Bank Risk of Failure and the Too-Big-to-Fail Policy

Huberto M. Ennis and H. S. Malek

There seems to be a perception among participants in U.S. financial markets that if a large banking organization were to get in trouble, the government would, under most circumstances, intervene to prevent its failure (or limit the losses to uninsured creditors upon failure). This possibility of a government bailout is commonly referred to as the "too-big-to-fail" policy. The idea behind this belief is that, in general, policymakers will be inclined to bail out institutions which are considered to be of "systemic" importance; that is, institutions whose potential failure could threaten the stability of the entire financial system.

The expectation of contingent bailouts tends to create efficiency costs in the economy. In general, a bank tends to become larger and riskier if its uninsured creditors believe that they will benefit from too-big-to-fail (TBTF) coverage. In this article we provide a formal discussion to clarify the origin of these distortions and review empirical evidence on the relative importance of these distortions in the U.S. banking system.

The TBTF subject is a timely issue. Stern and Feldman (2004) argue that the problem of TBTF is actually getting worse. They identify the increasing concentration and complexity in banking as the main reason for this deterioration. Although their opinion is certainly not shared by everyone, the mere possibility of such a costly distortion is enough to justify further study of this issue.

The too-big-to-fail terminology sometimes can be misleading. While the systemic importance of an organization tends to be closely related to its size, this is not always the case. For example, a handful of U.S. banks are not

particularly large but are still often perceived as too big to fail because they perform an essential activity in the smooth functioning of financial markets and the payment system. Furthermore, the TBTF problem is not exclusive to banks. Other financial intermediaries like large clearinghouses and significant players in the mortgage securities market are often perceived as too big to fail. In this article, however, we will restrict our focus to traditional banking activities and, for simplicity, will consider size as the main variable associated with the likelihood of being bailed out.

U.S. banks face a complex regulatory environment that guides and modifies their behavior. The perception of a TBTF policy is just one of several features that characterizes this environment. Two other important features tend to interact with TBTF: deposit insurance and the failure-resolution policy.[1]

The Federal Deposit Insurance Corporation (FDIC) is an independent government agency that provides deposit insurance to U.S. banking institutions. The current insurance system protects a depositor's insured funds up to $100,000, including principal and interest. The FDIC administers two insurance funds: the Bank Insurance Fund (BIF), which is dedicated to commercial banks, and the Savings Associations Insurance Fund (SAIF) for the savings and loans banks. Member-banks contribute periodic payments to a common pool, which is then used to finance the insurance liabilities in case of a bank failure. Prior to 1993, all banks paid to the FDIC the same contribution per dollar of deposits. However, since 1993, the contributions are partially based on risk. Under this new system, institutions are grouped into nine risk categories according to their level of capitalization and the rating obtained during supervisory examinations. Banks belonging to the higher risk categories are required to pay higher premiums. The range of premiums is updated semiannually by the FDIC according to the funding needs of the insurance funds. Presently, the premiums range from 0 to 0.27 percent of deposits. Since 92 percent of banks satisfy the requirements for a 0 percent assessment, they do not contribute to the fund. The target size of the fund is 1.25 percent of total insured deposits in the system, and, in case of unexpected financial pressure, the current regulation allows for the fund to draw on a $30 billion line of credit from the U.S. Treasury (to be repaid with future premiums by member banks).

As part of a response to a pronounced crisis in commercial banking resulting in a BIF deficit of $7 billion, Congress passed the FDIC Improvement Act (FDICIA) in December 1991.[2] The Act introduced risk-based premiums and new regulations for bank-failure resolution. The new rules specify a course of action for regulators to enforce adjustments in undercapitalized banks and, in this way, mitigate the potential losses to the fund associated with bank failures. Before FDICIA, the power to close a failing insured bank rested

---

[1] See Hetzel (1991) for a discussion of TBTF and the timely closing of insolvent banks.

[2] For a comprehensive survey of FDICIA, see Benston and Kaufman (1997).

with the chartering authority (either the Comptroller of the Currency or state governments). Nowadays, an institution whose capital ratio falls below 2 percent faces closure by the FDIC if the shortfall is not corrected within 90 days (see Walter [2004] for details). While the regulatory reforms introduced in FDICIA limit the protection of uninsured creditors, Section 141 still considers the possibility of a TBTF bailout. This "systemic risk" exception attempts to increase scrutiny over bank bailouts by requiring that both the Federal Reserve and the Treasury sign off on a rescue.[3]

Evidently, the complex deposit insurance system—in combination with the potential for TBTF coverage—creates an intricate set of incentives that influences the decisions of U.S. banks. In the model we provide to analyze the banks' decision process, banks are competitive and must offer the best possible contract to attract potential creditors. We show that when the deposit insurance system involves premium payments that do not fully reflect risk, banks tend to become riskier to exploit the potential net transfer to their creditors under the contingency of failure. We also study partial coverage and the interaction between deposit insurance and a TBTF policy. In particular, we show that the TBTF policy creates not only a risk distortion but also a size distortion, and that one distortion tends to increase the value of the other (and vice versa), creating a perverse amplification effect.

We model risk in a simple yet useful way. We consider only the risk of failure in the decision of banks. This simplification is appropriate for the study of TBTF, which is linked only to the events in the distribution of outcomes that result in failure. Of course, in general, the risk of failure is a consequence of a set of risky decisions made by banks. These decisions also imply a complex distribution of returns when the bank does not fail. We abstract from this aspect of the risk involved in banking and assume that if the bank does not fail, it has a fixed return R.

Studying the cost and benefits of TBTF bailouts is difficult. Failures of large banks are low-probability events. As a consequence, we do not have sufficient data to fully identify the pattern of behavior (of bankers, policymakers, and creditors) linked to bailouts. Also, the indirect (moral hazard) effect of TBTF on the investment portfolio of banks is difficult to discern. At the same time, the decision to bail out a particular bank depends on a large number of circumstances, and reaching general conclusions based on specific events is not good practice. For example, observing that a relatively important failing bank is not bailed out may help elucidate the position of policymakers

---

[3] Regulators often argue that even if a troubled financial institution is not closed, this does not mean that all its major claimants are protected from losses. In general, the regulators of a troubled institution might have its management removed and its existing equity extinguished. Also, sometimes significant (partial) losses might be imposed on uninsured creditors and counterparties (Greenspan 2000). Clearly, all these instruments will contribute to limit the distortions created by the perception of a TBTF policy.

with respect to the TBTF policy. However, just one situation is probably not enough evidence to conclude that TBTF is not a problem. A different bank, in different situations, may actually be bailed out. In other words, it may be useful to think about the bailout event as probabilistic, which is the approach that we take in this article. In the next section, we present a model where the probability of a TBTF bailout is strictly between zero and one (for a relevant set of bank sizes), and (on this range) such a probability is increasing in the size of the bank.

In the second section of this article, we revisit some empirical evidence first presented by Boyd and Gertler (1994), who studied the relationship between bank performance and asset size in the United States and concluded that the evidence indicates the emergence of a TBTF problem in the late 1980s. We extend that analysis to the period 1991–2003, revealing that the patterns justifying Boyd and Gertler's concerns are no longer in the data. We provide some interpretations for this change.

