Volume Title: NBER Macroeconomics Annual 2000, Volume 15

Volume Author/Editor: Ben S. Bernanke and Kenneth Rogoff, editors

Volume Publisher: MIT PRess

Volume ISBN: 0-262-02503-5

Volume URL: http://www.nber.org/books/bern01-1

Publication Date: January 2001

Chapter Title: Re-Examining the Contributions of Money and Banking Shocks to the U.S. Great Depression

Chapter Author: Harold L. Cole, Lee E. Ohanian

Chapter URL: http://www.nber.org/chapters/c11057

Chapter pages in book: (p. 183 - 260)

Harold L. Cole and Lee E. Ohanian

FEDERAL RESERVE BANK OF MINNEAPOLIS; AND FEDERAL RESERVE BANK OF MINNEAPOLIS AND U.C.L.A.

Re-examining the Contributions of Money and Banking Shocks to the U.S. Great Depression

1. *Introduction*

Many economists argue that deflation can account for much of the Great Depression (1929–1933) in the United States. According to this story, a sharp decline in the money supply caused rapid deflation, which in turn reduced output. Empirical research has documented large decreases in money, prices, and output between 1929 and 1933. But there is much less work assessing whether this shock can plausibly account for the Depression within fully articulated general equilibrium models. This paper quantitatively evaluates the deflation hypothesis with dynamic, general equilibrium business-cycle models.

Evaluating the deflation hypothesis with general equilibrium models requires an explicit theory of why deflation reduced output so much in the 1930s. Since there are several explanations for this in the literature, we first narrow the field by requiring that any successful deflation theory of the Depression also be consistent with macroeconomic activity during other major deflations. We therefore determine which deflation theories satisfy this criterion by comparing the Great Depression with

We thank Andy Atkeson, Ben Bernanke, Narayana Kocherlakota, Stanley Lebergott, Axel Leijonhufvud, Ken Rogoff, Art Rolnick, and our discussants Michael Bordo and Mark Gertler for helpful comments. We thank Daniel Hammermesh for CPS wage data, and Jonathon Parker for estimates of compositional wage changes. Special thanks go to Ed Prescott for many discussions about the Great Depression. We have also benefited from the outstanding research assistance of Jesus Fernandez-Villaverde. The views expressed herein are those of the authors and not necessarily those of the Federal Reserve Bank of Minneapolis or the Federal Reserve System.

macroeconomic activity during the early 1920s, which is a period of comparable deflation, but a much less severe downturn in economic activity.

We find that two of the four most popular explanations are ruled out by this consistency criterion. These are the *surprise-deflation story* of Lucas and Rapping (1969), which argues that the Great Depression was severe because the deflation was unexpected, and the *debt–deflation story* of Irving Fisher (1933), which argues that the Great Depression was severe because deflation substantially raised the real value of private debt. The two stories that are not ruled out are the high-wage story and the banking story. According to the *high-wage story*, deflation, combined with imperfectly flexible wages, raised real wages and reduced employment and output. A number of economists report evidence in favor of this story, including Eichengreen and Sachs (1985), Bernanke and Carey (1996), and Bordo, Erceg, and Evans (2000). According to the *banking story*, deflationary money shocks contributed to bank failures and to a reduction in the efficiency of financial intermediation, which in turn reduced lending and output. Bernanke (1983) reports evidence in favor of this story.

Following this empirical analysis, we develop two general equilibrium models to separately evaluate the wage shock hypothesis and the banking shock hypothesis. We ask two questions: Can these shocks drive down output per adult nearly 40% relative to trend between 1929 and 1933? Are the other predictions of the theories consistent with the data?

Our main finding is that wage shocks and banking shocks account for a small fraction of the Great Depression. We also find that some other predictions of the theories are at variance with the data. We conclude that these results raise questions about the deflation and banking hypotheses as explanations of the Great Depression in the United States.

The paper is organized as follows. Section 2 presents the comparison between the Great Depression and the 1921–1922 Depression, and the evaluation of the four popular deflation stories for the Great Depression. We then go on to develop general equilibrium models for the two stories that are not ruled out by this comparison—the high-wage story and the banking story. Section 3 presents a general equilibrium model with above-market wages, and also presents a quantitative assessment of the wage hypothesis. Section 4 presents a general equilibrium model with an intermediation sector to assess the macroeconomic impact of bank failures. Since our results support neither the wage nor banking story, Section 5 briefly discuss two other possible contributing factors to the Great Depression: changes in asset prices and changes in productivity. Section 6 presents a summary and conclusion.

Table 1 DEFLATION AND OUTPUT—OUTPUT AND ITS COMPONENTS[a]

Depression of 1921–1922 (1920 = 100)					*Great Depression (1929 = 100)*				
Year	*P*	*Y*	*C*	*I*	*Year*	*P*	*Y*	*C*	*I*
1921	85.2	93.9	102.4	86.1	1930	97.5	86.9	90.0	73.2
1922	80.6	96.2	102.7	114.4	1931	88.5	77.6	84.3	48.5
					1932	79.5	64.0	74.3	26.7
					1933	77.5	60.9	70.8	23.0

[a]The price level is from Romer (1988) for 1921–1923, and from Bureau of the Census (1975) for 1929–1933. The output data for 1920–1922 are from Kendrick (1961, p. 294). Romer (1988) argues that the Kendrick series is a better output measure for the 1920s than the Commerce Department measure, which is based on preliminary work of Kuznets and Kendrick. The output data for 1929–1933 are from the National Income and Product Accounts. The population data are from Bureau of the Census (1975, p. 10).

2. *An Empirical Puzzle about the Deflation Hypothesis*

A successful theory of the Great Depression based on deflation should account for macroeconomic activity during 1929–1933 and should also be consistent with macroeconomic activity during other major deflations. This section empirically evaluates this consistency requirement by comparing changes in prices and real output during 1929–1933 to those during a period of comparable deflation: 1920–1922.

Table 1 shows the percentage change in the GNP deflator (*P*), real GNP (*Y*), real consumption (*C*), and real investment (*I*) during these two episodes. The three quantity variables are deflated by their specific deflators, are measured relative to the adult (16 and over) population, and are detrended.[1] Deflation is similar during these two periods: the price level fell about 20% between 1920 and 1922, and also fell about 20% between 1929 and 1932. Despite these similar deflations, however, output fell much more between 1929 and 1932 than between 1920 and 1922. Real GNP fell 36% between 1929 and 1932, but just 4% between 1920 and 1922.

These data raise a puzzle about the deflation hypothesis: If the 20% deflation of the 1930s caused the Great Depression, why didn't the 20% deflation of the 1920s also cause a major depression? Resolving this puzzle requires finding some other shock(s) that magnify the depressing effects of deflation and that were present in the 1930s, but not in the 1920s. There are several stories for why the 1930s deflation had such

1. We detrended these three quantity variables at a rate of 1.9% per year. We define this rate as normal growth, because it is the growth rate of output per adult both before the Great Depression (1919–1929), and after World War II (1947–1997), and because it is close to the 2% average growth rate between 1900 and 1997. It is also worth noting that output per adult in 1929 is very close to an OLS trend line fitted to this series between 1900 and 1997. This suggests that output was close to its normal trend value in 1929.

large, negative real effects. But can these stories explain why the Great Depression was so much worse than the 1921–1922 Depression? We address this question in the next section.

2.1 CAN THE STANDARD STORIES EXPLAIN THE SEVERITY OF THE GREAT DEPRESSION?

Four popular deflation stories for the Great Depression are: (1) the deflation was unexpected, (2) nominal debt levels were high, (3) nominal wages were imperfectly flexible, and (4) there were many bank failures in addition to the deflation. We consider each of these stories in turn and ask whether they might be consistent with both the Great Depression and the 1921–1922 Depression. For each story, this consistency requires that the shock that magnified the real effect of deflation in the 1930s not be present in the 1920s.

2.1.1 Differences in Deflation Predictability between the 1920s and the 1930s Some theories predict that only unanticipated deflation depresses real economic activity. Lucas and Rapping (1969) argue that the 1930s deflation was unexpected and that this was an important factor behind the severity of the Great Depression. Can differences in the predictability of the 1920s and 1930s deflations explain the difference in the severity of these two depressions? We address this question by comparing nominal and ex post real interest rates between these two periods.[2] If differences in the predictability of deflation can explain both the Great Depression and the Depression of 1921–1922, we should observe very low nominal interest rates in the 1920s, but relatively high nominal and ex post real interest rates during the 1930s.

Table 2 shows average annual nominal and real interest rates on 3- to 6-month U.S. Treasury notes and certificates. The real rate is the nominal rate minus the percentage change in the annual GNP deflator. The most striking feature of these data is that both nominal and real interest rates are higher during the Depression of 1921–22. The average nominal rate on Treasury securities is 4.35% between 1921 and 1922, compared to an average of 1.1% between 1930 and 1933. The average real rate on these securities is 14.25% between 1921 and 1922, compared to an average of 7.21% between 1930 and 1930.[3] These data suggest that the 1930s defla-

2. There is some work addressing the predictability of the 1930s deflation (see Hamilton, 1992, and Cecchetti, 1992), but we are unaware of any studies of the predictability of the deflation of the early 1920s, or any comparison of the predictability of deflation between the two periods.
3. It may seem surprising that the deflation of the early 1920s was more unexpected, since monetary policy after wars traditionally produced deflation. However, the timing and rates of those deflations were probably much less certain.

Table 2 NOMINAL AND EX POST REAL INTEREST RATES: 1920s AND 1930s[a]

Depression of 1921			Great Depression		
	Interest rate (%)			Interest rate (%)	
Year	Nominal	Real	Year	Nominal	Real
1921	4.83	19.63	1930	2.23	4.73
1922	3.47	8.87	1931	1.15	10.38
			1932	0.78	10.95
			1933	0.26	2.78
Avg.	4.35	14.25	Avg.	1.10	7.21

[a]The data are from Board of Governors of the Federal Reserve System (1943). The results are very similar using 4–6-month prime commercial paper.

tion was *more* predictable than the 1920s deflation, rather than *less* predictable. We conclude from these data that unexpected deflation is not the key factor behind the relative severity of the Great Depression.[4]

2.1.2 Differences in Private Debt and Deflation between the 1920s and 1930s Fisher (1933) suggested that deflation and high private debt levels contributed to the Great Depression by reducing borrower wealth and constraining lending. This is known as the *debt–deflation* view of the Great Depression. Before asking whether this story is consistent with both depressions, it is important to note that there are two separate macroeconomic effects from this redistribution. We call one the *debt-burden effect* of debt and deflation, which is Fisher's original view. The other is the *wealth-transfer effect*, in which unexpected deflation transfers wealth from debtors to creditors. On average, creditors are older and borrowers are younger. This transfer increases the old generation's consumption, but changes their labor input little in absolute terms, since their labor endowment is low. The wealth transfer will tend to increase the hours of the young generation. Overall, the wealth-transfer effect should increase aggregate hours and output and thus will tend to offset the debt-burden effect. Thus, there is no theoretical presumption that wealth redistributions between debtors and creditors reduce aggregate employment and output.

If the debt–deflation story can explain the severity of the Great Depression, the debt-burden effect must be quantitatively much more impor-

4. Some economists have also suggested that high real interest rates were an important contributing factor to the Great Depression. The fact that real interest rates were substantially higher during the 1921–1922 Depression casts doubt on this explanation.

Table 3 INCREASE IN THE PRIVATE DEBT BURDEN DUE TO DEFLATION: THE DEPRESSION OF 1921–1922 VS. THE GREAT DEPRESSION[a]

Year	*Private debt relative to output*	*Δ (price level) in first 2 years of deflation (%)*	*Increase in debt burden in first 2 years of deflation*	*Δ GNP in first 2 years of deflation (%)*
1920	1.20	−19.4	0.29	−3.8
1929	1.56	−11.5	0.20	−22.4

[a]The increase in the debt burden is given by

$$\frac{100D/Y}{100 + \%\Delta P} - D/Y,$$

where D is the debt-to-output ratio. There are two basic sources of data on business liabilities in the Historical Statistics (Bureau of the Census, 1975). The first is the nominal debt series put out by the BEA, which we have used. The second is from IRS data on corporate tax returns (see series V 108–140). The IRS data only begin in 1926, and there appears to be a significant difference in the indicated increase in corporate debt levels between the two sources. The IRS data indicate that corporate debt in the form of bonded debt and mortgages rose 47% between 1926 and 1929. This figure seems too large and suggests that the coverage level was initially low when the IRS was first collecting the returns data. This view is supported by the observation that according to the IRS data the total debt of the corporate sector—including notes, accounts payable, bonded debt, and mortgages—was only $55.8 billion in 1926, while the net debt from the BEA for the total corporate sector in 1926 was $76.2 billion.

tant in the 1930s than in the 1920s. Two factors that affect the quantitative extent of the debt-burden effect are the size of the stock of debt at the start of the deflation and the pattern of deflation. A larger initial stock of debt and a rapid deflation will tend to increase the effect. We measure the increase in the debt burden as the increase in the real value of debt (relative to output) due to deflation over the first two years of each depression.

Table 3 shows the initial stock of debt relative to output at the price-level peak prior to each depression, as well as the percentage change in prices in the first two years of each depression, the implied percentage increase in the debt burden relative to initial output, and the percentage change in real output. The most striking feature of these data is that the debt-burden channel rises *more* in 1921–1922 than in 1929–1931. The more rapid 1920s deflation increased the debt burden by 0.29 between 1920 and 1922, compared to 0.20 between 1929 and 1931. This larger debt-burden increase, however, is associated with a much smaller decrease in output. Real GNP falls 3.8% between 1920 and 1922, but falls 22.4% between 1929 and 1931.

Explaining the severity of the Great Depression through debt and deflation thus requires a model in which an initial debt stock of 1.2, with 19% deflation, is associated with only a 4% decrease in output, while an initial debt stock of 1.56, with 11% deflation, drives down output by more than 22%.[5] We are unaware of any quantitatively plausible model that is consistent with these observations. We conclude from these data that the debt–deflation story does not explain why the Great Depression was worse than the 1921–1922 Depression.[6]

2.1.3 Differences in Wages between the 1920s and the 1930s Some economists believe that wage changes increased the depressing effects of deflation in the 1930s. Before addressing whether differences in wages can explain the difference between the Great Depression and the Depression of 1921–1922, it is important to recognize that there is disagreement over *how* wage changes may have contributed to the Great Depression. Some economists, for example Lucas and Rapping (1972) and Lucas (1983), argue that the Great Depression was severe because nominal wages fell so *much*. Others, for example Bernanke and Carey (1996), Bordo, Erceg, and Evans (2000), and Eichengreen and Sachs (1985), argue that the Great Depression was severe because nominal wages were imperfectly flexible and did not fall *enough*.

Since Lucas and Rapping's view is based on unexpected deflation, and it is unlikely that unexpected deflation is responsible for the severity of the Great Depression, we focus on the inflexible-wage hypothesis. According to this hypothesis, inflexible nominal wages, combined with deflation, raised real wages, which reduced employment and output.

Explaining the relative severity of the Great Depression through high wages requires: (1) real wages well above the trend in the 1930s, and significantly higher than wages in 1921–1922, and (2) a theory of labor market failure during the 1930s—if the Great Depression was caused by high real wages, there would have been enormous competitive pressure for wages to fall.

We begin by examining wages between the two depressions. Unfortunately, there are few survey wage data that are both of reasonable quality and consistently available during both the 1920s and the 1930s. Two sectors for which such data are available are agriculture and manufacturing. Tables 4 and 5 show that detrended wage changes are fairly similar

5. Olney (1999) argues that high consumer debt levels and extreme default penalties help account for the large drop in consumption in 1930. If this indebtedness was key, we would expect a larger than normal decrease in consumer durables spending in 1930. However, the decrease in the ratio of durables to output in 1930 is small relative to postwar recessions. The major decrease in consumption in 1930 is due to nondurables and services.

6. It is worth noting that the difference in debt levels between the two periods—1.2 vs. 1.56—may overstate the actual difference in the debt-burden channel, since financial markets were probably more sophisticated in the 1930s, and as a consequence might have managed larger debt levels more efficiently.

Table 4 FARM WAGES[a]

Depression of 1921–1922		*Great Depression*	
Year	*Real wage (1920 = 100)*	*Year*	*Real wage (1929 = 100)*
1921	71.9	1930	93.0
1922	73.1	1931	76.8
		1932	64.7
		1933	60.2

[a]Source: Bureau of the Census (1975, p. 468). The farm wage rate is the daily wage without room and board. It is deflated by the GNP deflator and is detrended at 1.4%/yr, as that is the average growth rate of real hourly compensation between 1947 and 1997.

Table 5 MANUFACTURING AVERAGE HOURLY EARNINGS[a]

Depression of 1921–1922		*Great Depression*	
Year	*Real wage (1920 = 100)*	*Year*	*Real wage (1929) = 100)*
1921	101.5	1930	102.1
1922	101.2	1931	106.8
		1932	106.5
		1933	104.2

[a]These data are deflated by the GNP deflator. We detrended manufacturing wages at a 1.4% annual rate, as that is the average growth rate of real hourly compensation between 1947 and 1997. The average growth rate of real manufacturing wages between 1923 and 1929 was slightly higher at 1.6%/yr.

between the two episodes and that wage changes differed significantly across sectors of the economy. Some real wages fell substantially during both depressions, while others remained near trend. The wage in the farm sector is an example of one real wage that fell significantly during both depressions. Table 4 shows that, on average, it is about 28% below trend during both periods.

In contrast, the real manufacturing wage rose modestly during the Great Depression and remained near trend in 1921–1922. Table 5 shows the manufacturing wage during these two depressions. The basic data for the Great Depression are from surveys conducted by the National Industrial Conference Board, and are considered to be among the best wage measures during the Great Depression.[7] The real manufacturing

7. The 1930s data are from Hanes (1996). The 1920s data are from the National Industrial Conference Board (1928) and Beney (1936) and include average hourly earnings of all wage earners in 25 industries plus anthracite mining, railroads, and building trades. Industries include metal, textiles, leather, paper, furniture, lumber, meat, and rubber. The data are on p. 25, Table 2.

wage, on average, was roughly 5% above trend during 1930–1933 and about 1% above trend during the Depression of 1921–1922.

These manufacturing wage differences between the 1920s and 1930s do not seem large enough to account for the relative severity of the Great Depression. But without a formal model we do not know how much of the Great Depression these differences can explain. We therefore construct a two-sector general equilibrium model in Section 3 to assess the quantitative contribution of high wages in some sectors to the Great Depression.

2.1.4 Differences in Bank Closings between the 1920s and the 1930s Many banks either temporarily suspended operations or failed during the early 1930s. Bernanke's (1983) widely cited work shows that the number of banks that either closed temporarily or failed is a significant predictor of output during the Great Depression. Bernanke's work has led a number of economists to conclude that bank closings were an important contributing factor to the Great Depression. For example, Romer's (1993) survey of the Great Depression argues that these closings were responsible for much of the fall in output between 1930 and 1933. According to the bank-closing hypothesis, bank suspensions and failures destroyed private information about borrowers, which in turn reduced the efficiency of financial intermediation (see Romer, 1993).

Can bank closings explain the difference between the Great Depression and the 1921–1922 Depression? Table 6 presents a comparison of bank closings in the 1920s and 1930s. Since the importance of a bank suspension or failure depends on the size of the bank, we measure bank closings not by the number of banks that closed, but rather by the fraction of deposits in banks that either suspended operations or failed. The table thus shows the fraction of total deposits in commercial banks that either suspended operations or failed, and shows the fraction of total deposits lost by depositors.[8]

Bank suspensions and failures were higher during the Great Depression. About 0.5% of banks, measured by deposits, either suspended operations or failed during the Depression of 1921–1922, and about 0.2% of total deposits was ultimately lost. In comparison, an average of 2.6% of banks either suspended operations or failed between 1930–1932, and an average of 0.4% of total deposits was ultimately lost during that

8. Since deposits at failed and suspended banks are only available for commercial banks, we show this ratio relative to commercial deposits. Commercial deposits accounted for over 85% of total deposits during 1919–1923 and over 80% during 1929–1934. We include failures and suspensions together because we are unaware of any data that separate these two categories.

Table 6 BEHAVIOR OF COMMERCIAL BANK DEPOSITS[a]

Depression of 1921–1922				*Great Depression*			
Year	*Susp./total (%)*	*Loss/total (%)*	*Total/output*	*Year*	*Susp./total (%)*	*Loss/total (%)*	*Total/output*
1921	0.5	0.2	0.52	1929	0.4	0.1	0.58
1922	0.3	0.1	0.55	1930	1.7	0.5	0.64
				1931	4.3	0.8	0.65
				1932	2.0	0.4	0.78
				1933	11.0	1.3	0.75

[a]Source: Board of Governors of the Federal Reserve System (1943).

period. Both of these ratios rose significantly in 1933 when President Franklin Roosevelt declared a bank holiday. An explicit economic model is needed to determine the quantitative importance of these differences for the severity of the Great Depression. We develop a model for this purpose in Section 4.

The final data we present are on the ratio of total commercial bank deposits to output during these two depressions. This ratio rises significantly during the Great Depression. We present these data because they will be a key in the model that we develop for assessing the macroeconomic impact of bank closings.

