

Nonmonetary Effects of the Financial Crisis in the Propagation of the Great Depression

By BEN S. BERNANKE*

During 1930–33, the U.S. financial system experienced conditions that were among the most difficult and chaotic in its history. Waves of bank failures culminated in the shutdown of the banking system (and of a number of other intermediaries and markets) in March 1933. On the other side of the ledger, exceptionally high rates of default and bankruptcy affected every class of borrower except the federal government.

An interesting aspect of the general financial crises—most clearly, of the bank failures—was their coincidence in timing with adverse developments in the macroeconomy.¹ Notably, an apparent attempt at recovery from the 1929–30 recession² was stalled at the time of the first banking crisis (November–December 1930); the incipient recovery degenerated into a new slump during the mid-1931 panics; and the economy and the financial system both reached their respective low points at the time of the bank “holiday” of March 1933. Only with the New Deal’s rehabilitation of the financial system in 1933–35 did the economy begin its slow emergence from the Great Depression.

A possible explanation of these synchronous movements is that the financial system simply responded, without feedback, to the declines in aggregate output. This is contradicted by the facts that problems of the financial system tended to lead output de-

clines, and that sources of financial panics unconnected with the fall in U.S. output have been documented by many writers. (See Section IV below.)

Among explanations that emphasize the opposite direction of causality, the most prominent is the one due to Friedman and Schwartz. Concentrating on the difficulties of the banks, they pointed out two ways in which these worsened the general economic contraction: first, by reducing the wealth of bank shareholders; second, and much more important, by leading to a rapid fall in the supply of money. There is much support for the monetary view. However, it is not a complete explanation of the link between the financial sector and aggregate output in the 1930’s. One problem is that there is no theory of monetary effects on the real economy that can explain *protracted* nonneutrality. Another is that the reductions of the money supply in this period seems quantitatively insufficient to explain the subsequent falls in output. (Again, see Section IV.)

The present paper builds on the Friedman-Schwartz work by considering a third way in which the financial crises (in which we include debtor bankruptcies as well as the failures of banks and other lenders) may have affected output. The basic premise is that, because markets for financial claims are incomplete, intermediation between some classes of borrowers and lenders requires nontrivial market-making and information-gathering services. The disruptions of 1930–33 (as I shall try to show) reduced the effectiveness of the financial sector as a whole in performing these services. As the real costs of intermediation increased, some borrowers (especially households, farmers, and small firms) found credit to be expensive and difficult to obtain. The effects of this credit squeeze on aggregate demand helped convert the severe but not unprecedented downturn of 1929–30 into a protracted depression.

*Stanford Graduate School of Business and Hoover Institution. I received useful comments from too many people to list here by name, but I am grateful to each of them. The National Science Foundation provided partial research support.

¹This is documented more carefully in Sections I C and IV below.

²This paper does not address the causes of the initial 1929–30 downturn. Milton Friedman and Anna Schwartz (1963) have stressed the importance of the Federal Reserve’s “anti-speculative” monetary tightening. Others, such as Peter Temin (1976), have pointed out autonomous expenditure effects.

It should be stated at the outset that my theory does not offer a complete explanation of the Great Depression (for example, nothing is said about 1929–30). Nor is it necessarily inconsistent with some existing explanations.³ However, it does have the virtues that, first, it seems capable (in a way in which existing theories are not) of explaining the unusual length and depth of the depression; and, second, it can do this without assuming markedly irrational behavior by private economic agents. Since the reconciliation of the obvious inefficiency of the depression with the postulate of rational private behavior remains a leading unsolved puzzle of macroeconomics, these two virtues alone provide motivation for serious consideration of this theory.

There do not seem to be any exact antecedents of the present paper in the formal economics literature.⁴ The work of Lester Chandler (1970, 1971) provides the best historical discussions of the general financial crisis extant; however, he does not develop very far the link to macroeconomic performance. Beginning with Irving Fisher (1933) and A. G. Hart (1938), there is a literature on the macroeconomic role of inside debt; an interesting recent example is the paper by Frederic Mishkin (1978), which stresses household balance sheets and liquidity. Benjamin Friedman (1981) has written on the relationship of credit and aggregate activity. Hyman Minsky (1977) and Charles Kindleberger (1978) have in several places argued for the inherent instability of the financial system, but in doing so have had to depart from the assumption of rational economic behavior.⁵ None of the above authors has emphasized the effects of financial crisis on

the real costs of credit intermediation, the focus of the present work.

The paper is organized as follows: Section I presents some background on the 1930–33 financial crisis, its sources, and its correspondence with aggregate output movements. Section II begins the principal argument of the paper. I explain how the runs on banks and the extensive defaults could have reduced the efficiency of the financial sector in performing its intermediary functions. Some evidence of these effects is introduced.

Possible channels by which reduced financial efficiency might have affected output are discussed in Section III. Reduced-form estimation results, reported in Section IV, suggest that augmenting a purely monetary approach by my theory significantly improves the explanation of the financial sector-output connection in the short run. Section V looks at the persistence of these effects.

Some international aspects of the financial sector-aggregate output link are briefly discussed in Section VI and Section VII concludes.

I. The Financial Collapse: Some Background

The problems faced by the U.S. financial system between October 1930 and March 1933 have been described in detail by earlier authors,⁶ but it will be useful to recapitulate some principal facts here. Given this background, attention will be turned to the more central issues of the paper.

The two major components of the financial collapse were the loss of confidence in financial institutions, primarily commercial banks, and the widespread insolvency of debtors. I give short discussions of each of these components and of their joint relation to aggregate fluctuations.

A. The Failure of Financial Institutions

Most financial institutions (even semipublic ones, like the Joint Stock Land Banks) came under pressure in the 1930's. Some,

⁶See especially Chandler (1970, 1971) and Friedman and Schwartz.

³See Karl Brunner (1981) for a useful overview of contemporary theories of the depression. Also, see Robert Lucas and Leonard Rapping's article in Lucas (1981).

⁴This is especially true of the more recent work, which tends to ignore the nonmonetary effects of the financial crisis. Older writers often seemed to take the disruptive impact of the financial breakdown for granted.

⁵I do not deny the possible importance of irrationality in economic life; however, it seems that the best research strategy is to push the rationality postulate as far as it will go.

such as the insurance companies and the mutual savings banks, managed to maintain something close to normal operations. Others, like the building-and-loans (which, despite their ability to restrict withdrawals by depositors, failed in significant numbers) were greatly hampered in their attempts to carry on their business.⁷ Of most importance, however, were the problems of the commercial banks. The significance of the banking difficulties derived both from their magnitude and from the central role commercial banks played in the financial system.⁸

The great severity of the banking crises in the Great Depression is well known to students of the period. The percentages of operating banks which failed in each year from 1930 to 1933 inclusive were 5.6, 10.5, 7.8, and 12.9; because of failures and mergers, the number of banks operating at the end of 1933 was only just above half the number that existed in 1929.⁹ Banks that survived experienced heavy losses.

The sources of the banking collapse are best understood in the historical context. The first point to be made is that bank failures were hardly a novelty at the time of the depression. The U.S. system, made up as it was primarily of small, independent banks, had always been particularly vulnerable. (Countries with only a few large banks, such as Britain, France, and Canada, never had banking difficulties on the American scale.) The dominance of small banks in the United States was due in large part to a regulatory environment which reflected popular fears of large banks and "trusts"; for example, there were numerous laws restricting branch banking at both the state and national level. Com-

petition between the state and national banking systems for member banks also tended to keep the legal barriers to entry in banking very low.¹⁰ In this sort of environment, a significant number of failures was to be expected and probably was even desirable. Failures due to "natural causes" (such as the agricultural depression of the 1920's upon which many small, rural banks foundered) were common.¹¹

Besides the simple lack of economic viability of some marginal banks, however, the U.S. system historically suffered also from a more malign source of bank failures; namely, financial panics. The fact that liabilities of banks were principally in the form of fixed-price, callable debt (i.e., demand deposits), while many assets were highly illiquid, created the possibility of the perverse expectational equilibrium known as a "run" on the banks. In a run, fear that a bank may fail induces depositors to withdraw their money, which in turn forces liquidation of the bank's assets. The need to liquidate hastily, or to dump assets on the market when other banks are also liquidating, may generate losses that actually do cause the bank to fail. Thus the expectation of failure, by the mechanism of the run, tends to become self-confirming.¹²

An interesting question is why banks at this time relied on fixed-price demand deposits, when alternative instruments might have reduced or prevented the problem of runs.¹³ An answer is provided by Friedman and Schwartz: They pointed out that, before the establishment of the Federal Reserve in 1913, panics were usually contained by the practice of suspending convertibility of bank deposits into currency. This practice, typically initiated by loose organizations of urban

⁷Hart describes the problems of the building-and-loans. An interesting sidelight here is the additional strain on housing lenders caused by the existence of the Postal Savings System; see Maureen O'Hara and David Easley (1979).

