CHAPTER 7
WHAT CAUSED THE GREAT DEPRESSION?

The Great Depression of 1929-33 stands out as one of the watersheds in the history of the United States. In Chapter 2 we described the course of this great contraction and its effects on production, employment, and living standards. In Chapter 6 we examined the role of monetary and fiscal policies in countering the depression and spurring recovery, though, as we saw, there is considerable doubt as to whether the policies actually did this. What we will now do is try to understand why the Great Depression occurred, but it will become evident that this is no easy task. Though it has been extensively analyzed, there continues to be considerable disagreement on the explanation of the contraction.¹ A good place to begin is by briefly describing an economic contraction’s general characteristics.

A Depression is a period of at least two consecutive quarters of declining real gross national product.² The last quarter prior to the beginning of the decline in real output is the peak of the expansion. The recovery begins when there are at least two consecutive quarters with a rising real gross national product. The previous quarter is then designated as the trough of the contraction. From peak to trough, real gross national product declines and is invariably accompanied by rising unemployment rates, declines in the average number of hours worked each week, and a declining utilization of industrial capital and of wholesale and retail distribution facilities and equipment. The Great Depression of 1929 to 1933 conforms to these definitions. The contraction began in the second quarter of 1929 and troughed in the first quarter of 1933. During this time real output and the overall level of prices both fell, though not always together or at the same rate over time. The unemployment rate rose until, for the entire year of 1933, it averaged 25 percent of the total labor force and nearly 38 percent of the nonfarm labor force. Manufacturing facilities all over the United States were idled, and storefronts were shuttered as businesses closed.

To understand an economic contraction, particularly of the magnitude of the Great Depression, it must be noted that the contraction is not initiated by a loss of knowledge or a noticeable loss of resources. Rather, a depression is a collapse in the volume of economic activity. During the Great Depression neither workers, plants, equipment, transportation facilities, nor raw materials disappeared—many just stood idle and unused because they were not demanded. People did not desire to go hungry, wear hand-me-down clothes, give up their phones, and stop purchasing “luxuries” like fresh fruit, new curtains, shoes, suits, and perfume. The loss of much or all of their income, due to reduced employment, forced them to reduce their demands for these and other goods.

A market system uses prices to coordinate economic activities. Consumers use market prices, given their individual preferences, to decide how to allocate their incomes among the possible goods and services they could purchase. Those consumer decisions lead individual producers to demand resources, given the prices and productivity of the resources, to produce the goods and services that consumers are demanding. The prices of the resources and the amount of resources that consumers supply determine their incomes. And these incomes become the basis of the consumers’ effective demands.

The task of coordinating demands and supplies for this multitude of goods and resources falls primarily to the money prices of those resources and goods. If some markets have an excess supply while others have an excess demand, then relative money prices will change to induce resources to move toward markets with excess demands and induce consumers to shift their demands toward markets with excess supplies. Money prices provide low-cost information, facilitating specialization through the division of labor and coordinating the use of individuals’ differing knowledge of the particular circumstances of time and place. Costly information and costly movement through space make perfect economic coordination unachievable.³ However, generally the coordination process works rather well and is able to adjust to changes in the parameters underlying individual demand and supply functions in the various markets.⁴

A depression occurs when something disrupts this coordination process, and it is here that the disagreements begin. Most of the sources of such disruptions can be conveniently classified under the headings of the Keynesian explanation, the Monetarist explanation, or the Austrian explanation, though such classifications are rather arbitrary and do not do justice to their complexity.⁵

The Keynesian explanation generally assumes that demands for products and services in one or more private sectors decline, and this initially disrupts market coordination. Some contend that in time adjustments will again coordinate markets, but
this period of time is so long that the government must intervene to guide the economy back to coordinated full employment. Others argue that, in fact, the market process is fundamentally flawed and government must continually guide the economy toward full employment to avert continuous depresesions.

The Monetarist explanation generally assumes that markets work well to coordinate economic activities. Proponents of this view see the problem in money. Because of its crucial role as a medium of exchange, measure of value, and way to hold wealth, disruptions in the stock of money or in its acceptance for goods, services, or resources disrupt market coordination. Most Monetarists believe that the supply of money is too important to be left to the private sector and must become a government monopoly. The problem then becomes the wise or appropriate use of monetary policy; they argue that depressions, when government controls the stock of money, can be traced to failures of monetary policy.

Those who favor the Austrian explanation also argue that market processes work well to coordinate economic activity. However, they maintain that the knowledge problem makes it impossible for government to direct the economy. The individuals in the government who must actually make the decisions can never have the knowledge that is available to all of the market participants, particularly the knowledge of time, place, and circumstances available to individual market transactors. “Austrians” also argue that money is too important to be monopolized or controlled by government authorities because they can never have the knowledge necessary to wield monetary policy wisely, and there is too much temptation to let political agendas determine it. In general Austrians argue that depressions occur because governments disrupt market processes through such things as grants of monopoly, tariffs and trade barriers, and wage and price controls or through policies that alter the stock of money in ways that discoordinate economic activities.

The Austrian Explanation of The Great Depression

The Austrian explanation of a depression or “business cycle” was developed by Ludwig von Mises and Friedrich A. Hayek in the twenties and early thirties prior to the development of either the Keynesian or Monetarist explanations. Overshadowed by the Keynesian approach since the publication of the General Theory in 1936, in the 1970s interest in the Austrian approach to economic analysis began to revive. The Austrian explanation of the Great Depression can be summarized as follows.

