Why Is There Debt?

Jeffrey M. Lacker

The striking feature of debt contracts is that over a wide range of circumstances the payment is fixed and invariant, although occasionally, as in a default, less than the full payment is made. In this article I offer a simple explanation for why such arrangements are widely observed. The explanation relies on recent advances in the theory of financial arrangements under imperfect information. I will argue that the opportunity for borrowers to hide their future resources sharply constrains the degree to which loan repayment can be made contingent on the borrower's future resources.

From one point of view it is not obvious that the ubiquity of debt contracts is a puzzle. A borrower acquires a sum of money today that will be repaid in the future, along with an additional payment, called interest. The interest rate is the price for the temporary use of resources. It seems perfectly natural that this amount is predetermined.

Modern economic theory has taught us to view matters differently. When a loan is made, the lender acquires a contingent claim, a promise by the borrower to pay an amount that can depend in any arbitrary, prespecified way on future events. Many familiar contracts actually do involve future payments that are contingent in significant ways. Insurance contracts are promises to make a payment contingent on some particular future loss. Partnership agreements and profit-sharing arrangements make future payments contingent on the uncertain profits of the firm. Traded securities such as stocks, bonds, options, and related derivative products have returns that are highly sensitive to future events. But in a debt contract, the payment is generally noncontingent in that the amount does not vary with future circumstances, such as the borrower's wealth. Of course a debt contract is contingent to the extent that the lender does not receive full repayment if the borrower defaults. But although default is an important feature of the arrangement, it occurs relatively rarely.

Finding plausible models in which people agree to debt contracts, although they are allowed to agree to any possible contingent repayment schedule, has proven surprisingly difficult. In fact, in many models people are much better off with a contingent contract than they are with a debt contract. In Section I, I present a simple, two-agent model that shows why standard economic theory predicts that contracts generally should be contingent. The model also serves as a useful starting point for further analysis.

In Section II, I present a model in which the borrower and lender agree to a loan repayment that is noncontingent because the borrower can conceal future resources. Section III points out that this model is deficient because nothing resembling default ever occurs, and then argues that collateral, broadly defined, is an important omitted feature of the model. Next, in Section IV, I present a model in which an implicitly collateralized debt contract, with occasional default, is the chosen arrangement. Three brief sections conclude the paper: Section V surveys literature that has addressed the same question; Section VI briefly discusses some policy implications; and Section VII summarizes the explanation offered here for the ubiquity of debt contracts and notes two remaining unsolved puzzles.

I. A SIMPLE MODEL OF CONTINGENT CLAIMS

To begin, consider an economy with only two people: a borrower and a lender. The economy lasts for just two time periods; call them periods 1 and 2. Imagine that the two people are farmers, and that the two periods represent the spring and fall of a given year. In the spring the lender harvests a crop: wheat, say. The lender's land produces no crop in the fall. The borrower's land produces no crop in the spring, but will produce a crop in the fall. Both agents would like to consume wheat in both the spring and
the fall. To do so, the borrower must obtain a loan of wheat in the spring, to be repaid from the proceeds of the fall harvest. For simplicity, I ignore the use of wheat in planting, and assume that the crops have already been planted. I also ignore the possibility of storing wheat from the spring to the fall; allowing storage would not affect the results. No other goods are available to these two agents.

To make the contingent nature of the contract of interest, some random event has to occur between spring and fall. I assume that in the spring the amount of the borrower's fall wheat harvest is uncertain. In the fall the harvest is realized, and both agents learn the exact value of the harvest. The payment contract is contingent if it depends on the amount of the borrower's crop. Other sources of uncertainty could have been considered—shocks to the preferences of the two agents for example—but in many ways, uncertainty concerning the borrower's ex post resources is the archetypal setting for financial contracting. If the borrower is a wage earner, for example, future income or employment might be uncertain. If the borrower is an individual entrepreneur, future returns from the venture might be uncertain. If the borrower is an incorporated firm, future liquid resources of the firm might be uncertain.

To proceed, the borrower's harvest in the fall is \( \theta \), and can take on one of \( N \) values: \( \theta_1, \theta_2, \ldots, \theta_N \), where these are ordered so that \( 0 < \theta_1 < \theta_2 < \ldots < \theta_N \). In the spring, both people believe that the probability that \( \theta \) takes on the value \( \theta_n \) is \( \pi_n \), where \( \pi_n > 0 \) for \( n = 1, 2, \ldots, N \), and \( \sum_{n=1}^{N} \pi_n = 1 \). The lender has a harvest of \( e_1 \) in the spring. The lender makes a loan advance of \( q \) in the spring, and receives a payment of \( y_n \) in the fall if the harvest is \( \theta_n \). In the spring the lender's consumption is \( e_1 - q \), while the borrower's spring consumption is \( q \). When the borrower's harvest is \( \theta_n \), the lender's fall consumption is \( y_n \) and the borrower's fall consumption is \( \theta_n - y_n \). A contract is a set of payments \( \{ q, y_1, y_2, \ldots, y_N \} \), and these completely determine the consumptions of the two agents.

I assume that the borrower evaluates the contract \( \{ q, y_1, y_2, \ldots, y_N \} \) according to the expected utility function

\[
U_B(q) + \beta \sum_{n=1}^{N} U_B(\theta_n - y_n) \pi_n,
\]

where \( \beta \) is a discount factor satisfying \( 0 < \beta < 1 \). This is the ex ante expected utility of the borrower in the spring. Similarly, the lender evaluates the contract according to the expected utility function

\[
U_L(e_1 - q) + \beta \sum_{n=1}^{N} U_L(y_n) \pi_n.
\]

The within-period utility functions \( u_B \) and \( u_L \) are assumed to be strictly increasing, continuous, concave and smoothly differentiable.

Contracts cannot require payments that exceed the available resources. Stated formally, contracts must satisfy the following resource feasibility constraints:

\[
q \geq 0,
\]

\[
e_1 \geq q,
\]

\[
y_n \geq 0, \quad n = 1, 2, \ldots, N,
\]

\[
\theta_n \geq y_n, \quad n = 1, 2, \ldots, N.
\]

Optimal Contracts

To obtain predictions in this simple environment about the arrangements that the two agents will choose, I restrict attention to optimal contracts. A contract is optimal if it is feasible and no other feasible contract exists that makes one agent better off, in terms of ex ante expected utility, without making the other agent worse off. Because of the simple nature of the environment, an easy way of finding optimal contracts is by maximizing the weighted average of the two agents' utility functions, subject to the resource feasibility constraints. The weights, sometimes called "Pareto weights," are arbitrary positive numbers, and varying their relative size traces out a range of contracts that gives more utility to one agent and less to the other. If a contract is optimal in this environment, then it is the solution to the constrained maximization problem for some Pareto weights, and vice versa.

The programming problem that finds optimal contracts, then, is the following.

Problem 1:

Maximize, by choice of \( q, y_1, y_2, \ldots, y_N \),

\[
\lambda_B \left[ U_B(q) + \beta \sum_{n=1}^{N} U_B(\theta_n - y_n) \pi_n \right]
\]

\[
+ \lambda_L \left[ U_L(e_1 - q) + \beta \sum_{n=1}^{N} U_L(y_n) \pi_n \right]
\]

subject to the resource feasibility constraints (3)-(6).
The Pareto weights $\lambda_B$ and $\lambda_L$ are arbitrary positive constants.

To show the properties of optimal contracts, I examine the set of first-order conditions that are necessary and sufficient for a contract to be a solution to Problem 1. If both the borrower and the lender are enjoying positive consumption in the fall for a given state $\theta_n$, so that $0 < y_n < \theta_n$, then the first-order condition for $y_n$ is

$$\lambda_L u'_L(y_n) = \lambda_B u'_B(\theta_n - y_n).$$

Condition (7) requires that the marginal utility of the lender's fall consumption, scaled by $\lambda_L$, must equal the marginal utility of the borrower's fall consumption, scaled by $\lambda_B$. This condition determines $y_n$ in a manner illustrated in Figure 1. The width of the box in Figure 1 is $\theta_n$, the realized harvest outcome to be divided between the two agents. The payment $y_n$ is measured horizontally from left to right, and the lender's marginal utility, measured vertically, falls as $y_n$ rises. Similarly, the consumption of the borrower is measured horizontally from right to left, and the borrower's marginal utility rises as $y_n$ rises. The optimality condition (7) dictates that the payment is determined by the intersection of the two weighted marginal utilities. Identical conditions apply for every other possible harvest outcome; the horizontal dimensions of the box vary with $\theta_n$, but otherwise the analysis is the same.

The Nonoptimality of Debt Contracts

I can now demonstrate that in this simple environment, the payment varies positively with the harvest, and a debt contract will be optimal only under special circumstances. Consider Figure 2, in which the determination of the payments is illustrated, just as in Figure 1, but for two possible harvest outcomes, $\theta_n$ and $\theta_m$, where $m > n$. The box for the larger harvest, $\theta_m$, is drawn with the same left edge, so that the origin from which the payments are measured does not move. Consequently, the lender's marginal utility schedule is the same for both harvest outcomes. The origin from which the borrower's consumption is measured shifts to the right, because $\theta_m > \theta_n$, so the borrower's marginal utility shifts to the right. If both marginal utility schedules slope down, the point of intersection moves down and to the right going from harvest $\theta_n$ to $\theta_m$. Therefore, the payment $y_m$ for the larger harvest is larger than $y_n$, the payment for the smaller harvest.

Under a debt contract there is a set of harvest outcomes over which the payment made by the borrower in the fall is a constant. It is easy to see what is required for such an arrangement to satisfy the optimality conditions. The lender's optimal consumption must remain the same, and this requires that the borrower be risk neutral, meaning that the borrower's utility is linear, not strictly concave. In this case a shift to the right in the borrower's marginal utility leaves the point of intersection, and thus the payment to the lender, unchanged. For a debt contract to be optimal in this environment, the borrower must
of the classical model, at least when there are no imperfections in the availability of information. Thus, the ubiquity of debt contracts is puzzling, at least from the viewpoint of classical general equilibrium models.

II. DEBT CONTRACTS IN A MODEL WITH LIMITED INFORMATION

Apparently, then, to explain debt contracts one must depart from the assumptions of the classical model. In the example above, both the lender and the borrower are fully aware of the realized value of the borrower’s harvest; in other words, there is perfect information. Suppose instead that the lender is uncertain of the borrower’s harvest at the time the payment must be made in the fall. The lender might be forced to rely on the borrower’s report about the harvest, especially if there is no independent information available to the lender. In this case the payment might have to be noncontingent, because otherwise the borrower would have reason to make a misleading and self-serving report.

To explore this notion, I now modify the model described above by assuming that the borrower is capable of hiding any amount of the harvest. The hidden crop can be consumed secretly, and hiding is itself costless. The remaining crop, the part not hidden, is displayed to the lender. This is less stringent than assuming that the lender is incapable of observing the harvest at all (pure private information), but still implies that the amount displayed provides only a lower bound on the actual amount of the harvest.

This appears to be a fairly realistic imperfection in information. Often a borrower can divert resources for private benefit that would otherwise be available to repay an obligation. A consumer, for example, can spend freely on current consumption and then default on debts. Similarly, a firm’s managers can divert...
resources in a variety of ways, through direct and indirect managerial compensation, wasteful investment, exploitation of discretion over accounting choices, or favored treatment of particular creditor classes. Often a lender has no direct knowledge of a borrower's total resources and thus must rely on the borrower's own financial statements. At the same time, it often seems as if lenders have or can obtain some information about a borrower. If a borrower claims to have a certain quantity of resources, the lender can ask the borrower for proof of his bank balance or other readily verifiable assets. The borrower is incapable of proving that he does not control additional assets; he can display less than his true resources but not more. This informational imperfection is consistent with the observation that parties to financial arrangements are often observed exchanging information at relatively little apparent cost.