⁸According to Raymond Goldsmith (1958), commercial banks held 39.6 percent of the assets of all financial intermediaries, broadly defined, in 1929. See his Table 11.

⁹Cyril Upham and Edwin Lamke (1934, p. 247). Since smaller banks were more likely to fail, the fraction of deposits represented by suspended banks was somewhat less. Eventual recovery by depositors was about 75 percent; see Friedman and Schwartz, p. 438.

¹⁰Benjamin Klebaner (1974) gives a good brief history of U.S. commercial banking.

¹¹Upham and Lamke, p. 247, report that approximately 2-3 percent of all banks in operation failed in each year of the 1920's.

¹²Douglas Diamond and Philip Dybvig (1981) formalize this argument. For an alternative analysis of the phenomenon of runs, see Robert Flood and Peter Garber (1981).

¹³For example, equity-like instruments, such as those used by modern money-market mutual funds, could have been used as the transactions medium. See Kenneth Cone (1982).

banks called clearinghouses, moderated the dangers of runs by making hasty liquidation unnecessary. In conjunction with the suspension of convertibility practice, the use of demand deposits created relatively little instability.¹⁴

However, with the advent of the Federal Reserve (according to Friedman-Schwartz), this roughly stable institutional arrangement was upset. Although the Federal Reserve introduced no specific injunctions against the suspension of convertibility, the clearinghouses apparently felt that the existence of the new institution relieved them of the responsibility of fighting runs. Unfortunately, the Federal Reserve turned out to be unable or unwilling to assume this responsibility.

No serious runs occurred between World War I and 1930; but the many pieces of bad financial news that came in from around the world in 1930–32 were like sparks around tinder. Runs were clearly an important part of the banking problems of this period. Some evidence emerges from contemporary accounts, including descriptions of specific events precipitating runs. Also notable is the fact that bank failures tended to occur in short spasms, rather than in a steady stream (see Table 1, col. 2, for monthly data on the deposits of failing banks). The problem was not arrested until government intervention became important in late 1932 and early 1933.

We see, then, that the banking crises of the early 1930's differed from earlier recorded experience both in magnitude and in the degree of danger posed by the phenomenon of runs. The result of this was that the behavior of almost the entire system was adversely affected, not just that of marginal banks. The bankers' fear of runs, as I shall argue below, had important macroeconomic effects.

B. Defaults and Bankruptcies

The second major aspect of the financial crisis (one that is currently neglected by historians) was the pervasiveness of debtor

insolvency. Given that debt contracts were written in nominal terms,¹⁵ the protracted fall in prices and money incomes greatly increased debt burdens. According to Evans Clark (1933), the ratio of debt service to national income went from 9 percent in 1929 to 19.8 percent in 1932–33. The resulting high rates of default caused problems for both borrowers and lenders.

The "debt crisis" touched all sectors. For example, about half of all residential properties were mortgaged at the beginning of the Great Depression; according to the *Financial Survey of Urban Housing* (reported in Hart), as of January 1, 1934,

The proportion of mortgaged owner-occupied houses with some interest or principal in default was in none of the twenty-two cities [surveyed] less than 21 percent (the figure for Richmond, Virginia); in half it was above 38 percent; in two (Indianapolis and Birmingham, Alabama) between 50 percent and 60 percent; and in one (Cleveland), 62 percent. For rented properties, percentages in default ran slightly higher. [p. 164]

Because of the long spell of low food prices, farmers were in more difficulty than homeowners. At the beginning of 1933, owners of 45 percent of all U.S. farms, holding 52 percent of the value of farm mortgage debt, were delinquent in payments (Hart, p. 138). State and local governments—many of whom tried to provide relief for the unemployed—also had problems paying their debts: As of March 1934, the governments of 37 of the 310 cities with populations over 30,000 and of three states had defaulted on obligations (Hart, p. 225).

In the business sector, the incidence of financial distress was very uneven. Aggregate corporate profits before tax were negative in 1931 and 1932, and after-tax retained earnings were negative in each year from 1930 to 1933 (Chandler, 1971, p. 102). But the subset

¹⁴Diamond and Dybvig derive this point formally, with some caveats.

¹⁵Finding an explanation for the lack of indexed debt during the deflationary 1930's—as in the inflationary 1970's—is a point on which I stumble.

of corporations holding more than \$50 million in assets maintained positive profits throughout this period, leaving the brunt to be borne by smaller companies. Solomon Fabricant (1935) reported that, in 1932 alone, the losses of corporations with assets of \$50,000 or less equalled 33 percent of total capitalization; for corporations with assets in the \$50,000–\$100,000 range, the comparable figure was 14 percent. This led to high rates of failure among small firms.

Although the deflation of the 1930's was unusually protracted, there had been a similar episode as recently as 1921–22 which had not led to mass insolvency. The seriousness of the problem in the Great Depression was due not only to the extent of the deflation, but also to the large and broad-based expansion of inside debt in the 1920's. Charles Persons surveyed the credit expansion of the predepression decade in a 1930 article: He reported that outstanding corporate bonds and notes increased from \$26.1 billion in 1920 to \$47.1 billion in 1928, and that non-federal public securities grew from \$11.8 billion to \$33.6 billion over the same period. (This may be compared with a 1929 national income of \$86.8 billion.) Perhaps more significantly, during the 1920's, small borrowers, such as households and unincorporated businesses, greatly increased their debts. For example, the value of urban real estate mortgages outstanding increased from \$11 billion in 1920 to \$27 billion in 1929, while the growth of consumer installment debt reflected the introduction of major consumer durables to the mass market.

Like the banking crises, then, the debt crisis of the 1930's was not qualitatively a new phenomenon; but it represented a break with the past in terms of its severity and pervasiveness.

C. *Correlation of the Financial Crisis with Macroeconomic Activity*

The close connection of the stages of the financial crisis (especially the bank failures) with changes in real output has been noted by Friedman and Schwartz and by others. An informal review of this connection is

facilitated by the monthly data in Table 1. Column 1 is an index of real industrial production. Columns 2 and 3 are the (nominal) liabilities of failing banks and nonbank commercial businesses, respectively.

The industrial production series reveals that a recession began in the United States during 1929. By late 1930, the downturn, although serious, was still comparable in magnitude to the recession of 1920–22; as the decline slowed, it would have been reasonable to expect a brisk recovery, just as in 1922.

With the first banking crisis, however, there came what Friedman and Schwartz called a "change in the character of the contraction" (p. 311). The economy first flattened out, then went into a new tailspin just as the banks began to fail again in June 1931.

A lengthy slide of both the general economy and the financial system followed. The banking situation calmed in early 1932, and nonbank failures peaked shortly thereafter. A new recovery attempt began in August, but failed within a few months.¹⁶ In March 1933, the bottom was reached for both the financial system and the economy as a whole. Measures taken after the banking holiday ended the bank runs and greatly reduced the burden of debt. Simultaneously aggregate output began a recovery that was sustained until 1937.

The leading explanation of the correlation between the conditions of the financial sector and of the general economy is that of Friedman and Schwartz, who stressed the effects of the banking crises on the supply of money. I agree that money was an important factor in 1930–33, but, because of reservations cited in the introduction, I doubt that it completely explains the financial sector-aggregate output connection. This motivates

¹⁶Judging by Table 1, the failure of this recovery seems to be unrelated to financial sector difficulties. However, accounts from the time suggest that the banking crisis of late 1932 and early 1933 (which ended in the banking holiday) was in fact quite severe; see Susan Kennedy (1973). The relatively low reported rate of bank failures at this time may be an artifact of state moratoria, restrictions on withdrawals, and other interventions.

