort to enhance noninterest revenues and interest margins.[22] Moreover, the financial performance of pure play Internet banks captured here is reminiscent of the financial performance often observed for many non-financial Internet firms: rapid growth but low (or negative) profits.

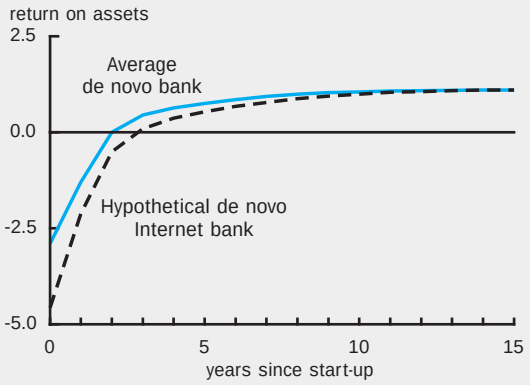
Limitations of this article

Proponents of e-commerce typically respond that virtual business models are financially sound, but that the path to profitability is simply longer than in brick and mortar business models. It is certainly possible that the pure play Internet banks and thrifts in this data set are too young to have fully exploited the advantages of the business model. As illustrated in figure 3, the learning curve for de novo Internet banks may simply be longer and flatter than for traditional de novo banks. If this is the case, then in the long run, the average bank examined in this article will eventually earn profits equal to or exceeding those generated by more traditional distribution strategies. For example, the CEO of Principal Bank recently reaffirmed his bank's intentions to remain a pure play Internet bank and not add branches: "If you have a solid business plan, you can remain focused and it's not necessary to change."[23]

Furthermore, this article evaluates the financial performance of pure play Internet banks to the exclusion of other Internet banking strategies. As such, the results generated here may have limited implications for the financial performance of Internet banks that use, say, the click and mortar strategy. By combining several distribution channels, banks might offer Internet banking while at the same time avoiding some of that channel's biggest problems, like the high costs of marketing, the costs associated with ATM subsidies, and the reluctance of customers to commit to primary relationships with a purely Web-based bank. Regulatory authorities may give banks and thrifts an additional push in this direction: Concerned about high growth rates but low profitability, chartering authorities are requiring increasingly higher levels of initial capitalization for applicants that seek charters for pure play start-ups.

FIGURE 3

Hypothetical time trends for ROA at newly chartered Internet and non-Internet banks



The Internet remains an emerging technology in the banking industry. In the short period between the preparation and publication of this article, it is likely that new technologies will have become available to banks; new Internet strategies will have been invented, launched, and perhaps abandoned; and the results of new studies will have been announced. As time passes, and more than six banks and thrifts meet the definitional requirements of a "pure play Internet bank" used in this article, the empirical testing performed here can be expanded to include more institutions, as well as a greater number of quarterly observations for each institution.

APPENDIX

Data and regression details

The regression tests used an unbalanced data panel of 3,263 quarterly observations of 528 banks and thrifts for the 13 quarters between 1997:Q2 and 2000:Q2. This includes 38 observations of six pure play Internet banks and thrifts and 3,225 observations of 522 benchmark banks and thrifts. Table A1 displays selected descriptive statistics for these two groups of banks and thrifts.

All banks and thrifts were located in urban markets (metropolitan statistical areas), and all began their operations with new bank or thrift charters in either 1997, 1998, or 1999. Flow variables are measured as quarterly values, stock variables are measured as quarter-end values, and all dollar values are measured in year-end 1999 dollars. I excluded from the data any bank or thrift that did not make loans, did not hold deposits, or held large concentrations of

TABLE A1

Summary statistics for data in regression tests

	Benchmark		Pure play	
	Mean	Standard deviation	Mean	Standard deviation
AGE (quarters)	4.74	3.03	4.45	2.79
ASSETS ($000s)	64,780	232,567	267,491	396,002
EMPLOYGROWTH (%)	11.18	3.84	10.94[a]	—[a]
THRIFT	0.0719	0.2584	0.7894	0.4132
OCC	0.2232	0.4165	0.0526	0.2263
MBHC	0.1854	0.3887	0.0526	0.2263
REALESTATE	0.6127	0.2212	0.7838	0.1883
BUSINESS	0.2767	0.1874	0.0912	0.1193

[a]For the pure play banks and thrifts, EMPLOYGROWTH is set equal to the national average during the sample period.
Note: There are 3,225 quarterly observations (522 banks) in the benchmark sample and 38 quarterly observations (six banks) in the pure play sample.

credit card loans that exceeded 25 percent of total loans. I deleted selected quarterly observations if they had unrealistic values for any of the financial performance variables. In addition, because quarterly accounting revenue and expense data can fluctuate in patterns that are not representative of actual financial performance, I truncated the value of all financial performance variables used on the left-hand-side of the regressions at the 1st and 99th percentiles of their sample distributions.