In the twenties the Federal Reserve System engineered increases in the money stock beyond what was justified by the increases in the gold stock. The Fed, concerned with speculation in the securities markets and enormous increases in securities prices, stopped increasing the monetary base (the nonbanking public’s currency holdings and the banks’ reserves) for most of 1928 and most of 1929. As long as the Federal Reserve System was increasing the monetary base, the banks had been able to make more business and consumer loans and purchase more financial securities, thereby expanding the stock of money. In doing this, interest rates were driven lower than they otherwise would have been. The lower interest rates induced individuals and firms to increase their borrowing and investment. As the interest rates declined, the market values of assets, particularly capital goods, increased, and the longer the life of an asset, the greater the rise in its market. Entrepreneurs were also led by the lower interest rates to shift resources to productive uses farther removed from final consumption uses. These changes in the values of capital goods induced businesses to alter production techniques to use more capital goods with longer lives, because these processes had became more profitable. Businesses also altered the ways in which they financed their investments to take advantage of those relative price changes and, as a result, relied more on equity financing and the issuance of longer-term debt. The result was to alter the pattern of production and financing.

Given the prices resulting from the expansion of the stock of money through the banking system, industrialists were making what appeared to be correct decisions. The investment boom induced them to divert resources from the production of consumer goods (whose prices had not changed as much because they have relatively short lives) toward investment goods (whose prices had risen more because they tend to have longer lives) and toward stages of production farther removed from final consumption. However, consumer preferences between current and future consumption had not altered in such a manner, and as the additional money was spent by the initial borrowers and began to circulate in the economy, consumer spending on currently produced goods rose, and consumer saving was insufficient to sustain the lower rates of interest. These demand changes and the rise in interest rates (with the resulting relative price changes) were postponed as the monetary authorities continued to expand the stock of money. However, once the
monetary authorities stopped increasing the money stock in 1928, the restoration of consumers’ preferences between spending and saving caused interest rates to rise and the market value of capital assets and the prices of raw materials to fall. The same investments and resources whose market values had risen the most with the lower interest rates experienced the largest price declines. Business firms found that they had undertaken unprofitable investments, and some partially completed investment projects could not be profitably completed. During the monetary expansion markets had become “discoordinated” because relative prices had been altered and had sent signals to industrialists to take actions that were actually inconsistent with consumer preferences. When prices adjusted with the end of the monetary expansion, the discovery of the malinvestments led to a drop in investment demand and in the demand for raw materials used in the production processes—particularly in stages of production farther removed from final consumption uses.

According to Austrian economists, this is how the downturn in the middle of 1929 originated. A “depression” of economic activity of some length was inevitable because it takes time to correct the previous malinvestments. However, a secondary depression created by further government intervention and a severe contraction of the stock of money associated with several banking crises transformed this into the protracted and extraordinarily severe Great Depression. At the end of 1929, President Hoover called several well-publicized conferences of leaders of major firms, asking them to maintain money wages because “labor was not a commodity to be liquidated.” Hoover insisted that the first shock of the depression had to fall on profits and not upon wages. On November 18 the presidents of the major railroads were summoned, on November 21 the great industrial leaders were called in, on November 22 the leading representatives of the building and construction industries came, and on November 27 the leading public utility executives appeared. On December 4 a larger group of industrial leaders were called to Washington, D.C., where they agreed to Hoover’s program, particularly the maintenance of money wage rates. The press and the labor unions hailed these “historic” new agreements.

Through the first half of 1930, nominal wage rates in major industries remained virtually constant. During this period product prices were falling so that real wages and real firm labor costs were rising. Firms began reducing production and employment as business profits were precipitously declining. By the end of 1930 and the beginning of 1931, firms could no longer resist the pressures and began cutting wage rates. However, real wage rates in many industries continued to rise through 1931. The constancy of money wages increased involuntary unemployment and stopped equilibrating price adjustments.\(^9\) Lowell Gallaway and Richard Vedder point out that the failure to adjust wage rates in 1930 had serious second-round effects that made the decline more severe in 1931 and 1932. As prices and productivity fell while businesses held wages constant, profits were squeezed. By mid-1930, before the first banking crisis occurred, the financial community was noticing this severe profit squeeze, leading to a decline in the demand for corporate equities and debt. Corporate savings declined from $2.8 billion in 1929 to -$2.6 billion in 1930 and -$4.9 billion in 1930. The decreasing market value of business loans made by banks wiped out much of the net worth of many financial institutions and may have been a major factor in the banking crisis.\(^10\)

This was not the only action taken during this period. In December of 1928, the Hawley-Smoot tariff proposed to provide some protection for American farmers, received Hoover’s blessing. During 1929 Congress worked on the tariff, and numerous special interests succeeded in gaining new or additional protection for all types of domestically produced commodities. Economists from across the United States denounced the tariff, and well before Hoover signed the tariff bill in June of 1930, other nations were threatening retaliatory actions. With the signing of the tariff, many other nations attempted to counteract it by new or altered quotas on U.S. exports and/or by sharply rising tariffs on commodities exported out of the United States. The Smoot-Hawley tariff and retaliatory tariffs by other nations sharply altered the supplies and demands for many products, necessitating a complex and extensive set of price and resource adjustments. The export-producing sectors in most countries are the most dynamic and innovative in the economy, so the trade war initiated by the Smoot-Hawley tariff was particularly harmful to the pace of development.\(^11\)

Beginning in 1931 the initiation of a sharp decline in the stock of money and the intermittent banking panics, both of which ended in March of 1933, required further relative and absolute price adjustments, including interest rates on bank loans and yields on securities that were higher than they otherwise would have been, leading to greater decreases in the prices of long-term assets.

In the Austrian view, then, the Great Depression was the inevitable result of the boom of the twenties that was led by the Fed’s creation of new money. The depression commenced when the mistakes made during the boom—the malinvestments—were discovered, and industrialists
began to make adjustments. The shocks to the economic system during the depression—the tariff, banking panics, and declines in the stock of money—and the initial failure of wage rates to fall created a secondary depression that prolonged and deepened what already was a serious contraction.