TABLE 1—SELECTED MACROECONOMIC DATA, JULY 1929—MARCH 1933

Month	IP	Banks	Fails	$\Delta L/IP$	L/DEP	DIF
1929J	114	60.8	32.4	.163	.851	2.31
A	114	6.7	33.7	.007	.855	2.33
S	112	9.7	34.1	.079	.860	2.33
O	110	12.5	31.3	.177	.865	2.50
N	105	22.3	52.0	.121	.854	2.68
D	100	15.5	62.5	-.214	.851	2.59
1930J	100	26.5	61.2	-.228	.837	2.49
F	100	32.4	51.3	-.102	.834	2.48
M	98	23.2	56.8	.076	.835	2.44
A	98	31.9	49.1	.058	.826	2.33
M	96	19.4	55.5	-.028	.820	2.41
J	93	57.9	63.1	.085	.818	2.53
J	89	29.8	29.8	-.055	.802	2.52
A	86	22.8	49.2	-.027	.800	2.47
S	85	21.6	46.7	.008	.799	2.41
O	83	19.7	56.3	-.010	.791	2.73
N	81	179.9	55.3	-.067	.777	3.06
D	79	372.1	83.7	-.144	.775	3.49
1931J	78	75.7	94.6	-.187	.763	3.21
F	79	34.2	59.6	-.144	.747	3.08
M	80	34.3	60.4	-.043	.738	3.17
A	80	41.7	50.9	-.104	.722	3.45
M	80	43.2	53.4	-.133	.706	3.99
J	77	190.5	51.7	-.120	.707	4.23
J	76	40.7	61.0	-.013	.704	3.93
A	73	180.0	53.0	-.103	.706	4.29
S	70	233.5	47.3	-.050	.713	4.82
O	68	471.4	70.7	-.310	.716	5.41
N	67	67.9	60.7	-.101	.726	5.30
D	66	277.1	73.2	-.120	.732	6.49
1932J	64	218.9	96.9	-.117	.745	4.87
F	63	51.7	84.9	-.138	.757	4.76
M	62	10.9	93.8	-.183	.744	4.91
A	58	31.6	101.1	-.225	.718	6.78
M	56	34.4	83.8	-.154	.696	7.87
J	54	132.7	76.9	-.170	.689	7.93
J	53	48.7	87.2	-.219	.677	7.21
A	54	29.5	77.0	-.130	.662	4.77
S	58	13.5	56.1	-.091	.641	4.19
O	60	20.1	52.9	-.095	.623	4.44
N	59	43.3	53.6	-.133	.602	4.79
D	58	70.9	64.2	-.039	.596	5.07
1933J	58	133.1	79.1	-.139	.576	4.79
F	57	62.2	65.6	-.059	.583	4.09
M	54	3276.3 ^a	48.5	-.767 ^a	.607 ^a	4.03

Notes: IP = seasonally adjusted index of industrial production, 1935–39 = 100; *Federal Reserve Bulletin*.

Banks = deposits of failing banks, \$millions; *Federal Reserve Bulletin*.

Fails = liabilities of failing commercial businesses, \$millions; *Survey of Current Business*.

$\Delta L/PI$ = ratio of net extensions of commercial bank loans to (monthly) personal income; from *Banking and Monetary Statistics and National Income*.

L/D = ratio of loans outstanding to the sum of demand and time deposits, weekly reporting banks; *Banking and Monetary Statistics*.

DIF = difference (in percentage points) between yields on Baa corporate bonds and long-term U.S. government bonds; *Banking and Monetary Statistics*.

^aA national bank holiday was declared in March 1933.

my study of a nonmonetary channel through which an additional impact of the financial crisis may have been felt.

II. The Effect of the Crisis on the Cost of Credit Intermediation

This paper posits that, in addition to its effects via the money supply, the financial crisis of 1930–33 affected the macroeconomy by reducing the quality of certain financial services, primarily credit intermediation. The basic argument is to be made in two steps. First, it must be shown that the disruption of the financial sector by the banking and debt crises raised the real cost of intermediation between lenders and certain classes of borrowers. Second, the link between higher intermediation costs and the decline in aggregate output must be established. I present here the first step of the argument, leaving the second to be developed in Sections III–V.

In order to discuss the quality of performance of the financial sector, I must first describe the real services that the sector is supposed to provide. The specification of these services depends on the model of the economy one has in mind. We shall clearly not be interested in economies of the sort described by Eugene Fama (1980), in which financial markets are complete and information/transactions costs can be neglected. In such a world, banks and other intermediaries are merely passive holders of portfolios. Banks' choice of portfolios or the scale of the banking system can never make any difference in this case, since depositors can offset any action taken by banks through private portfolio decisions.¹⁷

As an alternative to the Fama complete-markets world, consider the following stylized description of the economy. Let us suppose that savers have many ways of transferring resources from present to future, such as holding real assets or buying the liabilities of

governments or corporations on well-organized exchanges. One of the options savers have is to lend resources to a banking system. The banks also have a menu of different assets to choose from. Assume, however, that banks specialize in making loans to small, idiosyncratic borrowers whose liabilities are too few in number to be publicly traded. (Here is where the complete-markets assumption is dropped.)

The small borrowers to whom the banks lend will be taken, for simplicity, to be of two extreme types, "good" and "bad." Good borrowers desire loans in order to undertake individual-specific investment projects. These projects generate a random return from a distribution whose mean will be assumed always to exceed the social opportunity cost of investment. If this risk is nonsystematic, lending to good borrowers is socially desirable. Bad borrowers try to look like good borrowers, but in fact they have no "project." Bad borrowers are assumed to squander any loan received in profligate consumption, then to default. Loans to bad borrowers are socially undesirable.

In this model, the real service performed by the banking system is the differentiation between good and bad borrowers.¹⁸ For a competitive banking system, I define the *cost of credit intermediation (CCI)* as being the cost of channeling funds from the ultimate savers/lenders into the hands of good borrowers. The *CCI* includes screening, monitoring, and accounting costs, as well as the expected losses inflicted by bad borrowers. Banks presumably choose operating procedures that minimize the *CCI*. This is done by developing expertise at evaluating potential borrowers; establishing long-term relationships with customers; and offering loan conditions that encourage potential borrowers to self-select in a favorable way.¹⁹

Given this simple paradigm, I can describe the effects of the two main components of

¹⁷It should be noted that the phenomena emphasized by Friedman and Schwartz—the effects of the contraction of the banking system on the quantity of the transactions medium and on real output—are also impossible in a complete-markets world.

¹⁸To concentrate on credit intermediation, I neglect the transactions and other services performed by banks.

¹⁹See Dwight Jaffee and Thomas Russell (1976) and Joseph Stiglitz and Andrew Weiss (1981) on the way banks induce favorable borrower self-selection.

the financial crisis on the efficiency of the credit allocation process (i.e., on the *CCI*).

A. *Effect of the Banking Crises on the CCI*

The banking problems of 1930–33 disrupted the credit allocation process by creating large, unplanned changes in the channels of credit flow. Fear of runs led to large withdrawals of deposits, precautionary increases in reserve-deposit ratios, and an increased desire by banks for very liquid or rediscountable assets. These factors, plus the actual failures, forced a contraction of the banking system's role in the intermediation of credit.²⁰ Some of the slack was taken up by the growing importance of alternative channels of credit (see below). However, the rapid switch away from the banks (given the banks' accumulated expertise, information, and customer relationships) no doubt impaired financial efficiency and raised the *CCI*.²¹

It would be useful to have a direct measure of the *CCI*; unfortunately, no really satisfactory empirical representation of this concept is available. Reported commercial loan rates reflect loans that are actually made, not the shadow cost of bank funds to a representative potential borrower; since banks in a period of retrenchment make only the safest and highest-quality loans, measured loan rates may well move inversely to the *CCI*. I obtained a number of interesting results using the yield differential between Baa corporate bonds and U.S. government bonds as a proxy for the *CCI*; however, the use of the Baa rate is not consistent with my story that bank borrowers are those whose liabilities are too few to be publicly traded.

While we cannot observe directly the effects of the banking troubles on the *CCI*, we can see their impact on the extension of bank credit: Table 1 gives some illustrative data. Column 4 gives, as a measure of the flow of

bank credit, the monthly change in bank loans outstanding, normalized by monthly personal income.²² One might have expected the loan-change-to-income ratio to be driven primarily by loan demand and thus by the rate of production. Comparison with the first two columns of Table 2 shows, however, that the banking crises were as important a determinant of this variable as output. For example, except for a brief period of liquidation of speculation loans after the stock market crash, credit outstanding declined very little before October 1930—this despite a 25 percent fall in industrial production that had occurred by that time. With the first banking crisis of November 1930, however, a long period of credit contraction was initiated. The shrinkage of credit shared the rhythm of the banking crises; for example, in October 1931, the worst month for bank failure before the bank holiday, net credit reduction was a record 31 percent of personal income.²³

The fall in bank loans after November 1930 was not simply a balance sheet reflection of the decline in deposits. Column 5 in Table 1 gives the monthly ratio of outstanding bank loans to the sum of demand and time deposits. This ratio declined sharply as banks switched out of loans and into more liquid investments.