The full regression specification used in these tests is:

$$
\begin{aligned}
2)\quad PERFORMANCE_{i,t} &= \alpha \\
&+ \beta \times PUREPLAY_i + \gamma \times \ln(AGE_{i,t}) \\
&+ \delta_1 \times \ln(ASSETS_{i,t}) + \delta_2 \times EMPLOYGROWTH_i \\
&+ \delta_3 \times THRIFT_i + \delta_4 \times OCC_i + \delta_5 \times MBHC_i \\
&+ \delta_6 \times REALESTATE_{i,t} + \delta_7 \times BUSINESS_{i,t} \\
&+ \delta_8 \times YEAR98_t + \delta_9 \times YEAR99_t \\
&+ \delta_{10} \times YEAR00_t + \delta_{11} \times QTR1_t \\
&+ \delta_{12} \times QTR2_t + \delta_{13} \times QTR3_t + \varepsilon_{i,t}.
\end{aligned}
$$

Equation 2 is simply a detailed specification of equation 1 from the text. *PERFORMANCE* represents the financial performance of bank *i* at time *t*, based on a different financial ratio in each of the 17 estimated regressions. (These 17 financial performance variables are displayed in table 2.) *PUREPLAY* is a dummy variable equal to one if bank *i* is a pure play Internet bank. *AGE* measures bank *i*'s age in quarters at time *t*. *ASSETS* measures the size of bank *i* at time *t* in terms of assets. *EMPLOYGROWTH* measures the cumulative percentage increase in employment in the bank *i*'s home state between 1996 and 1999. (This variable is set equal to the national average for the pure play banks, which have no home geographic market.) *THRIFT* is a dummy variable equal to one if bank *i* is a thrift institution. *OCC* is a dummy variable equal to one if bank *i* is a commercial bank with a national bank charter. *MBHC* is a dummy variable equal to one if bank *i* is an affiliate in a multibank holding company. *REALESTATE* and *BUSINESS* measure that percentage of bank *i*'s loan portfolio comprised of real estate loans and business loans, respectively, at time *t*. *YEAR98*, *YEAR99*, and *YEAR00* are dummy variables equal to one for observations that occurred in 1998, 1999, and 2000, respectively. *QTR1*, *QTR2*, and *QTR3* are dummy variables equal to one for observations that occurred in the first, second, and third calendar quarters, respectively. The residual term $\varepsilon_{i,t}$ is assumed to be distributed normally with zero mean.

NOTES

[1]The information in this paragraph came from conversations with Federal Deposit Insurance Corporation (FDIC) staff; The Financial Times, Limited (2000) (based on data from ING); and The McGraw-Hill Companies Inc. (2000) (based on data from McKinsey & Company).

[2]The source for the numbers of banks and branches is the FDIC website. The sources for the number of ATMs are the American Bankers Association, Bank Network News, Ernst & Young, and Dove Associates.

[3]For example, Bank of America recently announced that most of its 14,000 ATMs will be retrofitted to become "automated banking machines," or ABMs.

[4]The sources for these numbers are Nathan (1999) (based on a survey by Booz, Allen & Hamilton) and The Economist Newspaper Limited (2000) (based on data from Jupiter Communications).

[5]Thomson Financial Media (2000) (based on data from McKinsey & Co.).

[6]There is an alternative interpretation of the third column: Because the hypothetical Internet bank has fewer physical locations to service its customers, it has to pay a higher rate to retain its existing depositors or it has to charge a lower rate to retain its existing borrowers. Note that this would leave the bank with the same profitability levels as the brick and mortar bank, but without the higher growth opportunities envisioned by advocates of the Internet model.

[7]Thomson Financial Company (2000h).

[8]Thomson Financial Company (2000g) (based on data from Coldwell Banker and Forrester Research).

[9]The sources are Thomson Financial Company (2000d) (based on data from FinanCenter.com) and Forbes Inc. (2000).

[10]Thomson Financial Company (2000f).

[11]Thomson Financial Company (2000i) (based on data from First Manhattan Consulting Group).

[12]Thomson Financial Company (2000b).