Incentive Constraints

Although the implication of this informational assumption is straightforward, I display the results more formally, since in a more complicated setting examined later the intuition will be less clear and the formalities more important. The borrower now has a choice to make in the fall: if the harvest is \( \theta_n \), the borrower can display an amount \( \theta_m \), where \( \theta_m \) can take on the values \( \theta_1, \theta_2, \ldots, \theta_n \). If the borrower displays \( \theta_n \) when the harvest is \( \theta_n \), nothing is being hidden, while if the borrower displays less than \( \theta_n \), an amount \( \theta_n - \theta_m \) is being hidden and consumed without the knowledge of the lender.

As before, a contract, \( \{q, y_1, y_2, \ldots, y_N\} \), specifies the spring payment to the borrower, \( q \), and the fall payment from the borrower, \( y_n \), contingent on the harvest. Also as before, a contract must satisfy the resource feasibility constraints (3)-(6). Now I impose the further condition, called incentive feasibility, that the borrower never has an incentive to hide any of the harvest. If the borrower does not hide any harvest when the harvest is \( \theta_n \), his utility is \( u_B(\theta_n - y_n) \). If the borrower displays \( \theta_m \) when the harvest is \( \theta_n \), hiding the amount \( \theta_n - \theta_m \), his utility is \( u_B(\theta_n - y_m) \). For the borrower to have no positive incentive to hide harvest, it must be true that

\[
 u_B(\theta_n - y_n) \geq u_B(\theta_n - y_m) \text{.}
\]

Therefore, the set of incentive feasibility constraints are

\[
 u_B(\theta_n - y_n) \geq u_B(\theta_n - y_m) \text{ for } n=2, \ldots, N, \text{ and for } m=1, \ldots, n-1. \quad (8)
\]

The incentive feasibility constraints stated here can be derived from a deeper formulation that allows contracts that might give the borrower incentive to hide some of the harvest. It can be shown, however, that the results of any arbitrary contract can be replicated by a contract that satisfies the incentive feasibility constraints.

The incentive feasibility constraints can be simplified. To use (8) in finding an optimal contract, the utility of not hiding any harvest must be compared to the utility of displaying any amount less than \( \theta_n \). It turns out that if the constraint for harvest \( \theta_n \) is satisfied for \( m=n-1 \), then the constraints are satisfied for all \( m < n \). As a result (8) can be reduced to

\[
 u_B(\theta_n - y_n) \geq u_B(\theta_n - y_{n-1}) \text{ for } n=2, \ldots, N. \quad (9)
\]

In other words, the utility of telling the truth only needs to be compared to the temptation of displaying the next smallest possible harvest \( \theta_{n-1} \).

An immediate implication of (9) is that a contract is incentive feasible if and only if \( y_n \) is constant or decreasing as \( \theta_n \) increases. For any given harvest outcome, the borrower can make the payment corresponding to that harvest or to any smaller harvest; a given payment is feasible for the corresponding harvest and for any larger harvest outcome. Because

---

7 There are many constraints on these abilities, of course. Boards of directors, or other representatives of creditors, monitor some aspects of managerial choice. Discretion over accounting is limited in myriad ways by accounting standards, legal requirements for certification by outside auditors and the like. The fraudulent conveyance provision of the bankruptcy code allows the bankruptcy court to "unwind" distributions to creditors made 90 days or less prior to filing. Nonetheless, managers retain considerable discretion and can often take actions for personal benefit to the detriment of creditors.

8 In contrast, such readily available observation is difficult to reconcile with pure private information, where it is assumed that the information is completely unavailable to the lender.

9 The proof is in Appendix A and uses "The Revelation Principle." The terminology is due to Myerson (1979). For an exposition of the Revelation Principle in similar settings see Townsend (1988). A warning is in order here, however; the display of harvest is an "action" and not a "message." In the present setting the distinction is immaterial, but in more general settings in which actions involve real costs, the distinction is important. See Laacker and Weinberg (1989), p. 1350.

10 To prove this note that (9) implies that for \( n=2,3,\ldots,N \), \( y_n \leq y_{n-1} \), so \( y_n \leq y_m \) for \( m=1,2,\ldots,n-1 \). This in turn implies that (8) is satisfied. The property that only immediately adjacent incentive constraints need to be checked arises in a wide variety of settings.
hiding the harvest is costless, the borrower will make the smallest possible payment and will display the corresponding amount of harvest. Thus the borrower will never have to make a larger payment for a larger harvest than for a smaller harvest.

Optimality of a Debt Contract

I am now in a position to show that something resembling a debt contract is optimal in this model. As before, a programming problem is solved, but with the addition now of the incentive-feasibility constraints (9). Specifically, I solve

Problem 2:
Maximize, by choice of \( q, y_1, \ldots, y_N \),

\[
\lambda_B \left[ u_B(q) + \beta \sum_{n=1}^{N} u_B(\theta_n - y_n) \pi_n \right] \\
+ \lambda_L \left[ u_L(c^1 - q) + \beta \sum_{n=1}^{N} u_L(y_n) \pi_n \right]
\]

subject to the resource feasibility constraints (3)-(6), and the incentive feasibility constraints (9).

To see why a completely noncontingent contract is optimal, compare the incentive feasibility constraints with the contract that was optimal in Section I. First, recall that incentive-feasible contracts can never be increasing with respect to the harvest \( \theta_n \), only constant or decreasing, because if \( y_n > y_{n-1} \) and the harvest is \( \theta_n \), then the borrower would lie in order to make a smaller payment. Second, recall that in Section I, without incentive constraints, risk-sharing alone determined the optimal contract and it had a strictly increasing payment schedule. But such a contract is not incentive feasible, and would always give the borrower an incentive to claim that the smallest possible harvest outcome had occurred. Among the set of contracts that are nonincreasing—and thus incentive feasible—the constant payment schedule is the one that is closest to the optimal contract from Section I in the sense that it has the largest slope. Thus a contract with a constant payment schedule is optimal.\(^{11}\)

A noteworthy feature of this model is that the range of contracts available to the two agents is severely restricted. Because the payment, call it \( R \), is constant across harvests, the payment can never be greater than the smallest possible harvest \( (R \leq \theta_n) \); otherwise the fixed payment is not feasible for small harvests. This is potentially a quite severe restriction, since the smallest possible harvest could be very different from the expected realization, and could imply a maximum loan repayment that is very small. In this situation, the borrower might be left desiring more credit than he can obtain via any incentive-feasible contract. To see this, one can combine the first-order conditions from Problem 2 to obtain the following equation linking the expected intertemporal marginal rates of substitution of the two agents:

\[
\beta \sum_{n=1}^{N} u_B(\theta_n - y_n) \pi_n = \sum_{n=1}^{N} u_L(c^1 - q) + \frac{\mu_1}{\lambda_B u_B(q)}
\]

The left side of (10) is the borrower's expected intertemporal marginal rate of substitution, the expected value of the ratio of marginal utility in period 2 to marginal utility in period 1. Similarly, the first term on the right side of (10) is the lender's expected intertemporal marginal rate of substitution. The second term contains \( \mu_1 \), the Lagrange multiplier on the constraint, \( y_1 - \theta_1 \leq 0 \). If \( \mu_1 > 0 \), this constraint is binding, and the borrower's expected marginal intertemporal rate of substitution is strictly less than the lender's. This means that the borrower would like to obtain more period 1 consumption in exchange for period 2 consumption, but cannot do so in any feasible contract. In this sense, one might describe such a borrower as constrained or rationed.\(^{12}\)

---

\(^{11}\) To complete a proof, I need to show that a contract with \( y_n < y_{n-1} \) for some \( n \) cannot be optimal. Suppose that \( y_n < y_{n-1} \), for some particular \( n \). Then the incentive constraint that relates \( y_n \) and \( y_{n-1} \) is not binding, and \( \phi_n = 0 \), where \( \phi_n \) is the Lagrange multiplier on the \( n \)th constraint in (9). Therefore, from the first-order conditions we have

\[
\beta u_L(c^1 - q) - \lambda_B u_B(\theta_n - y_n) \pi_n = 0
\]

and

\[
\beta u_L(c^1 - q) - \lambda_B u_B(\theta_n - y_n) \pi_n = 0
\]


---

\(^{12}\) This feature of the model is an exact restatement of an argument made by Irving Fisher (1930, pp. 210-11). He noted that a borrower's collateral will limit the amount he can borrow, and "In consequence of this limitation upon his borrowing power, the borrower may not succeed in modifying his income stream sufficiently to bring his rate of preference for present over future income down to agreement with the rate or rates of interest ruling in the market" (p. 211).
As a model of debt contracts, there is an obvious deficiency in the optimal contract just described: nothing ever occurs that resembles a default, a state in which something less than the fixed payment, \( R \), is made. The optimal contract is a constant payment, and thus is perfectly risk-free. Many debt contracts are virtually risk-free, but it seems that in most debt arrangements there appears to be at least a remote possibility of default. This possibility is an important feature of the contractual arrangement, even if the probability is small, because a borrower will always be tempted to simulate default. Apparently the environment described above is incompatible with payments that are almost always a fixed amount but occasionally are less.\(^{13}\) Can economic environments be found that display such contracts?

To guide a search for such environments, let us begin by asking what happens when an individual defaults on an actual debt contract. First, and obviously, the borrower pays less at a given date than was stipulated under the original contract.\(^{14}\) This is not all that happens, however. If the loan is explicitly collateralized the borrower may be forced to surrender the collateral. Under an "unsecured" obligation the borrower may agree to a restructured payment schedule, promising to make future payments in lieu of the current payment. Sometimes the borrower is forced into legal bankruptcy proceedings, which often involve liquidating assets and using the proceeds to repay claims. For managers of incorporated businesses, bankruptcy involves at least temporary surrender of some control rights associated with the business, because the bankruptcy court or the trustee can assume substantial power over management decisions. The bankrupt that is not liquidated often must agree to a set of restructured claims, as in Chapter 11 reorganizations or Chapter 13 "wage-earner plans" under the U.S. Bankruptcy Code. These outcomes are obviously interrelated, but the salient point is that usually the borrower surrenders something distinct from the originally promised payment: either money at a later date or some other asset or right.

One is thus led to consider contracts in a multiple-good environment, one in which the borrower has more than one good to sacrifice. In such a setting, a contract specifies a payment schedule for each good the borrower will later have available. In principle each payment schedule can be an arbitrary function of future circumstances. A debt contract in this setting is a set of payment schedules with special properties. First, in almost all circumstances fixed noncontingent amounts of a set of goods are paid, fixed sums of money at prespecified dates, for example. Second, in some circumstances less of these goods is paid and positive quantities of some other goods are surrendered, where "other goods" must be interpreted broadly to include legal claims and the like, as described above.

Under what circumstances would such a debt contract be optimal? Let us abstract from multiperiod debt contracts that stipulate a series of payments and focus attention on an obligation to make a single specified payment at a single future date. Consider first the set of states in which the borrower pays the fixed amount of the good—call it money for now. Perhaps the noncontingent nature of the payment schedule over these states can be motivated in exactly the same way as the noncontingent contract of the previous section; if the borrower could hide resources ex post, the payment schedule would have to be constant to avoid giving the borrower an incentive to hide.