It is important to point out that we are not discussing why a TBTF policy may be in place. Rather, we assume that there is a TBTF policy and then identify its potential effects on the size and risk decisions of banks. This assumption simplifies the exposition and allows us to focus exclusively on the distortions introduced by TBTF. But the simplification does not come without cost. In particular, we do not discuss two important issues related to the existence of TBTF bailouts: the potential benefits of avoiding spillovers and bank runs and the time inconsistency problem faced by policymakers. We refer the interested reader to the excellent discussion in Chapter 2 of Stern and Feldman (2004). However, we would like to stress here that we consider the study of those issues essential for a full understanding of the TBTF problem.

The remainder of the article is organized as follows. In Section 1 we present a simple model of the size and investment decision of competitive banks and study this decision under different explicit and implicit deposit insurance schemes. The model allows us to identify the distortions that the different possible schemes create on the level of risk taken by banks and the size of their operations. In Section 2, we review empirical evidence aimed at determining if the U.S. banking system functions under the perceptions of an implicit TBTF government insurance scheme. The last section provides concluding remarks.

1.   A SIMPLE MODEL

Consider an economy with a large number of banks and a large number of agents that play the role of potential depositors. Each agent has 300 units of funds available, and they can either deposit some (or all) of their funds at a bank or invest them in a safe asset which provides a gross rate of return, given by r. The banks make risky investments and may fail with a certain

probability, π. In the case that a bank does not fail, depositors get R units per unit deposited at that bank. We assume that R can take values in the interval $[0, \overline{R}]$, where $\overline{R}$ is an upper bound of the set of possible gross returns on bank deposits. Furthermore, we assume that banks can charge a fee, F, to each depositor.

Assume that the probability of bank failure, π, is increasing in R. This assumption captures the idea that taking higher risks is necessary to obtain higher returns. For simplicity, we assume that $\pi(R)$ is linear in R with slope, a. When the bank fails, we assume that no resources are left at the bank to pay depositors. In other words, without government intervention, depositors will get zero from the bank in case of failure. For reasons that will become clear shortly, we assume that r and $\overline{R}$ satisfy the following conditions:

$$r = \frac{1}{4a}, \text{ and } \frac{3}{4a} < \overline{R} < \frac{1}{a}.$$

Also, for simplicity we will assume that depositors can deposit an amount, x, of funds in the bank, where x can take one of three possible values: 50, 100, or 300. Furthermore, all depositors want to have at least 50 units deposited at the bank. We do not model explicitly the reasons for this minimum deposit, but the idea is that all agents wish to have at least some bank balances for settlement of "essential" payments.[4]

Finally, banks can choose their size. Let ξ be the proportion of the total population of agents making a deposit in a particular bank. To make the choice of ξ interesting, we assume the cost c per depositor of running a bank is convex in ξ with a minimum at ξ^o. The idea behind this assumption is that an optimal size of operation for banks exists and is associated with the size ξ^o. Running a bank that is too small (i.e., smaller than ξ^o) increases the operational cost per depositor; and running a bank that is too large (i.e., larger than ξ^o) also increases the cost.

We assume that banks compete to attract depositors. In equilibrium, banks earn zero profits and choose R and ξ so as to make the expected payoff to a depositor as high as possible. If a bank were not to follow such a strategy, some other bank would arrange its choices of R and ξ in order to attract all the depositors from the first bank. This equilibrium concept is standard in the banking literature. All agents and banks are identical, and in equilibrium they behave symmetrically. As a consequence, the equilibrium value of ξ is a good proxy for the size of the representative bank.

We now study different banking arrangements and their effects on the risk of failure and the size chosen by the banks.

---

[4] The discreteness in the size of deposits is assumed only for the sake of simplicity. It allows us to capture the main reasons driving agents' decisions without complicating the calculations.

Figure 1 Optimal Return-Risk Combination
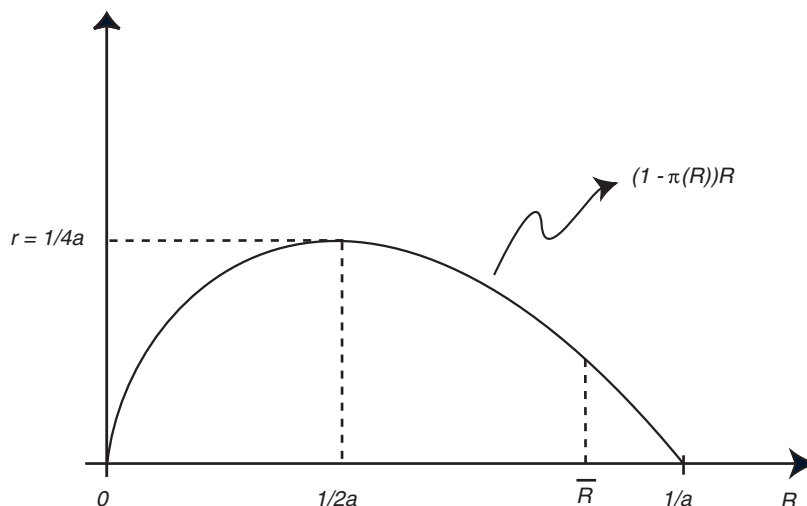
Notes: In a laissez-faire system, banks set the return $R^L = 1/2a$, which maximizes the expected return per unit deposited (net of fees) given by $(1 - \pi(R))R$.

Laissez-faire System

Consider first the case of a laissez-faire banking system—that is, one without any government intervention. The laissez-faire equilibrium provides an important benchmark for our evaluation of alternative explicit and implicit deposit insurance systems in the following subsections. Under laissez faire, the expected payoff to a depositor is given by

$$(1 - \pi(R)) \, xR + \pi(R)0 - F,$$

where the equilibrium fee, F, will cover the operational costs per depositor, $c(\xi)$. Let us call R^L and ξ^L the laissez-faire equilibrium values of R and ξ. These values maximize the payoff to depositors and, hence, must satisfy the following necessary and sufficient conditions:

$$\frac{d\pi}{dR} x R^L - \left(1 - \pi(R^L)\right) x = 0,$$

and

$$\frac{dc(\xi^L)}{d\xi} = 0,$$

which imply that R^L equals $1/2a$ and ξ^L equals ξ^o. Note that R^L is the value of R that maximizes the payoff, $(1 - \pi(R))xR$ (see Figure 1).

To complete the analysis, we need to determine if the depositors would find it beneficial to deposit in these banks any amount in excess of 50 units. If an agent deposits the minimum 50 units of its funds in a bank and the remaining 250 in the safe asset, then its expected payoff will be given by

$$\left(1 - \pi(R^L)\right) 50R^L - c(\xi^L) + 250r.$$

We need to compare this alternative with that of depositing any other feasible amount, x, greater than 50 (in particular, $x = 100$ or 300). The net benefit of increasing the amount deposited at a bank to $x > 50$ is given by

$$\left(1 - \pi(R^L)\right)(x - 50)R^L - (x - 50)r.$$

Recall that we assumed that $r = 1/4a$. Then, since $(1 - \pi(R^L))R^L = 1/4a$, we obtain that the net benefit is zero, and for any amount in excess of 50, depositors would be indifferent between making an investment or a deposit.