**Criticsims of the Austrian Explanation**
The Austrian explanation of the Great Depression has been criticized on a number of points, not all of which can be presented here. One criticism involves the extent of the investment maladjustments that would have arisen due to the inflation of the stock of money in the 1924-28 period. The general contention is that with relatively constant consumer and wholesale prices in the decade, it is difficult to believe that very large maladjustments in the structure of production and consumption could have occurred—certainly not ones large enough to have required massive resource reallocations capable of explaining the exceptional depth and length of the Great Depression. Austrians reply that rising prices are not necessary; all that is required is that the creation of money through bank lending drive the market rate of interest below that rate consistent with consumer preferences. In addition, Austrians also admit that the secondary depression, as Hayek termed it, was much more severe and responsible for the extraordinary length and severity of the depression as a whole, but, they argue, this should not blind us to the fact that the primary depression, which started the process, was caused by malinvestments due to the monetary expansion.

A second major point has been the importance of the constancy of money wage rates in 1930 and the resulting rise in real wage rates. Some have asserted that the economic structure changed after the 1920-21 depression, and real price and wage flexibility were lost. Others argue that there was nothing unusual about the wage stickiness, and it was not a major factor in the length or severity of the depression. However, then and now many have considered wages a major factor in the depression.

Some have argued that declining wage rates would not have promoted recovery because the declining wage rates would have reduced consumer incomes and thus spending. Austrian responses to this involve two points. First it is wage rates that decline and not necessarily total wages paid. As wage rates decline, businesses will tend to hire more workers. Thus, even at lower wages the relatively greater employment may lead to nearly the same, or perhaps even greater, total wages paid. The second point is that Austrians do not argue that all wage rates have to fall equally or even that all wage rates have to fall. What is necessary is that a pattern of wage rates consistent with other prices and with consumer preferences be established. In the boom preceding a contraction, wage rates in the production of capital goods further removed from final consumption uses and in the production of raw materials for industrial uses will tend to rise more. In the contraction, as the malinvestments are liquidated, capital values reduced, and the pattern of production shortened, wage rates, particularly in the capital-goods producing industries, have to decline.

The Austrian explanation for the Great Depression received a considerable acceptance in the early 1930s, much of which was due to Hayek’s arguments in the late 1920s that, given the Austrian theory of the business cycle, the boom, particularly in the United States, was not sustainable and a depression was coming. However, the Austrian policy proposals of letting the depression run its course to correct malinvestments and restructure wages and prices was considered by many to be an overly harsh solution. When Keynes proposed his theory in 1936, many quickly accepted it, some for its theoretical apparatus that seemed to them to better explain the Great Depression, and others for its policy implications that government could step in and replace private investment as a means to get depressed economies moving toward full employment. This eager acceptance cast the Austrian explanation aside and led to the dominance of the Keynesian explanation through the 1950s.

**The Keynesian Explanation**
The Keynesian, or “spending,” explanation asserts that the Great Depression began with a fall in one or more autonomous components of aggregate expenditures. The fall in spending led to a fall in incomes, and this brought on a further fall in induced investment spending, creating a general contraction of income and employment. Early Keynesian economists who examined the Great Depression assumed that the only way that it could have been stopped prior to its natural ending would have been to directly stimulate commodities spending. They were convinced that in the thirties the American economy was operating in the “liquidity trap,” making monetary policy ineffective.

In the Keynesian framework the key to analyzing and understanding the Great Depression is to find the component or components of planned autonomous expenditures that declined and set off the cyclical contraction. Several types of spending have been suggested as the culprits. The earliest explanations centered on declining housing construction and declining investment in the automobile industry at the end of the twenties. The
most recent candidate is an autonomous fall in consumption expenditures during 1930.\footnote{17}

**Excess Capacity and Saturation in the Automobile Market.**

The automobile industry had grown rapidly during the twenties. Though the number of firms fell, the surviving firms, especially Ford, General Motors, and Chrysler (after its creation in 1924) had expanded as the production of cars rose from 1,518,000 in 1921 to 4,587,400 in 1929. This expanded production required a substantial investment by the surviving producers, an investment that Keynesians contend began declining at the end of the twenties.

There were two interrelated aspects of the automobile industry’s decline in investment spending. First, it was suggested that by the end of the twenties investment in the automobile industry had created excess capacity. Second, the market for automobiles was also supposed to be “saturated” so that at the existing incomes of families and the distribution of that income and given the preferences of families for automobiles compared to other commodities and services, the rate of sales of automobiles per period (or the rate of growth of the sales per period) was at a maximum. Thus, even if there were no current excess capacity, this would still have led to a fall in investment spending compared to what had been occurring, reducing aggregate demand and leading into the Great Depression.

In 1934 Edwin Nourse examined the extent of excess capacity in the automobile industry in the late twenties and found that it was substantial. However, Nourse’s analysis was based on engineering and not economic concepts.\footnote{18} In a recent reexamination Lloyd J. Mercer and W. Douglas Morgan point out that “economic capacity” is given by the minimum point on the U-shaped short-run average cost curve, not by the maximum physical output the plant should be able to produce.\footnote{19} To evaluate the excess capacity argument, they estimated the capacity output where costs were minimized.\footnote{20} The utilization of the capital stock in the automobile industry during each period was then found by comparing the actual output in any year with the capacity output. When the ratio of actual to capacity output exceeded one it meant that unit costs were higher, providing an incentive for investment to lower the unit costs of production. When the ratio was less than one, it meant that the firms had excess capacity, leading to falling investment.

Figure 7.1 represents Mercer and Morgan’s estimates of current capacity utilization rates in the American automobile industry quarterly from 1921 through 1940 and indicates that by the late twenties there was some excess capacity in the automobile
industry. However, the amount of excess capacity is much less than the 1934 estimate of Edwin Nourse and appears to be much smaller than that described by many historians. In addition, the quarterly utilization rates suggest that there was some basis for increased expenditures on investment through the second quarter of 1929. It does not appear that “excess capacity” in the American automobile industry, and a resulting decline in investment, would do more than account for some downturn in economic activity in 1929.