Now consider the default states in which some other goods are paid. The fixed payment might be larger than the smallest possible amount of money the borrower could have available. When the borrower does not have enough money to make the fixed payment, the actual payment is obviously limited by the amount of money the borrower has. What is to keep the borrower from always feigning these outcomes so as to make the smallest possible payment? With other goods available, the contract could require that if the borrower makes less than the fixed money payment, then some other goods of equal value to the borrower must be transferred to the lender as well. The other goods sacrificed are enough to dissuade the borrower from pretending to be destitute. Thus the transfer of other "collateral" goods ensures that the borrower will not falsely claim to be unable to make the full payment.

Such an arrangement could expand dramatically the set of feasible contracts available to the borrower and lender. In the environment described in Section III, where no other goods were present, the lowest
possible harvest, $\theta_1$, placed an upper bound on the size of the fixed payment. With other goods available, a contract can be written with a fixed payment of money that is larger than the smallest amount the borrower might possess ex post. The other goods provide a way of relaxing the sharp constraint imposed by the value of the smallest possible harvest outcome in the environment of Section II.

But other puzzles arise in this story. Consider first the set of states in which a noncontingent amount of money is paid. Why, in these circumstances, pay money rather than some other goods, such as those paid in the default-like states? It must be because money, at least in those states, is more valuable to the lender than the other goods, or, equivalently, the other goods are more valuable to the borrower, relative to money, than they are to the lender. This seems like a reasonable condition, one that might be satisfied in many of the circumstances in which debt contracts appear. When a consumer buys a house or a car, say, it is less valuable to the lender, relative to money, than it is to the borrower; the lender would obtain less money by repossessing the collateral and selling it than the borrower would spend to retain it. Consumers quite plausibly could value a good at more than its market price if it is indivisible and consumers only buy one. Similarly, the value to the borrower of all that is forfeited in bankruptcy settlements of various types is usually less than the value of what is received by lenders. Indeed, the difference, regarded as a “deadweight loss,” seems to motivate a wide array of arrangements—both in and out of formal bankruptcy proceedings—designed to minimize this loss.

The other goods serve as collateral. This is most plain in loans explicitly collateralized by physical goods such as land, structures, chattels, automobiles, or inventories. Often a loan is collateralized by financial instruments such as accounts receivable, warehousing receipts or negotiable securities. Many debts are implicitly collateralized, as when income or profits in the more distant future stand behind a promise to make a payment out of income or profits in the near future, or when claims to a portion of the proceeds of liquidation stand behind an unsecured corporate obligation. Even an unsecured creditor can obtain a judgement against a defaulting debtor, allowing the creditor to have the debtor's assets seized to satisfy the claim. While the distinctions between these various means of collateralizing an obligation can be quite important, they are fundamentally similar. Indeed, in almost all instances the nonpayment of a contractual obligation provides the lender with a legal claim, the content of which is jointly determined by the terms of the original contract and the existing body of contract and bankruptcy law. While the resulting claim can have a wide range of characteristics, it provides the borrower with an incentive to make the stipulated payment whenever possible, to “keep his heart right” in the words of a practitioner. The role of collateral is not necessarily to indemnify the lender against potential loss, although it certainly does so to a degree. Rather, collateral is a means of satisfying incentive constraints that ensure voluntary compliance with the terms of the loan agreement.

The main legal distinction between an explicitly collateralized debt and an uncollateralized debt is how the claim stands vis-à-vis third parties such as other creditors or a bankruptcy trustee. For example, under the current U.S. law governing secured transactions the difference between secured and unsecured creditors is minor when there is only one creditor. A creditor with a collateralized debt can obtain the collateral to satisfy the claim, rather than see the collateral added to the pool of assets divided among all of the creditors in bankruptcy. This suggests that the essential role of explicit, as opposed to implicit, collateral is related to multilateral financial arrangements, and that uncollateralized lending has much in common with explicitly collateralized lending.

**IV. COLLATERALIZED DEBT**

In this section I describe a two-good economic environment, and I find conditions under which a collateralized debt contract is optimal for the reasons described above. The environment, an extension of the previous example, captures the essential elements of the argument outlined above.18

---

15 This role was noted by Barro (1976). The quoted practitioner is Chris Carlson, Richmond, VA.

16 One exception is when the collateral is an “exempt asset” under bankruptcy law, and is thus out of reach of any unsecured creditor but can be recovered under a collateralized loan. Exempt assets include the debtor’s “tools of trade,” some of the debtor’s household goods, and an interest in the debtor’s residence. Another exception is when the collateral is an asset that will not pass to the bankruptcy estate, such as the personal assets of the manager of a corporation.

17 Standard terminology in the theoretical finance literature, unfortunately, is that a debt contract like the one described in Section II is “uncollateralized” while a debt contract like the one described below is “collateralized.” The literature treats collateral as if it were exempt assets in a personal bankruptcy, or the personal assets of a manager in a corporate bankruptcy.

18 The model is a simplified version of my 1991 working paper.
The borrower is assumed to have two goods with which to conceivably repay a loan in the fall. One good is the fall harvest, as before, and the other good can be thought of as chattels: durable, portable, personal property such as clothes, furniture or perhaps tools. In a collateralized debt contract, when the harvest is sufficient the borrower pays a fixed, non-contingent amount of the harvest (the payment good) and none of the chattels (the collateral good). When the fall harvest is less than the fixed payment, the entire harvest is given to the lender along with a positive quantity of chattels. For good harvests the borrower does not hide the harvest because some of the chattels would have to be surrendered as well.

To proceed formally, then, let good 1 in each period be the wheat harvest, as in the previous models, and let good 2 be the collateral good, the borrower’s chattels. The borrower is endowed with k units of chattels, and k is known ahead of time to both the borrower and the lender. In the event that the fall harvest is \( \theta_n \), the borrower makes payments of \( y_{1n} \) of good 1 and \( y_{2n} \) of good 2, and consumes \( \theta_n - y_{1n} \) of good 1 and \( k - y_{2n} \) of good 2. As before, the spring loan advance is \( q \), so that the consumption of the borrower is \( q + y_{1n} + y_{2n} \) minus the loan advance. For simplicity, I assume that the borrower derives utility from consumption of chattels only in the fall.

The expected utilities of the two agents are now

\[
U_B(q) + \beta \sum_{n=1}^{N} [U_B(\theta_n - y_{1n}) + V_B(k - y_{2n})] \pi_n
\]

and

\[
U_L(e^\top_1 - q) + \beta \sum_{n=1}^{N} [U_L(y_{1n}) + V_L(y_{2n})] \pi_n.
\]

I have assumed here that both agents have additively separable utility in the fall. The functions \( v_n \) and \( v_L \) are the utilities of the borrower and the lender, respectively, with respect to chattels. I assume that both are continuous, concave and smoothly differentiable. A natural assumption to make is that \( v_B \) is strictly increasing, but \( v_L \) need not be increasing. The function \( v_L \) might be decreasing if the collateral good is worthless to the lender and disposing of it is costly, or if \( y_{2n} \) is viewed as a costly punishment, such as debtor’s prison.

The resource feasibility constraints extend naturally to this case:

\[
q \geq 0,
\]

\[
e_1^\top \geq q,
\]

\[
y_{1n} \geq 0, \quad n = 1, 2, \ldots, N,
\]

\[
\theta_n \geq y_{1n}, \quad n = 1, 2, \ldots, N,
\]

\[
y_{2n} \geq 0, \quad n = 1, 2, \ldots, N,
\]

\[
k \geq y_{2n}, \quad n = 1, 2, \ldots, N.
\]

A contract is now a loan advance, \( q \), and a pair of payment schedules, \( \{y_{11}, y_{12}, \ldots, y_{1N}\} \) and \( \{y_{21}, y_{22}, \ldots, y_{2N}\} \), that determine the payments of good 1 and good 2, respectively, for each harvest. As before, contracts might in general give the borrower an incentive to hide some of the harvest in some states, but, as before, we can restrict attention to contracts for which the borrower never has an incentive to hide. Contracts that have this property satisfy the following incentive feasibility constraints:

\[
U_B(\theta_n - y_{1n}) + V_B(k - y_{2n}) \geq U_B(\theta_n - y_{1m}) + V_B(k - y_{2n})
\]

\[
\text{for } n = 2, \ldots, N, \quad m = n - 1.
\]

As in the previous model, I have written the constraint only in terms of the temptation of displaying the next smallest possible harvest \( \theta_{n-1} \).

Optimal contracts can again be found as the solution to a programming problem, parallel to Problem 2. An optimal contract is a set of numbers \( \{q, y_{11}, y_{12}, \ldots, y_{1N}, y_{21}, y_{22}, \ldots, y_{2N}\} \) that solve

**Problem 3:**

Maximize, by choice of \( q, y_{11}, y_{12}, \ldots, y_{1N}, y_{21}, y_{22}, \ldots, y_{2N}, \)

\[
\lambda_B U_B(q) + \lambda_B \sum_{n=1}^{N} [U_B(\theta_n - y_{1n}) + V_B(k - y_{2n})] \pi_n
\]

\[
+ \lambda_L U_L(e^\top_1 - q) + \lambda_L \beta \sum_{n=1}^{N} [U_L(y_{1n}) + V_L(y_{2n})] \pi_n
\]

\[20\] The simple argument used in Section II to justify restricting attention to adjacent incentive constraints does not apply in the two-good environment here. The approach is valid nonetheless because I merely want to show that a particular candidate contract is optimal. The set of contracts that satisfy global incentive feasibility constraints is contained in the larger set of contracts that satisfy the weaker local constraints in (19). The candidate contract can be shown to satisfy global incentive feasibility, so if it is optimal relative to contracts satisfying (19), then it is optimal relative to the smaller set of contracts that satisfies global incentive feasibility constraints.
subject to the resource feasibility constraints (13)-(18) and the incentive feasibility constraints (19). 

The Collateralized Debt Contract

Under what conditions, then, does a collateralized debt contract solve Problem 3? To answer this question I need to state precisely what constitutes a collateralized debt contract. To start, the payment schedule for good 1 is

\[ y_1 = y_1(R) = \text{MIN}[\theta_n, R]. \] (20)

A fixed, noncontingent amount \( R \) is transferred, unless the harvest is too small and \( \theta_n < R \), in which case the entire crop is transferred. I have in mind contracts in which \( R > \theta_1 \), contracts that were not incentive feasible in the earlier model with just one repayment good. Figure 4(a) portrays a typical payment schedule for good 1. For future reference, define \( r \) as the largest index number for which \( \theta_r < R \). 

To complete the description of a typical collateralized debt contract, I need to specify the schedule of transfers of chattels, the collateral good. I apply two guiding principles: first, ensure incentive feasibility of the resulting contract; and second, minimize the consumption of the borrower’s chattels by the lender. Thus, the chattels payment schedule is the minimal schedule that ensures that the borrower does not have an incentive to cheat and hide some of the harvest.