It is important to note that the model presented here has no inherent interaction between size and risk, even though in reality there may be reasons to believe that a bank's size and risk of failure can be associated in some fundamental way. This simplification is useful because it allows us to concentrate on the interactions between size and risk that may originate in specific banking policies.
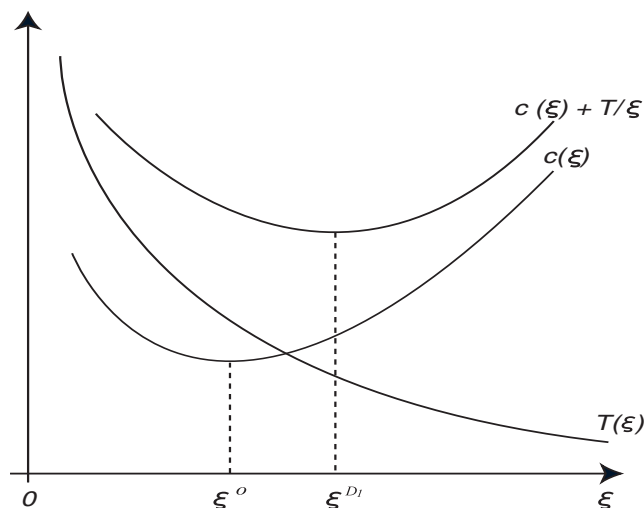
Deposit Insurance

We will consider four different deposit insurance systems. The systems differ from one another in the structure of premiums and the coverage that they provide.

We start with a deposit insurance system that provides full coverage of losses and in which banks pay to the insurance fund a lump-sum fixed premium, T, independent of bank size. While this kind of fixed premium seems unrealistic, such an extreme assumption is useful to illustrate how misalignments in the premium structure can create size distortions. In this simple model, designing the right premium structure to avoid this kind of size distortion is straightforward, and we describe such a structure below.

Under this system, banks choose the values of R and ξ that solve the following problem:

$$\max_{R,\xi} (1 - \pi(R))\, xR + \pi(R)xR - F, \tag{1}$$

where $F = c(\xi) + T/\xi$. Let us denote the solution to this problem with (R^{D1}, ξ^{D1}). It is then clear that under full coverage the banks will choose $R^{D1} = \overline{R}$, the maximum value of the possible (risky) returns. Recall that the probability of failure of a bank is increasing in R and, hence, by setting R^{D1} equal to $\overline{R}$, banks will be indirectly maximizing the probability of failure. Banks follow this strategy because the insurance premium that a bank pays

Figure 2 Optimal Bank Size

Notes: The size of a bank that minimizes the per-depositor cost of operation $c(\xi)$ is given by ξ^o. When the bank is paying a fixed lump-sum premium T to the deposit insurance fund, it will increase its size to ξ^{D1}.

does not depend on the risk taken by the bank, and, furthermore, full insurance coverage implies that neither the bank nor its depositors face any downside from selecting higher levels of risk.

With respect to the equilibrium size of banks, we have that the value of ξ that maximizes the objective in problem (1) solves the following (first order) condition:

$$\frac{dc(\xi)}{d\xi} = \frac{T}{\xi^2} > 0.$$

It is then straightforward to see that $\xi^{D1} > \xi^o$. Recall that ξ^o was the size of the bank that minimizes the cost of operation $c(\xi)$. Here, however, by becoming large, the bank reduces the per capita cost of deposit insurance for depositors. Hence, the optimal size of the bank is larger than the one that minimizes operational costs. In other words, the lump-sum premium distorts the optimal-size decision by banks (see Figure 2).

To avoid the size distortion, the deposit insurance fund could make the premium, T, dependent on the size of the bank. This structure of premiums makes sense to the extent that, for a given level of risk, larger banks will impose higher costs to the insurance fund. Suppose, for example, that $T = b\xi$. Then,

it is straightforward to show that the bank will choose to be of the optimal size ξ^o.

While this kind of premium scheme will solve the size distortion, there still remains the risk distortion. In fact, under this structure of premiums, banks would still choose to maximize the probability of failure. Of course, the fund could implement alternative regulations to limit the amount of risk taken by banks. For example, it could restrict the types of investments allowed to banks so that the bank would not be able to choose a level of R as high as $\overline{R}$. However, this model is too simple to study these more sophisticated regulations.

One other possibility would be to make the premium contingent not only on size, but also on risk. In fact, by choosing T to equal $\pi(R)xR\xi$, the insurance fund would give banks the necessary incentives to choose $R = R^L$, the same rate that banks would choose under laissez faire. In general, though, precisely assessing the risk taken by banks is difficult, and we can expect that the observed premium payments will not fully correct the risk distortion introduced by deposit insurance (Prescott 2002). For simplicity, in what follows we will assume the extreme case in which the premium only corrects the size distortion and is given by $T = b\xi$.

The last feature of deposit insurance that we wish to study is partial coverage. To be precise, suppose that in the case when a bank fails, the deposit insurance fund covers only up to 100 units of funds per depositor. Then, banks will choose the risk and size that solve the following problem:

$$\max_{R,\xi} (1 - \pi(R))\, xR + \pi(R)\min\{x, 100\}R - c(\xi) - b.$$

Let us call the solution (R^{D2}, ξ^{D2}). Since the total premium, T, is increasing with size, there will not be a size distortion in the decision of banks and therefore, ξ^{D2} equals ξ^o. With respect to the level of risk-return, R, the choice of banks will depend on whether the typical depositor has more or less than 100 units deposited at a bank.

For $x \leq 100$, the insurance provided is effectively full insurance. Then, as we saw before in the full-coverage case, depositors would find it most beneficial if banks maximize the risk-return combination.

Only if depositors have $x > 100$ does the partial coverage provide incentives to reduce risk at banks. In the banking literature, these depositors have been named "uninsured depositors." This terminology is not completely precise to the extent that all depositors receive insurance for the funds below the 100 limit. However, the terminology does convey the idea that these depositors are the ones susceptible to the risk of failure of their bank.