The perception that the banking crises and the associated scrambles for liquidity exerted a deflationary force on bank credit was shared by writers of the time. A 1932 National Industrial Conference Board survey of

²²In the construction of the bank loans series, data from weekly reporting member banks (which held about 40 percent of all bank loans) were used to interpolate between less frequent aggregate observations. Note that, for our purposes, looking at the change in loans is preferable to considering the stock of real loans outstanding: In a regime of nominally contracted debt and sharp unanticipated deflation, stability of the stock of real debt does *not* signal a comfortable situation for borrowers.

²³The effect of bank failures on credit outstanding is somewhat exaggerated by the fact that the credit contraction measure includes the loans of suspending banks that were not transferred to other banks; however, I estimate that this accounting convention is responsible for less than one-eighth of the total (measured) credit contraction between October 1930 and February 1933.

²⁰For an interesting contemporary account of this process, see the article by Eugene H. Burris in the *American Banker*, October 15, 1931.

²¹Since intermediation resources could have been shifted out of the beleaguered banking sector (given enough time), mine is basically a costs-of-adjustment argument.

credit conditions reported that “During 1930, the shrinkage of commercial loans no more than reflected business recession. During 1931 and the first half of 1932 (the period studied), it unquestionably represented pressure by banks on customers for repayment of loans and refusal by banks to grant new loans” (p. 28). Other contemporary sources tended to agree (see, for example, Chandler, 1971, pp. 233–39, for references).

Two other observations about the contraction of bank credit can be made. First, the class of borrowers most affected by credit reductions were households, farmers, unincorporated businesses, and small corporations; this group had the highest direct or indirect reliance on bank credit. Second, the contraction of bank credit was twice as large as that of other major countries, even those which experienced comparable output declines (Klebaner, p. 145).

The fall in bank loans outstanding was partly offset by the relative expansion of alternative forms of credit. In the area of consumer finance, retail merchants, service creditors, and nonbank lending agencies improved their position relative to banks and primarily bank-supported installment finance companies (Rolf Nugent, 1939, pp. 114–16). Small firms during this period significantly reduced their traditional reliance on banks in favor of trade credit (Charles Merwin, 1942, pp. 5 and 75). But, as argued above, in a world with transactions costs and the need to discriminate among borrowers, these shifts in the loci of credit intermediation must have at least temporarily reduced the efficiency of the credit allocation process, thereby raising the effective cost of credit to potential borrowers.

B. *The Effect of Bankruptcies on the CCI*

I turn now to a brief discussion of the impact of the increase in defaults and bankruptcies during this period on the cost of credit intermediation.

The very existence of bankruptcy proceedings, rather than being an obvious or natural phenomenon, raises deep questions of economic theory. Why, for example, do the

creditor and defaulted debtor make the payments to third parties (lawyers, administrators) that these proceedings entail, instead of somehow agreeing to divide those payments between themselves? In a complete-markets world, bankruptcy would never be observed; this is because complete state-contingent loan agreements would uniquely define each party's obligations in all possible circumstances, rendering third-party arbitration unnecessary. That we do observe bankruptcies, in our incomplete-markets world, suggests that creditors and debtors have found the combination of simple loan arrangements and *ex post* adjudication by bankruptcy (when necessary) to be cheaper than attempting to write and enforce complete state-contingent contracts.

To be more concrete, let us use the “good borrower-bad borrower” example. In writing a loan contract with a potential borrower, the bank has two polar options. First, it might try to approximate the complete state-contingent contract by making the borrower's actions part of the agreement and by allowing repayment to depend on the outcome of the borrower's project. This contract, if properly written and enforced, would completely eliminate the possibility of either side not being able to meet its obligations; its obvious drawback is the cost of monitoring which it involves. The bank's other option is to write a very simple agreement (“payment of such-amount to be made on such-date”), then to make the loan only if it believes that the borrower is likely to repay. The second approach usually dominates the first, of course, especially for small borrowers.

A device which makes the cost advantage of the simpler approach even greater is the use of collateral. If the borrower has wealth that can be attached by the bank in the event of nonpayment, the bank's risk is low. Moreover, the threat of loss of collateral provides the right incentives for borrowers to use loans only for profitable projects. Thus, the combination of collateral and simple loan contracts helps to create a low effective *CCI*.

A useful way to think of the 1930–33 debt crisis is as the progressive erosion of borrowers' collateral relative to debt burdens.

As the representative borrower became more and more insolvent, banks (and other lenders as well) faced a dilemma. Simple, noncontingent loans faced increasingly higher risks of default; yet a return to the more complex type of contract involved many other costs. Either way, debtor insolvency necessarily raised the *CCI* for banks.

One way for banks to adjust to a higher *CCI* is to increase the rate that they charge borrowers. This may be counterproductive, however, if higher interest charges increase the risk of default. The more usual response is for banks just not to make loans to some people that they might have lent to in better times. This was certainly the pattern in the 1930's. For example, it was reported that the extraordinary rate of default on residential mortgages forced banks and life insurance companies to "practically stop making mortgage loans, except for renewals" (Hart, p. 163). This situation precluded many borrowers, even with good projects, from getting funds, while lenders rushed to compete for existing high-grade assets. As one writer of the time, D. M. Frederiksen, put it:

We see money accumulating at the centers, with difficulty of finding safe investment for it; interest rates dropping down lower than ever before; money available in great plenty for things that are obviously safe, but not available at all for things that are in fact safe, and which under normal conditions would be entirely safe (and there are a great many such), but which are now viewed with suspicion by lenders.
[1931, p. 139]

As this quote suggests, the idea that the low yields on Treasury or blue-chip corporation liabilities during this time signalled a general state of "easy money" is mistaken; money was easy for a few safe borrowers, but difficult for everyone else.

An indicator of the strength of lender preferences for safe, liquid assets (and hence of the difficulty of risky borrowers in obtaining funds) is the yield differential between Baa corporate bonds and Treasury

bonds (Table 1, column 6). Because this variable contains no adjustment for the reclassification of firms into higher risk categories, it tends to understate the true difference in yields between representative risky and safe assets. Nevertheless, this indicator showed some impressive shifts, going from 2.5 percent during 1929–30 to nearly 8 percent in mid-1932. (The differential never exceeded 3.5 percent in the sharp 1920–22 recession.) The yield differential reflected changing perceptions of default risk, of course; but note also the close relationship of the differential and the banking crises (a fact first pointed out by Friedman and Schwartz). Bank crises depressed the prices of lower-quality investments as the fear of runs drove banks into assets that could be used as reserves or for rediscounting. This effect of bank portfolio choices on an asset price could not happen in a Fama-type, complete-markets world.

Finally, it is instructive to consider the experience of a country that had a debt crisis without a banking crisis. Canada entered the Great Depression with a large external debt, much of it payable in foreign currencies. The combination of deflation and the devaluation of the Canadian dollar led to many defaults. Internally, debt problems in agriculture and in mortgage markets were as severe as in the United States, while major industries (notably pulp and paper) experienced many bankruptcies (A. E. Safarian, 1959, ch. 7). Although Canadian bankers did not face serious danger of runs, they shifted away from loans to safer assets. This shift toward safety and liquidity, though less pronounced than in the U.S. case, drew criticism from all facets of Canadian society. The *American Banker* of December 6, 1932, reported the following complaint from a non-populist Canadian politician:

The chief criticism of our present system appears to be that in good times credit is expanded to great extremes... but, when the pinch of hard times is first being felt, credit is suddenly and drastically restricted by the banks... At the present time, loans are only being made when the banks have a very wide

margin of security and every effort is being made to collect outstanding loans. All our banks are reaching out in an endeavor to liquefy their assets....

[p. 1]

Canadian lenders other than banks also tried to retrench: According to the *Financial Post*, May 14, 1932, "Insurance, trust, and loan companies were increasingly unwilling to lend funds with real estate and rental values falling, a growing number of defaults of interest and principal, the increasing burden of property taxes, and legislation which adversely affected creditors" (quoted in Safarian, p. 130).

More careful study of the Canadian experience in the Great Depression would be useful. However, on first appraisal, that experience does not seem to be inconsistent with the point that even good borrowers may find it more difficult or costly to obtain credit when there is extensive insolvency. The debt crisis should be added to the banking crises as a potential source of disruption of the credit system.