The schedule is constructed recursively starting with the payment \( y_2^N \), for the largest harvest, and working down to \( y_1 \), the payment for the smallest harvest, with the payment set at each step to ensure that the incentive feasibility constraint for that harvest outcome is met with equality. First, the payment \( y_2^N \) can be set freely, so to minimize the payment for this harvest outcome set \( y_2^N = 0 \). Now for any arbitrary \( m < N \), assume that the chattels payment schedule has already been determined for \( n = m+1, \ldots, N \). The incentive feasibility constraint relating the payments \( y_2^m \) and \( y_2^n \), for \( n = m+1 \), is

\[ u_B(\theta_n - y_1) + v_B(k - y_2) = u_B(\theta_n - y_1) + v_B(k - y_2), \quad n = m+1. \] (21)

If \( y_2^m \), payment of chattels from the borrower to the lender, is so small that (21) is violated, then when the harvest is \( \theta_n \) the borrower has an incentive to lie to make the payments \( y_1^m \) and \( y_2^m \) rather than \( y_1^m \) and \( y_2^m \). If \( y_2^m \) is so large that (21) is a strict inequality, then the chattels payment could be reduced without violating incentive feasibility. Choose for \( y_2^m \) the smallest value of \( y_2^m \) that satisfies (21). For each harvest outcome, the chattels payment is the smallest possible amount that does not give the borrower an incentive to lie.

The specific shape of a typical chattels payment schedule is shown in Figure 4(b). For \( \theta_n > R \), the crop payment is the constant, \( R \), so for \( m > r \), (21) reduces to \( v_B(k - y_2^m) \geq R \). Because \( y_2^N = 0 \), we can set \( y_2^m = 0 \) for all \( m > r \). In other words, for harvests greater than \( R \), the chattels payment is zero. For \( m = r \), (21) as an equality is

\[ u_B(\theta_{r+1} - R) + v_B(k) = u_B(\theta_{r+1} - \theta_r) + v_B(k - y_2^m). \] (22)

This equation determines \( y_2^r \). For harvests \( \theta_m < \theta_r \) (so that \( m < r \)), (21) as an equality is

\[ u_B(\theta_{m+1} - \theta_m) + v_B(k - y_2^m). \]

\[ u_B(\theta_{m+1} - \theta_m) + v_B(k - y_2^m), \quad n = m+1. \] (23)

The left side of (23) is the borrower’s utility when the harvest is \( \theta_m = \theta_{m+1} \) and he pays \( y_2^m = y_2^m \) and \( y_2^m \). The right side of (23) is the borrower’s utility when the harvest is \( \theta_m = \theta_{m+1} \) and he instead pretends \( \theta_m \) has occurred and pays \( y_2^m = \theta_m \) and \( y_2^m \). The chattels payment for harvest \( \theta_m \), \( y_2^m \), is set so that the borrower is just indifferent between these two alternatives. Note that the largest transfer of the collateral good is \( y_2^1 \), and occurs for the smallest possible harvest, \( \theta_1 \).

To summarize, a collateralized debt contract is described by (20) and (21). For harvests greater than \( R \), the borrower transfers a fixed amount, \( R \), of the crop, and none of the chattels. For harvests less than \( R \), the borrower transfers all of the crop and some amount of the chattels; just enough, for each harvest, to dissuade the borrower from falsely claiming that that harvest has occurred if the harvest is actually larger.

---

21 Problem 3 is convex under the additional assumption, which I now make, that \( -u_B(\theta_1)/u_B(\theta_1) = \) the coefficient of absolute risk aversion of the borrower with respect to good 1, is non-increasing. Under this condition the set of utilities that satisfy feasibility constraints is convex (even though the constraints are not convex in the choice variables).

22 Therefore, \( R \) is contained in the half-open interval \( [\theta_r, \theta_{r+1}) \).

23 This equation uses the facts that for \( m < r \) and \( n = m+1 \), \( \theta_n - y_1 = \theta_n - \theta_n = 0 \), and \( \theta_n - y_1 = \theta_n - \theta_m = \theta_{m+1} - \theta_m \).
The borrower's collateral, $k$, can sharply constrain feasible contracts, because a contract cannot require transfer of more collateral than the borrower actually has. The constraint that the largest collateral transfer $y_{21}$ not exceed $k$ is analogous to the constraint in the model of Section II that the fixed payment not exceed the smallest possible harvest; it places an upper limit on the amount of the fixed payment $R$. Although collateral can allow payment schedules which would otherwise be infeasible, feasible payment schedules could still be constrained.

**Optimality of a Collateralized Debt Contract**

The next task is to examine the first-order necessary and sufficient conditions for Problem 3, and to see whether the collateralized debt contract just described can satisfy those conditions. The objective is to identify conditions on the agent's utility functions, the endowments, and the probability distribution governing the harvest that allow the collateralized debt contract to satisfy the first-order conditions.

One condition that is required for a collateralized debt contract to be optimal is that the collateral good must be more valuable at the margin to the borrower than to the lender:

$$\frac{v_1'(y_{21})}{u_1'(y_{21})} \leq \frac{v_1'(y_{21})}{u_1'(y_{21})},$$

or, upon rearranging,

$$\frac{u_1'(y_{1n})}{u_1'(y_{1n})} - \frac{v_1'(y_{2n})}{v_1'(y_{2n})} \geq 0. \quad (25)$$

The two ratios in (24) measure the marginal value of the collateral good relative to the harvest good for each agent. The inequality (24) states that the marginal rate of substitution between chattels and wheat is larger for the borrower than for the lender. Indifference curves that satisfy (24) are shown in an Edgeworth Box in Figure 5. Imagine increasing the crop payment, $y_{1n}$, and decreasing the chattels payment, $y_{2n}$, by infinitesimal amounts in a way that keeps the borrower on the same indifference curve. If (24) holds then such a move along the borrower's indifference curve (to the northwest in Figure 5) increases the lender's utility. Thus condition (24) states that, ceteris paribus, giving more of the crop to the lender and more of the chattels to the borrower can make one of them better off without making the other worse off.

A second condition for a collateralized debt contract to be optimal is actually a strengthening of the first condition; the direct benefit of giving more crop to the lender and more chattels to the borrower must be greater than the cost of the second-order effect on incentive constraints:

$$\left[\frac{u_1'(y_{1n})}{u_1'(y_{1n})} - \frac{v_1'(y_{2n})}{v_1'(y_{2n})}\right] \lambda_1 \beta \pi_n - \Delta_n \geq 0, \quad (26)$$

where $\Delta_n \equiv \left[\frac{u_1'(y_{1n})}{u_1'(y_{1n})} - \frac{u_1'(y_{1n} - y_{1n})}{u_1'(y_{1n} - y_{1n})}\right] \phi_{n+1}. \quad (27)$
and where $\phi_{n+1}$ is the nonnegative Lagrange multiplier on the incentive feasibility constraint for harvest $\theta_{n+1}$. The bracketed term in (26) is identical to the left side of (25), and measures the benefit of giving the borrower more of the collateral good and less of the crop when the harvest is $\theta_n$. Such a reallocation affects the incentive constraint for harvest $\theta_{n+1}$, and the term $\Delta_n$ is the cost associated with this effect. If the benefit term in (26) is greater than the cost term $\Delta_n$, then the debt contract is optimal.

To understand the cost term $\Delta_n$, again imagine increasing the crop payment $y_1$ and decreasing the chattels payment $y_2$ by infinitesimal amounts, giving more of the payment good to the lender and more of the collateral good to the borrower, in a way that keeps the borrower on the same indifference curve. In particular, increase $y_1$ to $y_1 + \epsilon$, for some very small $\epsilon > 0$, and decrease $y_2$ to $y_2 - \delta$, so as to keep the borrower on a constant indifference curve. This change affects the borrower’s incentive to tell the truth when the harvest is $\theta_{n+1}$, making it more tempting to display $\theta_n$ and make the corresponding payments, $y_1 + \epsilon$ and $y_2 - \delta$. Specifically, $u_B(\theta_{n+1} - y_1 - \epsilon) + v_B(k - y_2 + \delta) > u_B(\theta_n - y_1) + v_B(k - y_2)$, even though $u_B(\theta_n - y_1 - \epsilon) + v_B(k - y_2 + \delta) = u_B(\theta_n - y_1) + v_B(k - y_2)$ by construction. The change in the right side of the incentive constraint for harvest $\theta_{n+1}$ [see condition (19)] is approximately

$$u_B(\theta_n - y_1) - u_B(\theta_{n+1} - y_1) \epsilon. \quad (28)$$

a nonnegative quantity. The term $\Delta_n$ is just (28), the amount by which the state $n+1$ incentive constraint is tightened, multiplied by $\phi_{n+1}$, the Lagrange multiplier, or “shadow value” for that constraint. (The denominator of $\Delta_n$ rescales $\phi_{n+1}$ into units of state $n$ utility.) The term $\Delta_n$ represents the cost of a move toward the northwest boundary of the Edgeworth Box for state $n$. Therefore, condition (26) states that the gap between the borrower’s and the lender’s marginal rate of substitution between chattels and wheat must exceed the cost of an indirect effect on incentive constraints.\(^{25}\)

There are two intuitive ways to think about condition (26). First, it can be thought of as a lower bound on the gap between the borrower’s and the lender’s valuation of the collateral good—the bracketed term in (26)—for a given value of the cost term $\Delta_n$. If the gap is not large enough, the debt contract is not optimal and the best arrangement involves more frequent transfer of the chattels to the lender. Alternatively, condition (26) can be viewed as an upper bound on the borrower’s risk aversion, because the cost term $\Delta_n$ is approximately proportional to the borrower’s coefficient of absolute risk aversion. The

\(^{25}\) Notice that if $u_B$ is linear, so that the borrower is risk neutral with respect to good 1, then the derivative $u_B$ is a constant, $\Delta_n$ is zero, and (26) is equivalent to (24). For very small values of $\theta_{n+1} - \theta_n$, $\Delta_n$ is approximately proportional to $-u_B(c_{n+1})u_B' c_{n+1}$, the coefficient of absolute risk aversion of the borrower with respect to the payment good. Thus $\Delta_n$ is larger, ceteris paribus, the more risk averse is the borrower.
incentive constraints prevent the borrower from sharing as much risk as he would like with the lender, so if the borrower is very risk averse the value of relaxing an incentive constraint is large. If the borrower is too risk averse, the cost of indirectly tightening the incentive constraints outweighs the benefit of giving the borrower the chattels, and the debt contract is not optimal.26

As mentioned above, collateral is often described as a means of compensating the lender for possible losses in default, but its main role in this model is to secure compliance with the debt agreement. The amount of collateral transferred for a given harvest is just enough to discourage the borrower from pretending a low harvest has occurred when it actually has not. Thus to the borrower, the amount of collateral transferred is equal in value to the shortfall in the crop. The lender is actually worse off when he receives collateral than when the full payment is made, because the collateral is worth less to the borrower than to the lender, relative to the crop. The value of collateral to the lender does matter for the arrangement because, the more the lender values the collateral, the lower the interest rate the lender will require.27 However, the primary function of collateral here is to keep the borrower honest.

V. RELATED LITERATURE ON DEBT CONTRACTS

Kenneth Arrow (1974), in his 1973 Presidential Address to the American Economics Association, first suggested that private information might be why noncontingent contracts are widely observed. This idea arose in the early economics literature on markets for insurance, particularly medical insurance, in which the absence of insurance arrangements was traced to the nonobservability of some key aspect of future outcomes (see Arrow, 1963, and Spence and Zeckhauser, 1971). This observation has long been taken for granted in the insurance industry itself. For example, an insurance textbook (Angell, 1959) states that one requirement for a hazard to be insurable is that “[i]t must be difficult or impossible for the insured to pretend that he has suffered a loss when he has not done so.”