2.2 SUMMARY

This section has assessed whether four popular deflation stories for the Great Depression can explain why the 20% deflation of the 1930s produced the Great Depression, and why the 20% deflation of the 1920s produced a much milder downturn. For any of these stories to be consistent with both depressions requires that the story be quantitatively important during the 1930s, but quantitatively unimportant during the 1920s. We found that two of these four stories—unexpected deflation, and debt plus deflation—do not satisfy this criterion, and therefore do not seem capable of explaining the relative severity of the Great Depression. For the other two stories—imperfectly flexible wages and bank failures—we did find some differences between the 1920s and 1930s. We now develop two models—one for assessing the role of inflexible nominal wages, and one for assessing the role of banking shocks—to evaluate quantitatively how much these two factors contributed to the Great Depression.

3. *How Much of the Great Depression Was Due to High Wages?*

3.1 A TWO-SECTOR GENERAL EQUILIBRIUM MODEL

This section presents a general equilibrium model to quantitatively assess the macroeconomic effects of high wages. Since wages in some sectors, such as agriculture, were flexible, we develop a two-sector model in which the wage in one sector is fixed above the market-clearing level, and the wage in the other sector is flexible. We assume that the fixed wage in the distorted sector is equal to the manufacturing wage; this assumption is discussed in detail below. All labor hired in that sector must be paid the above-market wage. This approach captures the basic distorting effects of above-market wages but allows us to abstract from other monetary features that would complicate the environment. All other prices in the economy, including the wages in the nondistorted sector, adjust to equate supply and demand in the other markets.

We first summarize the physical environment. We then analyze the pure market-clearing version of the model with no wage distortions, and then analyze the model with above-market wages in the manufacturing sector.

3.1.1 Environment Time is denoted by $t = 0,1,2, \ldots$. There is a representative family with many members. Family members supply labor, consume a single physical good, and accumulate physical capital. There are two distinct types of physical goods: *Final goods* are the numeraire, and can be either consumed or invested to augment the capital stock. These final goods are produced using two types of *intermediate goods.* Each intermediate good is produced from a distinct sector. We denote the sector that will be distorted by the above-market wage as *sector m,* and the nondistorted sector as *sector n.* We denote the output of the final good by Y, and the output of the two intermediate goods by Y_i, where $i = m, n$. These two intermediate goods are produced using identical Cobb–Douglas technologies with capital, denoted by K_i, and labor, denoted by H_i, for $i = m, n$. The parameter A is labor-augmenting technological change.

Capital and labor are both sector-specific—neither labor nor capital can move from one sector to the other. Thus, workers who are unable to work as much as they wish in the distorted sector are not permitted to move to the nondistorted sector. This assumption amplifies the distorting effects of the high wage.

3.1.2 Technologies The technology for producing the intermediate good m is

$$Y_m = (AH_m)^{1-\theta}K_m^{\theta}.$$

The same technology is used to produce the intermediate good n:

$$Y_n = (AH_n)^{1-\theta}K_n^{\theta}.$$

The technology for final goods is a CES aggregate of the two intermediate goods:

$$Y = [\alpha Y_m^{\phi} + (1 - \alpha)Y_n^{\phi}]^{1/\phi}. \tag{1}$$

3.1.3 The Market-Clearing Model

THE HOUSEHOLD'S PROBLEM There is a representative household with many members. At date 0, it is assumed that half of the family members work in the m sector, and half work in the n sector. The household's preferences over sequences of consumption of the final good c_t and market time in the two sectors is given by[9]

$$\max \sum_{t=0}^{\infty} \beta^t\{\log (c_t) + B\,[\log(1 - h_{mt}) + \psi \log(1 - h_{nt})]\}. \tag{2}$$

The household owns the capital stock and chooses consumption (c_t), work effort in the two sectors (h_{mt} and h_{nt}), and investment (x_{mt} and x_{nt}) to maximize (2) subject to the following present-value budget constraint, capital accumulation constraint, and time constraint:

$$\sum_{t=0}^{\infty} Q_t[w_{mt}h_{mt} + w_{nt}h_{nt} - c_t + r_{mt}k_{mt} + r_{nt}k_{nt} - x_{mt} - x_{nt}] \geq 0, \tag{3}$$

$$k_{it+1} = x_{it} + (1 - \delta)k_{it}, \qquad i \in \{m, n\}. \tag{4}$$

The wage rates in the m and n sectors are denoted w_m and w_n, respectively, and the rental prices of capital in the two sectors are analogously

9. This preference specification with different utility weights on leisure permits us to retain the tractability of a representative-agent formulation. The different utility weights are required when employment is different between the two sectors (e.g. $\alpha \neq 0.5$). It can be shown that this specification is equivalent to an environment with agents who work in either sectors m or sector n, and who are perfectly insured against idiosyncratic shocks to their specific sectors.

denoted r_m and r_n. Note that the parameter ψ captures the relative size difference in employment for the household. The date-*t* price of the physical good in terms of date-0 goods is denoted by Q_t.

THE INTERMEDIATE-GOOD FIRMS' PROBLEM We assume that there is a single producer of the *m* intermediate good, and a single producer of the *n* intermediate good, both of whom behave competitively.[10] The intermediate-good producer in sector *i*, $i \in \{m, n\}$, maximizes profits given (p_i, w_i, r_i):

$$\max_{n_i, k_i} p_i k_i^{\theta} h_i^{1-\theta} - w_i h_i - r_i k_i. \tag{5}$$

The first-order conditions for hiring the inputs imply that factor prices are equated to the value of marginal products.

THE FINAL-GOOD FIRMS' PROBLEM The final-good producer also is competitive. The maximization problem is:

$$\max_{Ym, Yn} [\alpha Y_m^{\phi} + (1 - \alpha) Y_n^{\phi}]^{1/\phi} - p_m Y_m^d - p_n Y_n^d. \tag{6}$$

EQUILIBRIUM CONDITIONS A competitive equilibrium for this economy consists of sequences of allocations and a price system such that the allocations solve the household's problem subject to its budget constraint and given prices; that the allocations solve the firm's problem, given prices; that the labor market, the capital-services market, and the intermediate-good markets all clear; that the resource constraint is satisfied; and that prices are equal to marginal productivities.

3.1.4 The Model with Some Wages above the Market-Clearing Level We now modify our model so that the wage in sector *m* is set above its market-clearing level. Rather than develop a monetary model with fixed nominal wages and deflation, we adopt a much simpler specification that captures the distorting effects of above-market wages. At the start of period *t* the wage is fixed exogenously for that period at a level above its normal market-clearing level. We denote this fixed wage by $\bar{w}_{mt}$. All labor hired in this sector at date *t* must be paid this wage. The above-market wage is a completely unexpected shock each period.[11]

10. We assume a single firm that behaves competitively, rather than a large number of competitive firms, to economize on notation.

11. There are many ways to model household beliefs about future distortions to manufacturing wages. Our approach, in which households believe that the fixed manufacturing wage does not recur, treats each wage shock as a completely unexpected event. As we

The fixed wage changes our model in one key way: labor input in this sector is no longer a choice variable for the household.[12] The households are rationed in terms of their labor supply to this sector:

$$\frac{B}{1 - h_{mt}} > \frac{\bar{w}_{mt}}{c_t}.$$

Labor input in the distorted sector is determined by firms' labor demand. The representative firm hires labor until the fixed wage is equated to labor's value of marginal product:

$$(1 - \theta)p_{mt}(K_{mt}/H_{mt})^{\theta} = \bar{w}_{mt}. \tag{7}$$

The high wage has *direct* and *indirect* effects on aggregate output. We define the direct effect as the change in aggregate output from the increase in the distorted wage, holding all other prices fixed. This effect is measured by solving for y_m from (7), given $\bar{w}_{mt}$ and holding p_{mt} fixed, and then solving for aggregate output, holding y_n fixed. The indirect, or general equilibrium, effects of the high wage operate through changes in prices and the other wage. These indirect effects depend not only on $\bar{w}_{mt}$, but also on all the model parameters. Assessing the quantitative effects of the high manufacturing wage on the economy thus requires choosing parameter values and numerically computing the equilibrium path of the model economy.

3.2 CHOOSING PARAMETER VALUES AND COMPUTING AN EQUILIBRIUM

3.2.1 Technology and Preference Parameter Values Several of the parameters in our model are commonly used in the equilibrium-business-cycle literature. We choose values for these parameters that are similar to values in other studies. Since the data are available at an annual frequency, we define the unit of time in the model to be one year.

The common parameters in our model are β, A, B, δ, and θ. We set $\beta = 0.96$, which is comparable to values used in other studies. We assume that the level of technological progress, A, is given by $A_t = (1 + g)^t$, and choose $g = 0.02$. Our values for β and g imply a steady-state interest rate of about

show later, this approach simplifies computing the equilibrium considerably. This approach is also consistent with the prevailing view that the Great Depression was the result of unexpected shocks.

12. Since no other markets are distorted, all other equations in the model will continue to be satisfied.

6%. We choose B such that the household works about one-third of its discretionary time in the steady state. The additional leisure parameter ψ is chosen so that in the undistorted version of the model, the household chooses to allocate the appropriate fraction of labor to each sector at a common wage. We set $\theta = 0.33$, and the depreciation rate to 7%.

The final parameter we discuss in this section is ϕ, which governs the substitution elasticity between the two sectors in final-good production. Since manufacturing appears to be a key sector distorted by the high wage during the Depression, we use postwar data on changes in manufacturing's expenditure share and relative price to choose a value for ϕ. Manufacturing's expenditure share and relative price have both fallen over the postwar period, which is consistent with a substitution elasticity between manufacturing and nonmanufacturing of less than one. We choose a benchmark value of $\phi = -1$, which implies a substitution elasticity of 0.5. We also conduct our analysis with a low substitution elasticity of 0.1 to assess the robustness of our results.

3.2.2 The Distorted Wage and the Relative Size of the Distorted Sector Finally, we need to choose a measure of how much real wages rose in the distorted sector, and we need to choose a value for the fraction of the economy distorted by the high wage.

We use Hanes's (1996) compilation of the Conference Board's manufacturing wage data as the measure of the wage for the distorted sector. This wage is shown in Table 5 for each year of the Great Depression. The Conference Board wage data have also been used in some other analyses of the Great Depression, including O'Brien (1989), Lebergott (1991), Bernanke (1986), and Bernanke and Carey (1996). This wage is the most natural choice for a distorted wage in this study, because the data are of relatively high quality, and because there is a plausible economic explanation for why manufacturing wages were above market clearing despite the downturn in economic activity: government intervention. This intervention comes from President Herbert Hoover's belief that maintaining nominal wages would *prevent* a major depression by keeping demand high. In a White House meeting, Hoover asked the CEOs of major manufacturing corporations to not cut their wages. They agreed to maintain wages, and seem to have honored that agreement during the first two years of the Great Depression—manufacturing wages fell only 4.4% between December 1929 and September 1931. [See Lamont (1930) for a description of the meeting.[13]]

13. The effect of this intervention, however, weakened during the last two years of the Depression. By late 1931, Gerard Swope, CEO of General Electric, circulated an industrial plan that would have cartelized much of the U.S. economy. Hoover denounced

It is worth noting that there are also manufacturing-wage surveys produced by the Bureau of Labor Statistics (BLS) that could be used to measure the distorted wage, but these surveys do not cover all manufacturing industries, and they suffer from sampling problems. In particular, large firms, which tend to pay higher wages than small firms, were oversampled.

We now turn to choosing the fraction of the economy distorted by the high wage. In our model, this fraction is governed by the parameter α. Unfortunately, we do not know of any established measures of the fraction of the economy distorted by the high wage. The data we presented earlier suggest that on average, manufacturers paid high wages, but farmers did not. But since we do not have wage measures across the entire economy of the same quality as the Conference Board's wage data, it is difficult to estimate how much of the economy was subject to high wages.[14]

To address this uncertainty over the fraction of the economy distorted by the high wage, we conduct our analysis for two values of the parameter α. We first assume that the entire manufacturing sector was subject to the distorted wage. Given Hoover's view about the importance of maintaining high wages, we also assume the federal government paid the high wage. These two sectors account for about 28% of employment in 1929. We therefore choose a benchmark value for α such that this sector accounts for 28% of employment in the deterministic, flexible-price steady state of the model. We also conduct the analysis for $\alpha = 0.50$, which implies that the distorted sector was 50% of the economy. This choice seems to be a plausible upper bound on the fraction of the economy distorted by the high wage. This is because at least 30% of workers were not paid the high wage (farming and sole proprietors), and because there do not seem to be direct measures of wages of sufficient quality that indicate that half of all workers were paid wages above trend values.

3.2.3 Computing the Equilibrium Computation of the equilibrium of the model with high manufacturing wages is facilitated by our assumption

this plan and refused to recommend it to Congress. Nominal manufacturing wages began to fall significantly after Hoover's condemnation of the Swope plan.

14. There are wage measures in some nonmanufacturing sectors, and there are also BLS payroll and employment data outside of manufacturing that can be used to construct average employee compensation. A difficulty with these BLS payroll data is that the coverage is narrow in some sectors, the data do not include hours, and in some sectors the data combine all classes of workers, including executives. This last fact suggests that constructing measures of compensation per employee from these data is subject to significant compositional bias. We discuss compositional bias, and how it may have affected different wage measures during the Depression, at the end of this section.

that each wage shock is a completely unexpected, one-time event—the household expects at each date that the economy will return to pure market clearing the following period. This permits us to compute the equilibrium for each year of the Depression (1930–1933) recursively.

Since households expect the economy to return to market clearing in the following period, the value of capital next period is a function of the single state variable in the economy, the aggregate capital stock. To compute the equilibrium at date t when the manufacturing wage is higher than its competitive level, we use a log-linear approximation of the right-hand side of the Euler equation from the pure market-clearing model around its steady state. This approximation allows us to estimate the marginal value of an additional unit of capital and is used with the static first-order conditions of the model to compute the equilibrium for each year of the Depression. This involves solving N nonlinear equations in N unknowns for each year. We feed our measures of the manufacturing wage for 1930–1933 into the model and compute the equilibrium path of the economy for these years. Our findings are presented in the next section.

3.3 MACROECONOMIC EFFECTS OF HIGH WAGES: 1930–1933

Tables 7–9 show the predicted path of the U.S. economy between 1930 and 1933 for our model with benchmark parameter values and alternative parameter values. We find that the predicted depression for all these parameter values is much less severe than the actual U.S. Great Depression.

Table 7 shows the equilibrium path of output, consumption, and investment from our benchmark model with about 28% of the economy distorted by the high wage. Predicted real output is about 1% below trend in 1930 and about 2% to 3% below trend between 1931 and 1933. Most of the decrease in economic activity occurs in the distorted sector. The high wage reduces employment in the distorted sector about 7% below trend. In contrast, employment in the nondistorted sector falls

Table 7 PREDICTED GREAT DEPRESSION (1929 = 100), BENCHMARK MODEL

Year	Y	C	I	h_m	h_n
1930	99.2	99.8	96.9	97.8	99.3
1931	97.3	99.3	90.4	93.1	97.8
1932	97.2	98.9	91.1	93.3	98.0
1933	97.8	98.7	94.6	95.4	98.8

Table 8 DECOMPOSITION OF PREDICTED OUTPUT: DIRECT AND INDIRECT EFFECTS, BENCHMARK MODEL

		Effect (%)	
Year	$\%\Delta Y$	*Direct*	*Indirect*
1930	−0.8	−1.5	0.7
1931	−2.7	−6.4	3.7
1932	−2.8	−6.3	3.5
1933	−2.2	−4.2	2.0

Table 9 PREDICTED GREAT DEPRESSION: LARGE DISTORTED SECTOR (1929 = 100)

Year	Y	C	I	h_m	h_n
1930	98.7	99.7	95.0	97.2	98.9
1931	95.6	98.9	84.2	91.2	96.4
1932	95.2	98.2	84.9	91.1	96.5
1933	96.1	97.8	90.1	93.6	97.8

only about 2% to 3% below trend. These predicted decreases in economic activity are much smaller than the observed decreases in output, consumption, investment, and employment that occurred between 1929 and 1933.

There are two reasons why predicted economic activity falls so little compared to the actual decrease in economic activity. First, the distorted sector is relatively small, which means that the direct effect of the high wage on aggregate output is small. Second, the indirect, general equilibrium effects tend to reduce, rather than amplify, the direct effects.

The most important indirect effect is the increase in the relative price of manufactured good, which rises 3% to 4% above its steady-state level after 1930. The relative price rises because the manufactured good is in relatively scarce supply and is not highly substitutable with the nonmanufactured good. This increase offsets some of the distorting effects of the high manufacturing wage. Equation (7) shows that each percentage-point increase in the relative price of the manufactured good effectively reduces the fixed wage by one percentage point. Thus, the 4.4% increase in the relative price of manufactured goods in 1931 effectively reduces the manufacturing wage from 6.8% above trend to just 2.4% above trend.

Table 8 shows a decomposition of the change in output due to the direct and indirect effects. This decomposition shows that the negative direct effects are partially offset by the indirect effects.

The effects of the high wage depend on all the model parameters, but in particular depend on the share parameter α. We therefore assess the robustness of the results by increasing the distorted share of the economy to 50%, which in our view is a reasonable upper bound on the distorted share of the economy.

Table 9 shows the equilibrium path of the model economy with $\alpha = 0.5$. This higher value in the model produces a larger decrease in economic activity, but the decrease is still much smaller than the actual Great Depression. Real output is predicted to be 4.8% below trend in 1932 with $\alpha = 0.5$, compared to 2.8% below trend in the benchmark version of the model. We thus find that raising the share of the economy that must pay the high wage to 50% does not materially change the findings.

We also conducted the analysis by reducing the elasticity of substitution between the two sectors from 0.5 to 0.1. We do not present these results, because this change did not significantly affect the results. Output falls about one percentage point more than in the benchmark model, and the relative price of the good from the distorted sector rises more.

These results suggest that the high wage was not the primary cause of the Great Depression. Given our measure of the wage from the manufacturing sector, our benchmark model shows that this wage accounts for about a 3% decline in output at the trough of the Great Depression, compared to an actual 38% decline. Increasing the size of the distorted sector to 50% or reducing the substitution elasticity to 0.1 did not significantly change the results.

This simple model focused on the basic distorting effects of an above-market wage through two channels—the direct reduction in sectoral labor input, and the general equilibrium effects of the high wage through prices to the other sectors of the economy. One reason why the model doesn't generate a large depression is that the general equilibrium effects offset some of the distortion of the high wage. In particular, the sectoral high wage reduces output primarily in the distorted sector, and this drives up that sector's relative price and reduces the macroeconomic effect of the distortion.

This result raises the possibility that the wage story might have a better chance if the theory could be modified to eliminate the relative price increase. This approach is not likely to be successful, however. Eliminating the relative price increase arising from the wage distortion requires substantially reducing the demand for the output of that sector. This reduction in demand requires a second shock. In our model, this second shock is a decline in the parameter α, which governs the distorted sector's share of aggregate output. Reducing α would reduce the demand for goods from the distorted sector and would prevent the relative price of the distorted good from rising. But this higher real wage won't generate a major depression, because the reduction in α also reduces the quantitative importance of the distorted sector and thus reduces the macroeconomic effect of that sector.[15]

Finally, our model indicates another difficulty with the wage hypothesis: the timing of the depression and the timing of wage increases. With the exception of 1931, real wage increases do not occur at the same time as output declines. Real output fell 13% in 1930, yet the real manufacturing wage remained close to trend. Similarly, real output fell more than 17% between 1931 and 1932, yet the real manufacturing wage was roughly unchanged between 1931 and 1932. This lack of coincidence between the timing of output changes and wage increases suggests that some other shock reduced output in those years.

Accounting for the Depression through imperfectly flexible manufacturing wages is difficult—the real wage increase is too small and affects too little of the economy, and wage increases coincide with lower output only in 1931. The hypothesis would have a better chance if wages were significantly higher and affected more of the economy, and if there were more coincidence between the timing of wage increases and the Great Depression. But as the next section describes, these factors are unlikely.

3.4 MEASURED WAGES ARE PROBABLY BIASED UPWARDS

We are skeptical that actual wages were as high as the manufacturing-wage measures suggest. This is because the composition of employees changed during the Depression, and this compositional shift likely induces upward bias in the wage measures. Researchers who analyze the cyclical pattern of real wages argue that cyclical changes in the composition of employment leads to wage measures that are biased upward during recessions and biased downward during expansions. This is because hours of low-wage earners tend to be much more sensitive to the business cycle than hours of high-wage earners. Consequently, the average employed worker during a recession tends to be a higher-wage earner than the average employed worker during an expansion.