The coverage limit helps reduce the risk distortion but in general will not be enough to fully correct it. To see this, suppose that the typical depositor deposits 300 units of funds at the bank. Then, the bank will choose a level of

R that solves the following first order condition:

$$\frac{d\pi}{dR}R^A - \left(1 - \pi\left(R^A\right)\right) = \frac{1}{3}\left[\frac{d\pi}{dR}R^A + \pi\left(R^A\right)\right] > 0. \qquad (2)$$

Recall that $R^L = 1/2a$ is the value of R that makes the left-hand side of equation (2) equal to zero (see Figure 1). Hence, since the right-hand side of this equation is positive, R^A must be greater than R^L, and the risk distortion is still present. For most cases, R^A will be smaller than $\overline{R}$, and we can say that, in the presence of uninsured depositors, the insurance limit can partially resolve the risk distortion introduced by deposit insurance.[5]

From the previous discussion we can then conclude that R^{D2} is either equal to R^A (if $x > 100$) or to $\overline{R}$ (if $x \leq 100$) and, hence, greater than R^L in either case. To determine the actual value that R^{D2} will take in equilibrium, we need to establish whether the typical depositor would be willing to deposit more than 100 units in a bank. The payoff from depositing more than 100 units is given by

$$\left(1 - \pi(R^{D2})\right)300R^{D2} + \pi(R^{D2})100R^{D2} - c\left(\xi^o\right) - b.$$

Alternatively, suppose that the agent deposits only 100 units at a bank and invests the rest in the safe investment with return r. In this case, the payoff is given by

$$100R^{D2} - c\left(\xi^o\right) - b + 200r.$$

Since $(1-\pi(R^{D2}))R^{D2} < r$ (see Figure 1), it is easy to see that depositing 100 units at a bank and the rest in the safe investment is the best strategy. Another alternative for the agent is to hold three deposit accounts of 100, each one at a different bank. This alternative will dominate both the 300-unit deposit and the alternative involving the safe asset described above. In fact, if a depositor can open any number of these accounts, then the 100-unit limit would never be relevant. It should be said, though, that opening accounts in several different banks involves transaction costs that are not being explicitly modeled here. One possibility for reducing these transaction costs is for the depositor to delegate this activity to a broker. However, in the U.S. system, brokered deposits are subject to regulations enforced by the supervisory agencies. For the sake of simplicity, in what follows we will assume that depositors can only have one bank account in the system.

Summarizing, the typical depositor in this banking system will have only deposits for 100 units or less, and banks will choose $R^{D2} = \overline{R}$—that is, the rate of return that corresponds to the highest feasible risk of failure. In other words, even though partial coverage has the potential for limiting risk-taking

---

[5] If x is greater than 100 (but less than 300), R^A may still equal $\overline{R}$. Here, then, the discreteness of the size of deposits simplifies calculations.

behavior by banks, it also creates incentives for depositors to stay below the limit, thereby undermining the disciplining mechanism.

Too Big to Fail

Suppose now that with probability, p, the bank is bailed out upon failure. To show that the bailout is spurred by the fear that a large organization's failure will disrupt the entire financial sector, we assume that p is increasing in the bank's size, ξ. This is a simple way to capture the too-big-to-fail policy. We still consider the case where a deposit insurance system with partial coverage is in place. Hence, the too-big-to-fail policy has consequences for the payoff of only those depositors with deposits above the limit. The payoff to depositors in the event of a bank failure is given by the function:

$$\Phi(R, \xi) \equiv \min\{x, 100\}R + p(\xi)\max\{0, x - 100\}R.$$

Competitive banks choose the values of R and ξ that solve the following problem:

$$\max_{R, \xi} (1 - \pi(R)) xR + \pi(R)\Phi(R, \xi) - c(\xi) - b, \qquad (3)$$

where the objective function is the expected payoff to the representative depositor. Let us call the solution to this problem (R^T, ξ^T). It is useful to start with the extreme case of banks that are so large that the probability of a bailout is unity (i.e., $p(\xi^T) = 1$). Then, problem (3) reduces to the full-coverage deposit insurance system we studied at the beginning of the previous subsection, and banks in equilibrium chose $R^T = \overline{R}$, which implies that the risk of failure would be maximized.

In the general case when the probability of bailout, p, is between zero and one, the solution to problem (3) suggests some interesting insights about the distortions introduced by the too-big-to-fail policy. This policy is relevant only for those agents that have uninsured deposits. Suppose then, that the typical depositor of the bank has $x > 100$. The partial derivatives of the payoff function, Φ, are given by:

$$\Phi_R(R, \xi) \equiv \frac{\partial \Phi(R, \xi)}{\partial R} = 100 + p(\xi)(x - 100)$$

and

$$\Phi_\xi(R, \xi) \equiv \frac{\partial \Phi(R, \xi)}{\partial \xi} = \frac{dp(\xi)}{d\xi}(x - 100)R;$$

and the solution (R^T, ξ^T) to the bank problem must satisfy the following first order conditions:

$$\frac{d\pi}{dR}xR^T - \left(1 - \pi(R^T)\right)x = \left[\frac{d\pi}{dR}\Phi(R^T, \xi^T) + \pi(R^T)\Phi_R(R^T, \xi^T)\right], \quad (4)$$

and

$$\frac{dc(\xi^T)}{d\xi} = \pi(R)\Phi_\xi(R, \xi). \tag{5}$$

Since $\Phi_R(R, \xi)$ and $\Phi_\xi(R, \xi)$ are both positive, $R^T > R^L$ and $\xi^T > \xi^o$. In other words, the too-big-to-fail policy induces banks to become larger and riskier than in a laissez-faire system. Furthermore, by comparing expression (4) with expression (2) (in the previous subsection) we see that, in general, R^T will be greater than R^A, which was the return chosen by a bank with uninsured depositors under no contingent-bailout policy.

One remaining question is whether depositors would want to deposit funds in excess of 100 in a banking system like the one we study in this subsection. The (net of fees) payoff to an agent depositing 300 units of funds at the bank is given by

$$\left(1 - \pi(R^T)\right) 300R^T + \pi(R^T)\left(1 + 2p(\xi)\right) 100R^T.$$

Comparing this payoff with the payoff from depositing only 100 units of funds (and the rest at the safe interest rate, r) we see that the difference is given by

$$\left[\left(1 - \pi(R^T)\right) R^T - r\right] 200 + \pi(R^T)p(\xi) 200R^T. \tag{6}$$

Since R^T will generally be greater than R^L, we know that the first term in expression (6) is negative. However, the second term is positive, and for a large enough bailout-probability, p, it would compensate for the loss in the first term. It is then possible in this banking system for agents to find it beneficial to deposit all 300 units of funds at the bank.

Another interesting observation that results from expressions (4) and (5) is the interaction that exists between size and risk under the too-big-to-fail policy. Note that the right-hand side of expression (4) is increasing in p (which, in turn, is increasing in ξ). Then, the larger the bank, the larger the value of R the bank will wish to implement. Similarly, the right-hand side of expression (5) is increasing in R, and, hence, the higher the risk taken by a bank, the higher the incentives to increase its size. The reason for this complementarity between size and risk is that riskier banks are more likely to benefit from the possibility of bailouts (they are more likely to fail). Therefore, those banks are the ones that would like to increase the bailout probability, p, an objective that can be pursued by increasing the size of the banking organization.

This interaction captures the idea of a "virtuous circle" induced by an autonomous reduction on the probability of bailout (Stern and Feldman 2004, 21). Suppose that the appointment of a "conservative" regulator reduces the value of p for all values of ξ. This reduction in p will reduce the value R^T chosen by banks according to expression (4), which, in turn, will reduce the equilibrium size, ξ^T. A smaller ξ^T further lowers the risk taken by banks, reducing the failure probability and creating a virtuous circle that significantly reduces the likelihood of failure and bailout events.