Regardless of whether there was excess capacity in the late twenties, investment in capital to produce automobiles might have declined if the automobile industry saw, or suspected, that the market for automobiles was “saturated” by the late twenties. Mercer and Morgan also evaluate this proposition. They find that saturation in the level of the stock of automobiles did not exist prior to 1930. However, in dynamic terms using either total cars or cars per capita, saturation in the growth of the demand for automobiles existed from 1924 on (except for the single year of 1928) confirming the tendency toward saturation in the automobile market. However, these results have been questioned by two Canadian economists, P. J. George and E. H. Oksanen, who argue that Mercer and Morgan use an inappropriate income variable and fail to consider the possibility of a structural change around 1929. Making these two corrections, George and Oksanen reexamine the saturation of the American automobile market in the late twenties and find that there were important shifts in the demand for automobiles after the start of the depression in 1929. When this is considered, there is then little evidence that the market for automobiles was saturated in the late twenties.

**Housing Overinvestment In The Twenties**

Many economists have suggested that the declining investment demand caused by falling housing construction was a more important factor in initiating the Great Depression, and that a “housing surplus” in the late twenties was a strong factor causing the decline in housing construction. In 1960 Bert Hickman and Richard Muth separately published studies that reached opposite conclusions on the existence of a housing surplus in the late twenties. In 1971 Ben Bolch, Rendig Fels, and Marshall McMahon produced a study that reexamines the concept of a housing surplus, and in 1973 Bolch and John Pilgrim undertook a more extensive analysis of the effects of housing investment in the twenties on the Great Depression. In the 1960 studies, Hickman had found that the housing stock was “excessive” by the late twenties, whereas Muth had found that the actual stock of housing was below the desired stock in 1929, 1931, 1932, and 1935. Bolch, Fels, and McMahon say that Hickman was discussing too many housing units for the number of families, whereas Muth was discussing underbuilding in the sense of too little real investment per individual house, and their own study attempted to see which effect was more important.

Real gross housing expenditures on new units can be separated into two multiplicative components; housing starts and the real average investment per unit. After estimating the determinants of each, Bolch, Fels, and McMahon examined the building hypothesis by simulating what would have been the case in each situation if a “normal” situation had occurred. The result was overbuilding in the number of housing units from 1922 through 1929. They also found “overinvestment” per individual housing unit from 1921 to 1928 and conclude that excessive housing starts relative to family formation led to the sharp decline in housing construction at the end of the twenties. In their view the demographic changes in the twenties were the proximate causes of this. In a 1978 study Clarence Barber was even more emphatic on the importance of demographic changes, suggesting that these were the ultimate explanation of the Great Depression. Basing his explanation on a Harrod-Domar growth model, Barber argues that the rapid decline in the population’s growth rate disturbed that equilibrium, and this rapid fall in the natural growth rate initiated the depression. The importance of the falling housing investment in relation to the Great Depression was further stressed in Bolch and Pilgrim’s 1973 study. Using their econometric model of the American economy from 1919 to 1938, they conclude that the basic cycle of the interwar period was traceable to the residential construction cycle in the twenties, and this cycle was largely explained by demographic changes, specifically the declining birth rate of the domestic population and the sharp decline in immigration due to the passage of immigration laws.

Lloyd J. Mercer and W. Douglas Morgan argue that when Bolch, Fels, and McMahon undertook their estimates, they incorrectly omitted the separate effect of the stock of houses on housing starts, and this led to their finding of excessive gross construction expenditures on housing. Mercer and Morgan find overbuilding of houses for 1921 through 1925 but a deficiency of housing starts, or underbuilding, for 1926 through 1933. Using the alternative static and dynamic saturation framework that they had developed, they find that dynamic saturation of the housing market began in 1926 and lasted until 1934.
Peter Temin holds that their conclusion as to the importance of the residential construction cycle (and by direct association, population growth) followed from the structure of the model they used.\(^2\)

He argues that ultimately the rate of family formation was the primary variable determining income movements in the model. However, if the rate of family formation also depended on income, then the decline in housing construction and the onset of the contraction could not be traced directly to population growth through the rate of family formation. A 1973 study by Bert Hickman found that the depression reduced housing construction by slowing the rate of family formation.\(^3\)

Barber argues that Temin did not interpret Hickman’s model and results correctly.\(^4\)

However, this raises questions about the causal relationship between housing construction and the onset of the Great Depression. In addition, given the earlier declines in housing construction, it is not clear why this did not initiate a contraction prior to the late spring of 1929.

**An “Unexplained” Decline In Consumption Spending**

The most recent candidate for the cause of the spending decline that initiated the Great Depression is an “unexplained” decline in consumption spending in 1930 that Peter Temin has discovered.\(^5\) He finds that the decline in consumption spending from 1929 to 1930 was much larger than declines in the similar contractions of 1920-21 and 1937-38. Because he could find no satisfactory reason for the bulk of the decline in consumption spending from 1929 to 1930, Temin was forced to conclude that the Great Depression was started by an unexplained decline in consumption spending.

Frederick S. Mishkin’s 1978 study provides a potential explanation for the large decline in consumption spending in 1930.\(^6\) His “liquidity hypothesis” stresses the importance of the composition of the household balance sheet. During financial distress, households will prefer to have more financial assets compared to tangible assets. Rapid sales of tangible assets generally result in wealth losses. The likelihood that a household will suffer financial distress increases in direct proportion to the degree of the indebtedness and the relative percentage of holdings of tangible assets compared to financial assets. Consumers built up an unusually large indebtedness prior to the stock market crash. With the crash and falling prices during the contraction, the real value of household asset holdings fell and continued to fall through 1932. As a result household balance sheets deteriorated noticeably. Mishkin suggests that much of the drop in consumer spending on housing and other tangible assets can be explained by the changes in household balance sheets.