III. Credit Markets and Macroeconomic Performance

If it is taken as given that the financial crises during the depression did interfere with the normal flows of credit, it still must be shown how this might have had an effect on the course of the aggregate economy.

There are many ways in which problems in credit markets might potentially affect the macroeconomy. Several of these could be grouped under the heading of "effects on aggregate supply." For example, if credit flows are dammed up, potential borrowers in the economy may not be able to secure funds to undertake worthwhile activities or investments; at the same time, savers may have to devote their funds to inferior uses. Other possible problems resulting from poorly functioning credit markets include a reduced feasibility of effective risk sharing and greater difficulties in funding large, indivisible projects. Each of these might limit the economy's productive capacity.

These arguments are reminiscent of some ideas advanced by John Gurley and E. S. Shaw (1955), Ronald McKinnon (1973), and others in an economic development context. The claim of this literature is that immature or repressed financial sectors cause the "fragmentation" of less developed economies, reducing the effective set of production possibilities available to the society.

Did the financial crisis of the 1930's turn the United States into a "temporarily underdeveloped economy" (to use Bob Hall's felicitous phrase)? Although this possibility is intriguing, the answer to the question is probably no. While many businesses did suffer drains of working capital and investment funds, most larger corporations entered the decade with sufficient cash and liquid reserves to finance operations and any desired expansion (see, for example, Friedrich Lutz, 1945). Unless it is believed that the outputs of large and of small businesses are not potentially substitutes, the aggregate supply effect must be regarded as not of great quantitative importance.

The reluctance of even cash-rich corporations to expand production during the depression suggests that consideration of the aggregate demand channel for credit market effects on output may be more fruitful. The aggregate demand argument is in fact easy to make: A higher cost of credit intermediation for some borrowers (for example, households and smaller firms) implies that, for a given *safe* interest rate, these borrowers must face a higher effective cost of credit. (Indeed, they may not be able to borrow at all.) If this higher rate applies to household and small firm borrowing but not to their saving (they may only earn the safe rate on their savings), then the effect of higher borrowing costs is unambiguously to reduce their demands for current-period goods and services. This pure substitution effect (of future for present consumption) is easily derived from the classical two-period model of savings.²⁴

²⁴The classical model may be augmented, if the reader desires, by considerations of liquidity constraints, bankruptcy costs, or risk aversion; see my 1981 paper.

Assume that the behavior of borrowers unaffected by credit market problems is unchanged. Then the paragraph above implies that, for a given safe rate, an increase in the cost of credit intermediation reduces the total quantity of goods and services currently demanded. That is, the aggregate demand curve, drawn as a function of the safe rate, is shifted downward by a financial crisis. In any macroeconomic model one cares to use, this implies lower output and lower safe interest rates. Both of these outcomes characterized 1930–33, of course.

Some evidence on the magnitude of the effect of the financial market problems on aggregate output is now presented.

IV. Short-Run Macroeconomic Impacts of the Financial Crisis

This section studies the short-run or “impact” effects of the financial crisis. For this purpose, I use only monthly data on the relevant variables. In addition, rather than consider the 1929–33 episode outside of its context, I have widened the sample to include the entire interwar period (January 1919–December 1941).

Section I.C above has already given some evidence of the relationship between the troubles of the financial sector and those of the economy as a whole. However, support for the thesis of this paper requires that nonmonetary effects of the financial crisis on output be distinguished from the monetary effects studied by Friedman and Schwartz. My approach will be to fit output equations using monetary variables, then to show that adding proxies for the financial crisis substantially improves the performance of these equations. Comparison of financial to totally nonfinancial sources of the Great Depression, such as those suggested by Temin, is left to future research.

To isolate the purely monetary influences on the economy, one needs a structural explanation of the money-income relationship. Lucas (1972) has presented a formal model in which monetary shocks affect production decisions by causing confusion about the price level. Influenced by this work, most recent empirical studies of the role of money

have related national income to measures of “unanticipated” changes in money or prices.²⁵

The most familiar way of constructing a proxy for unanticipated components of a variable is the two-step method of Robert Barro (1978), in which the residuals from a first-stage prediction equation for (say) money are employed as the independent variables in a second-stage regression. I experimented with both the Barro approach and some alternatives.²⁶ Since my conclusions were unaffected by choice of technique, I report here only the Barro-type results.

In the spirit of the Lucas-Barro analysis, I considered the effects of both “money shocks” and “price shocks” on output. Money shocks ($M - M^e$) were defined as the residuals from a regression of the rate of growth of $M1$ on four lags of the growth rates of industrial production, wholesale prices, and $M1$ itself; price shocks ($P - P^e$) were defined symmetrically.²⁷ I used ordinary least squares to estimate the effects of money and price shocks on the rate of growth of industrial production, relative to trend.

The basic regression results for the interwar sample period are given as equations (1) and (2) in Table 2. These two equations are of interest, independently of the other results of this paper. The estimated “Lucas supply curve,” equation (2), shows an effect of price shocks on output that is statistically and economically significant. As such, it complements the results of Thomas Sargent (1976), who found a similar relationship for the postwar. The relationship of output to money surprises, equation (1), is a bit weaker. The fact that we discover a smaller role for money in the monthly data than does Paul Evans (1981) is primarily the result of our inclusion of lagged values of production on the right-hand side. This inclusion seems justified both on statistical grounds and for

²⁵A notable exception is Mishkin (1982).

²⁶Principal alternatives tried were 1) the use of anticipated as well as unanticipated quantities as explanatory variables; and 2) reestimation of some equations by the more efficient but computationally more complex method of Andrew Abel and Mishkin (1981).

²⁷The first-stage regressions were unsurprising and, for the sake of space, are not reported.

TABLE 2—ESTIMATED OUTPUT EQUATIONS

$$(1) \quad Y_t = \begin{matrix} .623 & Y_{t-1} & - & .144 & Y_{t-2} & + & .407 & (M - M^e)_t & + & .141 & (M - M^e)_{t-1} \\ (10.21) & & & (-2.37) & & & (3.42) & & & (1.16) \end{matrix}$$

$$+ \begin{matrix} .051 & (M - M^e)_{t-2} & + & .144 & (M - M^e)_{t-3} \\ (0.42) & & & (1.19) \end{matrix}$$

s.e. = .0272 *D.W.* = 2.02 Sample: 1/19-12/41

$$(2) \quad Y_t = \begin{matrix} .582 & Y_{t-1} & - & .118 & Y_{t-2} & + & .533 & (P - P^e)_t & + & .350 & (P - P^e)_{t-1} \\ (9.50) & & & (-1.76) & & & (5.33) & & & (3.33) \end{matrix}$$

$$+ \begin{matrix} .036 & (P - P^e)_{t-2} & + & .069 & (P - P^e)_{t-3} \\ (0.34) & & & (0.66) \end{matrix}$$

s.e. = .0260 *D.W.* = 2.01 Sample: 1/19-12/41

$$(3) \quad Y_t = \begin{matrix} .613 & Y_{t-1} & - & .159 & Y_{t-2} & + & .332 & (M - M^e)_t & + & .113 & (M - M^e)_{t-1} & + & .110 & (M - M^e)_{t-2} \\ (9.86) & & & (-2.63) & & & (2.92) & & & (0.99) & & & (0.96) \end{matrix}$$

$$+ \begin{matrix} .156 & (M - M^e)_{t-3} & - & .869E-04 & DBANKS_t & - & .406E-04 & DBANKS_{t-1} \\ (1.38) & & & (-4.24) & & & (-1.93) \end{matrix}$$

$$- \begin{matrix} .258E-03 & DFAILS_t & - & .325E-03 & DFAILS_{t-1} \\ (-1.95) & & & (-2.47) \end{matrix}$$

s.e. = .0249 *D.W.* = 1.99 Sample: 1/21-12/41

$$(4) \quad Y_t = \begin{matrix} .615 & Y_{t-1} & - & .131 & Y_{t-2} & + & .455 & (P - P^e)_t & + & .231 & (P - P^e)_{t-1} & - & .004 & (P - P^e)_{t-2} \\ (9.76) & & & (-2.13) & & & (3.99) & & & (1.97) & & & (-0.03) \end{matrix}$$

$$+ \begin{matrix} .024 & (P - P^e)_{t-3} & - & .799E-04 & DBANKS_t & - & .337E-04 & DBANKS_{t-1} \\ (0.22) & & & (-4.03) & & & (-1.66) \end{matrix}$$

$$- \begin{matrix} .202E-03 & DFAILS_t & - & .242E-03 & DFAILS_{t-1} \\ (-1.52) & & & (-1.83) \end{matrix}$$

s.e. = .0246 *D.W.* = 1.98 Sample: 1/21-2/41

Notes: Y_t = rate of growth of industrial production (*Federal Reserve Bulletin*), relative to exponential trend.
 $(M - M^e)_t$ = rate of growth of $M1$, nominal and seasonally adjusted (Friedman and Schwartz, Table 4-1), less predicted rate of growth.