As we have seen, the existence of a TBTF policy has two effects: it creates a size distortion in the banking industry, and it tends to accentuate the risk distortion that was already present under deposit insurance (i.e., R^T is greater than R^A). A commonly proposed policy to limit the effects of perceived implicit government guarantees is to limit the size of banks so that the probability p is equal to zero. Suppose, for example, that there is a bank size, ξ_p, such that $p(\xi) = 0$ for all $\xi \leq \xi_p$. Then, by limiting banks to be no larger than ξ_p, the government can eliminate the risk distortion originated in the TBTF perception. In general, however, limiting the size of banks will increase operational cost unless $\xi^o \leq \xi_p$. When the value of ξ is restrained by regulation to be below ξ_p, the value of R that banks choose solves a problem equivalent to the last problem studied in the previous subsection. It is somewhat ironic then that, in our model, limiting the size of banks to be smaller than ξ_p implies that banks will choose $R^{D2} = \overline{R}$, which could increase the riskiness in banking.

Another possible policy to limit the size of these distortions is to implement a system of "coinsurance" (Feldman and Stern 2004). The idea is that whenever a bank fails and gets bailed out, uninsured depositors will obtain only a proportion $\theta < 1$ of their deposits in excess of the insurance limit. The payoff in the event of a bank failure is now given by the function,

$$\Phi(R, \xi, \theta) = \min\{x, 100\}R + p(\xi)\,\theta \max\{0, x - 100\}R.$$

The bank problem is the same as in expression (3) but where $\Phi(R, \xi, \theta)$ replaces $\Phi(R, \xi)$. The solution to this problem will be a function of the parameter θ. Let us call such a solution (R^C, ξ^C). It is easy to see that for $\theta = 1$ we have $(R^C, \xi^C) = (R^T, \xi^T)$. However, for θ lower than unity, R^C is lower than R^T, and ξ^C is lower than ξ^T.[6] In other words, the coinsurance system reduces the incentives for banks to become bigger and riskier under a TBTF policy.

The deposit insurance premium, T, could be designed to reduce the size distortion induced by the TBTF policy. In particular, if the premium per unit deposited, b, is made increasing in the size of the bank, banks will have less incentive to become large, which, in turn, would limit the influence of the TBTF perception. The idea behind this strategy is important and can be restated in more general terms: whenever the TBTF problem is present, designing the structure of the deposit insurance premium to be neutral with respect to size (that is, in our model, $T = b\xi$) may not be optimal.

Finally, another way to control the risk-taking behavior of banks in the presence of a TBTF distortion is to directly limit the bank's activities via supervisory exams. In our simple model, this strategy amounts to reducing

---

[6] The solution R^C is lower than R^T as long as the coinsurance system does not make the optimal size of deposits equal 100 units or less.

the acceptable values of R that the bank may choose, or in other words, to lower the upper bound on returns, $\overline{R}$, a parameter in the model.[7]

2.   THE ELUSIVE EVIDENCE

Boyd and Gertler (1994) look back at the banking troubles of the 1980s and find that "large banks were mainly responsible for the unusually poor performance of the overall industry" (p. 2). They attribute this feature of the data to the combination of two main factors: deregulation and too-big-to-fail. In particular, they argue that after the collapse of Continental Illinois Bank in 1984, it became clear that large banks were subject to a TBTF policy.[8] Using a panel of U.S. bank data for the period 1984–1991 they conclude that a robust negative correlation exists between size and performance and suggest that this correlation may be indicative of an increased perception of a TBTF subsidy.

The idea behind this strategy is that banks that are riskier ex ante, are also more likely to perform poorly ex post. Moreover, riskier banks, as a consequence of having more spread distribution of returns, tend to have a higher probability of failure.[9] Combining these two hypotheses implies that poorly performing banks have a higher probability of failure. Then Boyd and Gertler (1994, 15) postulate that "by examining ex post returns we can get some feel for the outer tails of the distributions." As we saw in the previous section, under the influence of a TBTF policy, banks will tend to increase the probability of failure. It is, of course, not obvious that increasing the probability of failure is always associated with an increase in the overall risk of the bank. Similarly, riskier banks do not always perform poorly, on average, relative to less risky banks. However, data limitations suggest that, in principle, the proposed link between risk, poor performance, and likelihood of failure may be a useful working strategy.

Boyd and Gertler use the decreasing trend in U.S. bank profitability during the 1980s as a starting point for their study. Specifically, they stress the fact that profitability was significantly below its 1970s average by the late 1980s. Our Figure 3 illustrates this fact. We plot the annual net income as a percentage of total assets for U.S. insured commercial banks. We divide banks in two groups, those with more than $10 billion in total assets (large banks) and those with less than that amount. The decline in profitability during the 1980s is

---

[7] Our model does not allow us to study another form of controlling the risk-taking behavior by banks: capital requirements. See Prescott (2001) for a good formal introduction to the subject.

[8] In September 1984, the Comptroller of the Currency testified to the U.S. Congress that 11 bank holding companies were too big to fail (see O'Hara and Shaw 1990).

[9] In the previous section we did not allow for general distributions of returns which are an integral part of the interpretation for the evidence in this section. The link between the distribution of returns and the probability of failure is a technical issue that is not essential for understanding the incentives distortion introduced by TBTF, which was the main subject of the previous section.

Figure 3 Bank Performance and Size

U.S. Insured Commercial Bank Annual Return on Average Assets
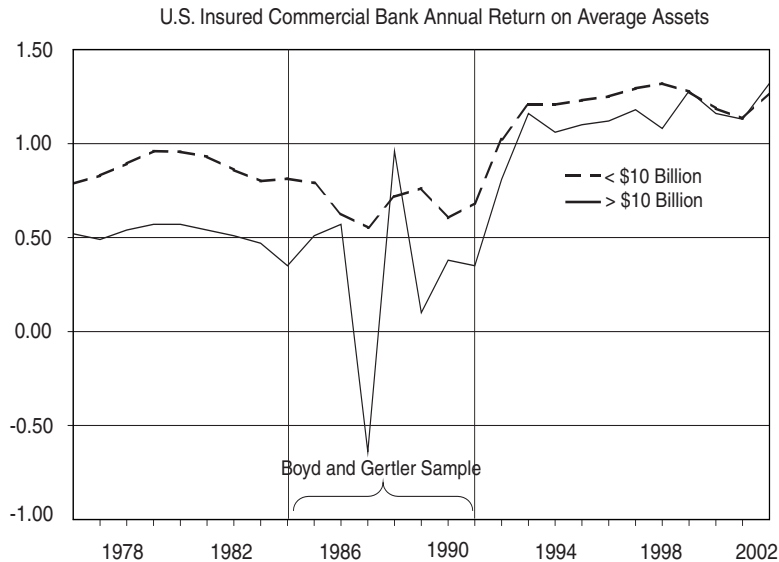
Source: FDIC Quarterly Banking Profile.

common for the two groups.[10] However, it is clear from the figure that large banks experienced an especially turbulent time during the second half of the 1980s. What is even more interesting is that after 1991, bank profitability recovered across the board to levels above those in the 1970s, staying fairly stable since then.

In summary, Figure 3 puts in perspective the sample period used by Boyd and Gertler and may cast some doubt on the robustness of their results. For this reason, we extend Boyd and Gertler's empirical analysis to include the data from 1992 to 2003.