Lebergott (1991) and Margo (1993) argue that compositional effects may have been particularly important during the Great Depression. Lebergott argues that compositional shifts in employee quality and in the quality of operating establishments may result in measured wages substantially

15. This discussion highlights the problems associated with focusing on the product wage instead of the real wage. In particular, high product wage results from a combination of a positive shock to real wages and a negative shock to product demand.

overstating actual wages. He indicates that layoffs were concentrated among low-wage, young workers, which tends to increase the average measured wage of those individuals remaining employed. He also notes that relatively young firms, rather than older established firms, failed during the Depression, and that these younger firms tended to pay significantly lower wages. This compositional change also raises the average measured wage of those individuals remaining employed. Margo makes a very similar point regarding compositional bias.[16]

How large are these biases? Lebergott cites some microeconomic evidence which, he argues, points to significant upward bias arising from changes in employee quality. He notes that Westinghouse and General Electric retained their most productive employees during the Depression, and also cut these employees' wages by 10% between 1929 and 1931. However, the Conference Board's wage survey for this industry, which was heavily influenced by these two firms, shows that wages were unchanged during this period. This deviation between the wages paid by these two firms and the survey wage is likely due to changes in the composition of employees at the two firms.[17] While this microeconomic example suggests the possibility of important compositional biases, we do not have the necessary individual wage and employment data to measure aggregate compositional effects. To obtain a rough idea of how compositional shifts may have affected measured wages more broadly, we compute estimates of compositional bias from the Current Population Survey (CPS) and the Panel Study of Income Dynamics (PSID). We estimate the bias using two separate computations. The first computation is motivated by Lebergott's argument that employment loss was concentrated among the lowest-wage earners. Determining how this compositional shift affects the wage requires specifying how

16. There is also evidence that some firms reclassified workers down (e.g., a foreman works as an assembly-line worker); see Bernanke and Carey (1996) and Lebergott (1991). This would tend to bias wages in the opposite direction if the individual's wage was unchanged, but the value of the individual's marginal product fell. It is unclear, however, whether reclassified workers' wages were changed as a consequence of the reclassification.

17. Lebergott notes that these two firms laid off low-productivity workers, reassigned some higher-skilled workers, and assigned the retained workers to either 2-, 3-, or 4-day workweeks, depending on worker ability, with the most productive workers receiving 4-day workweeks. Lebergott clearly interprets these personnel decisions and their impact on the measured wage as an example of upward compositional wage bias. As we noted above, this interpretation is clearly warranted provided that those reclassified employees who performed different tasks were paid their value marginal product. If these employees were paid in excess of their value marginal product, however, this effect would tend to offset the upward wage bias resulting from the change in the composition of employees and the allocation of work towards the most productive employees.

employment loss was distributed during the Depression. To capture Lebergott's argument, we assume that the bottom 20% of wage earners lost employment and that the remaining employment loss was evenly distributed across all other workers. Using CPS data from 1998 for all full-time workers, we find that the average wage for the top 80% of wage earners is about 15% higher than for all full-time wage earners. This implies that the average wage during the Great Depression may have been overstated by 15% if the distribution of employment loss was concentrated among low-wage earners in this fashion, and if the wage distribution in the 1930s was similar to the wage distribution today.[18]

Our second computation uses measures of cyclical compositional wage bias from postwar data to estimate the compositional bias in the Depression. Solon, Barsky, and Parker (1994) estimate the difference between the response to fluctuations in output relative to trends between aggregate wages and individual wages from the PSID. This difference is a direct measure of the compositional bias from using aggregate wages as a measure of an average wage, and the bias is an increasing function of the magnitude of the decrease in output. Applying their estimates to the Depression suggests that compositional shifts biased measured wages up by about 18%.[19]

While we cannot draw a firm conclusion about the quantitative magnitude of compositional wage bias during the Depression, these estimates suggest that measured wages may be substantially upward biased.[20] This suggests that manufacturing wages may have been significantly below trend at the trough of the Great Depression after correction for compositional bias.[21]

18. We thank Daniel Hamermesh for performing this computation. The data are from the CPS-ORG 1998. Full-time workers are defined as those working 35 or more hours per week.
19. Solon, Barsky, and Parker (1994) only reported the differences in the coefficient between the fluctuations in the coefficient on the unemployment rate relative to trend. We thank Jonathon Parker for computing their estimates using real chain-weighted GDP rather than unemployment. The measure of the compositional bias is (0.558–0.0896) log[*d*GDP (1933)/*d*GDP (1929)], where *d*GDP is the deviation of real GDP per adult from trend.
20. It is interesting to note that the cross-sectional differences in employment and wages between manufacturing and farming are consistent with significant compositional bias. Since the bias should be most severe for sectors in which employment fell substantially, we should observe relatively high wages associated with low employment. Manufacturing hours fell more than 40%, and measured wages were about 5% above trend. In contrast, farm hours remained near trend, and measured wages fell substantially.
21. Bordo, Erceg, and Evans (BEE, 2000) construct a measure of hourly employee compensation that rises about 4 percentage points more than the Conference Board's measure of hourly manufacturing wages, and use changes in this measure as a proxy for changes in the average wage during the Great Depression. There are two reasons why the change in their average compensation measure may deviate considerably from the

4. *How Much of the Great Depression Was Due to Banking Shocks?*

This section asks how much banking shocks contributed to the Great Depression. Unfortunately, there is no standard version of the neoclassical growth model with financial intermediation to use for this purpose, nor is there a standard definition of the banking shock—at least not as an explicit shock to primitives (technologies or endowments) that can be used in a general equilibrium model. We therefore develop a simple benchmark neoclassical model in which banking output, which is produced with deposits and information capital, is an input into production of the economy's final good. We define the banking shock to be the stock of information capital lost as a consequence of bank closings. This definition is consistent with the literature which associates the banking shock with bank failures and the destruction of information capital. We use the model to address three questions: How much did bank closings reduce intermediation capital? How much did this loss of intermediation capital reduce output? Are the predicted effects of bank closings on other variables consistent with the data?

4.1 A MODEL WITH FINANCIAL INTERMEDIATION

Our model extends the standard neoclassical growth model by requiring that some investment be *intermediated.* This modifies the standard model to include both internally and externally financed investment. In our model, a fraction of the capital stock is transferred from households to firms by an intermediation technology that uses real resources. This technology gives rise to borrowing and lending rates. The model allows us to analyze the effects of shocks to the intermediation technology on output, intermediated and internally financed investment, and borrowing and lending rates.

---

change in the average person's wage during the Depression. First, as we noted before, it is difficult to infer *individual* wage changes from an aggregated compensation measure because of compositional shifts in employment. Thus, their compensation measure is also subject to upward bias under the assumption that layoffs were concentrated among low-wage earners. Second, there is an inconsistency in their construction of total hours worked which is used in measuring average hourly compensation. In particular, their measure of total hours worked is equal to the number of full-time equivalent employees (from the NIPA) multiplied by Kendrick's (1961) average hours worked for full-time equivalent workers, which includes not only employees, but also proprietors and unpaid family workers. These latter two groups are quantitatively important, accounting for about 38% of Kendrick's full-time equivalent workers in 1929 (see p. 304). For BEE's calculation, this measure of hours would be correct only if fluctuations in proprietor and unpaid family hours were identical to fluctuations in employee hours.

We now describe the model in detail. There are two plants that produce a single physical good using capital. At the beginning of the period there are three types of capital: *installed physical capital* at each plant, which we denote by K_1 and K_2, respectively; *uninstalled physical capital,* which is held by households and is denoted by D; and *intermediation capital,* which we denote by Z. Intermediation capital is in fixed supply.

The capital stocks at each plant can be increased during the period with uninstalled capital. We denote by x_1 and x_2 the amounts that are installed during the period. This uninstalled capital must be intermediated, and some of this capital is used up during the intermediation process. The capital available for production is thus $K_j + x_j$. At the end of each period, some output is used to costlessly augment the capital stock at each plant, and the remainder is distributed to households who either consume it or hold it as uninstalled capital for the following period.

The plant technologies are subject to an i.i.d. shock, which is realized at the beginning of each period. The production shock can take on two levels: ϵ_h and ϵ_l, where $\epsilon_h > \epsilon_l > 0$. One plant receives the high shock ϵ_h, and one plant receives the low shock ϵ_l. Each plant has an equal probability of receiving the high productivity level, and we normalize the shocks so that $0.5(\epsilon_h + \epsilon_l) = 1$.

After the idiosyncratic plant productivity shock has been realized, uninstalled capital is allocated to the two production plants according to

$$\sum_{j=1}^{2} x_j \leq G(D, Z).$$

We will assume that G exhibits constant returns to scale (CRS) and that $G(D,Z) \leq D$. The resources used in the intermediation process are the quantity $D - G(D,Z)$.

Plant output is produced from a CRS Cobb–Douglas technology that uses capital and labor. For simplicity, we assume that there is one unit of labor at each plant, and that labor is in fixed supply. Plant output is given by

$$y_j = A\epsilon_j(K_j + x_j)^{\gamma}.$$

Plant output is used for either consumption or investment. Investment from retained output has no intermediation cost. The resource constraint for this economy is

$$\sum_j \left[A\epsilon_j(K_j + x_j)^{\gamma} - K_j' \right] \geq c + D',$$

where D' denotes the next period's level of uninstalled capital and K_j' the amount of capital installed at plant i at the beginning of the next period. We require that output net of retained investment be nonnegative.

The social planning problem for this economy is given by

$$\text{P1:} \qquad \max_{\{c_t, x_{i,t}, K_{i,t+1,Dt+1}\}} \sum_{t=0}^{\infty} \beta^t u(c_t) \tag{8}$$

subject to

$$G(D_t, Z) \geq \sum_j x_{j,t}, \tag{9}$$

$$\sum_j [A\epsilon_{t,j}(K_{t,j} + x_{t,j})^\gamma - K_{t+1,j}] \geq c_t + D_{t+1}, \tag{10}$$

$$A\epsilon_{jt}(K_{jt} + x_{jt})^\gamma - K_{jt}' \geq 0 \qquad \text{for each} \quad j = 1, 2 \text{ and } t, \tag{11}$$

$$x_{jt} \geq 0 \qquad \text{for each} \quad j = 1, 2 \text{ and } t. \tag{12}$$

We assume that the difference in the ϵ_h and ϵ_l is small enough that the nonnegativity constraint on retained earnings given in equation (11) never binds. Since the productivity shocks are i.i.d., it is optimal to set $K_1 = K_2 = K/2$. Thus, we aggregate plant capital and define the state variables to be (K, D).

The solution to this planning problem can be decentralized as a competitive equilibrium. This allows us to solve for equilibrium borrowing and lending rates. We assume competitive profit-maximizing firms operate each plant. We also assume that there is a competitive profit-maximizing intermediary who operates the intermediation technology. This intermediary receives funds from the household at the savings rate $1 + r_s$ and loans it out at the borrowing rate $1 + r_b$.

In equilibrium, the marginal cost of additional capital to the high productivity plant, $1 + r_b$, must be equal to its marginal productivity:

$$1 + r_b = \gamma A \epsilon_h \left(\frac{K}{2} + G(D_t, Z) \right)^{\gamma - 1}.$$

Similarly, the interest rate on savings must be just equal to the return on uninstalled capital:

$$1 + r_s = \gamma A \epsilon_h \left(\frac{K}{2} + G(D_t, Z) \right)^{\gamma-1} G_D(D_t, Z).$$

The *spread* between these two rates is

$$r_b - r_s = \gamma A \epsilon_h \left(\frac{K}{2} + G(D_t, Z) \right)^{\gamma-1} G_Z(D_t, Z).$$

Note that this spread is a decreasing function of the level of intermediation capital, Z. Thus, a decrease in Z will raise the spread between these two rates. It can also be shown that a decrease in Z will reduce output and the quantity of intermediated capital, but will increase the quantity of internally financed capital as firms substitute out of intermediation into internal finance. These results are presented in the Appendix.

4.2 HOW MUCH DID BANK CLOSINGS REDUCE INTERMEDIATION CAPITAL?

Our model provides a measure of the banking shock—the loss of intermediation capital as a consequence of bank closings. Assuming that intermediation capital is in fixed supply and is bank-specific, the fraction of intermediation capital lost due to bank closings is equal to the fraction of deposits in suspended or failed banks. This implication follows directly from the CRS intermediation technology. We therefore infer from the deposit data presented in Table 6 that bank closings cumulatively reduced intermediation capital about 8% between 1930 and 1932, and about 19% between 1930 and 1933.

4.3 HOW MUCH DID THE BANKING SHOCK REDUCE OUTPUT?

We now use our model to evaluate the contribution of this decrease in intermediation capital to the Depression. Fixing (K_t, D_t), the elasticity of output with respect to intermediation capital is given by

$$\frac{dY_t}{dZ_t}\frac{Z_t}{Y_t} = \frac{A\epsilon_h[K_t/2 + G(D_t, Z_t)]^{\gamma-1}\gamma G_Z(D_t, Z_t)Z_t}{Y_t}.$$

The numerator of the right-hand side is the total return to intermediation. Therefore, the left-hand side of this equation is the intermediation sector's share of value added. This value-added-share elasticity result is not specific to our model. In fact, *any* model with a CRS technology for producing final goods has the feature that, to a first-order approximation, the elasticity of the final good with respect to any intermediate good is equal to that good's share of value added.

Banking's share of value added was about 1% in the 1930s. In fact, the value-added share of an entire finance, insurance, and real estate (FIRE) sector was only about 13% in 1929, and dropped to 11% in 1933.[22] Note that this value-added measure actually overstates the elasticity, since our model attributes all of banking's value added to intermediation capital. Some of this sector's value added will be paid to labor, which means that the elasticity of output with respect to intermediation capital is actually lower than the share of value added. With this small elasticity, our model predicts that the decrease in intermediation capital caused by bank closings reduced output less than 1% between 1929 and 1933.

4.3.1 Can a Low Substitution Elasticity Plausibly Magnify the Shock? The macroeconomic effect of destroyed intermediation capital would be larger if bank finance and alternative forms of finance or other inputs were poor substitutes. A low substitution elasticity, however, is inconsistent with the data. If banking shocks were an important contributing factor to the Depression *and* this substitution elasticity was very low, the cost share of banking and of FIRE should have increased considerably during the 1930s. In contrast, the cost share of FIRE falls from 13% in 1929 to 11% in 1933, and banking's cost share falls from about 1.4% to about 1% over the same period.[23]

4.3.2 Can Externalities Magnify the Impact of the Shock? Evidence from State-Level Data An externality associated with intermediation capital could increase the economic impact of an intermediation shock. One drawback to the externality story is that there are many different ways of putting externalities into models, but often these externalities do not have strong micro foundations, nor are they straightforward to evaluate quantitatively. The banking/depression literature, however, suggests a specific type of externality that is straightforward to assess. This literature argues that bank failures reduced output by destroying local bank information, and thus suggests a productive externality associated with intermediation capital that affects local production. We therefore consider a version of our model in which there are N regions, and aggregate output is the sum of regional outputs.

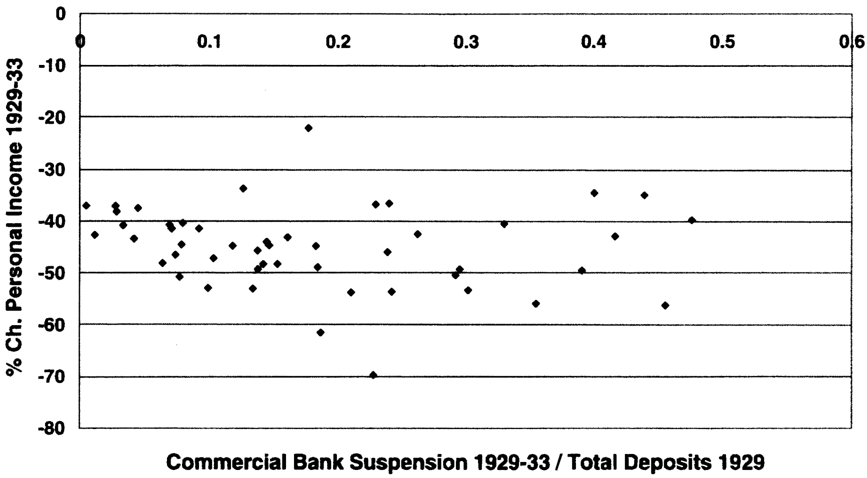
Suppose that output in region i is given by

$$Y_i = Z_i^{\delta} \sum_{j=1}^{2} A\epsilon_{t,ij}(K_{t,ij} + x_{t,ij})^{\gamma} H_{ij}^{1-\gamma},$$

22. Banking accounted for 10% of value added in FIRE in 1947. Kuznets (1941) reports a similar number for the period 1919–1938.

23. The data on banking's cost share are from Kuznets (1941, p. 731).

Figure 1 PERSONAL INCOME VS. SUSPENSIONS BY STATE DURING THE DEPRESSION

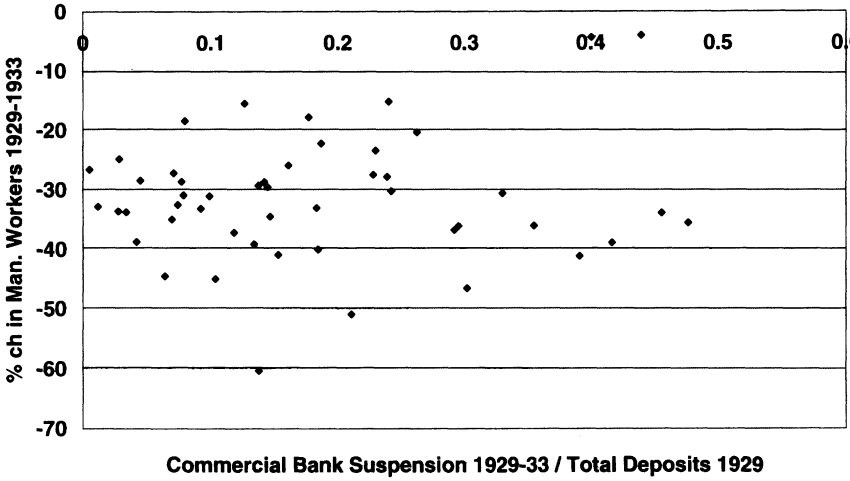


Where Z_i^δ is the productive externality from intermediation capital in region i. This version of our model predicts that regions that experience many bank closings should also experience relatively large depressions. We assess this prediction by first defining a region as a state and then computing the correlation between bank suspensions or failures and economic activity across the 48 U.S. states during the Great Depression. Note that this comparison is a regional extension of Bernanke's (1983) influential paper, which found that *aggregate* bank suspensions and failures were negatively correlated with *aggregate* output.

Figure 1 shows a scatterplot of the sum of suspended and failed deposits from 1929 to 1933 relative to total deposits in 1929 vs. the percentage change in nominal personal income between 1929 and 1933 by state. The most striking feature of these data is that the significant negative correlation between bank closings and output documented by Bernanke (1983) at the aggregate level does not emerge at the state level.[24] The plot shows no systematic relationship between the concentration of banking shocks and the severity of the Depression across states. The correlation between suspended deposits and nominal income is -0.15 and is not significantly different from zero. A regression of the percentage change in personal income divided by the aggregate GDP deflator on the fraction of deposits in suspended and failed banks yields an R^2 of 0.014 and a slope

24. Temin (1989) also notes that some bank-failure episodes were very regionally concentrated.

Figure 2 MANUFACTURING WORKERS VS. BANK SUSPENSIONS BY STATE DURING THE DEPRESSION



coefficient that is not significantly different from zero.[25] We also examined the relationship between the same measure of deposits and an alternative statewide measure of real economic activity—the percentage change in manufacturing employment between 1929 and 1933.[26] Figure 2 shows a scatterplot between these two variables. The correlation between these data is, in fact, positive rather than negative: 0.12.[27]

These data do not support the standard banking story for the Great Depression: that bank closings reduced output by destroying local information capital. The relatively small bank shock, combined with banking's small share in the production function, and the lack of any correlation between state-level bank closings and economic activity indicate that if banking was an important contributing factor during the Great Depression, it must have operated through some alternative mechanism in which the shock was much larger and was operative at the aggregate

25. We estimated two other versions of this equation. To control for level affects, we defined a dummy variable that takes the value of 1 if a state's per capita income was above the median. We used this dummy variable to analyze (1) an intercept shift and (2) an intercept shift and a slope coefficient shift. The results were quite similar to those for the simpler specification.

26. These data are from the biannual Census of Manufacturers.

27. The lack of a systematic pattern between bank closings and economic activity at the state level raises the possibility that the correlation between aggregate bank closing and aggregate output may indicate that aggregate bank closings are proxying for another variable. This is consistent with Green and Whiteman (1992).

level rather than the regional level. We analyze an alternative mechanism in the following section.

4.4 OTHER SHOCKS TO BANK CAPACITY

An alternative banking story is that depositors were afraid of bank runs and consequently withdrew deposits from all banks. This alternative story would have a better chance than the bank-failure story if the decrease in deposits resulting from depositor fear was substantially larger than the decrease in deposits at closed banks. This story is difficult to evaluate, however, because it is unclear how much of the decrease in total deposits was due to depositor fear and how much was an endogenous response to the large decrease in overall economic activity. Consequently, we can't measure the size of this shock associated with depositor fear.

Despite this measurement problem, our model makes one specific prediction about this story that can be evaluated. According to this story, banking services are in relatively scarce supply because of deposit withdrawal. The model predicts that an exogenous decrease in deposits will decrease the deposit/output ratio. This result is not specific to our model, but follows directly from CRS in production and the relative scarcity of deposits. The actual deposit/output ratio, however, differs considerably from this prediction. Table 6 shows that the ratio *rises* from 0.58 to 0.78 between 1929 and 1932. This increase implies that deposits were not relatively scarce during the Great Depression.