Temin and Mishkin’s theory of a consumption spending decline as the cause of the Great Depression may explain too much. Thomas Mayer argues that Temin should have compared 1930

| TABLE 7.1 PERCENTAGE CHANGES IN REAL MACROECONOMIC VARIABLES |
|------------------|------------------|------------------|
|                  | 1920-21          | 1929-30          | 1937-38          |
| GNP              | -2.4 (-0.4)      | -8.9 (-11.1)     | -5.4 (-5.2)      |
| CONSUMPTION      | +6.4 (+10.8)     | -5.4 (-8.1)      | -1.6 (-2.6)      |
| INVESTMENT       | -41.7 (-62.4)    | -35.6 (-42.7)    | -53.1 (-58.6)    |
| EXPORTS          | -14.2 (-32.5*)   | -19.1 (-8.4*)    | +1.7 (+147.2*)   |
|                  | 1919-21          | 1928-30          | 1936-38          |
| GNP              | -3.5 (-2.0)      | -3.4 (-4.9)      | +2.3 (-0.1)      |
| CONSUMPTION      | +11.6 (+14.5)    | -0.2 (-2.8)      | +2.1 (+0.4)      |
| INVESTMENT       | -30.7 (-41.0)    | -27.8 (-28.4)    | -26.4 (-29.8)    |
| EXPORTS          | -19.7 (-78.0*)   | -15.7 (-26.3*)   | +29.8 (+234.7*)  |

*Source:* Peter Temin, *Did Monetary Forces Cause the Great Depression?* (New York: W. W. Norton, 1976), table 6, p. 64; and Joseph Swanson and Samuel Williamson, “Estimates of National Income and Product for the United States Economy, 1919-1941,” *Explorations in Economic History* 10 (Fall 1972). The data not in parentheses are from Temin. The data in parentheses are from Swanson and Williamson, with their current price data deflated by the implicit GNP price deflator. Investment and Exports refer to gross private domestic investment and merchandise exports, except for * which indicates that the figure is for percentage changes in the net exports of goods and services.
with all of the interwar years rather than just 1921 and 1938 and should also have examined the residuals from the estimated consumption function for each year during the period. In addition, Mayer suggests that Temin’s consumption function was badly specified. Making these adjustments Mayer reestimates Temin’s consumption function and finds that the results strongly reject Temin’s contention of an unusually large unexplained fall in consumption spending in 1930. Mayer also estimates two better specified consumption functions for the 1921-41 period. Estimates from these provided little support for Temin’s hypothesis. Mayer concludes that “the results can be best described as mixed, so that, as a whole, the evidence fails to establish Temin’s hypothesis.”

**The Magnitude of the Initial Decline**

Over the last several decades, an increasing number of economists and economic historians have expressed doubt as to the validity of the Keynesian explanation of the Great Depression. Recently Peter Temin has considered some aspects of this, and we can follow his examination. Temin compares the percentage changes in real GNP and its major components during the initial stages of each of the three interwar contractions. Table 7.1 contains Temin’s data as well as, in parentheses, the more recent interwar GNP estimates of Joseph Swanson and Samuel Williamson. If investment spending is the key to the contractions, then this suggests that the 1920-21 contraction should have been more severe than that of 1929-30.

Table 7.2 uses Swanson and Williamson’s data to examine the proportion of the change in investment in various components for these periods. In both the 1920-21 and 1937-38 (or 1919-21 and 1936-38) contractions, the bulk of the investment decline was in inventory investment, and the smallest was in construction; in 1929-30 (or 1928-30), the bulk of the decline was in construction.

Temin suggests that the decline in construction activity from 1929 to 1930 might have “depressed expectations of investment in 1931 more than an equivalent fall” in inventory or equipment investment. But he admits that it is difficult to know how large this effect would have been if it actually existed. Temin’s reasoning is that construction tends to be more stable than inventory investment. If it is low in one year, then individuals will generally expect it to be low in the next year. Thus, the relatively large decline in construction may have more sharply depressed expectations.

Beyond this, the Keynesian model does not suggest that income will decline more (or less) from a decline in construction as compared to an equivalent

### Table 7.2 Proportion Of The Change In Investment In Various Components

<table>
<thead>
<tr>
<th></th>
<th>1920-21</th>
<th>1929-30</th>
<th>1937-38</th>
</tr>
</thead>
<tbody>
<tr>
<td>CONSTRUCTION</td>
<td>0.06</td>
<td>0.42</td>
<td>0.10</td>
</tr>
<tr>
<td>EQUIPMENT</td>
<td>0.19</td>
<td>0.22</td>
<td>0.28</td>
</tr>
<tr>
<td>INVENTORIES</td>
<td>0.75</td>
<td>0.36</td>
<td>0.62</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>1919-21</th>
<th>1938-30</th>
<th>1936-38</th>
</tr>
</thead>
<tbody>
<tr>
<td>CONSTRUCTION</td>
<td>-0.09</td>
<td>0.82</td>
<td>-0.33</td>
</tr>
<tr>
<td>EQUIPMENT</td>
<td>0.20</td>
<td>0.18</td>
<td>0.28</td>
</tr>
<tr>
<td>INVENTORIES</td>
<td>0.89</td>
<td>0.00</td>
<td>1.06</td>
</tr>
</tbody>
</table>

The Absence of Equilibrating Adjustments
Declining housing construction in the late twenties due to overbuilding and declining investment in the automobile and related industries due to market saturation and excess capacity in the late twenties are the usual Keynesian explanations for the autonomous spending decline that set off the contraction; however, the studies cited above raise some doubts as to the importance of these explanations.

Certain other aspects of this should also be considered. If the overbuilding of housing was reducing the demands for housing at the same time as the costs of constructing housing were rising, this indicates that those resources were more valuable in alternative uses. If the automobile market was saturated, then there must have been other desirable goods that the consumers were turning toward. The role of the price system is to communicate this information to the individual transactors so as to promote these equilibrating adjustments in resources and the production of goods and services. It has often been pointed out that the Keynesian model tends to ignore this by proposing that if a component of aggregate demand declines, such as investment, this will cause, via the income multiplier, an associated decline in most other components of aggregate demand. In this scenario the equilibrating role of the price system is not mentioned.