$(P - P^e)_t$ = rate of growth of wholesale price index (*Federal Reserve Bulletin*), less predicted rate of growth.

$DBANKS_t$ = first difference of deposits of failing banks (deflated by wholesale price index).

$DFAILS_t$ = first difference of liabilities of failing businesses (deflated by wholesale price index).

Data are monthly; *t*-statistics are shown in parentheses.

the economic reason that costs of adjusting production can be presumed to create a serial dependence in output. Like Evans, I was not able to find effects of money (or prices) lagged more than three months.

While these regression results exhibit statistical significance and the expected signs for coefficients, they are disappointing in the following sense: When equations (1) and (2)

are used to perform dynamic simulations of the path of output between mid-1930 and the bank holiday of March 1933, they capture no more than half of the total decline of output during the period. This is the basis of the comment in the introduction that the declines in money seem "quantitatively insufficient" to explain what happened to output in 1930-33.

Given the basic regressions (1) and (2), the next step was to examine the effects of including proxies for the nonmonetary financial impact as explanators of output. Based on the earlier analysis of this paper, the most obvious such proxies are the deposits of failing banks and the liabilities of failing businesses.

A preliminary problem with the bank deposits series that needs to be discussed is the value for March 1933, the month of the bank holiday. As can be seen in Table 1, the deposits of banks suspended in March 1933 is seven times that of the next worse month. The question arises if any adjustment should be made to that figure before running the regressions.

We believe that it would be a mistake to eliminate totally the bank holiday episode from the sample. According to contemporary accounts, rather than being an orderly and planned-in-advance policy, the imposition of the holiday was a forced response to the most panicky and chaotic financial conditions of the period. The deposits of suspended banks figure for March, as large as it is, reflects not all closed banks but only those not licensed to reopen by June 30, 1933. Of these banks, most were liquidated or placed in receivership; less than 25 percent had been licensed to reopen as of December 31, 1936.²⁸ Qualitatively, then, the March 1933 episode resembled the earlier crises; it would be throwing away information not to include in some way the effects of this crisis and of its resolution on the economy.

On the other hand, the mass closing of banks by government action probably created less confusion and fear of future crises than would have a similar number of suspensions occurring without government intervention. As a conservative compromise, I assumed that the "supervised" bank closings of March 1933 had the same effect as an "unsupervised" bank crisis involving 15 percent as much in frozen deposits. This scales down the March 1933 episode to about the size of the events of October 1931. The sensitivity of the results to this assumption is as follows:

increasing the amount of importance attributed to the March 1933 crisis raises the magnitude and statistical significance of the measured effects of the financial crises on output. (It is in this sense that the 15 percent figure is conservative.) However, the bank failure coefficients in the regressions retained high significance even when less weight was given to March 1933.

I turn now to the results of adding (real) deposits of failing banks and liabilities of failing businesses to the output equations (see equations (3) and (4) in Table 2). The sample period begins in 1921 because of the unavailability of data on monthly bank failures before then. In both regressions, current and lagged first differences of the added variables enter the explanation of the growth rate of industrial production (relative to trend) with the expected sign and, taken jointly, with a high level of statistical significance. The magnitudes and significance of the coefficients of money and price shocks are not much changed. This provides at least a tentative confirmation that nonmonetary effects of the financial crisis augmented monetary effects in the short-run determination of output.

Some alternative proxies for the nonmonetary component of the financial crisis were also tried. For the sake of space, only a summary of these results is given. 1) To examine the direct effects of the contraction of bank credit on the economy, I began by regressing the rate of growth of bank loans on current and lagged values of suspended bank deposits and of failing business liabilities. (This regression indicated a powerful negative effect of financial crisis on bank loans.) The fitted series from this regression was used as a proxy for the portion of the credit contraction induced by the financial crisis. In the presence of money or price shocks, the effect of a decline in this variable on output was found to be negative for two months, positive for the next two months, then strongly negative for the fifth and sixth months after the decline. For the period from 1921 until the bank holiday, and with monetary variables included, the total effect of credit contraction on output (as measured by the sum of lag coefficients in a polynomial

²⁸ *Federal Reserve Bulletin*, 1937, pp. 866-67.

distributed lag) was large (comparable to the monetary effect), negative, and significant at the 95 percent level. For the entire interwar sample, however, the statistical significance of this variable was much reduced. This last result is due to the fact that the recovery of 1933–41 was financed by nonbank sources, with bank loans remaining at a low level.

2) Another proxy for the financial crisis that was tried was the differential between Baa corporate bond yields and the yields on U.S. bonds. As described in Section I.C, this variable responded strongly to both bank crises and the problems of debtors, and as such was a sensitive indicator of financial market conditions. The yield differential variable turned out to enter very strongly as an explainer of current and future output growth, overall and in every subsample. As much of this predictive power was no doubt due to pure financial market anticipations of future output declines, I also put the differential variable through a first-stage regression on the liabilities of bank and business failures. Assuming that these latter variables themselves were not determined by anticipations of future output declines (see below), the use of the fitted series from this regression “purged” the differential variable of its pure anticipatory component. The fitted series entered the output equations less strongly than the raw series, but it retained the right sign and statistical significance at the 95 percent confidence level.

In almost every case, then, the addition of proxies for the general financial crisis improved the purely monetary explanation of short-run (monthly) output movements. This finding was robust to the obvious experiments. For example, with the above-noted exception of the credit variable in 1933–41, coefficients remained roughly stable over subsamples. Another experiment was to include free dummy variables for each quarter from 1931:I to 1932:IV in the above regressions. The purpose of this was to test the suggestion that our results are only a reflection of the fact that both the output and financial crisis variables “moved a lot” during 1930–33. The rather surprising discovery was that the inclusion of the dummies *increased* the magnitude and statistical sig-

nificance of the coefficients on bank and business failures. Finally, the economic significance of the results was tested by using the various estimated equations to run dynamic simulations of monthly levels of industrial production (relative to trend) for mid-1930 to March 1933. Relative to the pure money-shock and price-shock simulations described above, the equations including financial crisis proxies did well. Equations (3) and (4) reduced the mean squared simulation error over (1) and (2) by about 50 percent. The other (nonreported) equations did better; for example, those using the yield differential variable reduced the *MSE* of simulation from 90 to 95 percent.

These results are promising. However, a caveat must be added: To conclude that the observed correlations support the theory outlined in this paper requires an additional assumption, that failures of banks and commercial firms are not caused by anticipations of (future) changes in output. To the extent that, say, bank runs are caused by the receipt of bad news about next month’s industrial production, the fact that bank failures tend to lead production declines does not prove that the bank problems are helping to cause the declines.²⁹

While it may not be possible to convince the determined skeptic that bank and business failures are not purely anticipatory phenomena, a good case can be made against that position. For example, while in some cases a bad sales forecast may induce a firm to declare bankruptcy, more often that option is forced by insolvency (a result of past business conditions). For banks, it might well be argued that not only are failures relatively independent of anticipations about output, but that they are not simply the product of current and past output performance either: First, banking crises had never previous to this time been a necessary result of declines in output.³⁰ Second, Friedman and Schwartz, as well as other writers, have identified

²⁹Actually, a similar criticism might be made of Barro’s work and my own money and price regressions.

³⁰Philip Cagan (1965) makes this point; see pp. 216, 227–28. The 1920–22 recession, for example, did not generate any banking problems.

specific events that were important sources of bank runs during 1930–33. These include the revelation of scandal at the Bank of the United States (a private bank, which in December 1930 became the largest bank to fail up to that time); the collapse of the Kreditanstalt in Austria and the ensuing financial panics in central Europe; Britain's going off gold; the exposure of huge pyramiding schemes in the United States and Europe; and others, all connected very indirectly (if at all) with the path of industrial production in the United States.