Even if deposits were relatively scarce because of depositor fear, however, there is no theoretical presumption that this would generate a massive depression, because banking's share of value added is small. In fact, these cost-share statistics suggest a presumption that banking shocks should tend to have small, rather than large, macroeconomic effects. The Irish bank strikes of the 1960–1970s provide evidence that is consistent with this latter view. Murphy (1978) reports that on three occasions between 1966 and 1976, industrial disputes led to the shutdown of the Associated Banks, which accounted for over 80% of Irish M2. These strikes, the longest of which was six months, represent negative, exogenous shocks to the banking sector that are larger than any plausible bank capacity shock that might have occurred during the U.S. Great Depression. The macroeconomic effects of these strikes, however, were small. During the longest strike, detrended retail sales fell about 4%, and real output rose over the full calendar year of 1970. Murphy argues that the strike did not have important effects because households and firms developed substitutes for bank services, including private trade credit. These "natural experiments" show that a long-term shutdown of most of a country's banking system—a shutdown much larger than that which occurred during the Great Depression—need not substantially reduce economic activity.

These data are inconsistent with the view that the Depression was caused by a large exogenous decrease in deposits. Instead, they are consistent with the view that the decrease in deposits may have been primarily an endogenous response to the overall decline in economic activity.

4.5 OTHER IMPLICATIONS OF A BANKING SHOCK

Our analyses of the banking story—through an explicit shock based on bank closings and through an alternative story based on a decrease in overall bank capacity—do not support the view that banking was an important contributing factor to the Great Depression. Of course, any explicit analysis along these lines depends on a definition and measure of the banking shock. Some other aspects of the banking story can be assessed without an explicit definition and measure of this shock. Our model makes two such predictions. The first prediction is that any reduction in banking capacity should increase the spread between deposit and loan interest rates. The second is that any reduction in the availability of intermediated loans, or any increase in the cost of intermediated loans, should lead firms to substitute out of external finance and into internal finance.

4.5.1 Impact of the Banking Shock on the Cost of Intermediation Our model predicts that a negative shock to the banking sector increases the spread between the interest rate on intermediated debt and the bank's cost of funds. Before examining changes in interest spreads, it is important to recognize that these spreads are affected not just by intermediation shocks, but also by changes in loan maturity, changes in the composition of borrowers, and changes in default risk. Since these other factors may have changed significantly during the Great Depression, it is very difficult to separately identify changes in interest spreads that are due to changes in the intermediation technology.

This identification problem leads us to make two comparisons of interest-rate spreads. We first examine an interest-rate spread between a collateralized, short-term obligation and short-term Treasuries during the Great Depression. This comparison permits us to reasonably control for some of the other factors affecting interest spreads: both securities have roughly constant maturities, and the collateralized nature of the private obligation limits the effect of changes in either default probability or the composition of borrowers.

Our second comparison presents the spread between long-term, quality-rated corporate securities and government bonds during the Great Depression. This analysis has been conducted in the previous literature for low-quality corporate debt. However, the change in this low-quality spread cannot be solely attributed to intermediation shocks, because the default risk on these lower-quality securities increased during the Great Depression. Consequently, it is unclear how much of the change in the spread was due to intermediation, and how much was due to higher default risk. To confront this identification problem, we present spreads on high-quality securities whose default risk may not have changed much during the Depression. If a negative intermediation shock was important, spreads on all types of securities would be expected to rise in the 1930s. Alternatively, if the spread on low-quality debt was higher largely because of changes in default risk, the spread should be roughly unchanged for the highest-quality securities, but should rise for lower-quality securities.

We first analyze our measure of the short-term spread. Table 10 presents the spread between 3- to 6-month banker's acceptances and 3- to 6-month Treasury notes. The banker's acceptances are collateralized, which controls for changes in default risk. Since the bank that originally discounted the bill stood as the guarantor of its ultimate payment, it is important to note that the bank performed an important intermediation function in the production of this asset. Consequently, a negative shock to the intermediation technology should have increased the spread between these two securities. The table shows that the spread between the rate on banker's acceptances and Treasuries does not change much during the Depression. The stability of this interest-rate spread therefore

Table 10 BANKER'S ACCEPTANCE RATES AND GOVERNMENT SECURITY YIELDS[a]

	Interest rate (%/yr)		
Year	*(1) Banker's acceptances*	*(2) Short-term govt. debt*	*(1) − (2)*
1928	4.09	3.97	0.12
1929	5.03	4.42	0.61
1930	2.48	2.23	0.25
1931	1.57	1.15	0.42
1932	1.28	0.78	0.50
1933	0.63	0.26	0.37

[a]The data are from Board of Governors of the Federal Reserve System (1943).

Table 11 INTEREST-RATE SPREADS BETWEEN CORPORATE AND GOVERNMENT BONDS[a]

	Spread (%/yr)			
Year	Aaa − Gov	Aa − Gov	A − Gov	Baa − Gov
1929	1.13	1.33	1.68	2.30
1930	1.26	1.48	1.84	2.61
1931	1.24	1.71	2.67	4.28
1932	1.33	2.30	3.52	5.62
1933	1.18	1.92	2.78	4.45
Avg.	1.25	1.85	2.70	4.24

[a]The data are from Board of Governors of the Federal Reserve System (1943).

indicates that the efficiency of this type of intermediation was not impaired during the Depression.[28]

We next examine the spread between the rates on corporate bonds, which are a substitute for bank finance for large firms, and U.S. government bonds. Table 11 shows the spread for corporate bonds of different qualities—Aaa (lowest default risk), Aa, A, and Baa. There are two striking features of these data. First, the average increase in interest spreads is fairly small. Second, the magnitude of the increases in the spread is directly related to the quality of the debt: the average spread changes very little for high-quality debt, but increases for lower-quality debt.

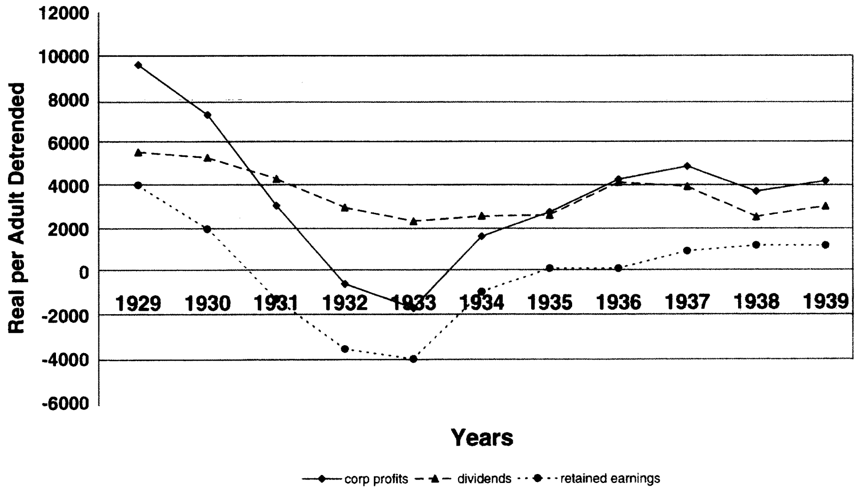
These data are consistent with the view that changes in default risk were an important contributing factor to higher spreads. To illustrate how these changes could have affected spreads, suppose that Baa securities pay off 60% of the principal if the firm defaults. With this assumption, the 230-basis-point spread between Treasuries and Baa bonds in 1929 implies that the default probability for Baa bonds was about 5% at that time. It also implies that the average 424-basis-point Baa spread during the Depression can be *completely* explained by an increase in this default probability from 5% to 8%. This increase does not seem implausible during this period.[29]

While we cannot draw a firm conclusion about the quantitative impor-

28. The gap between commercial loan rates and short-term government securities rose about 250 basis points during the Depression. The gap between commercial loans and government bonds, however, narrowed by about 120 basis points. Given the caveats mentioned above, plus a steepening in the yield curve, it is not clear how to interpret these changes.

29. Cole and Ohanian (2000) present a monthly analysis of those spreads, which permits a closer examination of changes in spreads with the onset of banking crises. We did not find much evidence of large increases in interest spreads around these periods.

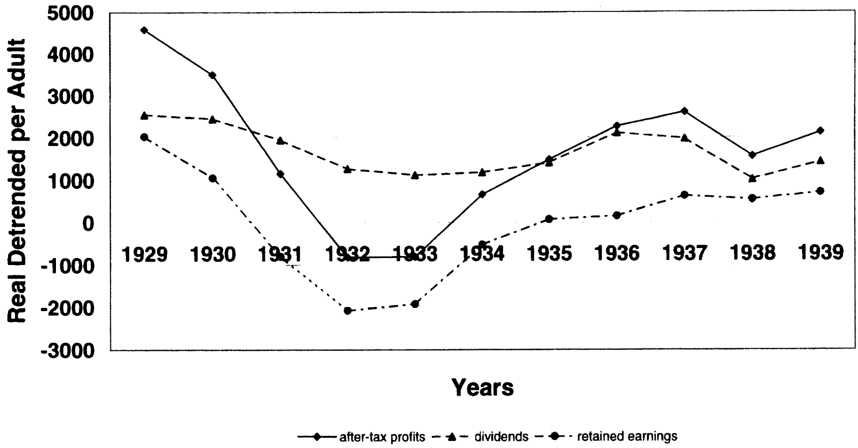
Figure 3 DOMESTIC INDUSTRIES: PROFITS, DIVIDENDS, AND RETAINED EARNINGS (WITHOUT IVA)



tance of changes in default risk, it is certainly true that default risk rose during the Depression and thus contributed to higher spreads. But even if we abstract from default risk and completely attribute these higher spreads to negative intermediation shocks, it seems unlikely that these increases—ranging from 12 to 194 basis points—can plausibly explain the Great Depression. If higher spreads were the key to understanding the Great Depression, they should have increased much more during the Depression than during milder recessions. But this is not the case. The average rise in the Baa–Treasury spread for all post-World War II recessions is more than 200 basis points. This includes several recessions in the 1970s and early 1980s in which this spread rose as much as 500 basis points. All of these recessions were much milder than the Great Depression, despite these much larger interest-spread increases.

In summary, interest spreads did not rise much outside of low-quality corporate securities, and it is unclear how much of this increase is due to intermediation shocks. Moreover, the average increase in spreads does not seem to be nearly large enough to account for the magnitude of the Great Depression. In the following section, we present the second prediction of our model that does not rely on an explicit definition of the banking shock. Our model shows that *if* a negative banking shock increased the cost of funds and disrupted economic activity, firms should have increased retained earnings.

Figure 4 MANUFACTURING INDUSTRIES: PROFITS, DIVIDENDS, AND RETAINED EARNINGS (WITHOUT IVA)



4.5.2 Impact of Banking Shocks on Other Sources of Finance The theory predicts that a reduction in the availability of intermediated finance, or an increase in the cost of intermediated finance, should lead firms to substitute out of intermediated finance and increase retained earnings. Figures 3 and 4 show real profits, dividends, and retained earnings per adult relative to trend in the entire corporate sector and in the manufacturing subsector, respectively. The most striking feature of these data is that firms were not increasing retained earnings as the theory predicts. In sharp contrast, retained earnings fell substantially as firms maintained high dividend payments. Corporate profits fell by nearly 40% between 1929 and 1930, but dividend payments fell by only about 4%. Profits decreased by over 70% between 1929 and 1931, but dividend payments fell by only 25% during that period. By 1932, corporations experienced substantial losses, but retained earnings fell even more as firms maintained dividend payments equal to 51% of their 1929 level. This pattern also emerges at the sectoral level. Figure 4 shows that a very similar pattern prevailed among manufacturing corporations, and Table 12 shows that this pattern continues among durable and nondurable manufacturers and among mining corporations.[30]

30. There was some variation in dividend payouts at the industry level. For example, dividends in the tobacco industry were particularly high during the Depression. These outliers did not affect the sectoral-level statistics much. Real nondurable manufacturing dividends in 1933 were 66% of their 1929 level. Excluding tobacco, these dividends in 1933 were 62% of their 1929 level.

Table 12 PROFITS AND DIVIDENDS IN KEY SECTORS[a]

	Mining		*Durable Mfg.*		*Nondurable Mfg.*	
Year	*Profits*	*Divs.*	*Profits*	*Divs.*	*Profits*	*Divs.*
1929	430	309	2247	1335	2332	1213
1931	−75	118	−155	811	1303	1133
1933	−115	66	−721	314	−85	803

[a]The data are from NIPA and are measured without inventory valuation adjustment. They are in real dollars per adult, and are detrended at the average rate of growth of output per adult: 1.9%. We thank Mark Gertler for suggesting this measure of cash flow (net of depreciation).

The maintenance of dividend payments at the expense of retained earnings throughout the Depression suggests that firms were liquidating their enterprises, rather than finding substitutes for costly bank finance. Reconciling this large drop in retained earnings with the banking story seems difficult. To do so requires explaining why firms drained their coffers and increased their exposure to negative banking shocks.

5. *Interactions between the Wage and Banking Shocks*

Even though we find that neither banking shocks nor wage shocks account for much of the Great Depression, is it possible that the interaction between these two shocks has a large macroeconomic effect? There are two reasons why we do not think this is very likely. If there was an important connection between the two types of shocks, we should observe a strong negative correlation between the incidence of banking crises and economic activity in sectors distorted by high wages. Manufacturing was ostensibly distorted by the high wage, but the correlation between manufacturing employment and bank closings was positive at the state level, rather than negative. Moreover, the correlation between state per capita income and bank failures in states with large manufacturing sectors—those with above-median ratios of manufacturing employment to population—is roughly the same as that for all the states, and is not significantly different from zero.

There are also theoretical reasons for doubting that an interaction between the two shocks would have large effects. To illustrate this point, consider the simplest possible method of incorporating the banking shock into the wage model. Suppose that intermediation capital is another input into production, and denote the sectoral level of intermediation capital by Z_i. Sectoral output is now given by the production function $Y_i = (AH_i)^{1-\theta-\gamma}K_i^{\theta}Z_i^{\gamma}$, where θ is unchanged and $\gamma = 0.01$ to match banking's value-added share. Given this specification, it is straightforward to show that the 18% decrease in Z that occurred between 1929 and 1933 would reduce output in the wage model an additional 0.18%. This result partially reflects the fact that the decrease in intermediation capital leads to general equilibrium changes in factor prices that moderate the impact of the factor change.

6. *What Else was Different about the Great Depression?*

The two candidate shocks we have considered—bank failures and imperfectly flexible wages—don't seem capable of plausibly explaining the Great Depression. So if it wasn't banking or wages, what other factors might have been responsible?[31]

6.1 LOWER ASSET PRICES

The first alternative shock we examine is lower asset prices. The stock-market crash of 1929 is considered by some economists to have contributed to the Great Depression (see Romer, 1993). It is difficult to evaluate this story, since there currently is no generally accepted theory of asset price fluctuations. Without such a theory, one cannot establish that asset price changes contributed significantly to the Great Depression.[32] But we can take a first step by empirically assessing whether other periods of large and prolonged decreases in asset prices also coincide with major depressions. One of the best known of these episodes is Japan in the 1990s. We therefore compare changes in stock prices and output in the U.S. in the 1930s with Japan in the 1990s. Tables 13 and 14 show real stock prices and output for these two countries. We find some important similarities in asset price changes between the two countries, but very different output changes after share prices fall.

Stock prices in both countries roughly doubled during the three-year period before their respective market peaks. Output growth relative to respective trends is also very similar in the two countries during these three-year periods of rising stock prices. Following their respective market peaks, stock prices fell sharply in both countries. U.S. share prices fell about 68%, and Japanese share prices fell about 55%. Despite these similar stock price patterns, output growth differs substantially after

31. One difference between these two episodes is that the deflation of 1921–1922 immediately followed a significant inflation, whereas that 1929–1933 followed a period of roughly stable prices. If nominal prices were more flexible during the early Depression, the deflation may have had smaller real effects. Little is known, however, about the differences in price flexibility during these two downturns.

32. Without a good theory of asset price fluctuations, it is unclear what shock drove down asset prices, or how asset prices interacted with employment, consumption, and investment decisions.

Table 13 REAL U.S. DETRENDED STOCK PRICES AND OUTPUT (1929 = 100)

Year	*S&P index*[a]	*Output index*
1926	50.4	102.8
1927	61.7	100.1
1928	78.2	97.7
1929	100.0	100.0
1930	81.4	86.9
1931	57.1	77.6
1932	31.6	64.0

[a]Source: Board of Governors of the Federal Reserve System (1943, Table X, pp. 492–498).

Table 14 REAL JAPANESE DETRENDED STOCK PRICES AND OUTPUT[a] (1989 = 100)

Year	*Nikkei index*	*Output index*
1986	55.1	96.2
1989	100.0	100.0
1990	81.6	101.4
1991	63.1	101.5
1992	44.6	98.9

[a]Quantities are *not* per adult. They have been detrended using a 3.7% growth rate, which is the average for real output between 1979 and 1989. Data are from the DRI International Database.

prices begin to fall. U.S. output is 36% below trend three years after its stock-market peak, whereas Japanese output remains on trend three years after its stock-market peak.[33]

These data show that large asset price decreases are not always associated with big depressions. Japanese stock prices fell nearly as much in the 1990s as U.S. share prices fell in the 1930s, but Japanese output remained close to trend while stock prices fell.[34] These Japanese data and

33. Japan did experience a *growth* slowdown after 1991, and by 1998 was 15% below trend. However, note that this decrease comes 9 years after the decrease in asset prices.
34. Land values in Japan also followed the same roller-coaster pattern as stock prices in the 1990s. Commercial real estate values doubled during the same period that stock prices doubled, and fell 35% three years after the market peak. These data are thus inconsistent with the view that Japan maintained high macroeconomic activity because other asset values remained high. (See commercial real estate prices in the six largest cities from the Japan Real Estate Institute: http://www.reinet.or.jp/index-e.htm.)

the pattern of retained earnings during the U.S. Great Depression raise questions about the asset-price story. First, if lower asset prices contributed to the U.S. Great Depression, why didn't a similar decrease produce a Great Depression in Japan? Second, if the macroeconomic impact of lower prices is through lower borrower net worth, as is often presumed in the literature, then why did firms continue to pay such dividends during the 1930s rather than increase retained earnings? Finally, if decreases in asset values have a substantial negative effect on output, through either borrower or consumer net worth, then why did the increase in asset prices have so little effect in either Japan or the United States? Any theory of the Depression based on lower asset values should be able to explain why lower asset prices don't always produce major depressions, and explain why retained earnings fell in the 1930s.[35]

6.2 THE FALL IN TOTAL FACTOR PRODUCTIVITY

The second alternative shock we consider is a total factor productivity (TFP) shock. This shock is much different during the Great Depression than other periods and in particular differs sharply from 1921–1922. TFP rose about 5% relative to trend in 1921, but fell about 14% below trend between 1929 and 1933.[36]

It is unlikely that this TFP decrease during the Great Depression reflects technological regress or is solely due to factor measurement error. To see this latter point, consider three types of measurement error: capital utilization, changes in labor quality, and changes in capital quality. The utilization of the capital stock was low during the Great Depression, and this overstatement of the capital input will bias down TFP measurement. But the other two sources of factor mismeasurement will tend to offset mismeasured capital input. The average quality of labor input probably rose during the Depression, as the least productive workers were probably the first to be laid off. This indicates that measures of labor input based on employment or hours worked will understate labor input in efficiency units. Similarly, the oldest, least efficient capital was idled during the Depression (Bresnahan and Raff, 1991). This "vintage effect" implies that measures of capital input based on the number of

35. These data cast doubt on the ability of theoretical models in which financial-market imperfections amplify the effects of macroeconomic shocks by reducing net worth to explain a significant portion of the Great Depression. [See Kiyotaki and Moore, 1997, and Bernanke, Gertler, and Gilchrist, 2000 (BGG).] According to these models, output should have expanded significantly when stock prices were rising. Moreover, these models predict that enterprises should have substantially increased internal cash when share prices began falling. Both of these predictions stand in contrast to the data.

36. Romer (1988) argues that there was a favorable supply shock during the 1921–1922 Depression, although she does not discuss TFP changes.

idle factories will understate capital input in efficiency units. Both of these compositional effects will tend to understate the true decline in TFP and tend to offset the effect of capital utilization.

Since labor's share is about twice as large as capital's share, considerable mismeasurement of capital utilization is required to bias the TFP measure. For example, if true capital input was 20% lower than measured capital input after correcting for vintage effects, and true labor input in efficiency units was 5% higher than measured labor input due to compositional shifts, TFP would have decreased by 11%, compared to the measured decrease of 14%.

Negative productivity shocks also show up in disaggregated data. Bernanke and Parkinson (1991) report negative productivity shocks in manufacturing and argue that the shocks reflect labor hoarding or increasing returns to scale. But there are good reasons to question these two explanations. Recent research indicates CRS in manufacturing, rather than increasing returns. And at least the traditional reason given for labor hoarding—the cost of laying off and subsequently rehiring a worker exceeds the cost of retaining the worker—seems unlikely during this period. Managers seem to have been liquidating their enterprises during the Great Depression, rather than planning for an upcoming expansion that would have productively utilized the hoarded labor.

The TFP decrease may not be adequately explained by technological regress, factor mismeasurement, or returns to scale. More research is needed to determine the sources of and reasons for this large change and how much it may have contributed to the Great Depression. Since a decrease in productivity reduces marginal productivity, this shock may represent the best chance for the wage hypothesis to account for a reasonable fraction of the output decrease.