Figure 4 presents the average return on assets for banks of different sizes. One of the main motivations for Boyd and Gertler's conclusions is the hump-

---

[10] Keeley (1990) argues that banks became riskier during the 1980s as a consequence of a generalized decrease in franchise value across the industry. Franchise value can help control risk-taking behavior by banks because bank owners fear losing this value upon failure. The evolution of banks' franchise value is an important determinant of their behavior, but, unfortunately, we will not have much to say about it in this article. See Demsetz, Saidenberg, and Strahan (1996) for further discussion of this issue.

Figure 4 Return on Assets and Size
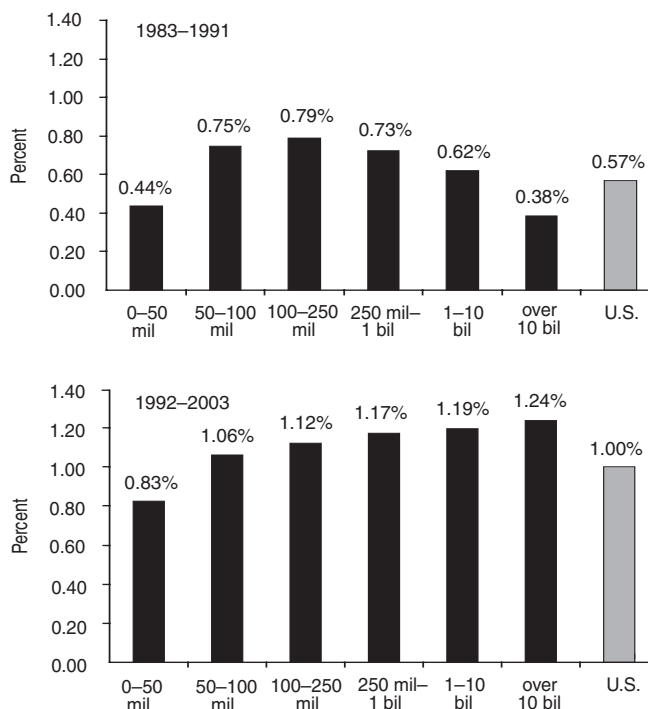
Notes: We use data for all insured commercial U.S. banks (except credit card banks). To construct return on assets we divide annual net income by total assets. We consider each annual observation for each bank as the basic data entry in the calculation of averages across sizes (i.e., we do not take time averages for each bank). The total number of observations is around 120,000 for 1983–1991 and 110,000 for 1992–2003.

Sources: Report of Condition and Income Data (Call Report); Federal Reserve Bank of Chicago Web page.

shaped pattern of the first panel of Figure 4. Large banks performed relatively poorly during that period, presumably because of the improper pricing of risk induced by the TBTF distortion.[11] However, the second panel shows that in the period after 1991, the return on assets experienced by banks was, in fact, a monotone-increasing function of size. There are two competing explanations

---

[11] Banks with less that $50 million in assets also performed worse than the middle-sized banks. This pattern may be the consequence of the inability of small banks to exploit economies of scale.

for this change in pattern. Perhaps the hump-shaped pattern observed in the 1983–1991 period was the result of a special event at the end of the 1980s that hit hardest the performance of large banks. In this case, the pattern may not be related to a TBTF perception, and when no special event took place in the 1992–2003 period, the pattern disappeared. The other interpretation for the change in the pattern is that after 1991, changes on banking regulation and other policies induced a decrease in the likelihood of TBTF bailouts.

Before discussing the relevance of these two alternative explanations, we follow Boyd and Gertler's methodology and check whether the change in pattern just discussed is robust to controlling for regional effects. The idea behind this exercise is that the performance of banks may be driven by regional economic shocks. For example, if most of the large banks in the country are in a region that experienced an especially unfavorable shock during the period under study, then it is possible to find that, on average, mid-sized banks outperformed large banks just as a consequence of "location" effects.

While this type of "robustness" check may have been important for the 1983–1991 period, there are a priori reasons to believe that the adjustment is bound to be insignificant for the sample period of 1992–2003. First, several large banks today have nationwide operations and, hence, are less exposed to business fluctuations in specific regions. Second, looking at bank performance across regions during the 1992–2003 period does not reveal any clear regional disparities. The situation was not the same in the sample period studied by Boyd and Gertler, when the west-central region of the South and the west-central region of the Midwest experienced severe regional banking shocks.

Let us denote by D_j^r, a dummy variable indicating that a bank is headquartered in region j; by D_k^s, a dummy variable indicating that a bank belongs to size class k; and by x_{ijk}, a time-average value of bank return on assets. We run the following regression to obtain estimates of size effects on performance, controlling for a region:

$$x_{ijk} = a_j D_j^r + b_k D_k^s + \varepsilon_{ijk}.$$

This is equation (1) in Boyd and Gertler (1994). We construct two sets of time-average return on assets, one for the period 1984–1991, and one for the period 1992–2003. Table 1 presents the estimated values of b_k for both sample periods. We can see that the hump-shaped pattern in the 1984–1991 period is robust to regional adjustments. Similarly, after 1991, bank performance becomes a monotone-increasing function of size even after controlling for regional factors.[12]

Boyd and Gertler (1994) also investigate the relationship between time-average loan chargeoffs and bank size. They find that for the period

---

[12] We also run a regression where we allowed the coefficients b_k to vary across regions (equation 2 in Boyd and Gertler [1994]). The results were very similar.

Table 1 Size-Performance, Controlling for Regional Effects

| Time Period	b_k Coefficient for Each Asset Size Class					
	$k = 1$	$k = 2$	$k = 3$	$k = 4$	$k = 5$	$k = 6$
1984–1991	-0.0009	0.0009	0.00	-0.0014	-0.0021	-0.0062
	(-2.78)	(2.36)	–	(-2.40)	(-2.52)	(-2.60)
1992–2003	-0.0019	-0.0002	0.00	0.0010	0.0018	0.0077
	(-7.90)	(-0.86)	–	(1.01)	(1.92)	(1.37)

Notes: We use data for all insured commercial U.S. banks (except credit card banks).
To construct return on assets we divide annual net income by total assets. We take time
averages for each bank that existed in the base year, 1983, across the eight years in
the period 1984–1991. For the period 1992–2003 we follow the same procedure using
1991 as the base year. The k size classes are the same as in Figure 4. The number
of observations in the regression for the period 1984–1991 is 13,964 and for the period
1992–2003 is 11,230. The values in parentheses are t-values.

Sources: Report of Condition and Income Data (Call Report); Federal Reserve Bank of
Chicago Web page.

1984–1991, the relationship has a U-shape. In other words, small and large
banks tend to have higher chargeoffs to assets than medium-sized banks. This
finding is taken as further evidence of the possible effects of the TBTF policy.
In Figure 5 we reproduce Boyd and Gertler's result and provide the same data
for the period 1992–2003. Once again, there has been a change in pattern
between these two periods. For the data after 1991, the relationship between
chargeoff and bank size is monotone increasing. Larger banks tend to have,
on average, riskier loans.