If it is accepted that bank suspensions and business bankruptcies were the product of factors beyond pure anticipations of output decline, then the evidence of this section supports the view that nonmonetary aspects of the financial crisis were at least part of the propagatory mechanism of the Great Depression. If it is further accepted that the financial crisis contained large exogenous components (there is evidence for this in the case of the banking panics), then there are elements of causality in the story as well.

V. Persistence of the Financial Crisis

The claim was made in the introduction that my theory seems capable, unlike the major alternatives, of explaining the unusual length and depth of the Great Depression. In the previous section, I attempted to deal with the issue of depth; simulations of the estimated regressions suggested that the combined monetary and nonmonetary effects of the financial crisis can explain much of the severity of the decline in output. In this section, the question of the length of the Great Depression is addressed.

As a matter of theory, the duration of the credit effects described in Section II above depends on the amount of time it takes to 1) establish new or revive old channels of credit flow after a major disruption, and 2) rehabilitate insolvent debtors. Since these processes may be difficult and slow, the persistence of nonmonetary effects of financial crisis has a plausible basis. (In contrast, persistence of purely monetary effects relies on the slow diffusion of information or unexplained stickiness of wages and prices.) Of

course, plausibility is not enough; some evidence on the speed of financial recovery should be adduced.

After struggling through 1931 and 1932, the financial system hit its low point in March 1933, when the newly elected President Roosevelt's "bank holiday" closed down most financial intermediaries and markets. March 1933 was a watershed month in several ways: It marked not only the beginning of economic and financial recovery but also the introduction of truly extensive government involvement in all aspects of the financial system.³¹ It might be argued that the federally directed financial rehabilitation—which took strong measures against the problems of both creditors and debtors—was the only major New Deal program that successfully promoted economic recovery.³² In any case, the large government intervention is *prima facie* evidence that by this time the public had lost confidence in the self-correcting powers of the financial structure.

Although the government's actions set the financial system on its way back to health, recovery was neither rapid nor complete. Many banks did not reopen after the holiday, and many that did open did so on a restricted basis or with marginally solvent balance sheets. Deposits did not flow back into the banks in great quantities until 1934, and the government (through the Reconstruction Finance Corporation and other agencies) had to continue to pump large sums into banks and other intermediaries. Most important, however, was a noticeable change in attitude among lenders; they emerged from the 1930–33 episode chastened and conservative. Friedman and Schwartz (pp. 449–62) have documented the shift of banks during this time away from making loans toward holding safe and liquid investments. The growing level of bank liquidity created an illusion (as Friedman and Schwartz pointed out) of easy money;

³¹See Chandler (1970), ch. 15, and Friedman and Schwartz, ch. 8.

³²E. Carey Brown (1956) has argued that New Deal fiscal policy was not very constructive. A paper by Michael Weinstein in Brunner (1981) points out counterproductive aspects of the N.R.A.

however, the combination of lender reluctance and continued debtor insolvency interfered with credit flows for several years after 1933.

Evidence of postholiday credit problems is not hard to find. For example, small businesses, which (as I have noted) suffered disproportionately during the Contraction, had continuing difficulties with credit during recovery. Lewis Kimmel (1939) carried out a survey of credit availability during 1933–38 as a companion to the National Industrial Conference Board's 1932 survey. His conclusions are generally sanguine (this may reflect the fact that the work was commissioned by the American Bankers Association). However, his survey results (p. 65) show that, of responding manufacturing firms normally dependent on banks, refusal or restriction of bank credit was reported by 30.2 percent of very small firms (capitalization less than \$50,000); 14.3 percent of small firms (\$50,001–\$500,000); 10.3 percent of medium firms (\$500,001–\$1,000,000); and 3.2 percent of the largest companies (capital over \$1 million). (The corresponding results from the 1932 NICB survey were 41.3, 22.2, 12.5, and 9.7 percent.)

Two well-known economists, Hardy and Viner, conducted a credit survey in the Seventh Federal Reserve District in 1934–35. Based on "intensive coverage of 2600 individual cases," they found "a genuine unsatisfied demand for credit by solvent borrowers, many of whom could make economically sound use of working capital.... The total amount of this unsatisfied demand for credit is a significant factor, among many others, in retarding business recovery." They added, "So far as small business is concerned, the difficulty in getting bank credit has increased more, as compared with a few years ago, than has the difficulty of getting trade credit." (These passages are quoted in W. L. Stoddard, 1940.)

Finally, another credit survey for the 1933–38 period was done by the Small Business Review Committee for the U.S. Department of Commerce. This study surveyed 6,000 firms with between 21 and 150 employees. From these they chose a special sample of 600 companies "selected because

of their high ratings by a standard commercial rating agency." Even within the elite sample, 45 percent of the firms reported difficulty in securing funds for working capital purposes during this period; and 75 percent could not obtain capital or long-term loan requirements through regular markets. (See Stoddard.)

The reader may wish to view the American Bankers Association and Small Business Review Committee surveys as lower and upper bounds, with the Hardy-Viner study in the middle. In any case, the consensus from surveys, as well as the opinion of careful students such as Chandler, is that credit difficulties for small business persisted for at least two years after the bank holiday.³³

Home mortgage lending was another important area of credit activity. In this sphere, private lenders were even more cautious after 1933 than in business lending. They had a reason for conservatism; while business failures fell quite a bit during the recovery, real estate defaults and foreclosures continued high through 1935.³⁴ As has been noted, some traditional mortgage lenders nearly left the market: life insurance companies, which made \$525 million in mortgage loans in 1929, made \$10 million in new loans in 1933 and \$16 million in 1934.³⁵ During this period, mortgage loans that were made by private institutions went only to the very best potential borrowers. Evidence for this is the sharp drop in default rates of loans made in the early 1930's as compared to loans made in earlier years (see Carl Behrens, 1952, p. 11); this decline was too large to be explained by the improvement in business conditions alone.

To the extent that the home mortgage market did function in the years immediately following 1933, it was largely due to the direct involvement of the federal government. Besides establishing some important new institutions (such as the FSLIC and the system of federally chartered savings and loans), the government "readjusted" existing debts, made investments in the shares of

³³See Chandler (1970), pp. 150–51.

³⁴U.S. Department of Commerce (1975), series N301.

³⁵U.S. Department of Commerce (1975), N282.

thrift institutions, and substituted for recalcitrant private institutions in the provision of direct credit. In 1934, the government-sponsored Home Owners' Loan Corporation made 71 percent of all mortgage loans extended.³⁶

Similar conditions obtained for farm credit and in other markets, but space does not permit this to be pursued here. Summarizing the reading of all of the evidence by economists and by other students of the period, it seems safe to say that the return of the private financial system to normal conditions after March 1933 was not rapid; and that the financial recovery would have been more difficult without extensive government intervention and assistance. A moderate estimate is that the U.S. financial system operated under handicap for about five years (from the beginning of 1931 to the end of 1935), a period which covers most of the time between the recessions of 1929–30 and 1937–38. This is consistent with the claim that the effects of financial crisis can help explain the persistence of the depression.

VI. International Aspects

The Great Depression was a worldwide phenomenon; banking crises, though occurring in a number of important countries besides the United States, were not so ubiquitous. A number of large countries had no serious domestic banking problems, yet experienced severe drops in real income in the early 1930's. Can this be made consistent with the important role we have ascribed to the financial crisis in the United States? A complete answer would require another paper; but I offer some observations:

1) The experience of different countries and the mix of depressive forces each faced varied significantly. For example, Britain, suffering from an overvalued pound, had high unemployment throughout the 1920's; after leaving gold in 1931, it was one of the first countries to recover. The biggest problems of food and raw materials exporters were falling prices and the drying up of

overseas markets. Thus we need not look to the domestic financial system as an important cause in every case.

2) The countries in which banking crises occurred (the United States, Germany, Austria, Hungary, and others) were among the worst hit by the depression. Moreover, these countries held a large share of world trade and output. The United States alone accounted for almost half of world industrial output in 1925–29, and its imports of basic raw materials and foodstuffs in 1927–28 made up almost 40 percent of the trade in these commodities.³⁷ The reduction of imports as these economies weakened exerted downward pressure on trading partners.

3) There were interesting parallels between the troubles of the domestic financial system and those of the international system. One of the Federal Reserve's proudest accomplishments had been the establishment, during the 1920's, of an international gold-exchange standard. Unfortunately, like domestic banking, the gold-exchange standard had the instability of a fractional-reserve system. International reserves included not only gold but also foreign currencies, notably the dollar and the pound; for countries other than the United States and the United Kingdom, foreign exchange was 35 percent of total reserves.