7. *Summary and Conclusion*

Our results suggest that two popular stories for the Great Depression—the inflexible-wage deflation story and the banking-shock story—account for only a small fraction of the output fall that occurred between 1929 and 1933. The problem with the inflexible-wage story is that measured wages were above trend in only a subset of the economy, and that a reasonable correction for shifts in the composition of employment would reduce those wage measures below trend. The problem with the banking-shock story is that the shock is small, and the elasticity of aggregate output with respect to a banking shock is also small. Moreover, three important auxiliary predictions of the banking story don't line up with the data. The theory predicts that states that had worse banking crises should have had worse depressions. But there is no systematic relationship between state economic activity and the number of bank closings. The theory also predicts that firms should have increased internal cash in response to the banking shock. In contrast, firms reduced retained earnings substantially during the Great Depression. The theory also predicts that the ratio of bank deposits to output should have decreased during the Depression. This ratio increased substantially during the Depression. Any successful financial intermediation theory of the Depression should be consistent with these three facts.

We conclude that the Great Depression remains a puzzle. The paper suggests two directions for future research. One direction is to analyze money (deflation) shocks through alternative channels. The second direction is to analyze real shocks. The fact that real output per adult fell 13% in 1930 without any significant deflation suggests the possibility that a real shock contributed to the initial downturn. And the large decrease in TFP suggests the possibility that some shock may have affected productivity during the Great Depression.

Appendix. Characterizing the Equilibrium of the Financial-Intermediation Model

In what follows, we will assume that the difference in the idiosyncratic productivity shocks is small enough that the nonnegativity constraint on retained earnings never binds. Under this assumption, the f.o.c.s that characterize a solution include

$$\beta^t u'(c_t) = \lambda_t, \tag{13}$$

$$\lambda_t \gamma A \epsilon_{t,i}(K_{t,i} + x_{t,i})^{\gamma-1} = \mu_t - \xi_{t,i}, \qquad \text{where} \quad x_{t,i}\,\xi_{t,i} = 0, \tag{14}$$

$$\lambda_{t+1}\gamma A E_t[\epsilon_{t+1,i}(K_{t+1,i}+x_{t+1,i})^{\gamma-1}] = \lambda_t, \tag{15}$$

$$\mu_{t+1} G_1(D_{t+1}, Z) = \lambda_t, \tag{16}$$

where μ_t, λ_t, and $\xi_{t,i}$ are the Lagrange multipliers on the constraints (9) and (10) and the nonnegativity constraints on x_i, respectively.

It is easy to see from the f.o.c. on plant capital, (15), that $K_1 = K_2$. Hence we can aggregate plant capital and treat (K, D) as the state variables, where $K/2$ is plant capital. It is easy to see that x_l cannot be positive, since condition (14) would imply that x_h was also positive, and hence at both plants the marginal product of capital would be greater than λ_{t-1}, which would contradict (15).

The steady state of this model will be given by (K, D), where

$$\beta\gamma A\left[\epsilon_h\left(\frac{K}{2}+G(D, Z)\right)^{\gamma-1}+\epsilon_l\left(\frac{K}{2}\right)^{\gamma-1}\right]=1, \tag{17}$$

$$\beta\gamma A\epsilon_h\left(\frac{K}{2}+G(D, Z)\right)^{\gamma-1}G_1(D,Z)\leq 1, \tag{18}$$

with strict equality if $D > 0$, and c is given by

$$c=\frac{1}{2}A\epsilon_h\left(\frac{K}{2}+G(D, Z)\right)^{\gamma}+\frac{1}{2}A\epsilon_l\left(\frac{K}{2}\right)^{\gamma}-(K-D).$$

We can develop the analysis further by assuming an explicit functional form for G. the Leontieff specification allows us to obtain closed-form solutions for the variables D and K:

$$G(D, Z)=\min(\alpha D, Z),$$

where $\alpha < 1$, and $(1-\alpha)D$ is the cost of intermediation.

If D is positive and interior, that is, less than Z, it is straightforward to show that

$$D=\frac{1}{\alpha}(\beta A\gamma)^{1/(1-\gamma)}\left[\epsilon_h^{1/(1-\gamma)}-\left(\frac{\alpha\epsilon_l}{2\alpha-1}\right)^{1/(1-\gamma)}\right]. \tag{19}$$

If the value of D is such that $D \in [0, Z/\alpha]$, then the steady-state level of K is

$$K=\left(\frac{\alpha\beta A\gamma\epsilon_l}{2\alpha-1}\right)^{1/(1-\gamma)}.$$

If the value of D implied by (19) is negative, then it is easy to show that in the steady state $D = 0$, and

$$K=\left(\frac{\beta\gamma A}{2}(\epsilon_h+\epsilon_l)\right)^{1/(1-\gamma)}.$$

If the value of D implied by (19) is greater than Z/α, then in the steady state $D = Z$ and K is the solution to (17) when we set $G(D, Z) = Z$.

This allows us to conduct some comparative statics on what happens to K and D when intermediation capital changes. If Z binds, then $dK/dZ < 0$

and $d(K + \alpha D)/dZ > 0$. Furthermore, if $D > 0$, then $d(\alpha D)/d\alpha > 0$, and hence $dK/d\alpha < 0$, while $d(K + \alpha D)/d\alpha > 0$. Our model predicts that a decrease in intermediation capital increases internally installed capital, but significantly reduces intermediated investment. Similarly, an increase in the cost of intermediation $(1 - \alpha)$ increases internally installed capital and reduces intermediated investment. It is also easy to see how the spread in the lending and borrowing rate is affected by a change in Z. In this example, the marginal cost of funds to the high-productivity plant must be

$$1 + r_b = \gamma A \epsilon_h (K/2 + \alpha D)^{\gamma-1}.$$

The interest rate on savings must be

$$1 + r_s = \gamma A \epsilon_h (K/2 + \alpha D)^{\gamma-1} \alpha.$$

This implies that the spread between these two rates is given by

$$r_b - r_s = \gamma A \epsilon_h (K/2 + \alpha D)^{\gamma-1} (1 - \alpha).$$

A decrease in intermediation capital that binds will lower the quantity of intermediated capital, αD, and raise the quantity financed out of retained earnings, K. It also raises both the borrowing and lending interest rates and the spread between them, since the marginal productivity of capital at the high-productivity plant is raised. The spread also is decreasing in α, which governs the fraction of capital consumed by the intermediation process.

Finally, assume that G_D, G_Z, $G_{DZ} > 0$ for all D, $Z > 0$. In this case, a reduction in Z works like an increase in intermediation costs. Since G is CRS, $G(D,Z) = g(Z/D)D$, where $g' > 0$. In response to a decrease in Z, the equilibrium level of Z/D will increase. This indicates that the relevant factor for intermediation costs is not the level of intermediation capital per se, but the level relative to the quantity of intermediated capital.

REFERENCES

Beney, A. (1936). *Wages, Hours, and Employment in the United States 1914–36.* New York: National Industrial Conference Board.

Bernanke, B. S. (1983). Nonmonetary effects of the financial crisis in propagation of the Great Depression. *American Economic Review* 73(June):257–276.

———. (1986). Employment, hours, and earnings in the Depression: An analysis of eight manufacturing industries. *American Economic Review* 76(March):82–107.

———, and Carey, K. (1996). Nominal wage stickiness and aggregate supply in the Great Depression. *Quarterly Journal of Economics* 111(August):853–883.
———, Gertler, M., and Gilchrist, S. (2000). The financial accelerator in a quantitative business cycle framework. *Handbook of Macroeconomics*, J. Taylor and M. Woodford (eds.). Amsterdam: North-Holland.
———, and Parkinson, M. L. (1991). Procyclical labor productivity and competing theories of the business cycle: Some evidence from interwar U.S. manufacturing industries. *Journal of Political Economy* 99(June):439–459.
Board of Governors of the Federal Reserve System. (1943). *Banking and Monetary Statistics, 1914–1941.* Washington.
Bordo, M., C. Erceg, and C. Evans. (2000). Money, sticky wages, and the Great Depression. *American Economic Review*, forthcoming.
Bresnahan, T., and D. Raff. (1991). Intra-industry heterogeneity and the Great Depression: The American motor vehicles industry, 1929–35. *Journal of Economic History*, June, pp. 317–331.
Bureau of the Census. (1975). *Historical Statistics of the United States.* Washington.
Cechetti, S. (1992). Prices and the Great Depression: Was the deflation of 1930–32 really unanticipated? *American Economic Review*, March, pp. 80–102.
Cole, H., and L. Ohanian. (2000). New Deal policies and the persistence of the Great Depression: A general equilibrium analysis. Federal Reserve Bank of Minneapolis. Discussion Paper.
Eichengreen, B., and J. Sachs. (1985). Exchange rates and economic recovery in the 1930s. *Journal of Economic History*, December, pp. 925–946.
Fisher, I. (1933). The debt deflation theory of the Great Depression. *Econometrica* 1:337–357.
Green, S., and C. Whiteman. (1992). A new look at old evidence: On "Nonmonetary effects of financial crisis in the propagation of the Great Depression. University of Iowa.
Hamilton, J. (1992). Was the deflation during the Great Depression anticipated? Evidence from the commodity futures market. *American Economic Review*, March, pp. 157–178.
Hanes, C. (1996). Changes in the cyclical behavior of real wage rates, 1870–1990. *Journal of Economic History* 56(December):837–861.
Kiyotaki, N., and J. Moore. Credit Cycles. *Journal of Political Economy*, 105(2):211–248.
Kendrick, J. W. (1961). *Productivity Trends in the United States.* Princeton, NJ: Princeton University Press.
Kuznets, S. (1941). *National Income and its Composition, 1919–1938.* New York: NBER.
Lebergott, S. (1991). "Wage rigidity" in the depression: Concept or phrase? Wesleyan University. Discussion Paper.
Lamont, R. (1930). The White House conferences. *Journal of Business* III(3):269–271.
Lucas, R. (1983). In *Conversations with Economists*, Arjo Klamer (ed.). Rowman and Allanheld.
———, and Rapping, L. (1969). Real wages, employment and inflation. *Journal of Political Economy.* 77(5):721–754.
———, and L. Rapping (1972). Unemployment in the Great Depression: Is there a full explanation? *Journal of Political Economy* 80(1):186–191.

Margo, R. (1993). "Employment and unemployment in the 1930s. *Journal of Economic Perspectives*, Spring, pp. 41–59.
Murphy, A. E. (1978). Money in an economy without banks: The case of Ireland. *The Manchester School of Economic and Social Studies*, pp. 41–50.
National Industrial Conference Board. (1928). *Wages in the United States, 1914–1927*. New York.
O'Brien, A. (1989). Behavioral explanation for nominal wage rigidity during the Great Depression. *Quarterly Journal of Economics* 104(November):719–735.
Olney, M. (1999). Avoiding default: The role of credit in the consumption collapse of 1930. *Quarterly Journal of Economics*, February, pp. 319–335.
Romer, C. D. (1988). World War I and the postwar depression. *Journal of Monetary Economics* 22:91–115.
———. (1993). The nation in depression. *Journal of Economic Perspectives* 7(Spring):41–59.
Solon, G., R. Barsky, and J. Parker. (1994). Measuring the cyclicality of real wages: How important is composition bias?" *Quarterly Journal of Economics* 109(1):1–25.
Temin, P. (1989). *Lessons from the Great Depression*. Cambridge, MA: The MIT Press.

Comment[1]

MICHAEL BORDO, CHRISTOPHER ERCEG, AND CHARLES EVANS
Rutgers University, Board of Governors of the Federal Reserve System, Federal Reserve Bank of Chicago

1. *Introduction*

Cole and Ohanian's paper is both ambitious and provocative. It is ambitious because it investigates the ability of several alternative and widely cited explanations of the Great Depression to explain the quantitative magnitude of the output downturn that occurred in 1929–1933. It is provocative because it concludes that none of the standard battery of explanations can account for more than a small fraction of the observed output decline.

One explanation that Cole and Ohanian test is the sticky-wage hypothesis: that the massive deflation of 1929–1933 depressed output by driving up real wages. Their results are very different from our findings in "Money, sticky wages, and the Great Depression" (Bordo, Erceg, and Evans, 2000), in which we find that the sticky-wage channel accounts for

1. This paper represents the views of the authors and should not be interpreted as reflecting the views of the Board of Governors of the Federal Reserve System, the Federal Reserve Bank of Chicago, other members of their staff, or the National Bureau of Economic Research.

about 70% of the output decline over 1929–1933. In this comment, we focus on several key problems with the authors' formulation that account for the divergence in our results. We conclude by discussing their comparison of the postwar disinflation in 1920–1921 with the disinflation over 1929–1933. In the subsequent comment, Mark Gertler focuses on Cole and Ohanian's test of the bank-failure explanation for the Depression.

2. *Evaluation of Cole and Ohanian's Model for Testing the Sticky-Wage Hypothesis*

In our paper "Money, sticky wages, and the Great Depression," we construct a dynamic general equilibrium model to evaluate whether the monetary contraction over 1929–1933 can account for the output decline. Our model has only one productive sector, with capital and labor the only inputs. The representative agent makes consumption–investment decisions based on the permanent-income hypothesis, and has rational expectations. Persistent negative shocks to money growth over 1930–1933 cause a largely unanticipated fall in the price levels. Because wages are sluggish to adjust to the employment gap, real wages rise, causing hours worked and output to contract progressively. The negative effects of the real-wage rise on hours worked are exacerbated by the decline in the capital stock.

Our simple model accounts surprisingly well for the joint behavior of output, hours worked, and our measure of the real wage over the downturn phase of the Depression, particularly over 1929–1932. We interpret our results as providing support for the null hypothesis that (unexpected) contractionary shocks to money operating through a sticky-wage channel played a substantial role in the output downturn. As always, "support" is taken to mean nonrejection: it is possible that other explanations may perform as well or better at accounting for the same stylized facts, in which case we would need an additional basis to differentiate between the models.

Cole and Ohanian develop a two-sector model in order to allow for potentially different real-wage behavior across the manufacturing and nonmanufacturing sectors. Aside from this formal difference, there are several key features that account for their rejection of the null that nominal wage stickiness accounts for a sizeable fraction of the output downturn. First, Cole and Ohanian assume that real wages adjust flexibly (in a spot labor market) in the nonmanufacturing sector, which comprises 72% of employment in their baseline calibration (government is also included in their manufacturing sector). Second, the shock to the real product wage in manufacturing implied by their model greatly understates the observed increase. Third, Cole and Ohanian simply assume that trend productivity grew at 2% per year over the 1929–1933 period. We consider each of these restrictions in turn below. We argue that they appear unjustified empirically, and thus strongly bias Cole and Ohanian's model against the sticky-wage hypothesis.

2.1 WERE WAGES IN THE NONMANUFACTURING SECTOR PERFECTLY FLEXIBLE?

Cole and Ohanian motivate modeling wage behavior outside the manufacturing sector as perfectly flexible, based on wage behavior in the farming sector. In the upper panel of Table 4 of their paper, they show that real farm wages (deflated by the GNP deflator) collapsed during the Great Depression, falling over 40% between 1929 and 1933.[2]

While it is probably reasonable to model the farming sector as having flexible wages, there is little support for extending this characterization to the remainder of private nonmanufacturing. First, the farming sector was quite small at the onset of the Depression, accounting for only 10% of national income, or about 14% of the national income attributable to the private nonmanufacturing sector. Moreover, according to Kendrick (1961), the farm sector constituted about 11% of the total labor input to the private nonmanufacturing sector on a quality-adjusted basis (Table A-5, p. 267).

Second, the farming sector appears to have behaved very differently from the remainder of the private nonmanufacturing sector. Table 1 compares employment in farming to employment in the (private) non-farm non-manufacturing sector. While employment in farming remained nearly stable, employment in nonfarm nonmanufacturing fell 30% by 1933. Table 2 considers the relative price of farm output, derived by deflating the wholesale price index (WPI) for farm products by the GNP deflator. Real farm prices collapsed during the 1929–1933 period, declining by somewhat over 40% by 1933 (about the same fall as in the real wage). On the other hand, the decline in relative farm prices and the relative price of manufactured goods over the 1929–1933 period (discussed below) imply that the real price of nonfarm nonmanufactured goods must have risen over the period.

Even without considering wage data directly, the data on employment and relative prices suggest very different wage behavior across the farm and nonfarm nonmanufacturing sectors. The fact that employment in farming remained almost stable despite a massive fall in the product

2. Data on real wages and output reported by Cole and Ohanian are adjusted by deterministic trends. In this note, we report all data without making any trend adjustments.

Table 1 EMPLOYMENT IN THE NONMANUFACTURING SECTOR (CHANGE FROM 1929)

Year	Change (percentage pts.) Farm[a]	Nonfarm[b]
1930	−2.1	−4.8
1931	−0.1	−14.2
1932	0.4	−26.9
1933	−0.2	−30.4

[a]Kendrick (1961, Table B-1).
[b]United States Department of Commerce (1981, Table 6.8A).

Table 2 REAL PRICE OF FARM OUTPUT[a] (CHANGE FROM 1929)

Year	WPI[b] (percentage points)
1930	−14.0
1931	−35.5
1932	−53.3
1933	−44.8

[a]Source: Bureau of Labor Statistics, *Wholesale Prices*, June 1934, Table 1.
[b]For farm products, divided by GNP deflator.

price would require a corresponding collapse in the real wage (measured relative to the GNP deflator). This is exactly what was observed. By contrast, the sharp decline in employment in the nonfarm nonmanufacturing sector despite a rise in the relative price of its output would suggest that the real wage in that sector rose.[3]

Table 3 compares a measure of the aggregate real wage that we constructed in our paper (2000) with the manufacturing real wage. The nonmanufacturing real wage behaves quite similarly to the manufactur-

3. Our argument implicitly assumes—as in Bernanke and Parkinson (1991)—that there was not a sizable fall in total factor productivity during the 1929–1933 period. In this case, the nonfarm nonmanufacturing sector's labor demand curve would be reasonably stable, except for movements in the capital stock. Thus, a large observed fall in sectoral employment (despite a relative price rise) would require a rise in that sector's real wage.

Table 3 REAL WAGES (USING GNP DEFLATOR)[a] (CHANGE FROM 1929, IN PERCENTAGE POINTS)

	1930	*1931*	*1932*	*1933*
Aggregate (all industries)	2.4	7.2	9.7	5.8
Manufacturing	3.5	9.2	10.7	10.6

[a]These real wage measures are derived by deflating average hourly earnings by the GNP deflator. The average hourly earnings series are described in Bordo, Erceg, and Evans (2000), footnote 2.

ing real wage (when wages are deflated by a common deflator), at least through 1932. We admit that there are important reasons to be cautious about drawing inferences from the aggregate wage data, particularly given that information about average hours worked is sparse in most sectors outside of manufacturing. Nevertheless, at the very least there is direct evidence suggesting that wages in the nonfarm nonmanufacturing sector rose somewhat, and the behavior of sectoral relative prices and that of employment seem consistent with the real wage movement. Thus, Cole and Ohanian's decision to assume flexible wages in the nonmanufacturing sector based on evidence of wage behavior in farming seems difficult to justify. This choice biases their model against finding an important role for the sticky-wage channel.

2.2 DOES THE IMPLIED BEHAVIOR OF THE PRODUCT REAL WAGE IN MANUFACTURING FIT THE DATA?

In Cole and Ohanian's model, manufacturing output depends on the product real wage, the capital stock in manufacturing, and the level of technology. The form of the dependence can be seen by taking the logarithm of the representative manufacturing firm's first-order condition for choosing hours worked:

$$h_{mt} = k_{mt} - \frac{1}{\theta}(w_{mt} - p_{mt} - a_t). \tag{1}$$

Here h_{mt} is the (natural) log of hours worked in manufacturing, k_{mt} is the log of the manufacturing capital stock, w_{mt} is the log of the manufacturing wage deflated by the GNP deflator (the "consumer" real wage), p_{mt} is the log of the price of the manufactured goods deflated by the GNP deflator, and a_t is the log of an index of technology. The product real wage in manufacturing is simply the difference between the consumer real wage in manufacturing and the real price of manufacturing output ($w_{mt} - p_{mt}$).

Cole and Ohanian interpret the sticky-wage model as implying a se-

Table 4 REAL WAGES AND RELATIVE PRICES IN MANUFACTURING[a] (CHANGE FROM 1929)

	Change (percentage points)		
Year	*Consumer real wage in manufacturing,* w_{mt}	*Real price of manufactured good,* p_{mt}	*Product real wage in manufacturing,* $w_{mt} - p_{mt}$
1930	3.5	−5.1	8.6
1931	9.2	−9.7	18.9
1932	10.7	−6.8	17.5
1933	10.6	−3.7	14.4

[a]The deflator for manufacturing output is the WPI for nonagricultural products, from the Bureau of Labor Statistics, *Wholesale Prices*, June 1934, Table 3.

quence of unanticipated shocks to the consumer real wage in manufacturing (w_{mt}). The shock in each period is assumed to equal the deviation of the observed consumer real wage from its 1929 value, except for an adjustment for "trend" productivity growth that we discuss below. The unadjusted consumer real-wage series is shown in the first column of Table 4 (it is identical to the second column of Table 3). While w_{mt} is taken as exogenous, the relative price p_{mt} is endogenously determined, depending inversely on the output of the manufactured relative to the non-manufactured good. Cole and Ohanian's model generates a substantial rise in the real price of the manufactured good, and thus implies a rise in the product real wage that is much smaller than the rise in the consumer real wage.