Many recent papers have proposed explanations for debt contracts with occasional default. Douglas Diamond (1984) described a model of debt contracts based on private information about the borrower’s resources, as here, and based on the idea that a lender can impose “nonpecuniary penalties” on a borrower in the event of default. The amount of the penalty varies with the borrower’s reported resources, and is set optimally to ensure that the borrower does not have an incentive to lie. Diamond’s model is virtually a special case of the model presented above; the surrender of collateral serves as a penalty in my model, and the collateral good can be interpreted quite broadly as any action that reduces the utility of the borrower. Thus the model presented above unifies the treatment of collateral and penalties in loan contracts, and highlights their essential similarity.

An alternative model of debt contracts was first proposed by Robert Townsend (1979) and is based on the idea that the lender might be able to verify the borrower’s report at a cost. If the borrower reports a small harvest, the lender verifies the amount of the harvest and the borrower makes an agreed-upon payment. When the borrower’s harvest is sufficient to make the full payment, no verification takes place. The borrower never cheats, because verification would occur and he would be discovered. The debt contract is optimal in such an environment because it minimizes the frequency of costly verification. The logic is closely parallel to that of the model presented in this article. In both models, default involves deadweight loss—the transfer of collateral to the lender in my model and verification in the costly verification model—and the optimal contract seeks to minimize the cost.

Unfortunately, debt contracts are only optimal in the costly verification model in the presence of an ad hoc restriction on contractual arrangements. For each possible report by the borrower, a contract specifies that the lender either verifies or does not. More generally, a contract could specify that for a given report the lender verifies with some probability, not necessarily equal to zero or one. A deterministic contract is one in which verification probabilities are all either zero or one, while a randomized contract is one in which some verification probabilities are between zero and one. In the costly verification model, debt contracts are optimal only when attention is restricted to deterministic

26 This reasoning is only heuristic, because independently varying, say, the lender’s valuation of the collateral good will affect the cost term as well via the multiplier $\phi_{\omega}$. Nonetheless, parametric examples can easily be constructed that match the intuition in the text. Also, one can easily obtain an explicit expression for $\phi_{\omega}$ in terms of the primitive elements of the environment.

27 The interest rate on a loan is just $R/q - 1$. I have in mind a setting in which the lender compares the total return from the loan contract to returns on alternative uses of funds, so the more valuable the collateral the smaller R has to be.
contracts. Agents in the model can usually improve upon the debt contract with a randomized contract, and when randomized contracts are allowed the optimal contract does not, in general, resemble debt. The reason is that when verification occurs with positive probability, payments can be contingent. Verifying with small probabilities over a wide range of harvest outcomes can provide sufficient incentives and allow improved risk-sharing, while incurring less verification costs on average.28

One might think that randomized economic arrangements are unrealistic, and that there must be some as yet undiscovered reason why such arrangements are undesirable, but the possibility of randomization must be taken seriously in this context. Many financial arrangements actually do involve randomized audits, especially when one firm acts as an agent for another and has the opportunity to hide resources. The models presented above do not rely on a restriction to deterministic arrangements.29

Michael Jensen and William Meckling (1976) observed that because debt contracts force the borrower to bear all of the risk, he has more incentive than he would under a risk-sharing arrangement to take costly, private, ex ante actions that affect his return. This has led some to suggest that perhaps debt is selected over other feasible contingent arrangements because it provides superior incentives to the borrower to take appropriate ex ante actions (see Innes, 1990). Unfortunately, if one assumes that the return is freely observable by the lender ex post, then the debt contract is optimal only for very special assumptions about preferences and technology, and under strong restrictions on available contracts.30 If instead one assumes that the return is unobservable, then, as in Section II above, risk-free debt contracts are optimal, independent of the ex ante action choice.

Two recent papers, by Oliver Hart and John Moore (1989) and by Charles Kahn and Gur Huberman (1989), focus on renegotiation in debt contracts. To motivate debt contracts as an optimal arrangement, they assume that the borrower’s resources are observed by both the borrower and the lender but are not verifiable by a third party such as a court, and thus “enforceable” contracts cannot be made contingent. One could object by noting that courts often ascertain litigants’ wealth, and often enforce highly contingent contracts such as partnership agreements.

Although a wide range of literature examines the effects of debt contracts or the choice between debt and some other particular contract, the form of the contracts available to agents is generally taken as given. Thus this literature often has little to say about why contracts are limited to particular forms.

VI. SOME POLICY IMPLICATIONS, BRIEFLY NOTED

Recommended public policies toward credit markets are often predicated on models in which debt contracts play a prominent role, and so a model that explains debt contracts might have novel policy implications. What novel prescriptions for government credit policy might be suggested by the model described here? A complete answer is beyond the scope of the paper and is the subject of continuing research, but some tentative conclusions are possible.

Many policy prescriptions are sensitive to the assumption that capital markets are “perfect,” meaning that people can borrow or lend as much as they like on the same terms. For example, the Ricardian Equivalence Theorem revived by Barro (1974), which states that under certain conditions government debt policy is irrelevant, depends critically, as Barro noted, on perfect capital markets. In the model I presented above, the capital market imperfection is derived endogenously from informational constraints, but a blanket endorsement of policy prescriptions that depend on capital market imperfections seems unwarranted. Rather, one needs to assess how the informational imperfection affects the policymaker’s ability to improve on private arrangements; in some cases the policymaker may be as sharply constrained as private agents.

One category of potentially useful measures might be termed “collateral enhancement.” I showed above how the quantity of collateral available to the borrower could sharply constrain the loan contract. Under current U.S. law, there are limits to the collateral a consumer can offer; one cannot offer to a
prospective lender one’s imprisonment for nonpayment of a debt, for example. Moreover, under the “fresh start” provision of the Bankruptcy Act one cannot waive the right to discharge unsatisfied debts in bankruptcy. Consumers presumably could obtain more credit if they could offer to be imprisoned or could waive the right to discharge a debt, because such stiff penalties would make larger repayments credible. Interestingly, debts arising from government guaranteed educational loans are not dischargeable in bankruptcy. Consumers presumably could obtain such stiff penalties would make larger repayments credible. Interestingly, debts arising from government guaranteed educational loans are not dischargeable in bankruptcy during the first five years following the date that the first payment becomes due [11 U.S.C. § 523(a)(8)]. The claim in bankruptcy represented by a guaranteed student loan is thus more burdensome than a dischargeable claim, and presumably allows improved loan terms for the borrower or the lender. The analysis of the present paper suggests that allowing borrowers to waive the right to discharge debts in bankruptcy might improve the functioning of credit markets. However, there might be compelling countervailing reasons for the prohibition of waivers of discharge that are not taken into account by the models presented above; see Jackson (1985) for a discussion.

Another possible rationale for government credit policy concerns the valuation of collateral. Suppose the borrower in the model described above faces two possible lenders who differ only in the value they place on the borrower’s chattels. The optimal arrangement is for the borrower to obtain a loan from the lender who values the collateral good most highly, since this will provide the borrower with a lower interest rate. If, for some reason, a borrower’s collateral has a social value that is higher than its private value to lenders, due to an externality of some type, then direct government lending or government loan guarantees might be warranted. To justify such policies one would have to argue that the public valuation of the collateral is higher than its highest private valuation, and one can legitimately question whether this condition holds for many current loan-guarantee programs.31

Beyond these simple observations, little is known as yet about the policy implications of models like the one presented above. On one hand, it is difficult to imagine policy interventions that make some people better off without making anyone worse off in this type of model, other than the two just mentioned. In particular, based on this model alone there does not seem to be an efficiency rationale for loan subsidies or more general interest rate manipulations. Such policies could have important consequences for the distribution of welfare, of course, but would have to be evaluated by criteria other than Pareto optimality. On the other hand, the model leaves out some features, such as ex ante private information, that some economists claim rationalize credit market intervention.32 The claims usually pertain to markets that are dominated by the use of debt contracts, and yet the claims are based on models in which debt contracts are imposed, rather than derived as optimal. It is not yet known whether the conclusions of those models would survive if they were modified so that debt contracts arise endogenously, as in the model I have presented here.

VII. CONCLUDING REMARKS

So why is there debt with occasional default? My answer has two components. First, borrowers can fool lenders about their circumstances, so having the borrower share risk with the lender gives the borrower an irresistible temptation to cheat. Thus payment schedules are noncontingent in such situations. Second, if the borrower is incapable of making the stipulated payment, the lender has recourse either to explicit or to implicit collateral. Such recourse is sufficient to dissuade the borrower from withholding payment.

It is worth pointing out that important puzzles concerning debt contracts remain unsolved. The sole source of uncertainty here is the borrower’s future resources, and it seems quite reasonable to assume that borrowers can hide resources from lenders. But much of the uncertainty that faces borrowers and lenders concerns widely observed events about which neither is able to lie. Examples include publicly known prices and published economic data. The theory of Section I predicts that repayment contracts ought to be contingent on many publicly observed events. For example, officially published data on average prices of consumer goods are widely available, and are closely correlated with the real value

31 William Gale (1990) has described credit market models in which borrowers have private information beforehand about the riskiness of their future resources. He shows that in such models government loan guarantees targeted to high-risk borrowers can improve efficiency. In his 1991 paper he applies this model to existing federal credit programs and calculates that policy is likely to be quite inefficient. Debt contracts are assumed in his model, rather than derived endogenously, and it is unclear how the analysis would be affected by the latter.

of the monetary payments made by debtors to creditors. Why are so few debt contracts indexed for inflation?

A second puzzle is perhaps related to the first. A vast literature in monetary economics is motivated by the observation that money is widely used in spot exchanges for goods. And yet, almost all debt contracts are repaid in money as well. Perhaps the widespread use of money to settle debts is an equally important puzzle. The model described above does not have an explicit role for money, but the logic of the model suggests a rudimentary answer. The borrower might have some sort of advantage relative to the lender in selling the crop, and therefore returns money rather than the crop itself to the lender. This answer is rudimentary because it does not explain just why the borrower would have such an advantage. Evidently, much remains to be learned about financial arrangements such as debt contracts.

APPENDIX A

A Derivation of the Incentive Feasibility Constraints

In this appendix I show that any pattern of consumptions by the two agents that can be achieved by any arbitrary contract, possibly giving the borrower an incentive to hide the harvest, can be achieved by a contract that satisfies the incentive feasibility constraints and does not give the borrower an incentive to hide any of the harvest. Therefore, a given consumption pattern can be achieved if and only if it results from a contract that satisfies the incentive feasibility constraints. The argument is presented in the model of Section II, but can easily be extended to cover the model of Section IV.

To begin, take as given an arbitrary contract \( \{q, y_1, y_2, \ldots, y_n\} \), that satisfies the resource feasibility constraints (13)-(18), and consider a given harvest \( \theta_n \), where \( n > 1 \). The borrower can display any harvest \( \theta_m \), where \( m \) can equal \( 1, 2, \ldots, n \), and \( m \) is chosen to maximize \( u_B(\theta_n - y_m) \). Define \( y_m^* \) as the payment the borrower actually makes after optimally choosing a utility maximizing display. It does not matter if the utility maximizing display is not unique, because the utility maximizing payment is always unique. The payment \( y_m^* \) clearly satisfies \( u_B(\theta_n - y_m^*) \geq u_B(\theta_n - y_m) \) for \( m = 1, 2, \ldots, n \).