[13]Thomson Financial Company (2000e) (based on data from AdRelevance).

[14]Thomson Financial Company (2000a).

[15]Thomson Financial Company (2000c).

[16]The contents of this informal FDIC database are not in the public domain. There are other lists of Internet banks in the public domain, but the institutions on these lists vary because there is no standard definition of an "Internet" bank. A public source that uses a very broad definition is *Online Banking Report* (www.onlinebankingreport.com), which lists over 1,500 "true Internet banks and credit unions."

[17]In the preliminary tests leading up to this article, I estimated a similar set of regressions using a larger set of nine banks and thrifts that were chosen using a less stringent set of filters. The results from those preliminary regressions were quite similar to the results reported here in table 2.

[18]I also estimated regressions using the inverse of bank age in place of the natural log of bank age. The basic results were unchanged in those regression tests.

[19]The 16 excluded banks and thrifts from the informal FDIC database are not included in either of these two samples.

[20]These significance tests were constructed using White's estimator for the standard errors.

[21]The call reports do not separate quarterly expenses on physical plant from quarterly expenses on equipment.

[22]Thomson Financial Company (2001).

[23]Thomson Financial Company (2000b).

REFERENCES

DeYoung, Robert, 2000, "Mergers and the changing landscape of the commercial banking, Part 2," *Chicago Fed Letter*, Federal Reserve Bank of Chicago, No. 150, February.

_____, 1999, "The birth, growth, and life or death of newly chartered commercial banks," *Economic Perspectives*, Federal Reserve Bank of Chicago, Vol. 23, No. 3, pp.18–35.

DeYoung, Robert, and Iftekhar Hasan, 1998, "The performance of de novo commercial banks: A profit efficiency approach," *Journal of Banking and Finance*, Vol. 22, May, pp. 565–587.

Economist Newspaper Limited, The, 2000, "Branching out," *The Economist: A Survey of Online Finance*, May 20, pp. 19–23.

Financial Times Limited, The, 2000, "Dutch group to launch Internet bank in US," *The Financial Times*, July 23, p. 1.

Forbes Inc., 2000, "Online banking's goodies," *Forbes*, June 12, p. 366.

Furst, Karen, William W. Lang, and Daniel E. Nolle, 2000, "Who offers Internet banking," *Quarterly Journal*, Office of the Comptroller of the Currency, Vol. 19, No. 1, June, pp. 1–21.

Hunter, William C., and Aruna Srinivasan, 1990, "Determinants of de novo bank performance," *Economic Review*, Federal Reserve Bank of Atlanta, pp. 14–25.

McGraw-Hill Companies Inc., The, 2000, "Online banks can't go it alone," *Business Week*, July 31, p. 86.

Nathan, Luxman, 1999, "Community banks are going online," *Communities and Banking*, Federal Reserve Bank of Boston, Fall, No. 27, pp. 2–8.

Rosen, Kenneth T., and Amanda L. Howard, 2000, "E-retail: Gold rush or fool's gold?," *California Management Review*, Spring.

Sullivan, Richard J., 2000, "How has the adoption of Internet banking affected performance and risk at banks?" *Financial Industry Perspectives*, Federal Reserve Bank of Kansas City, December, pp. 1–16.

Thomson Financial Company, 2001, "Online banking: First Internet Bank of Indiana cuts jobs in a quest for profit," *American Banker*, January 5, p. 11.

_____, 2000a, "Dimon: EBPP a 'killer app'," *American Banker*, October 13, p. 1.

_____, 2000b, "No branches for us, say 3 Internet banks," *American Banker*, September 21, p. 1.

_____, 2000c, "OTS finding Web banks a regulatory handful," *American Banker*, September 21, p. 1.

_____, 2000d, "Web-only banks' CD rates pressure bigger online rivals," *American Banker*, September 1, p. 1.

_____, 2000e, "Dollars moving only slowly to Web advertisements," *American Banker*, June 28, p. 10A.

_____, 2000f, "Wingspan whips up teaser rate of 10% checking account interest," *American Banker*, April 1, p.1.

_____, 2000g, "Studies: People unwilling to close loans on-line," *American Banker*, March 10, p. 11.

_____, 2000h, "Greenspan: Be temperate with new merger powers," *American Banker*, March 9, p. 6.

_____, 2000i, "Bank Web sites draw bargain hunters, study says," *American Banker*, March 4, p. 14.

Thomson Financial Media, 2000, "A future for bricks and mortar," *U.S. Banker*, October, pp. 76–80.