This implication is contradicted by the data. The second column of Table 4 shows the real price of manufacturing output, defined as a deflator for manufacturing output divided by the GNP deflator. The real price of manufacturing output had fallen by 10% in 1931 relative to its 1929 level. This implies that the product real wage (column 3) rose 10 percentage points *more* than the consumer real wage in that year. By contrast, Cole and Ohanian report that their model implies a rise in the product real wage that is 4.4% *less* than the rise in the consumer real wage in 1931.

Thus, Cole and Ohanian's model appears to seriously understate the shock to the product real wage in manufacturing. Even ignoring their subtraction of "trend productivity growth" from the product real wage implied by their model, their model understates the rise in the product real wage by 10–15 percentage points over the Depression downturn. Holding manufacturing capital constant, this would translate into a serious understatement of the effects of a rise in the real product wage on manufacturing output.

To gauge the effects of understating the shock to the product real wage, note from (1) that the elasticity of labor demand in manufacturing is $1/\theta$, where θ is the capital share in manufacturing. Since $\theta = \frac{1}{3}$ in their calibration, we can take this elasticity to equal 3. Thus, relative to a model that simply took the product real wage as exogenous and equal to our values in column 3, their model understates the decline in manufacturing output by 30–45%. Moreover, because this understatement of the rise in the product wage greatly reduces the effect of the wage shock on aggregate output in their model, it also mitigates the fall in the manufacturing capital stock. Thus, their model's understatement of the wage shock may have a considerably larger effect on manufacturing hours worked and output through the capital channel.

2.3 IS IT APPROPRIATE TO SCALE FOR TREND PRODUCTIVITY GROWTH?

Cole and Ohanian assume that productivity grows at a constant rate of 2% per year over the 1929–1933 period. Given this assumption, they take the wage shock to manufacturing to be the consumer wage (in column 1 of Table 4) scaled down for trend productivity growth, i.e., $w_{mt} - a_t$.

The assumption of trend productivity growth obviously makes it more difficult to account for an output downturn, particularly in 1932–1933. Given the preference specification, this means that output and real wages in each sector would grow 2% per year in the absence of any shocks. Moreover, the size of the exogenous shock to the consumer real wage in the manufacturing sector is scaled down from what is reported in column 1 of Table 4. As a result, the baseline parameterization of Cole and Ohanian's model implies that while GDP falls below trend in 1929–1933, the level of GDP rises continuously over the period. Even output in manufacturing, the sticky-wage sector, rises above its 1929 level by 1933!

We believe that the inclusion of this trend term lacks justification, at least over the period considered. It seems especially puzzling given that the authors argue that "[aggregate] total faster productivity fell about 14% percent below trend between 1929 and 1933." If total factor productivity in fact declined, it may be more appropriate to extract a negative trend, or to allow for negative shocks to productivity. Allowing for a negative trend would of course allow sticky wages in manufacturing to exert much larger output effects (even more so if the relative-price problem in the current model were rectified). The authors seem to acknowledge this near the conclusion of the paper when they state that "a decrease in productivity . . . may represent the best chance for the wage hypothesis to account for a reasonable fraction of the output decrease."

In our own model, we take a conservative approach and assume that productivity remained unchanged over 1929–1933. However, we agree with Cole and Ohanian that it would be interesting to further investigate the implications of a possible fall in total factor productivity, despite the obvious difficulties in constructing a convincing measure over this period.

2.4 SUMMARY OF KEY PROBLEMS WITH MODEL

Cole and Ohanian claim to be testing the null hypothesis that wage stickiness accounted for the Great Depression. A convincing test requires building a model that is favorable to the null, subject to the constraints imposed by the data. We argue that Cole and Ohanian construct a model that is unduly biased against the null hypothesis, and this accounts for their rejection.

It is important to emphasize that the different results they derive are primarily driven by the three features of the model discussed (in Sections 2.1–2.3) above, and not from other features of their setup, e.g., the fact that their model has two sectors instead of one. Some preliminary work that we have done suggests that a multisector model can account for a very substantial output downturn if it: (1) takes the product real wage in manufacturing as exogenous (fitting it to the observed product-real-wage series), (2) allows for some degree of rigidity in the product real wage in the nonmanufacturing sector, even if considerably less than in manufacturing, and (3) assumes total factor productivity growth is zero over the 1929–1933 period.

From a methodological perspective, the authors' introduction of a multisectoral model to study the Depression period is innovative and welcome. The authors highlight how the farm sector behaved differently over that period, and future research may identify sectoral differences that have important consequences for the effects of a shock on aggregate activity. Regarding the monetary transmission mechanism, our preference is to model it directly rather than take the reduced-form approach of assuming an exogenous real wage shock. The authors' approach involves an unpalatable assumption about how agents in the model assume the real wage will eventually adjust to equilibrate the labor market (viz., the manufacturing real wage is expected to adjust flexibly in the subsequent year). The transmission mechanism, including the process by which wages are expected to adjust, has critical implications for the response of the capital stock to a shock.

3. *Why Was the Disinflation of 1920–1921 Different from the Disinflation of 1929–1933?*

The authors emphasize that for an explanation linking the large output contraction of 1929–1933 to deflation to be plausible, it should be able to explain why a similar-sized deflation in 1920–1921 had different real effects. In our NBER Working Paper version of "Money, sticky wages, and the Great Depression" (1997), we proposed the same "consistency check" to evaluate the plausibility of the sticky-wage channel. It is worthwhile briefly restating our interpretation of the two periods, since it is strongly at variance with that proposed by Cole and Ohanian.

In our estimation, two key differences mainly account for why deflation in 1920–1921 induced a smaller and less prolonged downturn in real activity than the deflation of 1929–1933. First, the deflation in 1920–1921 was more predictable than in 1929–1933, and considerably shorter-lived. Second, wage-setting practices in the early 1920s were more flexible than later in the decade.

Our contention that the deflation of 1920–1921 was more predictable requires some clarification. The disinflation was predictable insofar as contemporaries of the period expected that the authorities would pursue a monetary policy that supported the gold standard at the prewar parity. Since prices had risen rapidly during both World War I and into the early interwar period, it was clear that tight monetary policy and deflation would be required to maintain the gold standard after the embargo on gold exports was lifted in June 1919. The main uncertainty involved the timing and speed of the eventual disinflation. The authorities compounded this uncertainty by pursuing an accomodative monetary policy through late 1919, despite substantial gold outflows. The authorities then abruptly tightened policy, inducing a price decline that was much sharper than in 1929–1933. The GNP deflator fell 24% between 1920:3 and 1921:2, and an additional 8% by 1922:1 before roughly stabilizing. Output began contracting in 1920:1, and fell by 17% over the subsequent year—slightly more than the 14% decline that occurred during the first year of the Great Depression.[4] Friedman and Schwartz (1963, p. 232) characterize the downturn as "one of the severest on record. Its brevity makes annual data misleading guides to its severity."

Thus, the 1920–1921 depression was in fact quite severe. According to Friedman and Schwartz, the real effects were exacerbated by a twofold error of the monetary authorities: first, their refusal to move to a tighter policy stance immediately following the end of the world war; and sec-

4. Our source for the quarterly real GNP and the GNP deflator series is Balke and Gordon (1986).

ond, by tightening policy too sharply once they finally decided to disinflate. However, the policy of disinflation rapidly gained credibility once in place. Agents realized that it was crucial for maintaining the gold standard, and was in fact similar to the policies being pursued by other central banks. Thus, the disinflation episode was relatively short-lived, prices stabilized, and a rapid recovery ensued. Output recovered to its predepression peak within a year and a half of the trough. Relatively flexible wage-setting policies—in which wages fell quickly with prices—aided in the quick recovery.

By contrast, the price deflation in the Great Depression was drawn out over a considerably longer period. It is much less plausible that the deflation was anticipated: it occurred after a long period of price stability, and was associated with a large drop in the money multiplier due to bank failures and policy inaction. Moreover, the real effects of the disinflation were likely exacerbated by the adoption of less flexible wage-setting practices in the late 1920s. As O'Brien (1989) has emphasized, this change in wage-setting is in part attributable to the mistaken belief that maintaining consumer purchasing power (through keeping nominal wages high) was the key to ameliorating the effects of business cycles.[5]

Thus, we disagree with the basic thrust of Cole and Ohanian's characterization of the two disinflation episodes: namely, that the 1920–1921 disinflation was both less anticipated than the disinflation of 1929–1933, and yet associated with a very mild output downturn. It is true that the disinflation of 1920–1921 had a surprise component, as the timing and speed of the monetary tightening weren't known ex ante. This contributed to a sharp downturn in real activity, the severity of which is understated by Cole and Ohanian (due to their reliance on annual data). However, we have argued that a much larger component of the overall disinflation was predictable in 1920–1921, that the disinflation in any event was short-lived, and that these factors helped output to bounce back quickly. While the authors use annual data on both nominal and ex post real interest rates as the basis for concluding that the disinflation of 1920–1921 was less anticipated, we believe that using such data to make inferences about inflation expectations is highly problematic. The 1920–1921 period was very turbulent, as a large inflation was followed by a

5. The theory that sharp wage cuts during the 1920–1921 depression had induced a more severe output downturn by reducing household purchasing power became quite popular in the 1920s, even among the business community. Individual firms were urged to sacrifice their private gain (cutting wages) to help secure the overall benefit of maintaining household purchasing power. Thus, while Cole and Ohanian are correct that President Hoover encouraged employers to avoid wage cuts, there was substantial support for such a policy during the first two years of the Depression. Moreover, pressure on employers to keep wages high appears to have extended well beyond the manufacturing sector.

quick deflation. Real interest rates swung wildly, with the (annualized) ex post short-term real interest rate fluctuating from roughly −20% in the first half of 1920 to 45–50% in the second half of that year. In such circumstances, the usual difficulties of disentangling the effects of shifts in inflation expectations on nominal rates from the effects of changing real rates and risk premia are greatly exacerbated.

REFERENCES

Balke, N. S., and R. J. Gordon. (1986). Appendix B: Historical data. In *The American Business Cycle: Continuity and Change,* R. J. Gordon (ed.). Chicago: University of Chicago Press, pp. 793–794.

Bernanke, B. S., and M. L. Parkinson. (1991). Procyclical labor productivity and competing theories of the business cycle: Some evidence from interwar U.S. manufacturing industries. *Journal of Political Economy* 99(June):439–459.

Bordo, M., C. Erceg, and C. Evans. (1997). Money, sticky wages, and the Great Depression. Cambridge, MA: National Bureau of Economic Research. NBER Working Paper 6071.

———, ———, and ———. (2000). Money, sticky wages, and the Great Depression. *American Economic Review,* forthcoming.

Friedman, M., and A. J. Schwartz. (1963). *A Monetary History of the United States, 1867–1960.* Princeton: Princeton University Press.

Kendrick, J. W. (1961). *Productivity Trends in the United States.* Princeton, Princeton University Press.

O'Brien, A. (1989). Behavioral explanation for nominal wage rigidity during the Great Depression. *Quarterly Journal of Economics* 104(November):719–735.

United States Department of Commerce, Bureau of the Census. (1981). *The National Income and Product Accounts of the United States, 1929–76.*

Comment

MARK GERTLER
New York University and NBER

1. *Introduction*

Cole and Ohanian have produced an interesting and provocative paper. On one point I am in complete argument: Any explanation of the Great Depression should ultimately involve writing down a quantitative model that captures the magnitude of the contraction. At the same time, it is important to note that in this paper the authors do not provide a model that rationalizes the Depression. Rather they present a set of particular models with the objective of rejecting certain monetary and financial theories. I will argue below that neither of the two models they develop is adequate for providing a robust evaluation of the theories in question.

In addition, despite the (welcome) appeal to formal modeling, much of the analysis is in fact based on informal descriptive evidence. The authors' empirical strategy is to evaluate theories on the basis of simple comparisons of two different episodes. The authors, for example, rule out debt deflation as a factor in the Great Depression by repeating the familiar argument that the deflation during 1920–1921 did not produce a contraction of similar magnitude in the early 1930s (see, e.g., Kindleberger, 1986). Here the entire strategy rests on controlling for other relevant differences across the two episodes. I will argue below that the authors have not done this control adequately. Indeed, the problem of omitted factors is an issue of concern in virtually the entire descriptive analysis.

To be clear, identification of causal factors during the Great Depression is a difficult task. However, recent literature, beginning with Choudri and Kochin (1980), Eichengreen (1992), and Bernanke and James (1991), has made considerable progress by focusing on cross-country evidence. It is puzzling that the authors completely ignore this literature. While the authors draw inferences from just two data points—the 1920–1921 and 1929–1933 downturns in the United States—the cross-country analysis instead exploits a panel of twenty to thirty observations. On this score, using the cross-country data, Bernanke and James show formally that debt-deflation was indeed associated with major contractions.

The paper is similarly silent on the well-known work of Eichengreen (1992) and others that emphasize the role of the gold standard. This work puts monetary factors at the center of the Depression by showing that the countries that suffered severe contractions were precisely those that constrained their monetary policy to defend the gold standard. Countries that freed up their monetary policies by abandoning gold early fared much better.

Beyond presenting a compelling case for monetary factors, the issue of the gold standard circles directly back to debt-deflation: The countries in Bernanke and James's sample that experienced simultaneously deflation, financial crisis, and severe depression were also those that stayed on the gold standard. That is, the attempt to maintain the gold standard by monetary tightening was apparently at least one of a number of possible forces (in conjunction with other factors, e.g. wage rigidity and a weak financial system) that helped propagate deflation and depression. I would certainly agree that to complete the argument a formal model is necessary. On the other hand, I don't see how the authors can dismiss debt-deflation and other monetary and financial forces as possible factors without confronting this research.

In Section 2 below I fill in some important missing context to the authors' descriptive analysis by providing a brief discussion of the main events of the Depression. One theme I wish to emphasize is that, in contrast to 1920–1921, the period 1929–1933 was one of sustained contractionary forces. The Great Depression was likely the cumulative effect of these forces, as opposed to being the consequence of any single factor in isolation.

The initial downturn of 1929–1930 was due in large part to a collapse in household spending, including residential investment. It is reasonable to infer that monetary and financial factors had an influence on this early spending contraction. However, as stressed by Friedman and Schwartz (1963), the most significant effect of monetary and financial factors in the Depression likely came after the economy had already been weakened substantially by the initial downturn.[1] At this time, the combined forces of debt–deflation, strains in the banking system, and asset price contractions, along with subsequent tightening of monetary policy, likely helped turn what had been a severe recession into a depression. By contrast, the debt deflation and high real rates during 1920–1921 occurred in the wake of a release of pent-up consumption demand following the end of World War I and in the midst of a more favorable international economic climate (see, e.g., Temin, 1989). Nor, during 1920–1921, as Eichengreen emphasizes, was U.S. monetary policy constrained by gold.

In Sections 3 and 4 I discuss the models of wage rigidity and banking. I argue that from the start the authors' model does not give the wage-rigidity hypothesis a fair hearing, because it does not allow for nominal-wage stickiness and, accordingly, precludes the possibility of the kind of contraction in aggregate demand that is the essence of this hypothesis. I argue similarly that the banking model is too specialized to provide a robust assessment and, among other things, discuss why in general the cost share of banking in GDP is unlikely to provide a measure of the effects of a banking crisis. In Section 5, I take issue with the authors' interpretation of the evidence on each of the following three topics: (1) risk spreads, (2) dividends, and (3) the Japanese stock-market collapse. Concluding remarks are in Section 6.

2. *Overview of the Depression*

I now provide a brief description of the events of the Depression. My goal here is to outline the case for monetary and financial factors and also to show that simple comparisons with 1920–1921 can be misleading.

1. This timing consideration is highly pertinent. Most historical analyses stress debt–deflation not as a causal factor in the 1929–1930 downturn, but rather as a factor that helped turn this downturn into a protracted depression. The authors, however, focus on its role in the early stages of the Depression. I elaborate in the next section.

Figure 1 INDUSTRIAL PRODUCTION

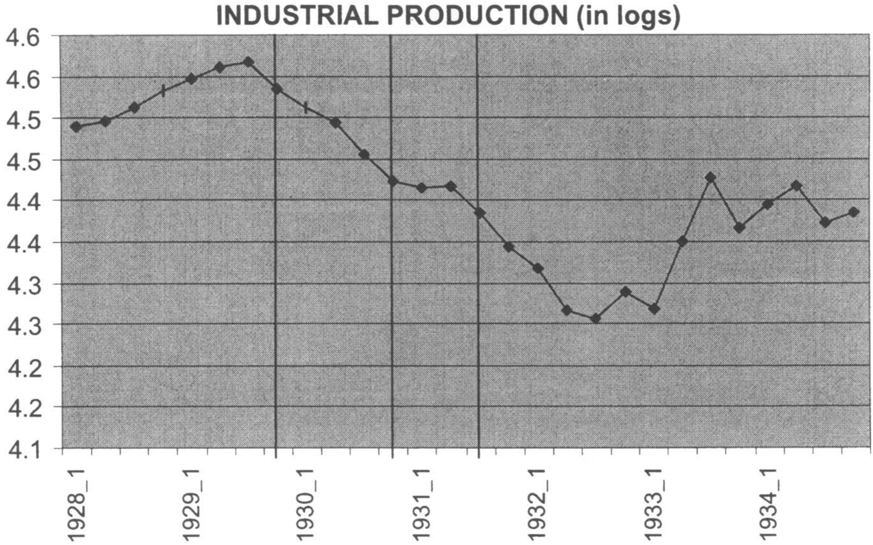


Figure 1 plots the behavior of log industrial production over the period 1928:1–1934:4. Figure 2 plots the behavior of the nominal commercial-paper rate along with two measures of inflation: the log difference of the GNP deflator and the log difference of the wholesale price index (WPI). In each figure, the three vertical lines mark dates associated with the three critical phases of the Depression, as described in Friedman and Schwartz: (1) October 1929 (the stock-market crash); (2) October 1930 (the beginning of the banking crisis); (3) September 1931 (Britain's abandonment of the gold standard). I discuss each phase in turn.

2.1 OCTOBER 1929

After a period of robust economic growth, a slowdown set in just prior to the stock-market crash. As argued by Hamilton (1987) and Romer (1993), tightening of monetary policy over the prior year was likely a contributing factor to this slowdown.[2] Following the crash, as Figure 1 indicates, there is a sharp slide in industrial production that does not level until the summer of 1930. A notable aspect of this initial output contraction was the sharp collapse in household spending, including residential investment as well as consumption demand. As Table 1

2. Hamilton cites the gold standard and a desire by the Federal Reserve to curb stock-market speculation as the motive for the tightening.

Figure 2 NOMINAL CP RATE VS. WPI AND GNPD INFLATION

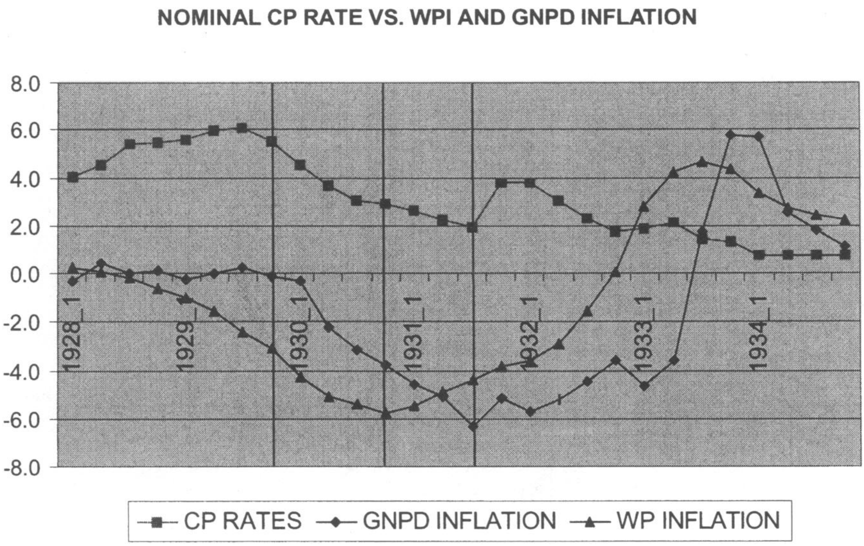


Table 1 COMPOSITION OF OUTPUT DROP

	GNP	*Nondur. cons. and services*	*Dur. cons.*	*Residential investment*	*Inventories*	*Nonresid. invest.*
% change, 1929–1930	−9.9	−4.8	−20.9	−39.4	−117.1	−18.1
% contribution to total: 1929–1930	100.0	29.4	16.9	20.4	20.4	23.9
% Ave. over post-war recessions	100.0	21.5	8.7	14.7	30.7	21.1

shows, between 1929 and 1930, durable-goods consumption dropped 20% and residential investment plummeted nearly 40%. Overall, the two spending components accounted for 37% of the output decline in 1930. Total household spending (obtained by adding in nondurables and services) accounted for nearly two-thirds of the output drop. As the table indicates, relative to the norm provided by postwar recessions, the relative contribution of household spending was unusually large across all categories (i.e., durable, housing and nondurable, and services).[3]

3. The average contribution of each GDP component to postwar recessions is taken from Romer (1996).

Several pieces of evidence suggest that monetary and financial factors played a role. First, as Figure 2 indicates, the short-term nominal interest rate rose steadily over the preceding year, reflecting the monetary tightening. Real interest rates climbed to roughly 6% based on the GDP deflator and 9% based on the WPI. While the former is a broader-based index, the latter is likely measured with more precision, especially at the quarterly frequency.[4] In addition to monetary tightening, credit conditions—particularly the weakness in consumer balance sheets and general household illiquidity in the wake of the stock-market crash—likely also contributed to the household spending drop, according to evidence presented by Mishkin (1978) and Olney (1999).