There is evidence, however, of wide variations in the behavior of various prices during the depression. Wheat prices could not cover the costs of harvesting some of the wheat crop; sheep prices were less than the costs of shipping sheep to market; cotton prices could not purchase enough food to keep pickers in the fields. The evidence indicates that there were massive disruptions in the price system and that prices failed to adjust rapidly enough to simultaneously clear the individual markets. One reason that such a disruption could occur is that there were large negative shocks (large sharp decreases) in the stock of money. It is this Monetarist explanation of the Great Depression to which we now turn.

The Monetarist Explanation

The Monetarist explanation of the Great Depression is mainly concerned with explaining why a “normal,” albeit relatively severe, contraction was transformed into the catastrophic depression and deflation from the end of 1930 through the first quarter of 1933. In this explanation a series of panics and banking failures initiated a sharp and continuing reduction in the stock of money. As individuals and firms attempted to restore the desired money balances, the decrease in spending caused production, employment, incomes, and prices to begin to fall. As long as the contraction of the stock of money continued, the real and nominal levels of economic activity also had to contract. Once the stock of money stopped falling at the end of the first quarter of 1933, the contraction of economic activity also ceased.

Ben Bernanke has extended this explanation by examining nonmonetary effects of the financial crisis. Bernanke argues that the financial crises during the 1929-33 period reduced the ability of financial institutions to function as credit intermediaries and raised the costs of credit intermediation. This was particularly damaging to smaller borrowers as loans were recalled, and new loans often became unavailable. Larger firms, which might have borrowed, were reluctant to do so. Because of the rise in the costs of borrowing relative to the “safe” rate on savings, Bernanke asserts that this unambiguously reduced the demands for current-period goods and services. He believes that these nonmonetary financial effects are important in explaining the length and depth of the Great Depression.

The Onset of the Depression in 1929

Peter Temin and Anna Schwartz have recently suggested that the Fed’s restrictive monetary policy in 1928 and 1929, adopted to dampen the stock market boom, initiated the contraction. Alexander Field has found that the growth of the stock market from the mid-twenties to 1929 significantly increased the transactions demand to hold money balances by brokers and their customers. Because Fed authorities failed to recognize, this they engaged in policies designed to dampen or break the boom in the stock market. As the amount of money balances demanded increased, along with the prices of stocks and the volume of stocks exchanged, the resulting liquidity crunch caused sharp increases in interest rates, as can be seen in Figure 7.2. These interest rate increases were an important factor in causing a further decline in housing construction in 1928 and, after April of 1929, a decline in sales of automobiles. He suggests that the Federal Reserve System might not have persisted in its antispeculative monetary
policy if it had recognized the greater demand for money balances. As it was, the policy was continued until the crash at the end of October of 1929.

The Effects of Panics and Bank Failures
By fall of 1930, the contraction was already noted as being quite severe. However, according to the Monetarists the contraction’s character changed in November and December of 1930. The reason for this can be seen in Figure 7.3, which presents Friedman and Schwartz’s monthly deposit-to-reserve ratios, deposit-to-currency ratios, and high-powered money stock for the period of 1928 through 1934. After being roughly constant through 1928, 1929, and most of 1930, the deposit-to-reserve and deposit-to-currency ratios begin to fall in late 1930. The deposit-to-reserve ratio fell through 1935, though at a somewhat slower rate in 1934 and 1935. The deposit-to-currency ratio fell sharply in 1931, much more slowly in 1932, and then began to rise in early 1933. As individuals and firms held more currency relative to deposits by converting their deposit balances into currency, this depleted banks’ reserves and caused a reduction in the total stock of money. As banks increased their reserve ratios, they decreased lending and deposit creation to build up the, reserves causing a further fall in the stock of money. Therefore, the onset of falling deposit-to-reserve and deposit-to-currency ratios in late 1930 initiated huge declines in the stock of money. Though the Federal Reserve System began to increase the high-powered money, the increases were far too small to stop the money stock from falling.

The Monetarists contend that the rising bank reserve ratios and the public’s increased holdings of currency relative to deposits indicated a general “liquidity scramble” initiated by the onset of banking panics and failures. The first panic in November of 1930 is crucial because it set off the liquidity scramble by the banks and the public and established the Federal Reserve System’s failure to act to stop the liquidity scramble then and later in the contraction. As the stock of money fell due to the liquidity scramble and with a stable demand for money balances, this caused individuals and firms to reduce spending to restore the desired money balances. The spending reductions, initiated by the fall in the stock of money beginning at the end of 1930, explain the extraordinary length and severity of the contraction. In addition, the recurring panics and bank failures brought into question the stability and soundness of the entire economic system, further depressing economic activity.

In general, the Monetarists contend that this was the result of inappropriate monetary policy by
the Federal Reserve System. If the Fed had vigorously pursued an expansionary monetary policy, then the stock of money need not have fallen with the liquidity scramble by the banks and public. Furthermore, if the money stock had been expanded sufficiently to meet the increased liquidity demands when the initial panics and rash of bank failures began occurring in late November of 1930, then the banks and the public would never have engaged in such a massive liquidity scramble, and the increase in the high-powered money which would have been necessary to stop the stock of money from falling would not have been as large. In this sense then, considering the period from the end of 1930 through the second quarter of 1933 as the Great Depression, the Monetarists contend that this was due to inept monetary policy by the governors of the Federal Reserve System.

Could the Fed Have “Put Out the Fire” in Late 1930?

According to the monetarists the first banking crisis in November and December of 1930 ignited the liquidity scramble and the decline of the money supply and set off the spiraling contraction of economic activity. This was due to the failure of the Fed to stop the banking crises, particularly that of 1930, which reduced the likelihood that they would act to stem later financial crises. Brunner and Meltzer imply that if the Fed had not allowed the money stock to decline during 1930, then the first banking crisis would not have occurred or would have been less severe.