Now consider an arbitrary harvest \( \theta_p < \theta_n \), and define \( y_p^* \) analogously as the utility maximizing payment for the harvest \( \theta_p \). Clearly, \( y_p^* = y_m \) for some \( m \) in the set \( \{1, 2, \ldots, p\} \). Since \( p < n \), it is also true that \( y_p^* = y_m \) for some \( m \) in the set \( \{1, 2, \ldots, n\} \); in other words, the utility maximizing payment for the harvest \( \theta_p \) is a payment that could have been made for the harvest \( \theta_n \). As a result, the payment \( y_m^* \) can provide no more utility when the harvest is \( \theta_n \) than the utility maximizing payment \( y_p^* \). Therefore, \( u_B(\theta_n - y_p^*) \geq u_B(\theta_n - y_m^*) \). Since both \( n \) and \( p \) are arbitrary, this condition holds for \( n = 2, \ldots, N \), and for \( p = 1, \ldots, n - 1 \). These are exactly the incentive feasibility constraints (19).

I have defined a set of payments \( \{y_1^*, y_2^*, \ldots, y_N^*\} \), the utility maximizing payments chosen by the borrower when the contract is \( \{q, y_1, y_2, \ldots, y_N\} \). Now define a new contract \( \{q, y_1^*, y_2^*, \ldots, y_N^*\} \), by substituting the actual payments for the originally stipulated payments. This new contract satisfies the incentive feasibility constraints, and thus does not provide any positive incentive to hide harvest. The new contract results in consumption patterns for both the borrower and the lender that are identical to those resulting from the original contract. Because the original contract is arbitrary, I have shown that any consumption patterns that can be achieved can also be achieved under a contract that provides no incentive to hide the harvest.
References


INTRODUCTION

In 1990 banks throughout the United States had total provision for loan losses of over $31 billion, an amount almost twice bank profits. Since the mid-1980s, provision for loan losses has been one of the most important factors affecting bank profitability. Headlines and narratives like those listed above demonstrate the interest of the financial press in banks' loss provisions. Yet for many banking students the subject generates questions: What types of accounts are being discussed? Is there a difference between loan loss reserves, loan loss provision, provision for credit losses, and allowance for loan losses? Where do these reserves come from? How do banks decide how much to add to the reserve? Why does increasing reserves produce losses for banks? And why do banks use reserves in the first place?¹

This paper seeks to answer these questions. In doing so it lists and defines the terminology frequently used in discussions of bank loan losses (see "Definitions of Terms" on p. 29) and examines the history and current use of the reserve for loan losses. It also discusses how and why methods for determining the level of reserve for loan losses have changed.

¹ For expositional simplicity leasing is ignored since it is handled in essentially the same manner as lending. Names of accounts are therefore shortened throughout the article. For example, provision for loan and lease losses as on bank Reports of Condition and Income is called provision for loan losses.

DESCRIPTION OF RESERVES FOR LOAN LOSSES

The primary business of banking is the collection and investment of depositors’ funds. As a part of this business banks bear credit risk, i.e., the possibility that the borrower will fail to repay as promised. The two major assets in which banks invest depositors’ funds are securities and loans. Credit losses on securities are minimal because the bulk of these holdings are government securities with little or no default risk. Loans are a different story. In 1990 banks throughout the United States wrote off over $29 billion in loans as uncollectible (net of recoveries), an amount almost twice total profits of all U.S. banks for the year.

The federal banking regulators (Federal Deposit Insurance Corporation, Office of the Comptroller of the Currency, and Federal Reserve) require that all banks include in their financial statements an account named allowance for loan losses (also known as reserves for loan losses). Figure 1 provides an illustrative example showing how the reserve for loan losses (line 4) is typically reported. The account absorbs loan losses both from loans the bank can currently identify as bad loans and from some apparently good loans that will later prove to be uncollectible. The reserve for loan loss account is established and maintained by periodic charges against earnings. The charges show up on the income statement as an expense category named provision for loan losses (see Figure 2, line 10). The reserve for loan losses is
Specific Reserves

At many banks, for analytical purposes or on internal books, the reserve is divided into two categories, specific or allocated reserves, and general reserves. Specific reserves are those that a bank views as being associated with some particular loan or group of loans. When a bank determines that a loan presents a greater-than-normal risk of loss it may either add to its reserves specifically for that loan or designate some portion of reserves to be allocated for the loan. Those reserves that are not allocated to particular loans or groups of loans are the general reserves. Division of the reserve account into these two categories allows the bank to analyze its loan loss reserve needs more precisely. On financial reports, however, general and specific reserves are summed and reported simply as reserves for loan losses.

When loan losses are recognized, that is, when a bank decides that some portion of a loan will not be collected and therefore must be charged off or written down, the amount of the loss is deducted from the asset category loans and also from reserves for loan losses. Suppose for example a bank had made a $100 loan but only expected to be able to collect $40 from the borrower. In Figure 1, $60 would be deducted from $64,000 on line 3 so as to reduce the loan portfolio by the uncollectible amount of the questionable loan. The $60 would also be deducted from $1,000 on line 4. If the bank had already anticipated a $60 loss on the loan and had added $60 to its reserve then the bank’s current income would not be affected by the write-off. On the other hand if the loan loss had not been anticipated before the loan was written down, then in all likelihood the bank would add $60 to its reserves following the write-down in order to maintain its reserve at a level sufficient to absorb future loan losses.

Why Banks Create Loan Loss Reserves

Displaying loans on a bank’s balance sheet as the amount of funds lent without an adjustment for expected but uncertain future losses would mislead the bank’s board of directors, creditors, regulators, and investors by overstating the bank’s assets. The income-earning potential of the bank and its capital would also be overstated, making the bank appear stronger than it really is. One would prefer the balance sheet to show as assets only that portion of loans that will be collected. It is difficult, however, for a bank’s management to determine before the fact which loans will not be repaid. The compromise

<table>
<thead>
<tr>
<th>Figure 1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Balance Sheet as of December 31, 1990</td>
</tr>
</tbody>
</table>

| Illustrative National Bank |
|-----------------|-----------------|
| (000) | Liabilities and Equity |
| Assets | |
| 1 Cash | $ 8,000 |
| 2 Securities | 20,000 |
| 3 Total loans | $ 64,000 |
| 4 Less: Reserves for loan losses | 1,000 |
| Equals: Net loans | 63,000 |
| 6 Other real estate owned | 400 |
| 7 Other assets | 8,600 |
| 8 Total assets | $ 100,000 |
| Liabilities and Equity |
| 9 Deposits | $ 74,000 |
| 10 Other liabilities | $ 19,000 |
| 11 Total liabilities | $ 93,000 |
| 12 Owners’ Equity | 7,000 |
| 13 Total liabilities and owners’ equity | $ 100,000 |

FEDERAL RESERVE BANK OF RICHMOND
Figure 2
Income Statement for Year Ending
December 31, 1990
Illustrative National Bank
(000)

<table>
<thead>
<tr>
<th>Interest income</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Interest and fees on loans</td>
<td>$ 7,000</td>
</tr>
<tr>
<td>2 Interest on securities</td>
<td>1,800</td>
</tr>
<tr>
<td>3 Other interest income</td>
<td>200</td>
</tr>
<tr>
<td><strong>Total income</strong></td>
<td><strong>$10,000</strong></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Noninterest income</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>4 Service charges</td>
<td>400</td>
</tr>
<tr>
<td>5 Other noninterest income</td>
<td>600</td>
</tr>
<tr>
<td><strong>Total income</strong></td>
<td><strong>$10,000</strong></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Interest expense</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>7 Interest on deposits</td>
<td>$ 4,000</td>
</tr>
<tr>
<td>8 Other interest expense</td>
<td>2,000</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Noninterest expense</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>9 Salaries and benefits</td>
<td>1,000</td>
</tr>
<tr>
<td>10 Provision for loan losses</td>
<td>300</td>
</tr>
<tr>
<td>11 Other noninterest expense</td>
<td>1,700</td>
</tr>
<tr>
<td><strong>Total expense</strong></td>
<td><strong>$ 9,000</strong></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Income before taxes</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>13 Income before taxes</td>
<td>$ 1,000</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Income taxes</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>14 Income taxes</td>
<td>250</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Net income</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>15 Net income</td>
<td>$ 750</td>
</tr>
</tbody>
</table>

**Calculation of Reserves for Loan Losses for 1990**
Illustrative National Bank
(000)

| 1 Reserves for loan losses, beginning of 1990 | $ 900 |
| 2 Less: Charge-offs during 1990               | 285   |
| 3 Plus: Recoveries during 1990 of loans       | 85    |
| 4 Plus: Provision for loan losses, 1990        | 300   |
| 5 Reserves for loan losses, end of 1990        | $ 1,000 |

**Informational Value of the Reserve for Loan Losses**

Depositors, bank stock investors, and bank analysts are not, in general, privy to information about the riskiness of banks' loans beyond that revealed by the amount of past due and nonaccrual loans which banks are required to report. In other words, the management of a bank has more information about the quality of the loan portfolio than do outsiders. Data on the amount of reserves a bank holds and additions made to reserves are useful to outsiders, since they provide additional information about the quality or riskiness of the loan portfolio. The value of this information is demonstrated by the strong reaction of bank stock prices to unexpected news about bank reserves.

The loan quality information or signal provided by the reserve should be most trustworthy immediately after regulators examine a bank. Examiners provide an independent, unbiased assessment of the quality of a bank's loan portfolio and also have the power to force the bank to restate loans and reserves when their values deviate from the regulator's best estimates. Financial reports coming out soon after a visit from examiners are, therefore, more likely to include an accurate statement of expected net realizable loan values.

**Loan Categories**

At any given time a bank is likely to have some loans in each of the following four categories:

1. **Good loans.** The borrower is making scheduled interest and principal payments and the bank has no reason to suspect that the borrower will not pay back the loan in full.
2. **Loans past due or otherwise in doubt.** Scheduled interest or principal payments have been missed or the bank has some other information indicating that repayment of the loan is in doubt.

3. **Written-down loans.** The bank has removed some of the face value of the loan from its books because it believes it will be able to collect only a portion of the loan.

4. **Charged-off loans.** The value of the loan has been completely removed from the bank’s books, because the bank believes it will be able to collect little or nothing from the borrower. The bank may continue to attempt to collect funds from the borrower though it has charged the loan off its books and may be carrying some collateral from the loan on its books.

Most loans stay in category 1 until repaid. Some loans however start off in category 1 but later travel through all three remaining categories before being closed out. Any loan in categories 2 or 4 is a problem loan. Loans in category 3 are often considered problem loans. In some cases, however, when a loan has been written down by an amount sufficient to lower its reported value to its collectible amount, it might be considered a good loan.

**The Problem Loan**

For most loans only the passage of time and scheduled interest and principal payment dates allow banks to distinguish good loans from problem loans. When the borrower is more than 30 days past due on a scheduled payment the loan is considered past due and the bank lists it as such in its financial statements. The bank probably will have made some effort to contact the borrower to secure payment before delinquency reaches this stage. As scheduled payments fall further in arrear, the likelihood of ultimate repayment diminishes.

When repayment of a loan becomes less likely most banks will add to the reserve in anticipation of a possible loss. Beyond setting aside additional reserves, past due or doubtful loans may be handled in one of several ways depending on the bank's policies. Some banks promptly charge past due or doubtful loans off their books and then attempt to recover from the borrower whatever funds possible. Other banks carry such loans on their books until the borrower recovers or until forced either by the passage of time or by regulators to charge off the loan. Banks will at times attempt to renegotiate the terms of a loan if renegotiation seems likely to encourage some repayment. In most cases if a loan is past due more than 180 days it will be charged off or at least written down. When a loan is charged off, interest income accrued but not received during the current accounting period is subtracted from current income, and interest accrued but not received in prior accounting periods is deducted from reserves for loan losses [Board of Governors (1984), Section 219.1, p. 4].