To be sure, accounting for the huge decline in household spending remains an unsolved problem. My point here is simply that there is sufficient evidence to believe that monetary and financial factors had at least some role of significance. In addition, frictions in financial markets seem a natural avenue to pursue to help account for the sizable drop in household spending, particularly the drop in consumer durables and residential investment.

The sharp drop in household spending stands in sharp contrast to the downturn between 1920 and 1921. As noted in the introduction, the end of World War I likely released a pent-up demand for household spending, particularly for durable goods. Temin provides more detail on the factors that contributed to strong household demand. The contractionary forces of debt deflation and high real rates were thus partly offset by this strong postwar consumption demand.[5] Some support for this general story comes from the authors' Table 1. Observe that consumption actually rises in 1921.

2.2 OCTOBER 1930

Around this time the drop in industrial output slows, but financial conditions steadily deteriorate. Deflation picks up momentum: As Figure 2 suggests, the GNP deflator begins a protracted decline. Due to the combined effects of the deflation and the initial economic downturn, the debt burden rises significantly: The ratio of private debt to output rises

4. Romer (1993) similarly finds that real rates reached roughly 9% at this time, by constructing a measure of expected inflation based on the producer price index.
5. Also relevant according to Temin were differences in the international economic climate: In the early 1920s the United States benefited from strong demand to facilitate reconstruction from the war. By contrast, export demand tanked during the Depression as the industrialized world fell into recession along with the United States in the early 1930s. Eichengreen (1992) further emphasizes that the absence of the gold standard in 1920–1921 reduced the synchronization of downturns in 1920–1921 across countries, relative to 1929–1931.

Table 2 ANNUAL INCREASE IN THE PRIVATE DEBT BURDEN DUE TO DEFLATION: THE DEPRESSION OF 1920–1921 VS. 1930–1931 AND 1931–1932

Years	*Initial private debt relative to output*	*Annual Δ (%) in*		
		Price level	*Debt burden*	*GNP*
1920–21	1.20	−14.80	20.85	−6.10
1930–31	1.78	−9.23	18.12	−10.70
1931–32	1.96	−10.17	22.16	−17.53

from 1.5 in 1929 to 2.0 in 1931. Loan defaults significantly weaken the capitalization of commercial banks. One manifestation of this distress is a rise in the number of bank failures. During this time, Friedman and Schwartz argue, the Federal Reserve could have taken action to stem the tide, but failed to do so.

As I noted in the introduction, unlike 1920–1921, which featured a transitory period of simultaneous falling prices and output, the deflation of 1929–1933 sets in largely when the economy has already weakened considerably, after the initial contraction of 1929–1930 described above. The authors instead focus on the role of the debt deflation in the first part of the Depression. To underscore the significance, I redid the authors' calculation of the impact of the declining price level on the debt burden (see their Table 3), this time beginning in 1930, after the initial downturn. Also, I consider 1920–1921 as the relevant period to analyze the deflation of that time, in keeping with conventional historical analysis.[6] To keep the period length consistent, I compare 1920–1921 with 1930–1931 and 1931–1932.

Table 2 reports the calculations. Note that the rise in the debt burden is roughly similar across years. The percentage output contraction in each of the Depression years exceeds the output contraction of 1920–1921—10.7% and 17.53% versus 6.1%. But the difference is far less stark than what obtains from the authors' analysis.[7] After allowing for other differ-

6. Virtually every historical account I have read refers to the contraction of 1920–1921, and not 1920–1922 as do the authors. Note from the authors' Table 1 that private spending is actually up relative to trend in 1922—overall output is down only because of a contraction in government spending due to the demilitarization following World War I. My guess is that the historical literature presumes that given that the decline in military spending likely reduced potential output, it does not seem right to treat 1922 as a recession year.

7. Also, Friedman and Schwartz (1963, p. 232) argue that the use of annual data greatly understates the severity of the 1920–1921 recession. Specifically, they state: ". . . although this contraction was relatively brief—the National Bureau dates the trough in

ences between the two episodes (e.g., vastly different initial behavior of household spending, different monetary policy regimes, different banking conditions), the relative experiences with deflation are far less anomalous than the authors' analysis suggests.[8]

An additional key difference from 1920–1921—one highlighted by the authors—is the development of the banking crisis. In the next section I address the authors' contention that banking problems were unimportant. In the meantime, I simply observe that there is considerable evidence to suggest that bank-dependent borrowers, including unincorporated businesses and corporations with imperfect access to credit, as well as households, were hit particularly hard during the first two years of the Depression. According to Fabricant (1934, 1935), business losses were concentrated mainly among small and medium-sized firms. Large firms on average made profits throughout the Depression.[9] Further, it is not the case that firms with imperfect access are "small potatoes." It is not unreasonable to suggest that unincorporated businesses and small and medium-sized corporations accounted for between half to two-thirds of GDP.[10] Thus, the disruption of credit markets affected a sizable component of the business sector along with households.

2.3 SEPTEMBER 1931

At this point, as noted above, Britain abandoned the gold standard. The Federal Reserve chose to defend, despite the severely weakened economy and despite very high ex post real interest rates. As a consequence, nominal interest rates rose 200 basis points and ex post real rates (using the GNP deflator) climbed to 10% (see Figure 2). Shortly afterward, as Figure 1 shows, industrial production began a free fall, and what had been a severe recession turned into a depression. Of course, trying to

July 1921—it ranks as one of the severest on record. Its brevity makes annual data misleading guides to its severity."

8. An additional difference was that the deflation in 1920–1921 was preceded by a large run-up in prices. To the extent debt contracts were long-term, a good fraction of the effect of the deflation on real debt burdens simply offset the effect of the earlier inflation. By way of contrast, the deflation during the Depression followed a long period of price stability.

9. The positive relationship between size and profitability during the Depression holds even after controlling for industry. See Table 6 in Fabricant (1935). Bernanke (1983) also emphasizes the heterogeneous performance of firms across size class during the Depression.

10. Unincorporated businesses accounted for roughly a third of GDP. Small and medium-sized corporations accounted for anywhere between a quarter and a half of overall corporate business. For example, firms under $50 million in assets—Fabricant's threshold for large firms—accounted for 53% of total corporate receipts in 1931 and 40% of the total corporate capital stock. Firms under $5 million in assets—clearly smaller firms—accounted for 40% of total corporate receipts and 25% of total corporate capital. Source: *Historical Statistics*.

infer causality with a single time-series observation is dangerous. It is precisely at this juncture, however, that the cross-country evidence provided in Eichengreen (1992) and Bernanke and James (1991) helps resolve the identification problem: Countries such as the United States that failed to abandon the gold standard early suffered more severe economic distress, greater deflation, and more severe banking and financial crises than countries such as Britain that moved relatively quickly to free up their monetary policy. These facts held not only for OECD countries; Campa (1990) showed that the connection between adherence to the gold standard and the severity of the Depression applied equally well to Latin American countries.

Again it is important to stress differences from 1920–1921. Unlike the tightening in this earlier period that came on the heels of an economic boom, the monetary tightening in late 1931 was the culmination of a series of contractionary shocks to the economy over the previous two years that had left both real and financial economic conditions in a highly fragile state. A tightening that follows a long period of duress may have a more potent effect than otherwise, since precautionary asset holdings and other insurance mechanisms that can help borrowers with imperfect access to credit weather bad times may have dried up. It is arguable that the tightening in late 1931 occurred exactly at this kind of point.

2.4 SUMMARY

To briefly recapitulate: To me, the evidence suggests that the authors' simple comparisons of 1920–1921 and the Depression are not adequate to rule out debt–deflation or other monetary and financial factors as having a role in the Depression. In addition to completely ignoring the international evidence, the authors do not take account of critical differences between 1920–1921 and 1929–1933, including the vastly different initial conditions influencing household spending and differences in the monetary policy regime, as well as the sustained and cumulative nature of the contractionary forces that was a feature of the latter period, but not the former.

3. *The Wage-Rigidity Model*

Here I argue that by not allowing for any kind of nominal rigidity the authors do not give the wage-rigidity hypothesis fair due. In particular, the authors consider a two-sector intertemporal general equilibrium model. Real wages are fixed exogenously in one sector, but flexible in the other. Otherwise, the model is completely frictionless. To capture the effect of rising real wages during the Depression, the authors consider a transitory increase in the real wage in the fixed wage sector. Not much happens in the aggregate, because the flexible wage sector soaks up a fair amount of the displaced workers.

It is not at all clear why the authors choose this particular model to evaluate the effect of wage rigidity. Because the model does not allow for nominal wage rigidity, it does not permit the kind of contraction in aggregate demand that advocates of the wage-rigidity hypothesis emphasize as a way to help explain the Depression. In contrast, Bordo, Erceg, and Evans (2000) show that by allowing for staggered nominal wage setting, a simple monetary model can help explain a good fraction of the output decline.

The authors also treat the increase in the real wages as if it came out of thin air. One might think *a priori* that the source of the wage increase should be relevant to the choice of model used to evaluate the issue. Here the international evidence sheds some light. The countries that experienced the largest increases in real wages were—as might be expected—those that stayed longest on the gold standard. One interpretation then is that contractionary monetary policy, interacting with nominal wage rigidities, produced the real wage increases. Specifically, contractionary monetary policy helped induce the contraction in aggregate activity and a corresponding deflation. With nominal wages a bit stickier than nominal prices, real wages increased. If this interpretation is indeed correct, then a model along the lines of Bordo, Erceg, and Evans would seem more appropriate than the authors' to study the effect of wage rigidity.[11]

4. *The Banking-Crisis Model*

I now address the authors' contention that the banking crises were unimportant. I have three basic concerns, involving: (1) potential measurement of the overall contraction in banking, (2) assumptions of the model that constrain its ability to produce a crisis, and (3) identification issues in the cross-state banking analysis.

4.1 MEASUREMENT OF THE DECLINE IN BANK LENDING CAPACITY

The authors use deposits of failed banks to measure the decline in banking services. A problem with this measure is that it does not take account of the decline in lending capacity of banks that continue to operate.

11. It may also be necessary to allow for countercyclical markups and/or some form of labor-market friction that produces real-wage rigidity in order to generate a sufficiently high elasticity of output with respect to the real wage.

Indeed, the convention in the banking literature is to use the decline in bank capital as an indicator of the contraction in lending capacity.

To see the relevance of capital, consider a typical bank balance sheet. On the asset side are loans and securities. Liabilities consist of deposits plus capital. Capital serves as a buffer to protect the return on deposits against loan losses. In practice, the quantity of bank capital influences a bank's ability to acquire uninsured deposits. In this way, it affects its lending capacity. Evidence from bank-level panel data from both the modern era (eg., Bernanke and Lown, 1991) and the Depression (e.g., Calomiris and Berry, 1998) suggests a quantitatively significant link between bank capitalization and bank lending. Accordingly the contraction of bank capital likely provides a better measure of the decline in bank lending services than the deposit measure the authors use.

One complication is that bank capital is usually measured in book- rather than market-value terms. However, recent work by Calomiris and Berry (1998) obtains evidence on the contraction of both market and book values during the Depression for a sample of New York City banks. If we use the New York City data as a guide to correcting the aggregate book-value numbers, then a crude estimate of the decline in the market value of bank capital is approximately 50%. Accordingly, using capital as a measure of lending capacity, as is consistent with the banking literature, suggests a much larger decline than the authors' deposit-based measure of 15%.

4.2 THE MODEL

While the authors' framework may be interesting as a model of financial intermediation, it is not clear that it is particularly useful for studying crises. Within the model, a particular input requires bank finance. A key limitation—if the model is to be used to study crises—is that all other factors (labor, etc.) remain fixed in the wake of the shock. This greatly constrains the ability of a banking crises to generate a contraction in real activity.

I illustrate this point with a very simple model. Let X denote a variable factor that requires bank finance (e.g., inventories), and θ be the service flow from this input. Let L (e.g. labor) be another variable input. Output Y is then given by a simple Cobb–Douglas production function, as follows:

$$Y = (\theta X)^{\alpha} L^{1-\alpha}.$$

Note that, holding L constant, the elasticity of output with respect to X is given by the cost share:

$$\left(\frac{\partial Y}{\partial X} \frac{X}{Y} \right) \bigg|_L = \alpha.$$

The authors then proceed to analyze a banking crisis, as follows. Suppose that B is the quantity of available bank loans. A shock to banking arises that reduces B below its frictionless equilibrium value. X, accordingly, is constrained to equal B, and the decline in B exactly matches the decline in X. Importantly, no other factors adjust during the crisis. Given this assumption, the percentage decline in output due to a percentage contraction in bank lines is simply given by the cost share; in simple terms, $X = B$ implies

$$\left(\frac{\partial Y}{\partial B} \frac{B}{Y} \right) \bigg|_L = \alpha.$$

To compute the overall decline in output from the banking crisis, the authors multiply the cost share α by their measure of the percentage decline in banking of x percent. Since the cost share of banking in the GDP is a tiny number, the authors conclude that the banking crisis did not have much effect.

I stress, however, that the assumption that all other factors are held constant is key to justifying the cost share as the measure of the output elasticity with respect to bank loans. Suppose instead that labor is perfectly elastic in supply at the wage w. Then it is easy to show that the effect of a banking crisis on output may be considerably larger. In this instance the relevant output elasticity is given by

$$\frac{\partial Y}{\partial B} \frac{B}{Y} = 1.$$

Here output drops proportionally with bank loans.[12] Further, if the elasticity of lending with respect to capital is roughly unity (which presumes that banks maintain a stable ratio of capital to loans), then given my estimate of a roughly 50% decline in bank capital and my (overly) simple model, the banking crisis could have produced a decline in output of up

12. The elasticity is lower if there are diminishing returns. It is higher, however, if the elasticity of substitution between the bank-financed input and the other variable inputs is lower. It will also be higher if there are overhead financing costs, since in this instance a given percentage reduction in banking lending will imply a proportionally greater decline in funds available to finance the variable input.

to 50% for bank-dependent firms. I would add that this calculation ignores the potential impact on household spending.

I am not suggesting that anyone take my model seriously. My point is only to illustrate that the authors' connection between the cost share and the impact of a banking crisis rests on the assumption that the crisis makes no impact on other factors of production. True, my assumption of perfectly elastic factor supply is extreme. On the other hand, during the Great Depression there was a huge contraction in employment with relatively little movement in real wages. This elastic-like behavior of employment was surely not a simple consequence of preferences. Rather, it likely reflected labor-market frictions in conjunction with other forces. The key point is that a proper evaluation of the banking crisis likely requires taking into account other frictions, such as labor-market rigidities, possibly including nominal as well as real rigidities, that open up the possibility of large output fluctuations. The mere fact that a banking shock doesn't do much in an otherwise frictionless framework does not imply that it will be unimportant once the frictions outside the banking sector are properly taken into account.

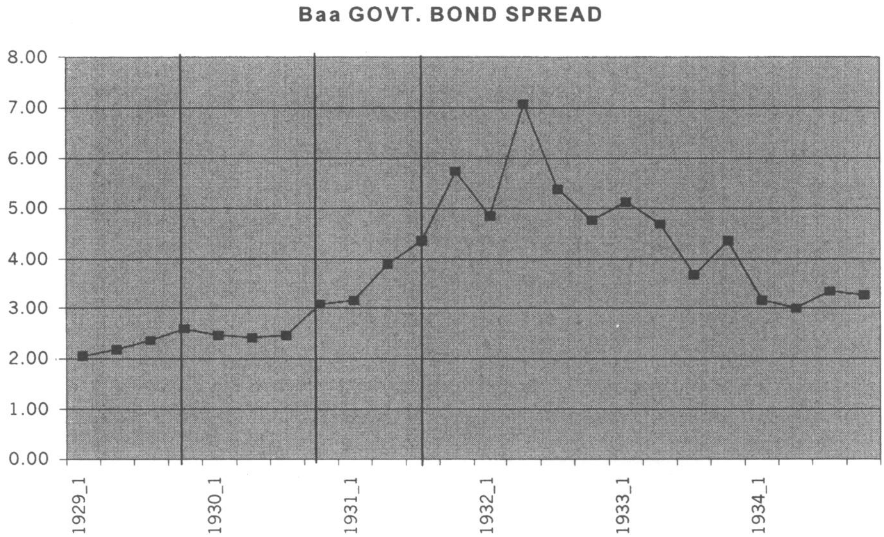
4.3 CROSS-STATE BANKING EVIDENCE

Examining the link between banking performance and output across states is in principle a good idea. Several problems confound the identification, however. As I have mentioned, the loss of bank capital is likely a better measure of the decline in lending capacity than are the deposits of failed banks. A state could have a banking industry in poor health due to low capitalization, but few banks that actually fail. (This is more likely to be true for states with large banks, since regulators are more likely to let small banks fail than large ones.) Accordingly, measurement error in the authors' independent variable (banking lending capacity) could be one factor responsible for the lack of explanatory power.

Second, there is likely unobserved heterogeneity across states. For example, midwestern states are dominated by durable-goods industries. Not controlling for this difference will bias the results. Finally, high integration of state economies also inhibits identification. A contraction in bank activity in Illinois that reduces demand for automobiles will lead to a contraction in Michigan output.

The authors' case would be more compelling if they set up the hypothesis test so that the argument they prefer requires a finding of statistical significance. But doing it the other way around—having victory depend on the absence of statistical significance—implies a test of low power: Absence of statistical significance could reflect a variety of factors having nothing to do with the authors' argument.

Figure 3 Baa GOVERNMENT-BOND SPREAD



5. *Risk Spreads, Dividends, and Japan*

I now tie up some loose ends regarding the authors' discussion of financial factors involving risk spreads and dividend behavior during the Depression and also the Japanese stock-market crash of the early 1990s.

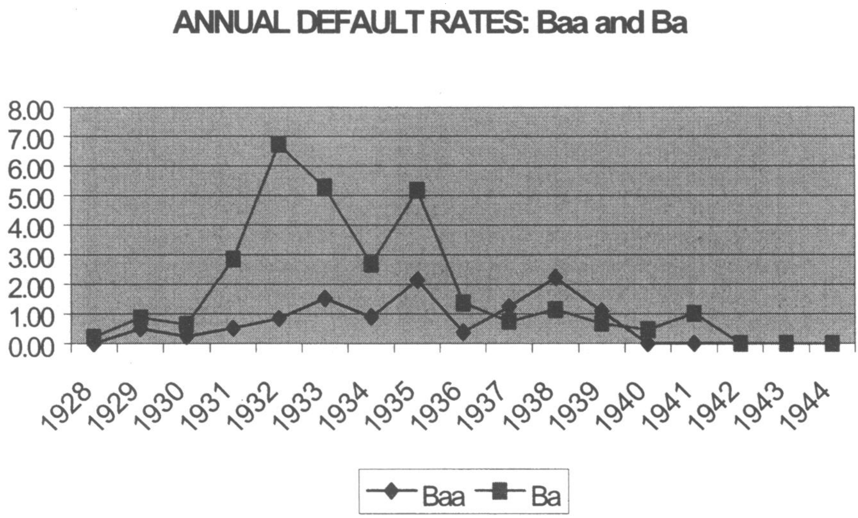
5.1 RISK SPREADS

One traditional indicator of the steady deterioration of financial conditions throughout the Depression is the behavior of the spread between Baa corporate bonds and long-term government bonds. As Figure 3 indicates, this spread rises from roughly 200 basis points in early 1929 to between 600 and 700 basis points in the wake of the September 1931 monetary tightening. The rise in the spread correlates well with the onset of debt deflation and the banking crises. Note further the nearly 300-basis-point jump in the spread in the interval between the banking crises and the wake of the subsequent monetary tightening.

The authors argue that the movement in the spread simply reflects expected default costs. This observation alone, however, does not rule out a role for financial factors. To the extent there are losses associated with bankruptcy, expected default costs entail an agency premium for external finance.[13]

13. In this instance, the agency cost of external finance equals the default probability times the deadweight cost associated with bankruptcy.

Figure 4 DEFAULT RATES (%)



Leaving aside the question of whether there are significant bankruptcy costs, however, it is very unlikely that expected default costs alone could account for the sharp rise in the spread. As Figure 4 indicates, default rates on Baa bonds were relatively low during the Depression, peaking at just 1.5% in 1933.[14] This low default rate should not be surprising: Even though Baa is not the highest rating possible, it is nonetheless an investment-grade classification, a status achievable only if the likelihood of default is quite small.[15] It is also important to recognize that the bonds are long-term—the average maturity in Moody's sample is between twenty and thirty years—implying that the spread depends not only on the expected default probability in any given year, but rather on the expected average annual default probability over the life of the bond.[16] The low average annual default probabilities portrayed in Figure 4 sug-

14. I thank Richard Cantor for supplying me with the Moody's default-rate data.

15. To gain some perspective, the default rate on Aaa bonds was effectively zero during the Depression. Defaults were concentrated mainly among speculative (non-investment-grade) securities.