For the Fed to have stemmed the first liquidity crisis, it would have had to provide the banks experiencing the panic with whatever currency was necessary to satisfy the depositors’ demands. The Fed had two methods to do this; rediscounting or open market operations. The rules of member bank rediscounting at that time required that banks rediscount short-term, “self-liquidating” commercial paper. Industrial bonds, government securities, stocks, and longer term assets of the member banks were not eligible. Rediscounting was used for very short periods of time, and nonmember banks could not rediscount except indirectly through a correspondent bank that was a member of the Fed. The localized nature of the banking crisis in November of 1930 (as shown later) and these rediscounting restrictions raise doubts as to whether the Fed would have been able to stem the late November banking crisis by providing the additional liquidity for the banks experiencing the deposit-to-currency conversions.

Open market operations were handled by the New York City Federal Reserve Bank. Purchases of...
securities on the open market did not immediately increase the money stock. More importantly, for open market purchases to have stemmed the first banking crisis, the purchases would have had to have been made with Federal Reserve System currency and directed only to those banks in the southeastern part of the United States where the banking crisis was occurring. It is doubtful if this could have been accomplished.

Recent research also suggests that a more expansionary monetary policy in 1930 prior to the banking crisis would not have prevented the crisis. An examination of the balance sheets of these national banks found that most were small and in agricultural areas. An examination of the balance sheets of these national banks found that through 1930 most of the failures could be predicted a year in advance, because failing banks usually had large amounts of poorly performing loans, many from the teens and twenties. The economic contraction in 1929 and 1930 led to sharp declines in the prices of agricultural crops, and farmers and local merchants were often unable to repay their loans.

Elmus Wicker also found that the 120 banking failures in the southeast in the last two weeks of November and the renewed failures in December were all linked to the failure of the large Nashville investment banking house, Caldwell and Company, and this “pinpoints the ‘contagion of fear’ that spread among depositors.” Caldwell and Company controlled the largest chain of banks in the south, with assets of over $200 million; the largest insurance group in the region, with assets totaling $230 million; and had controlling interests in other banks, insurance companies, industrial enterprises, investment trusts, and newspapers. From the middle of 1929 on, the Caldwell enterprises were in financial difficulty. In June 1930 they merged with a Kentucky bank-holding chain, Banco Kentucky, which also had financial problems. Apparently each thought the other would be its rescuer.

The failure of Caldwell and Company set off the crisis. In Arkansas, 54 of the suspended banks belonged to the A. B. Banks chain, whose stock was owned by the Caldwell-owned Home Insurance Co. The American Exchange Trust, the key chain bank, had lent the parent company $100,000 in the summer of 1930. When Caldwell and Company failed in November, a run began on the American Exchange Trust, and in four days it lost $4 million of deposits out of $15 million total. The bank closed, and within days the other 54 banks in the chain followed suit. In Kentucky, the 15 or more banks that closed were affiliated with Banco Kentucky. In Tennessee, 10 of the bank failures were tied to Caldwell and Company, and the 15 bank failures around Asheville, North Carolina, had business arrangements with Rogers Caldwell and his Bank of Tennessee. Therefore, it is doubtful that a more expansive monetary policy in 1930 prior to the panic—or quick action by the Fed during the panic—could have stopped the bank failures in the last half of November of 1930.

Friedman and Schwartz also place considerable emphasis on the December, 1930, failure of the Brooklyn-based Bank of the United States. They contend that this failure unduly diminished confidence in the banking system because of the bank’s official sounding name and that authorities allowed an essentially sound bank to fail. Temin takes issue with this and in his examination finds evidence that it was not a basically sound bank. White’s statistical analysis finds that the failure of the Bank of the United States did not “appear to have been unique or unexpected.” And in a 1985 study Joseph L. Lucia has found that at the time of its failure, the Bank of the United States was both insolvent and illiquid.

Wicker’s 1980 study casts more doubt on the significance of the failure of the Bank of the United States. If this failure was so destructive of confidence in the banking system, then the immediately following failures should have been in the region around New York City or at least geographically dispersed across the United States. Instead, the failures were concentrated in those areas where the November bank suspensions had occurred, the southeast. Wicker proposes that these failures were secondary effects from the November failure of Caldwell and Company. This suggests that the failure of Brooklyn’s Bank of the United States did not significantly accelerate the liquidity scramble.

**Criticisms of the Monetarists’ Explanation**

In 1976 Peter Temin led the way in a major reevaluation of the Monetarists’ explanation of the Great Depression. They contend that the turning point was November of 1930; the first banking panic and the failure of the Bank of the United States set off the liquidity scramble by the banks and the public. As the liquidity crisis accelerated, the second flurry of bank suspensions occurred between March and June of 1931. The falling deposit-to-currency and deposit-to-reserve ratios reduced the money supply and, through a stable money demand function, the stock of money. The falling stock of money led to faster price deflation and an accelerated decline in real output and employment.
A reduction in the supply of money relative to the demand for money will initially cause interest rates to rise and the prices of financial and real assets to fall, though given more time for adjustments to occur, interest rates will also fall. Temin wondered what assets the public will attempt to sell to restore money balances after the decline in the supply of money and found an answer in Friedman and Schwartz’s work.59 The effect of the reduction in the supply of money relative to the demand for money balances should be seen first in the interest rates or yields on those assets most like money, and they should rise soonerest and farthest.

Figure 7.2 presents some of the interest rates that Temin gathered to examine this. As can be seen, Temin found no evidence of rising interest rates in late 1930 or early 1931, the period directly following the initial banking panic. In 1919-20, 1928-29, the fall of 1931, and the first quarter of 1933 because of Fed monetary policy, the supply of money fell relative to the demand for money balances, and interest rates rose as predicted. Temin’s concludes that in late 1930 and early 1931, the demand for money balances fell more than the supply of money fell and that this effect led to the decline in the stock of money. The Monetarists’ explanation was, therefore, wrong.