The decision between charging off all or only a portion of a loan will depend on whether the bank believes any of the loan is collectible, on the bank’s normal procedures for handling losses, and on examiners’ opinions. Banks with very conservative loan loss procedures may choose to completely charge off any past due or doubtful loan even if it is likely to be partially repaid. Other banks may, when relatively certain that some portion of a loan will ultimately be collected, deduct only a portion of the face value of the loan from the asset category loans, meaning the loan is written down to its collectible amount. The amount of the write-down is also deducted from reserves for loan losses. If it is unlikely that any portion of a loan will be ultimately collectible then the loan normally will be charged off completely. Regulatory examiners may, following an examination, require a bank to set aside additional reserves for a loan, to write it down, or completely charge it off, depending on their opinions of the probability of repayment.

Collection of funds on a loan that has been completely or partially charged off can be a long and expensive process. Banks usually foreclose on or repossess available collateral. The amount a bank will ultimately recover from written-down or charged-off loans depends on the financial health of the borrower, the borrower’s willingness to pay, the value of any collateral, the strength of guarantors or cosigners, and the ability of the bank’s workout department or that of the individual loan officer assigned to the account. Any recovery of an amount previously charged off or charged down is added to reserves upon its collection (see Figure 3, line 3).

**DETERMINATION OF THE SIZE OF THE RESERVE FOR LOAN LOSSES**

Banks' use of the reserve for loan losses, and especially banks' decisions with respect to the size of the account, have changed since the 1940s. The main forces shaping the change have been tax policy.
regulators' instructions, and the growing loan losses of the 1980s. For the first 30 years of the routine use of the account, tax policy determined the amount of reserve held by banks. Then regulatory pressures and high loan losses became dominant determinants.

The Influences of Tax Policy

From 1947 until the mid-1970s or early 1980s, the amount of reserve for loan losses held by banks was largely based on tax considerations. Few banks employed the account before 1947. Most banks relied instead on the “specific charge-off method” since its tax treatment was straightforward [FDIC (1947), pp. 25-26, and Blake (1952), pp. 30-35]. That method of accounting for loan losses involved the subtraction of loan losses from current income or net worth when the loan was charged off.

On December 8, 1947, the Commissioner of Internal Revenue liberalized its policy for banks by ruling that banks' reserves for loan losses could be calculated in a manner that differed from that of other businesses [FDIC (1948), p. 43]. Banks were allowed to hold a reserve for loan losses equal to three times their average yearly loan loss experience of the past 20 years. Soon after the 1947 ruling most large banks and many small banks began holding reserves for loan losses (see Table 1). With some modifications, this policy continued until 1969. Banks could hold reserves exceeding the maximum specified by the IRS, but once the maximum was exceeded additions to the reserve were not tax deductible. This was the case for years before and since 1969. See Table 2 for details of tax laws and rulings.

The Tax Reform Act of 1969 broke with the most recent 20 years of IRS policy and gradually required banks to hold a reserve equal to their current and past five years' losses [U.S. Congress, House of Representatives (1969), pp. 464-75]. The 1969 act was passed in part to lower banks' tax advantage over other businesses. The change was to be phased in over the next 18 years (see Table 2, 1969 Tax Reform Act). During the phase-in period a bank could either add to reserves for loan losses until they equaled a percentage of loans specified by the act, or until they equaled the bank's average ratio of loan losses to loans of the past six years. The maximum ratio of reserves for loan losses to loans specified by the act declined every six years over the 18-year phase-in.

In 1986 the Tax Reform Act of 1986 was passed, eliminating, for banks with more than $500 million in assets, the opportunity to subtract, as a pre-tax expense, any provision for future loan losses beyond the amount of loans actually charged off during the year. Small banks continued to hold reserves based on the specifications of the Tax Reform Act of 1969 [U.S. Congress, Joint Committee on Taxation (1987), pp. 549-53].

The rapid growth in reserves following 1947 and the maintenance of levels close to the maximum allowed by the IRS until the early 1980s are apparent in the chart (see listing of IRS maximums in Table 2). While bank loan losses were small and on average fairly constant relative to total loans from 1947 through the early 1970s, banks held reserves throughout the period that greatly exceeded losses. Banks' best estimates of expected loan losses during most of the period were almost certainly considerably lower than the amount of reserves held. However, it was to the banks' advantage to hold reserves at the maximum allowed by the IRS since doing so resulted in lower taxes.

Tax Considerations Become Less Important

Until at least the early to mid-1970s, tax rulings and laws encouraged banks to hold reserves that greatly exceeded losses so that significant regulatory efforts aimed at influencing banks' holdings of reserves were not necessary. Beginning in 1976, however, federal regulators began to encourage banks to hold a reserve of at least 1 percent of total loans. By 1976 the maximum reserve allowed by the IRS had declined to 1.2 percent of loans.

Beginning in 1981 bank failures began to rise and in 1982 net loan losses relative to total loans began a fairly steady increase that would last through the 1980s and into the 1990s (see chart). Regulators and accountants were no longer willing to permit

---

Table 1

<table>
<thead>
<tr>
<th>Percentage of Banks with a Reserve Account in Selected Years</th>
</tr>
</thead>
<tbody>
<tr>
<td>1948</td>
</tr>
<tr>
<td>1950</td>
</tr>
<tr>
<td>1957</td>
</tr>
<tr>
<td>1963</td>
</tr>
<tr>
<td>1971</td>
</tr>
<tr>
<td>1975</td>
</tr>
</tbody>
</table>

Table 2
Tax Laws and Rulings Affecting Banks’ Reserves for Loan Losses

<table>
<thead>
<tr>
<th>Year</th>
<th>Type of decree</th>
<th>Effect on reserves</th>
</tr>
</thead>
<tbody>
<tr>
<td>1921</td>
<td>Revenue Act</td>
<td>Allowed all businesses to make additions to bad debt reserves from pre-tax income. Amount set aside was to be reasonable based on loss experience of individual businesses.</td>
</tr>
<tr>
<td>1947</td>
<td>Ruling</td>
<td>Allowed banks to cumulate reserves for loan losses from pre-tax income up to three times the banks’ average annual losses of the past 20 years.</td>
</tr>
<tr>
<td>1954</td>
<td>Ruling</td>
<td>Banks could choose any 20-year period after 1927 on which to calculate their maximum reserves.</td>
</tr>
<tr>
<td>1965</td>
<td>Ruling</td>
<td>All banks could accumulate reserves from pre-tax income up to 2.4 percent of total loans. Further additions must come from after-tax income.</td>
</tr>
<tr>
<td>1969</td>
<td>Tax Reform Act</td>
<td>Mandated the following phased reduction of maximum reserves percentage above which provisions could not be made from pre-tax income: 1969-75 maximum reserves/loans = 1.8 percent 1976-81, 1.2 percent 1982-87, 0.6 percent. Also specified eventual replacement of percentage-of-loans method with maximum reserves based only on bank’s loss experience. Between 1969 and 1987 banks could choose either the appropriate percentage or the “experience method” in which the maximum reserve equals the product of the average net charge-off to total loans ratio for the most recent six years times current outstanding total loans. Banks could switch between percentage-of-loans method and experience method from year to year between 1969 and 1987. After 1987 only the experience method could be used.</td>
</tr>
<tr>
<td>1986</td>
<td>Tax Reform Act</td>
<td>Banks with assets over $500 million must use “specific charge-off method” that permits no additions to reserves for loan losses from pre-tax income beyond current year’s charge-offs. For smaller banks, 1960 Tax Reform Act holds.</td>
</tr>
</tbody>
</table>

banks to base the size of their reserves either on a standard rule or on a shrinking arbitrary percentage set by the IRS (after 1982 banks were not taxed on additions to reserves when the reserve was less than .6 percent of loans). Regulators began to encourage banks to calculate reserves based on their own expectations of future losses in the loan portfolio. The chart shows that in the early 1980s banks, on average, responded to regulatory pressure, or at least to growing loan losses, by maintaining reserves well above the maximum .6 percent of total loans permitted by the IRS. The chart also demonstrates that the gap between reserves and loan losses (both expressed per dollar of loans) shrank from the early 1970s to 1987 but recently has returned to levels common in the 1950s and 1960s. The earlier gap developed in response to tax incentives, but the more recent gap reflects expected large losses from loans to less developed countries and from commercial real estate loans.

While regulators have been pushing banks to base reserves on expected loan losses, they have recently de-emphasized reserves somewhat as a component of regulatory capital. Traditionally reserves for loan losses have been counted in regulators’ measures of capital (see “Definitions of Terms” on p. 29 for the ratios regulators use currently in capital adequacy measures). Before 1988 all of a bank’s reserve for loan losses was included in the regulators’ main measure of bank capital, primary capital, and therefore was allowed to play an important role in adding to bank capital adequacy. Since 1988, reserves for loan losses have been de-emphasized somewhat in capital adequacy measures, since they are counted only in Tier 2 capital and only up to a specified
proportion of assets [Board of Governors (1991), pp. 3-474.1 and 3-474.2]. According to the capital guidelines agreed upon by all three federal regulators in 1988, capital adequacy is measured using Tier 1 capital and total capital (the sum of Tier 1 and Tier 2 capital). Total capital includes reserves for loan losses, up to a specified limit, and therefore is augmented by additions to reserves.

Determining the Size of the Loan Loss Reserve

Banks employ various techniques to set their reserve for loan loss levels. The amount of reserve maintained is scrutinized by bank regulators and is often modified following bank examinations. Banks maintain reserves at a constant ratio to loans, to past loan losses, or at levels comparable to those maintained by their peers. Alternatively they set reserves to advance income or tax management goals. Finally they set reserve levels by performing an analysis of potential loan losses in their portfolios. They may even use a blend of some or all of the preceding.

**Constant Percentage-of-Loans Rule** This technique requires that the bank decide on some target level for the ratio of reserves to total loans and then add to the reserve account whenever the ratio falls below target. The percentage-of-loans technique requires no determination of expected future loan losses. The method was used by the majority of banks before the mid-1970s with the target percent determined by the IRS and by tax laws. For large banks, since the passage of the Tax Reform Act of 1986, and for small banks, since 1988 and the beginning of the final phase of the Tax Reform Act of 1969, there is no tax incentive to base reserves on a percent of loans. Some small banks, however, may continue to use the rule, setting the ratio of reserves to loans at 1 to 2 percent.

Use of the technique limits the analysis a bank must perform to determine the size of its reserve account but can lead to several problems. First, regulators and a bank's outside accountants are likely to object to the technique at some point since both the Financial Accounting Standards Board (FASB) and federal regulators have stated plainly that the reserve is to be based on expected losses [FASB (1989), p. 35]. Therefore a bank may be required to show that there is a relationship between its reserves and expected loan losses. Second, using the technique may leave the reserve for loan losses too small to deal with several quarters of substantial loan losses. If instead the bank were performing a more sophisticated analysis of expected loan losses, loan losses might be better predicted and the reserve augmented in preparation.

**Peer Equivalent** In its most basic form the peer equivalent technique involves setting the reserve for loan losses equal to or near the level maintained by a bank's peers. Financial reports for banks are widely published, so determining the amount of reserves held by peer banks of equivalent size operating in equivalent markets is a simple matter.