In 1931, the expectations that the international financial system would collapse became self-fulfilling. A general attempt to convert currencies into gold drove one currency after another off the gold-exchange standard. Restrictions on the movement of capital or gold were widely imposed. By 1932, only the United States and a small number of other countries remained on gold.

As the fall of the gold standard paralleled domestic bank failures, the domestic insolvency problem had an international analogue as well. Largely due to fixed exchange rates, the deflation of prices was worldwide. Countries with large nominal debts, notably agricultural exporters (the case of Canada has been mentioned), became unable to pay. Foreign bond values in the United States were extremely depressed.

³⁶U.S. Department of Commerce (1975), N278 and N283.

³⁷U.S. Department of Commerce (1947), pp. 29–31.

As in the domestic economy, these problems disrupted the worldwide mechanism of credit. International capital flows were reduced to a trickle. This represented a serious problem for many countries.

To summarize these observations: the fact that the Great Depression hit countries which did not have banking crises does not preclude the possibility that banking and debt problems were important in the United States (or, for that matter, that countries with strong banks had problems with debtor insolvency). Moreover, my analysis of the domestic financial system may be able to shed light on some of the international financial difficulties of the period.

VII. Conclusion

Did the financial collapse of the early 1930's have real effects on the macroeconomy, other than through monetary channels? The evidence is at least not inconsistent with this proposition. However, a stronger reason for giving this view consideration is the one stated in the introduction: this theory has hope of achieving a reconciliation of the obvious suboptimality of this period with the postulate of reasonably rational, market-constrained agents. The solution to this paradox lies in recognizing that economic institutions, rather than being a "veil," can affect costs of transactions and thus market opportunities and allocations. Institutions which evolve and perform well in normal times may become counterproductive during periods when exogenous shocks or policy mistakes drive the economy off course. The malfunctioning of financial institutions during the early 1930's exemplifies this point.

REFERENCES

- Abel, Andrew and Mishkin, Frederic, "An Integrated View of Tests of Rationality, Market Efficiency, and the Short-Run Neutrality of Monetary Policy," Working Paper 726, National Bureau of Economic Research, 1981.
- Barro, Robert, "Unanticipated Money, Output, and the Price Level in the United States," *Journal of Political Economy*, August 1978, 86, 549-80.
- Behrens, Carl, *Commercial Bank Activities in Urban Mortgage Financing*, New York: National Bureau of Economic Research, 1952.
- Bernanke, Ben, "Bankruptcy, Liquidity, and Recession," *American Economic Review Proceedings*, May 1981, 71, 155-59.
- Brown, E. Carey, "Fiscal Policy in the 'Thirties: A Reappraisal," *American Economic Review*, December 1956, 46, 857-79.
- Brunner, Karl, *The Great Depression Revisited*, Boston: Martinus Nijhoff, 1981.
- Cagan, Philip, *Determinants and Effects of Changes in the Stock of Money, 1875-1960*, New York: National Bureau of Economic Research, 1965.
- Chandler, Lester, *America's Greatest Depression*, New York: Harper & Row, 1970.
- _____, *American Monetary Policy, 1928-1941*, New York: Harper & Row, 1971.
- Clark, Evans, *The Internal Debts of the United States*, New York: Macmillan Co., 1933.
- Cone, Kenneth, "Regulation of Depository Financial Intermediaries," unpublished doctoral dissertation, Stanford University, 1982.
- Diamond, Douglas and Dybvig, Philip, "Bank Runs, Deposit Insurance, and Liquidity," mimeo., University of Chicago, 1981.
- Evans, Paul, "An Econometric Analysis of the Causes of the Great Depression in the U.S.," mimeo., Stanford University, 1981.
- Fabricant, Solomon, *Profits, Losses, and Business Assets, 1929-1934*, Bulletin 55, National Bureau of Economic Research, 1935.
- Fama, Eugene, "Banking in the Theory of Finance," *Journal of Monetary Economics*, January 1980, 6, 39-57.
- Fisher, Irving, "The Debt-Deflation Theory of Great Depressions," *Econometrica*, October 1933, 1, 337-57.
- Flood, Robert and Garber, Peter, "A Systematic Banking Collapse in a Perfect Foresight World," Working Paper No. 691, National Bureau of Economic Research, 1981.
- Frederiksen, D. M., "Two Financial Roads Leading Out of Depression," *Harvard Business Review*, October 1931, 10, 131-48.
- Friedman, Benjamin, "Debt and Economic Activity in the United States," Working Paper

- No. 704, National Bureau of Economic Research, 1981.
- Friedman, Milton, and Schwartz, Anna J., *A Monetary History of the United States, 1867-1960*, Princeton: Princeton University Press, 1963.
- Goldsmith, Raymond, *Financial Institutions in the American Economy Since 1900*, Princeton: Princeton University Press, 1958.
- Gurley, John G. and Shaw, E. S., "Financial Aspects of Economic Development," *American Economic Review*, September 1955, 45, 515-38.
- Hart, A. G., *Debts and Recovery, 1929-1937*, New York: Twentieth Century Fund, 1938.
- Jaffee, Dwight, and Russell, Thomas, "Imperfect Information and Credit Rationing," *Quarterly Journal of Economics*, November 1976, 90, 651-66.
- Kennedy, Susan E., *The Banking Crisis of 1933*, Lexington: University Press of Kentucky, 1973.
- Kimmel, Lewis H., *The Availability of Bank Credit, 1933-1938*, New York: National Industrial Conference Board, 1939.
- Kindleberger, Charles P., *Manias, Panics, and Crashes*, New York: Basic Books, 1978.
- Klebaner, Benjamin, *Commercial Banking in the United States: A History*, Hinsdale: Dryden Press, 1974.
- Lucas, Robert E., Jr., "Expectations and the Neutrality of Money," *Journal of Economic Theory*, April 1972, 4, 103-24.
- _____, *Studies in Business Cycle Theory*, Cambridge: Massachusetts Institute of Technology Press, 1981.
- Lutz, Friedrich, *Corporate Cash Balances, 1914-43*, New York: National Bureau of Economic Research, 1945.
- McKinnon, Ronald J., *Money and Capital in Economic Development*, Washington: The Brookings Institution, 1973.
- Merwin, Charles L., *Financing Small Corporations*, New York: National Bureau of Economic Research, 1942.
- Minsky, Hyman P., "A Theory of Systematic Fragility," in E. I. Altman and A. W. Sametz, eds., *Financial Crises*, New York: Wiley-Interscience, 1977.
- Mishkin, Frederic, "The Household Balance Sheet and the Great Depression," *Journal of Economic History*, December 1978, 38, 918-37.
- _____, "Does Anticipated Money Matter? An Econometric Investigation," *Journal of Political Economy*, February 1982, 90, 22-51.
- Nugent, Rolf, *Consumer Credit and Economic Stability*, New York: Russell Sage Foundation, 1939.
- O'Hara, Maureen and Easley, David, "The Postal Savings System in Depression," *Journal of Economic History*, September 1979, 39, 741-53.
- Persons, Charles E., "Credit Expansion, 1920 to 1929 and Its Lessons," *Quarterly Journal of Economics*, November 1930, 45, 94-130.
- Safarian, A. E., *The Canadian Economy in the Great Depression*, Toronto: University of Toronto Press, 1959.
- Sargent, Thomas J., "A Classical Macroeconomic Model for the United States," *Journal of Political Economy*, April 1976, 84, 207-38.
- Stiglitz, Joseph E. and Weiss, Andrew, "Credit Rationing in Markets with Imperfect Information," *American Economic Review*, June 1981, 71, 393-410.
- Stoddard, W. L., "Small Business Wants Capital," *Harvard Business Review*, Spring 1940, 18, 265-74.
- Temin, Peter, *Did Monetary Forces Cause the Great Depression?*, New York: W. W. Norton, 1976.
- Upham, Cyril B. and Lamke, Edwin, *Closed and Distressed Banks: A Study in Public Administration*, Washington: The Brookings Institution, 1934.
- Board of Governors of the Federal Reserve System, *Banking and Monetary Statistics*, 1943.
- Federal Reserve Bulletin, various issues.
- National Industrial Conference Board, *The Availability of Bank Credit*, New York, 1932.
- Survey of Current Business, various issues.
- U.S. Department of Commerce, *Historical Statistics of the United States*, Washington: USGPO, 1975.
- _____, *National Income*, Washington: USGPO, 1954.
- _____, *The United States in the World Economy*, Washington: USGPO, 1947.