Another variable that can be used as a proxy for bank risk is the variance
of return on assets (see, for example, Berger and Mester [2003]). Boyd and
Gertler (1994) do not compute this variable for their period. We provide this
calculation for both subperiods in Figure 6. It is interesting to see that the
variance of (annual) return on assets has significantly decreased after 1991
for all size classes. Also, the variability of return on assets does not show a
monotonic relationship with the asset size of banks. In the 1984–1991 period,
banks with over $10 billion in assets had a variance of return on assets that
was higher than that for the previous size class (those banks with $1 to $10
billion in assets). However, this pattern is lost after 1991.

The data studied here for the period 1992–2003 are consistent with a
banking system that is not necessarily distorted by the perception of potential
TBTF subsidies. Under this interpretation, larger banks give riskier loans
(higher chargeoffs to loans) but have a larger size of operations that allows
them to better diversify those risks (lower variance on return on assets). A large

Figure 5 Chargeoffs to Loans and Size
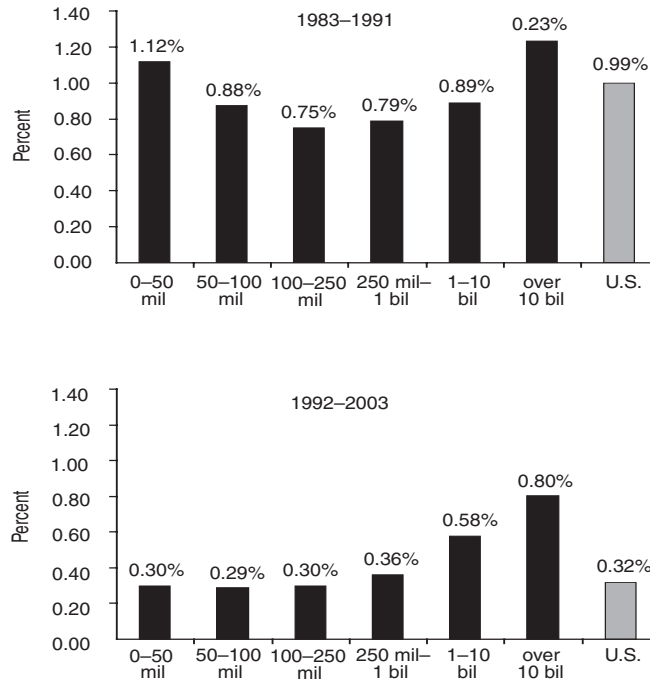
Notes: We use data for all U.S. insured commercial banks (except credit card banks). To construct chargeoffs to loans we divide annual net chargeoffs by total loans and leases. We consider each annual observation for each bank as the basic data entry in the calculation of averages across sizes (i.e., we do not take time averages for each bank). The total number of observations is around 110,000 for 1983–1991 and 120,000 for 1992–2003.

Sources: Report of Condition and Income Data (Call Report); Federal Reserve Bank of Chicago Web page.

size of operations may imply some extra cost, but the riskier loans also allow these large banks to obtain higher average returns. Ennis (2001) provides a model of banking where this kind of logic is formally studied.

At the same time, the data for 1984 to 1991 seem perhaps more consistent with the existence of a TBTF distortion. The natural question to ask then is, could it be that changes in banking regulation at the beginning of the nineties have solved the TBTF problem? The effectiveness of FDICIA in controlling TBTF has been a matter of controversy among experts. For example, Stern and Feldman (2004) argue that the post-FDICIA regime is not much different from the pre-FDICIA regime and, as a consequence, if TBTF was a problem

Figure 6 Variance of Return on Assets and Size
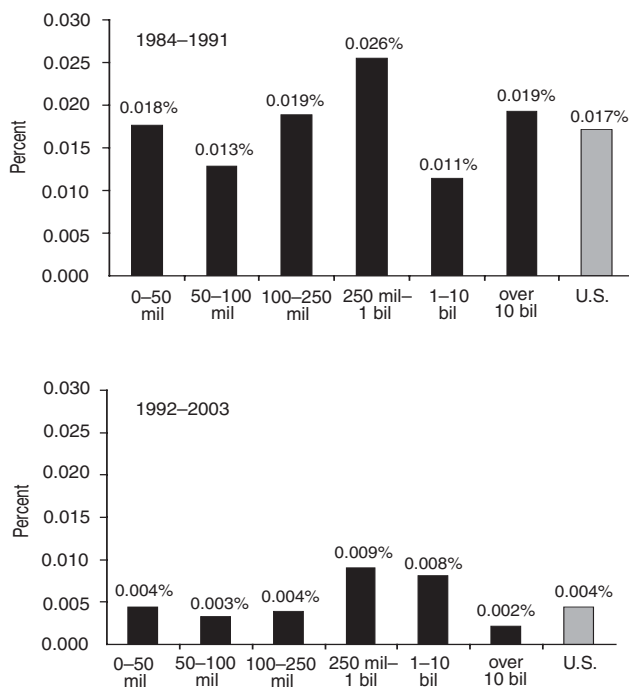
Notes: To construct the averages for 1984–1991, we compute the variance of annual return on assets for each bank existing in 1983 and organize the banks in size classes according to the average amount of assets they owned in that year. For the period 1992–2003 we use 1991 as the base year. We only use banks for which we have at least three annual observations. The total number of observations is around 13,000 for 1983–1991 and 10,000 for 1992–2003.

Sources: Report of Condition and Income Data (Call Report); Federal Reserve Bank of Chicago Web page.

before 1991, it is still a problem afterwards.[13]   No large bank has been in trouble since the enactment of FDICIA, and it is difficult to determine the ultimate effect of the change in the regulation.

    An alternative explanation for the change in the patterns observed in the data is that the late 1980s was an unusual period. The idea is that large banks

_____

[13] According to Stern and Feldman (2004), FDICIA made explicit a set of procedures that were implicit before 1991. They judge those procedures insufficient to stop TBTF bailouts. For a more favorable view of the reforms in FDICIA, see Benston and Kaufman (1997).

specialized in certain activities (loans to less-developed countries and large commercial real-estate loans) that performed poorly during the second half of the 1980s. Boyd and Gertler (1994) discuss this interpretation but are very skeptical about its merits. They argue that medium-sized banks participated in the same set of activities as large banks but performed much better. Then, Boyd and Gertler conclude that the reason why medium-sized banks outperformed large banks is that large banks were less risk-sensitive as a consequence of the TBTF distortion.

It is interesting to note that some of the findings in this article are in accordance with the findings in the empirical literature that investigates the viability of exploiting market discipline in banking regulation. A significant portion of this literature studies the extent to which bond yield spreads reflect the financial conditions of banks. Most of this work finds that, while during the early to mid-1980s the relationship between bond yield and bank risk was weak (presumably due to implicit government guarantees), during the late 1980s and the 1990s the relationship became much stronger (see, for example, Flannery and Sorescu [1996] and the review in Flannery and Nikolova [2004]). These findings have been taken as evidence that the TBTF problem has been mitigated since the beginning of the 1990s. However, Morgan and Stiroh (2002), using data for the 1993 to 1998 period, still find that the behavior of bond spreads for those banks most likely to be subject to a TBTF policy was significantly different from that of other smaller banks and other debt-issuing corporations.