The advantage of the technique is that, like the constant percentage-of-loans technique, it allows the bank to avoid any detailed and costly analysis of its loans. While a few small banks may make exclusive use of such a simple approach, most banks make use of peer information as one of several elements in their determination of appropriate reserve level. Banks compare their own reserves relative to loans to that of peers to determine if their reserve is in line with that of their peers. Regulators also encourage banks to compare themselves with peers but not to the exclusion of analysis of expected losses [see, for example, Board of Governors (1984), Section 219.1, p. 3; and OCC (1984), Section 217.3, p. 1].

**Loss History** Most banks use prior years' history of loan losses to help them determine current reserves for loan losses. Since the amount of each small bank's tax benefits available from provisions for loan losses is determined by a formula based upon past years' loan losses, some of these banks place considerable weight on such losses when deciding current reserves. For other banks, prior losses on fairly homogeneous loans such as credit card loans, auto loans, personal loans, and home mortgages can provide a reasonable guide to what can be expected in the future.

Since the regulatory agencies warn their examiners not to allow banks to rely too heavily on historical loss data, it is likely that most banks do not place an unwarranted emphasis on past experience when determining their appropriate reserve levels [see, for example, OCC (1984), Section 217.1, p. 2; Board of Governors (1984), Section 219.1, p. 2; and AICPA (1983), p. 62]. The problem with relying completely on loss history is that loan losses are affected by factors that change over time, such as the phase of the business cycle and management philosophy about the declaration of loan losses, so that the experience of the last several years may not always be a good predictor of future conditions.
**Income Management** Banks can smooth variations in reported income through their choices of when to take provisions for loan losses. By taking small provisions during periods of poor operating income and large provisions when income is high, a bank can shift reported income from prosperous to depressed times, thus smoothing its reported income stream. Choosing the size of provisions to dampen reported income fluctuations may, however, lead the bank’s auditors, regulators, or the Securities and Exchange Commission (SEC) to question the bank’s income or expenses reporting.

**Tax Management** When additions to the reserve for loan losses were tax deductible beyond actual charge-offs or loan loss experience, bank income taxes were lowered in high income years by taking larger provisions for loan losses. When income was down, and tax benefits were not as valuable, provisions were decreased. Banks can still produce some tax benefits through shrewd use of the reserve account. Large banks, for which tax deductions are limited to actual loan charge-offs, can to some extent concentrate charge-offs when income is high. Small banks, which since 1988 have been using the experience method of determining tax-deductibility, can set aside the maximum provisions allowed by past loss experience when income is high, and fairly low provisions in years when income is low. As with income management, these maneuvers are likely to produce questions from the IRS, regulators, and auditors.

**Loan Analysis** Regulators, in their efforts to promote more accurate reporting of banks’ income and net worth, have been encouraging banks to use careful loan analysis in the determination of reserve levels since the mid-1980s. When a bank sets its reserves for loan losses equal to an estimate—based on analysis of each loan or loan category—of the loss inherent in the loan portfolio, it determines its reserves using the loan analysis method. While there is considerable variation among banks in the specifics of the analysis, the basic procedures are similar.

Banks generally divide loans into categories and then apply differing analyses to each category to estimate the reserves needed for each category. These estimates are summed across categories to arrive at a total for the loan portfolio (see Figure 4). In general, loans are divided at a minimum into large classified loans, other large loans, and small commercial and consumer loans.

---

**Figure 4**

**Estimate of Needed Reserves for Loan Losses**

<table>
<thead>
<tr>
<th>Loan Category</th>
<th>Principal Amount</th>
<th>Estimated Needed Reserve</th>
</tr>
</thead>
<tbody>
<tr>
<td>Large classified loans</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Potentially weak</td>
<td>$ 5,000</td>
<td>$ 500</td>
</tr>
<tr>
<td>Substandard</td>
<td>4,000</td>
<td>800</td>
</tr>
<tr>
<td>Doubtful</td>
<td>2,000</td>
<td>1,000</td>
</tr>
<tr>
<td>Loss</td>
<td>500</td>
<td>500</td>
</tr>
<tr>
<td>Other large loans</td>
<td>1,250,000</td>
<td>12,500</td>
</tr>
<tr>
<td>Problem small commercial loans</td>
<td>8,000</td>
<td>1,600</td>
</tr>
<tr>
<td>Problem small consumer loans</td>
<td>10,000</td>
<td>2,500</td>
</tr>
<tr>
<td>Small commercial loans</td>
<td>900,000</td>
<td>9,000</td>
</tr>
<tr>
<td>Consumer loans</td>
<td>1,000,000</td>
<td>10,000</td>
</tr>
<tr>
<td><strong>Total estimated needed reserves</strong></td>
<td><strong>$38,400</strong></td>
<td></td>
</tr>
</tbody>
</table>

For most banks the majority of large loans, i.e., those that are significant in relation to bank capital or total loans, are found in the commercial loan portfolio. Classified loans are those that have been placed in higher-than-normal risk classes either by the bank’s internal loan review or by examiners. A bank’s entire portfolio of large loans is frequently reviewed to determine (1) which loans present greater-than-average risk and should therefore be classified and (2) whether those loans already classified should be unclassified or moved to a higher risk category. Classified loans are scrutinized more carefully than other loans when determining reserves for loan losses.

An expected loss or range of losses for all classified loans for each risk class may be estimated from past years’ losses and recoveries for that class of loans, from knowledge of the individual classified loans, or from a combination of both. A reserve need is computed for each loan or class of loans as the multiplicative product of the chance of expected loss for the loan or class times the dollar amount of the expected loss. Some of the factors banks typically consider when deciding the probability and amount of loss from a classified loan are the following: whether the loan is currently past due, and if so, how far past due; also, the financial condition of the borrower, the availability of responsible cosigners or...
guarantors, the availability of collateral and its value, national and regional economic trends, and, finally, industry trends.2

The losses inherent in the portfolio of other large loans, i.e., large loans that are not classified, must also be estimated to determine the amount of reserves needed for these loans. The estimate is based on (1) historical loss data for large loans with normal risk, classified by type of loan, (2) knowledge of the creditworthiness of the individual borrowers, and (3) economic and industry trends.

Expected losses on small commercial loans and consumer loans that are not past due or on nonaccrual status are estimated from loss histories of the various types of loans and from other considerations that may influence losses in the future. For example, a bank may have suffered losses ranging from 2 to 4 percent per year of its credit card portfolio over the past five years. It would be reasonable, therefore, for the bank to maintain reserves for credit card loans equal to 4 percent of the average amount of the bank's outstanding credit card loans, assuming conditions affecting losses on such loans to be unchanged in the coming year.3 If rising unemployment or some other factor that might increase losses is expected in the coming year, the amount of reserves needed for these loans would be higher. Small commercial loans and consumer loans that are past due or on nonaccrual status generally require larger reserves than current loans, since a borrower's failure to make scheduled loan payments is an indication that a future loss may be imminent.

CONCLUSION

Most banks no longer set their loan loss reserves at some fixed percentage of total loans as was customary until the early 1980s. Owing to (1) the elimination of most of the tax incentive to maintain excess loan loss reserves, (2) to regulators' abandonment of a fixed target reserve to loans ratio, (3) to the diminished role of reserves in regulatory capital measures, and (4) to regulatory pressure to use loan loss analysis in reserve determination, the reserve is now more likely to measure potential loan losses than in the past. Nevertheless, the desire to smooth reported profits, to lower taxes, and to limit the expenses of estimating future loan losses continues to provide an incentive for banks to hold reserves at levels that differ from their best estimates of the losses inherent in their loan portfolios.

During 1990 the Financial Accounting Standards Board, the primary accounting rule-making body, began considering a proposal that could, if implemented, result in a new accounting standard to be used by banks in their calculations of the amount of reserve needed for individual "impaired loans" (loans for which it is probable that the bank will not collect all principal and interest payments according to the terms of the loan contract). Under the new standard the amount of reserve considered adequate for an impaired loan would equal the difference between the book value of the loan and the present value of the expected cash flow generated by the loan. The new standard would apply only to impaired loans.

3 For low value, high volume loans regulators require banks to hold reserves only for the coming year's expected losses, rather than holding reserves for expected losses over the life of the loan, which may exceed one year.
**Definitions of Terms**

<table>
<thead>
<tr>
<th>Term</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Allocated transfer risk reserve</td>
<td>Balance sheet item, separate from loan loss reserve (LLR), that accounts for the risk that foreign borrowers will not be able to acquire sufficient foreign exchange to repay loans.</td>
</tr>
<tr>
<td>Charge-off</td>
<td>Completely removing a loan from the balance sheet by subtracting its book value from loans and from LLR. Also called write-off.</td>
</tr>
<tr>
<td>Default</td>
<td>Failure of borrower to satisfy provisions of loan agreement.</td>
</tr>
<tr>
<td>Experience method</td>
<td>Basing the amount of the addition to LLR on historical loan loss experience.</td>
</tr>
<tr>
<td>Foreclosure</td>
<td>Legal proceeding removing from the debtor all interest in mortgaged property when conditions of the mortgage have been violated.</td>
</tr>
<tr>
<td>Loan loss reserves (LLR)</td>
<td>Balance sheet account. Deducts from total loans the portion of loan principal not expected to be paid back. Also called allowance for loan losses or reserves for credit losses.</td>
</tr>
<tr>
<td>Loan workout</td>
<td>Process following default in which a bank attempts to recover whatever loan funds it can.</td>
</tr>
<tr>
<td>Net loans</td>
<td>Total loans less LLR and allocated transfer risk reserve.</td>
</tr>
<tr>
<td>Nonaccrual loan</td>
<td>A loan carried on the bank’s balance sheet that no longer accrues interest. Any payments received are deducted from principal but not booked as income.</td>
</tr>
<tr>
<td>Other real estate owned</td>
<td>Balance sheet account showing the book value of all real estate, other than bank premises, owned by the bank. Consists largely of repossessed real estate.</td>
</tr>
<tr>
<td>Past due loan</td>
<td>A loan more than 30 days behind in interest or principal payments.</td>
</tr>
<tr>
<td>Percentage method</td>
<td>Basing the amount of the addition to LLR on a percentage specified by regulators or tax policy.</td>
</tr>
<tr>
<td>Problem loan</td>
<td>A loan judged likely to produce a loss. Characterized by some occurrence such as late principal or interest payments. Includes any loan past due or on nonaccrual status. Also called a troubled loan.</td>
</tr>
<tr>
<td>Provision for loan losses</td>
<td>Income statement expense account showing amount added to LLR.</td>
</tr>
<tr>
<td>Recovery</td>
<td>Funds received on a loan previously charged off.</td>
</tr>
<tr>
<td>Restructured loan</td>
<td>A loan on which the bank has granted the borrower some concession because of the borrower’s financial difficulties.</td>
</tr>
<tr>
<td>Tier 1 capital</td>
<td>Stockholders’ equity + perpetual preferred stock + minority interest in consolidated subsidiaries.</td>
</tr>
<tr>
<td>Tier 2 capital</td>
<td>Limited-life preferred stock + subordinated debt + reserves for loan losses up to a specified maximum percent of risk-weighted assets (1.5 percent before 1993 and 1.25 percent after 1992).</td>
</tr>
<tr>
<td>Total capital</td>
<td>Tier 1 capital + Tier 2 capital. Tier 2 capital cannot exceed Tier 1 capital in Total capital.</td>
</tr>
<tr>
<td>Write-down</td>
<td>Reducing the book value of a loan by subtracting a portion of that value from the loan and from LLR.</td>
</tr>
</tbody>
</table>

References