16. To get some sense of the significance of maturity for the spread, consider a discount bond that pays either 1.0 with probability $1 - \pi$ after T periods or 0.5 with probability π. Suppose further that investors are risk-neutral and that the riskless rate is fixed at $R > 1$. Then it is straightforward to show that the rate spread between the risky bond and riskless security may be expressed as $\pi/T \times 0.5$. Note that π is the cumulative default probability and that π/T is the annual average default probability. For $T = 20$, for example, an annual average default probability of 1.5%—a number well above the annual average for Baa-rated bonds—would generate a spread of only 75 basis points.

gests that expected defaults cannot come close to explaining 300–400-basis-point jump in the spread.[17]

One possibility is that the spread might have reflected the likelihood of a Baa-rated firm being reclassified into a higher default-rate category. However, the risk associated with non-investment-grade Ba bonds (the quality-level below Baa) appears too low to rationalize downgrade risk being a factor in the Baa spread. As Figure 4 illustrates, the Ba default rate peaked at 7% in 1932, before settling back to a low level by 1936. Annual default rates on Ba bonds of this magnitude do not appear able to raise expected default costs on Baa bonds sufficiently to explain the spread for the highest-rated Baa bonds, especially given that the probability of being reclassified from Baa to Ba was likely not huge.[18] Finally, note that the issue of ratings downgrades also cuts the other way: The measured spread may significantly understate the true rise in the spread to the extent that firms in distress were downgraded from Baa and thus dropped from the sample used to construct the average Baa rate.

What then accounts for the sharp rise in the spread over this period? Friedman and Schwartz argue that much of it reflected an increasing liquidity premium that was due to capital-constrained banks unloading their holdings of Baa bonds on the open market, especially in the wake of the banking crises and subsequent monetary tightening.[19] Apparently, the same Baa bonds intermediated by banks commanded a larger premium when floated on the open market. This suggests that the movements in the spread indeed reflected financial distress and, among other things, that the contraction in banking was indeed affecting real credit costs.

5.2 DIVIDENDS

The authors argue that firms on the whole greatly smoothed dividends throughout the Depression, suggesting an absence of financial distress. I agree that dividend behavior is an important issue, but question the

17. The idea that expected default costs explain little of the movement in the risk spread on corporate bonds is consistent with recent evidence. See, for example, Elton et al. (2001).
18. The increase in the probability of default from downgrade risk equals the probability of downgrade times the difference in the default rate between the higher and the lower risk class. Elton et al. show that the annual probability of a downgrade from Baa to Ba was only 5.4% over 1987–1996, the period when the corporate default rate was the highest since the time of the Depression. While this transition probability was surely higher in the Depression, it is safe to say that it was still considerably less than unity, particularly for the bonds in the Moody's sample (since the latter tends to drop bonds selling at a deep discount.)
19. Friedman and Schwartz (1963, p. 312) state: "Interest rates clearly show the effects of the banking crisis. . . . The yield on corporate bonds rose sharply, the yields on government bonds began to fall. The reason is clear. In their search for liquidity, banks and others were inclined first to dispose of their lower grade bonds."

authors' claim, for three basic reasons: (1) Though there was considerable smoothing in the early stages of the Depression, firms did cut dividends significantly after 1930. (2) A simple comparison of total dividends with total corporate profits significantly overstates the degree of dividend smoothing, due to an aggregation bias. Firms with positive profits, mainly large firms, accounted for most of the aggregate dividend payments; small and medium-sized firms that experienced substantial losses could not cut dividends below zero. (3) The residual dividend smoothing (mainly by large firms) may have in part reflected pressures from shareholders who themselves faced financial distress.

To gain some perspective, Table 3 reports the behavior of nominal dividends relative to nominal personal income. Consistent with conventional wisdom, a buildup of retained earnings over the late 1920s permitted corporations to keep dividends relatively stable between 1929 and 1930. After 1930, however, dividends dropped sharply not only in absolute terms, but also in comparison with the overall drop in personal income. In 1930, dividends were roughly 7.1% of total personal income. By 1933 the ratio drops by roughly 40%, to 4.3%. Thus, the aggregate evidence does suggest significant dividend cuts.

Why didn't dividends fall to zero? First, the aggregation bias is relevant. Even though total profits became negative after 1930, a significant fraction of firms continued to earn positive profits. Table 4, for example, shows that corporate income-tax payments remained positive throughout the Great Depression, suggesting that a core of firms were indeed earning money over this entire time period. Fabricant (1934) presents direct estimates of total earnings of corporations with non-negative profits along with total losses by firms with negative profits. As the Fabricant data make clear, the drop in aggregate profits over the period in part reflected a fraction of firms each year drifting from positive to negative profits. Since dividends are bounded below at zero, total dividends may drop less than proportionally to total profits simply in part due to aggre-

Table 3 CORPORATE DIVIDENDS AND PERSONAL INCOME

Year	*Corporate dividends ($million)*	*Personal income ($million)*	*Ratio (%)*
1929	5801	85,905	6.75
1930	5468	77,015	7.10
1931	4066	65,896	6.17
1932	2544	50,150	5.07
1933	2038	47,004	4.34

Table 4 CORPORATE TAXES, PROFITS, AND DIVIDENDS[a]

Year	*Aggr. profits before taxes*	*Corp. taxes*	*Aggregate profits* $Y > 0$	*Aggregate profits* $Y < 0$	*Aggr. profits* $(Y > 0)$ −taxes −divs.
1929	9,990	1,369	13,841	−3,851	6,671
1930	3,697	842	7,987	−4,290	1,677
1931	−372	498	4,801	−5,173	237
1932	−2,309	385	2,800	−5,109	−129
1933	956	521	3,789	−2,833	1,230

[a]Millions of dollars.

gation bias (as opposed to everything being accounted for by individual firms actually smoothing dividends).

To gain some sense of the bias, I used the implied average corporate income-tax rate from the Fabricant data to construct estimates of earnings by corporations with positive profits.[20] Table 4 reports these estimates along with the difference between the after-tax earnings of these corporations and aggregate dividends. Note that throughout the Depression, (estimated) aggregate earnings by these corporations are sufficient to cover dividends, the one exception being a slight shortfall in 1932.[21] I don't mean to suggest that only firms with positive profits paid dividends, but rather that the bulk of dividend payments came from this group and not, as the authors imply, firms with highly negative earnings.

Indeed, according to Fabricant, it was mainly large firms with positive earnings that continued to pay dividends throughout the Depression.[22]

20. To construct the average tax rates I divided Fabricant's (1934) estimates of total corporate taxes (see his Table 1) by profits of corporations with positive earnings (see his Table 2). Since his data only go through 1932, I used for 1933 the same implied average tax rate as for 1932. The average tax rates I used accordingly for 1929–1933 are 9.89, 10.54, 10.37, 13.75, and 13.75. To then get the estimate of total profits for corporations with positive earnings I divided the national income and product accounts (NIPA) measure of corporate income taxes by the estimated average tax rates. Note that Fabricant's measures of total corporate income, dividends, and taxes differ a bit from the NIPA data, since the former do not eliminate double counting from cross-holdings of stock. Finally, my calculations do not adjust for inventory valuation adjustment, which implies that profits are understated somewhat for 1929–1932 and overstated somewhat for 1933.
21. Fabricant (1935) estimates that net saving of firms with positive profits in 1932 was $132 million, suggesting that a small portion of aggregate dividends was paid by firms making losses, but that on the whole firms with positive profits accounted for the bulk of dividend payments.
22. Examples of industries in which large firms were in a position to maintain a relatively steady flow of dividend payments include: food, tobacco, chemicals, public utilities, and communications.

The small and medium-sized firms that experienced heavy losses largely suspended dividend payments. A look at the disaggregated evidence, accordingly, suggests that firms in financial distress were indeed adjusting dividend behavior as one would expect.

Why didn't large companies reduce dividends to zero? First, the fact that these companies on average maintained positive profits throughout the Depression suggests that they were at least capable of making payouts without dipping into capital. Second, cutting dividends is not costless, especially during a period where shareholders have already experienced significant financial distress. Pressure to smooth consumption of liquidity-constrained shareholders may have affected dividend policy.

In sum, one cannot conclude, from the simple aggregate evidence on dividends and profits that the authors report, that financial constraints were unimportant.

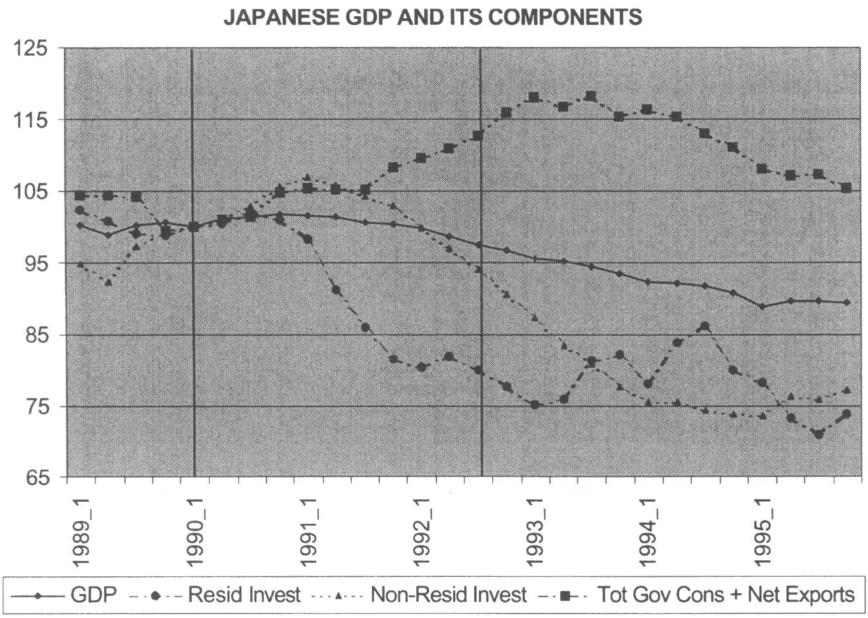
5.3 JAPAN

The authors present some Japanese data from the time of the collapse of the Nikkei index to suggest that theories which emphasize asset prices as a source of variation in financial conditions were not likely at work during the Depression. Here I argue that scratching just a bit below the surface leads one to exactly the opposite conclusion. Again, the issue boils down to taking account of all the relevant heterogeneity.

The authors argue at the time of the decline in stock prices—from early 1990 to mid 1992—Japanese output did not drop significantly; and they conclude accordingly that the asset price collapse did not make a large impact on the economy. However, extending the sample period just a few years and disaggregating the data yields a quite different scenario. Figure 5 plots the behavior of four Japanese series over the period 1989:1 to 1995:4: real output, residential investment, nonresidential investment, and the sum of government purchases and net exports. Each series is detrended using the authors' procedure. Each variable is normalized to be 100 at the beginning of the stock-market downturn. Finally, the two vertical lines denote the beginning and end of the Nikkei collapse: 1990:1 and 1992:2, respectively.

Note first that in the midst of the crash residential investment drops precipitously. By late 1992 it is down 25% relative to trend and remains in this rough vicinity for the next three years. By late 1991, nonresidential investment also begins a sharp decline. It is 15% below trend by early 1993 and bottoms out at 25% below trend by 1994:1, remaining at this level for the next two years. The behavior of investment overall is entirely consistent with financial theories.

Figure 5 JAPANESE GDP AND ITS COMPONENTS: 1989:1–1995:4



Given the investment collapse, why did Japanese output not fall more precipitously early on? Here it is important to account for a number of key differences from the Great Depression. First, in contrast to the United States during the 1930s, Japanese government spending rose significantly over the early 1990s, particularly public investment expenditures. Second, while U.S. monetary policy was constrained by the gold standard, depreciation of the yen over this time helped induce a rise in Japanese net exports. Figure 5 shows that the sum of government spending and net exports rose steadily from early 1990 to early 1993, reaching more than 15% above trend at this time. This additional source of stimulus—which did not arise in the Depression—helped moderate the overall impact of the contraction in domestic investment.

Finally, unlike the laissez-faire banking system in the United States during the 1930s, the Japanese banking system was heavily protected by public guarantees, which permitted lending to continue through the early 1990s, even though the collapse of both stock and land prices directly weakened bank balance sheets. By 1993, however, the severe problems in Japanese banking surfaced (see, e.g., Hoshi and Kashyap, 1999). Only at this point do significant constraints on banking activities begin. In contrast, U.S. banks during the Depression did not enjoy this period of protection from the initial financial crisis.

Once the banking problems were no longer contained, Japanese output began a steady contraction. By early 1993 output was 5% below trend, dropping to 10% below trend by the end of 1995. The stagnation continues, and many observers cite the weak financial system as a key factor.

The descriptive evidence I have just cited, of course, does not prove that financial factors were at work in Japan. But nor does the authors' simple evidence suggest otherwise. More systematic empirical work is required to resolve the issue.

6. *Concluding Remarks*

Elsewhere the authors have done interesting work on the Depression. Cole and Ohanian (2000) propose an interesting explanation of why the slump persisted for nearly six years after 1933, based on reduced competition induced by New Deal regulatory policies. The purpose of the current paper, however, is simply to try to dismiss monetary and financial theories of the 1929–1933 downturn, without offering a clear alternative to judge. In my opinion, the authors do not succeed.

It is critical, however, to develop a quantitative model to show that monetary and financial factors can indeed account for the key features of the Depression. I would guess that such a framework would likely have to incorporate all of the following features: (1) some form of nominal rigidity, and possibly also a real labor-market rigidity, to permit a significant contraction in aggregate demand in conjunction with a relatively small movement in real wages; (2) a central bank constrained by a gold standard; and (3) frictions in the credit market that can disrupt spending by households and small and medium-sized firms. Perhaps one new lesson from this discussion for modern business-cycle theories that feature monetary and financial factors is that more emphasis should be placed on how credit-market frictions may constrain household spending, given the key role of the collapse of consumer demand in the early stages of the Depression.

REFERENCES

Bernanke, B. (1983). Non-monetary effects of the financial crisis in the propagation of the Great Depression. *American Economic Review* 73(3):257–276.

———, and H. James. (1991). The gold standard, deflation and financial crises: An international comparison. In *Financial Markets and Financial Crises*, R. G. Hubbard (ed.). Chicago: University of Chicago Press, pp. 33–68.

———, and C. S. Lown. (1991). The credit crunch. *Brookings Papers on Economic Activity*, 2, 205–39.
Bordo, M., C. Erceg, and C. Evans. (2000). *American Economic Review*, forthcoming.
Calomiris, C. W., and B. Wilson. (1998). Bank capital and portfolio management: The 1930s capital crunch and scramble to shed risk. Cambridge, MA: National Bureau of Economic Research. NBER Working Paper 6649.
Campa, J. M. (1990). Exchange rates and economic recovery in the 1930s: An extension to Latin America. *Journal of Economic History* 50(September):677–682.
Choudri, E. U., and L. A. Kochin. (1980). The exchange rate and the international transmission of business cycle disturbances. *Journal of Money, Credit and Banking* 12(November):565–574.
Cole, H., and L. Ohanian. (2000). New Deal policies and the persistence of the Great Depression: A general equilibrium analysis. Federal Reserve Bank of Minneapolis. Mimeo.
Eichengreen, B. (1992). *Golden Fetters: The Gold Standard and the Great Depression, 1919–39*. New York: Oxford University Press.
Elton, E., M. Gruber, D. Agrawal, and C. Mann. (2001). Explaining the rate spread on corporate bonds. *Journal of Finance*, February.
Fabricant, S. (1934). Recent Corporate Profits in the United States. National Bureau of Economic Research. Bulletin 50.
———. (1935). Profits, losses and business assets, 1929–34. National Bureau of Economic Research. Bulletin 55.
Friedman, M., and A. J. Schwartz. (1963). *A Monetary History of the United States, 1867–1960*. Princeton: Princeton University Press.
Hamilton, J. (1987). Monetary factors in the Great Depression. *Journal of Monetary Economics*, 19(2):145–170.
Hoshi, T., and A. Kashyap. (1999). The Japanese banking crisis. In *NBER Macroeconomics Annual*, B. Bernanke and J. Rotemberg (eds.). Cambridge, MA: The MIT Press, pp. 129–201.
Kindelberger, C. (1986). *The World in Depression: 1929–39*. Berkeley: University of California Press.
Mishkin, F. S. (1978). The household balance sheet and the Great Depression. *Journal of Economic History* 38(4):918–937.
Olney, M. (1999). Avoiding default: The role of credit in the consumption collapse of 1930. *Quarterly Journal of Economics* 114(1):319–335.
Romer, C. (1993). The nation in depression. *Journal of Economic Perspectives* 7(2):19–39.
Temin, P. (1989). *Lessons from the Great Depression*. Cambridge, MA: The MIT Press.

Discussion

Hal Cole and Lee Ohanian began by replying to the discussants. There was agreement about the need for general equilibrium modeling of the Depression. In response to Mark Gertler, Cole agreed that small firms faced severe financial distress; but if many large firms were solvent, why didn't they take over the small firms, their markets, or both? In response to Michael Bordo, Ohanian noted that the small effects found by their wage-rigidity model were not due to a lack of intrinsic persistence in the model, but rather resulted from partial-equilibrium employment effects being partially undone by general equilibrium effects. Ohanian also re-emphasized the problems with 1930s wage data, citing Stanley Lebergott on the potential importance of compositional effects. Specifically, if low-paid workers (or employees of small firms, which paid lower wages) were more likely to lose their jobs, then aggregate wage data overstate the increase in the real wage of the typical worker.

Beginning the general discussion, Rick Mishkin argued for a broader interpretation of financial shocks, which would take into account the deterioration of balance sheets as well as banking problems. To the extent that households and firms were in financial distress in the 1930s, increased moral hazard in credit relationships would have reduced their access to credit. Cole argued that the Japanese experience of the 1990s, following the boom and bust in Japanese asset prices, suggests that a sustained slowdown in growth, rather than a collapse of output, is the likely result of balance-sheet problems. Ben Friedman criticized the assumption of the paper that the importance of bank credit to the economy could be measured by the share of banking in value added; he argued that credit may play an essential role that is not well captured by a smooth neoclassical production function. Diego Comin noted that total factor productivity declined sharply during the Depression, a fact that might be construed in support of a real-side interpretation of the collapse.

Robert Gordon emphasized the difficulty of disentangling cause and effect in an environment when all sectors of the economy are contracting simultaneously; he argued for more use of cross-national comparisons to identify causal factors. He also pointed out that, even though inventories are a small part of the economy, changes in inventory investment play a large role in fluctuations. The fact that banks finance a large portion of inventories suggests another possible channel of influence from banks to the real economy. On the cross-state evidence on bank failures, Gordon noted that "this is one economy"; that is, we would not necessarily expect the severity of the Depression to differ greatly across states even if the incidence of bank failures differed geographically.

Susanto Basu suggested an alternative banking model in which firms have access to two technologies, one that uses financial intermediation and is relatively efficient and one that does not use intermediation and is less efficient. If a firm is unable to get a loan, it uses the second technology. In this model, some part of the observed TFP decline reflects the loss of financial services; further, firms forced to use the less efficient technology will want less capital and may choose to liquidate part of the firm. Pierre Gourinchas cautioned that general equilibrium models of the Depression might not be able to employ the usual technique of approximating around the steady state, as the deviations from the steady state in the 1930s were presumably large and nonlinearities might be quite important. Ohanian noted that the relatively simple models used in their paper permitted exact solution and did not require approximations around the steady state.

Michael Woodford pointed out that the paper's sticky-wage model has the highly counterfactual implication that unemployment is zero, as workers displaced in the sticky-wage sector find work in the flexible-wage sector. He also noted that the effects of higher real wages on employment depend critically on the elasticity of marginal product with respect to employment. It may be that elasticity is low in the short run (that is, the labor demand curve is flat) due to factors such as variable capital utilization; if so, relatively small changes in real wages could have large employment effects.

Ben Bernanke criticized the paper for ignoring cross-country evidence. According to studies encompassing 20–30 countries, those countries that left gold earlier (and thus were able to reflate their money supplies and price levels) did better than those that remained on gold. There is also some cross-country evidence in favor of the banking hypothesis, e.g., in his work with Harold James. Bernanke also objected to the modeling of the effects of banking crises; instead of putting financial services in the production function, he prefers an approach that allows for increased agency costs of lending when financial conditions deteriorate. Commenting further on the model of banking, he pointed out that intermediation services affect spending as well as production; for example, if buffer-stock consumers face increased unemployment risk while simultaneously losing access to credit, they are likely to sharply reduce their spending. If financial distress affects spending more than production, the lack of correlation between bank failures and production at the state level is not surprising, as already suggested by Bob Gordon; for example, if financial distress reduces the demand for automobiles in Alabama, output in Michigan rather than in Alabama will be most affected. Finally, Bernanke noted several differences between the experiences of 1920–1921 and 1929–1933; these included (1) the fact that the 1920–1921 deflation, unlike the later deflation, followed a sharp inflation that was widely expected to be temporary and (2) institutional changes in labor markets that reduced wage flexibility and increased labor hoarding in the latter episode.