The purpose of this section was to provide some evidence to test the view that TBTF may be a latent problem in the U.S. banking system. Overall, however, it seems that looking at the data on performance across size classes does not allow any definite conclusion.

There are, of course, other ways to look for evidence of TBTF distortions. One methodology is to look at the effect of announcements about the existence of a TBTF policy over the equity value of banks. For example, O'Hara and Shaw (1990) used this strategy. They found that in September 1984, after the Comptroller of the Currency testified before Congress that certain banks were "too big to fail," the equity value of those banks increased significantly (relative to the rest of the industry).

Another way to approach the question is to study the effect of mergers on the value of the claims issued by the merging organizations. Benston, Hunter, and Wall (1995) study the prices that were bid to acquire target banks in the early to mid-1980s. They find little evidence of a TBTF-subsidy-enhancing motive in a sample of U.S. bank mergers during that period. On the other hand, Penas and Unal (2004) study changes in the return on nonconvertible bonds issued by merging banks during the 1991–1997 period. They find a significant increase in bondholder returns after a merger and that the increase is non-monotone with respect to the asset size of the bank. In particular, holders of

bonds issued by mid-sized banks (especially those that after merging became relatively large within the system) are the ones that benefit the most from a merger. The authors attribute this pattern to a TBTF perception in the market for bonds.

Yet another methodology is to look at the cost-savings implications of increases in bank size. Some empirical studies have found that economies of scale exhaust at fairly modest bank sizes ($ 200 million in assets). If this is the case, then the existence of larger banks may be the consequence of a TBTF distortion. However, the empirical literature on economies of scale in banking is far from a consensus. Wheelock and Wilson (2001), for example, find that economies of scale do not exhaust until banks have at least $500 million in assets and do not find evidence of significant diseconomies of scale for larger banks (see also Hughes, Mester, and Moon 2001).[14]

3. CONCLUSION

In this article we have formally identified some basic principles that guide the behavior of banks interacting under the coverage of a government safety net, and in particular, a TBTF policy. We also studied some empirical regularities of U.S. bank performance across size classes and evaluated the extent to which they provide evidence of a significant size and risk distortion originated in a perceived TBTF subsidy.

Our conclusion is a word of caution. While, in principle, the cost of the TBTF distortions could be large, the available evidence is far from conclusive. This is an important reality to acknowledge. Several policy measures are currently being considered to reduce the potential distortions induced by TBTF (Stern and Feldman 2004). To the extent that some of these policies create new inefficiencies in the economy (by, for example, limiting the behavior of banks in particular ways), we need to be able to assess better their potential benefits. In this respect, then, it seems necessary, if not urgent, to improve our knowledge of the actual magnitude of the TBTF problem in the U.S. economy. Our reading on this matter is that the available evidence is very preliminary and in no way definitive.

---

[14] Assessments by credit rating agencies provide another source of useful information. Stern and Feldman (2004, Chapter 4), for example, present extensive evidence suggesting that credit rating agencies are in agreement on the existence of a TBTF policy for large banks.

REFERENCES

Benston, George J. and George G. Kaufman. 1997. "FDICIA after Five Years." *Journal of Economic Perspectives* 11 (Summer): 139–58.

——————, William C. Hunter, and Larry D. Wall. 1995. "Motivations for Bank Mergers and Acquisitions: Enhancing the Deposit Insurance Put Option versus Earnings Diversification." *Journal of Money, Credit, and Banking* 27 (August): 777–88.

Berger, Allen N. and Loretta J. Mester. 2003. "Explaining the Dramatic Changes in Performance of U.S. Banks: Technological Change, Deregulation, and Dynamic Changes in Competition." *Journal of Financial Intermediation* 12: 57–95.

Boyd, John H. and Mark Gertler. 1994. "The Role of Large Banks in the Recent U.S. Banking Crisis." Federal Reserve Bank of Minneapolis *Quarterly Review* (Winter): 319–68.

Demsetz, Rebecca S., Marc R. Saidenberg, and Philip E. Strahan. 1996. "Banks with Something to Lose: The Disciplinary Role of Franchise Value." Federal Reserve Bank of New York *Economic Policy Review* (October).

Ennis, Huberto M. 2001. "On the Size Distribution of Banks." Federal Reserve Bank of Richmond *Economic Quarterly* 87 (Fall): 1–25.

Flannery, Mark J., and Sorin M. Sorescu. 1996. "Evidence of Bank Market Discipline in Subordinated Debenture Yields: 1983–1991." *Journal of Finance* 51: 1347–77.

Flannery, Mark J., and Stansilava Nikolova. 2004. "Market Discipline of U.S. Financial Firms: Recent Evidence and Research Issues." In *Market Discipline across Countries and Industries*, ed. William C. Hunter, George G. Kaufman, Claudio Borio, and Kostas Tsatsaronis. Cambridge: The MIT Press: 87–100.

Greenspan, Alan. 2000. "Banking Evolution." Speech at the 36th Annual Conference on Bank Structure and Competition of the Federal Reserve Bank of Chicago, Chicago, Illinois.

Hetzel, Robert L. 1991. "Too Big to Fail: Origins, Consequences, and Outlook." Federal Reserve Bank of Richmond *Economic Review* 77 (November/December): 3–15.

Hughes, Joseph P., Loretta J. Mester, and Choon-Geol Moon. 2001. "Are Scale Economies in Banking Elusive or Illusive? Evidence Obtained by

Incorporating Capital Structure and Risk-Taking into Models of Bank Production.” *Journal of Banking and Finance* 25 (December): 2169–208.

Keeley, Michael C. 1990. “Deposit Insurance, Risk, and Market Power in Banking.” *American Economic Review* 80 (December): 1183–1200.

Morgan, Donald P., and Kevin J. Stiroh. 2002. “Too Big to Fail and Market Discipline of Banks: A Cross-Sector Test.” Federal Reserve Bank of New York. Mimeo.

O’Hara, Maureen and Wayne Shaw. 1990. “Deposit Insurance and Wealth Effects: The Value of Being ‘Too Big to Fail’.” *Journal of Finance* 45 (December): 1587–601.

Penas, María Fabiana and Haluk Unal. 2004. “Gains in Bank Mergers: Evidence from the Bond Markets.” Forthcoming *Journal of Financial Economics*.

Prescott, Edward S. 2001. “Regulating Bank Capital Structure to Control Risk.” Federal Reserve Bank of Richmond *Economic Quarterly* 87 (Summer): 35–52.

Prescott, Edward S. 2002. “Can Risk-Based Deposit Insurance Premiums Control Moral Hazard?” Federal Reserve Bank of Richmond *Economic Quarterly* 88 (Spring): 87–100.

Stern, Gary H. and Ron J. Feldman. 2004. *Too Big to Fail: The Hazards of Bank Bailouts*. Brookings Institution Press. Washington, D.C.

Walter, John R. “Closing Troubled Banks: How the Process Works.” Federal Reserve Bank of Richmond *Economic Quarterly* 90 (Winter): 51–68.

Wheelock, David C. and Paul W. Wilson. 2001. “New Evidence on Returns to Scale and Product Mix Among U.S. Commercial Banks.” *Journal of Monetary Economics* 47: 653–74.