Temin’s study was quickly answered by a flood of additional studies and criticisms, some supporting and some rejecting his contentions. Thomas Mayer was one of the first to examine Temin’s analysis.60 Though he criticizes a number of Temin’s points, he indicates that Temin has pointed out a particularly crucial problem emphasizing the failure of interest rates to rise as predicted. Anna Schwartz criticizes Temin’s analysis, suggesting that in 1930 and 1931 the supply of commercial paper was reduced as much or more than the demand for it.61 This explained the failure of commercial paper prices to fall and rates to rise. In addition, the reduction in the federal debt through December of 1930 increased the supply of loanable funds and reduced market clearing interest rates.

Temin and Peter Lindert both question Schwartz’s analysis.62 Schwartz asserts that commercial paper rates did not rise because banks used commercial paper to borrow currency from the Fed to meet the depositors’ currency demands, but Temin notes that the failure of rates to rise meant that “Any deficiencies in money balances were not communicated to the investing and consuming public and therefore could not affect spending plans.”63 Brunner and Meltzer had earlier noted that rates had not risen and suggested that this was due to an inflow of gold.64 A similar criticism applies because if the gold inflow stopped rates from rising, it must have supplied money to those banks, individuals, and nonfinancial firms who were attempting to build up reserves and restore money balances. Thus, the supply of money did not contract, and the Monetarists’ explanation of the cause of the Great Depression, a falling supply of money in late 1930 and early 1931, is contradicted.65

It should be noted that few economists and economic historians disagree with the suggestion that the sharply falling stock of money beginning in the late spring of 1931 and the sharply restrictive Federal Reserve System monetary policy initiated in the fall of 1931 explain the contraction from 1931 through mid-1932. The disagreement revolves around the reasons for the onset of the sharply declining stock of money at the end of 1930. If the decline in the money supply relative to the demand for money was so pronounced following the late November and December 1930 banking crises, why did interest rates not rise as they clearly did when the money supply was reduced in 1919-20, 1928-29, the fall of 1931, and the first quarter of 1933? This is the major point of critics who suggest that for the late 1930-early 1931 period the Monetarists have placed too much emphasis on the decline in the supply of money and inadequate emphasis on the decline in the demand for money.

The Monetarist explanation, like the Keynesian and Austrian explanations, is essentially domestic in nature. All three assume that the contraction began in the United States and through international trade and monetary flows spread to other nations. Yet there is a long tradition that the worldwide depression began in a number of countries simultaneously rather than solely in the United States. We now turn to the international explanations of the Great Depression.

Disruptions in the International Sector

The view that the worldwide depression of the thirties was transmitted through changes in international trade and international lending has a long history, and no one would argue with the idea that this, in fact, did occur. The question is whether the depression began in the United States and was then transmitted to other nations or whether the depression began simultaneously in a number of countries and was then transmitted to the rest of the world through, say, the gold exchange standard, changes in international lending, or in the mechanisms for reparations and loan payments that arose out of the First World War. President Hoover, at the end of his term, blamed international forces and did not believe that the depression began in the United States. H. W. Arndt was one of the earlier
advocates of the idea that international disruptions transmitted the depression around the world. More recently J. R. Hicks and Charles P. Kindleberger have restated this view.

To Kindleberger the international economic system in the twenties was unstable because there was no dominant nation to consciously take responsibility for stabilizing the international economic system. The depression affected so many countries and was so long and severe because Great Britain could not and the United States would not provide stability by "(a) maintaining a relatively open market for distress goods; (b) providing countercyclical long-term lending; and, (c) discounting in crisis."68

The United States became the world's leading capital exporter in the twenties. Unlike Great Britain in the nineteenth century, U.S. overseas investment was positively related to the level of domestic economic activity. As economic activity expanded in the U.S., our overseas lending increased, stimulating economic activity in the recipient countries, and, through increased American exports, economic activity in the United States. When the American economy declined, as it did in 1929, overseas investment fell, foreign countries' economies began to decline, and American exporting fell. Therefore, according to Kindleberger, this procyclical pattern of American long-term overseas lending spread the economic contraction that began in the United States and European countries in 1929 and accelerated the decline in economic activity.69

Neither the United States nor Great Britain took responsibility among leading countries to maintain open markets for "distress goods."70 In fact, the United States closed markets by passing the Smoot-Hawley tariff, which brought forth retaliation leading into a contractionary spiral of world trade. With an independent world depression in agriculture in the twenties, this accelerated the declines in aggregate demand. Kindleberger maintains that if Great Britain and particularly the United States had led the way in maintaining relatively open markets for distressed goods, especially agricultural products, the decline in world trade would have been smaller and the depression shorter and less severe in all of the affected countries.

Finally the U.S. and British governments failed to discount during the banking crisis in Europe in 1931.71 The onset of a run at Austria's Kreditanstalt on May 11, 1931, elicited loans from various central banks. However, Kindleberger argues that delays in putting the loans together and the "niggardliness" of the sums doomed the bank and caused the banking crisis to spread across Europe. Great Britain's withdrawal from lending and the United States' failure to do so ultimately drove Great Britain off the gold standard, causing the collapse of the gold exchange standard, another huge negative shock to international trade. Therefore, according to Kindleberger, at the end of the twenties and the beginning of the thirties, these economic conditions and failures in world economic leadership by Great Britain and the United States combined to bring about the worldwide Great Depression.72

Many scholars disagree with Kindleberger’s contentions. Heywood Fleisig agrees that the world depression began with a decline in U.S. long-term international lending in 1928 and 1929 that was aggravated by a decline in American imports in 1930 and 1931.73 However, he argues that both were domestic in origin. The decline in international lending was the result of the booming stock market and the restrictive monetary policies adopted by the Federal Reserve System. Christian Saint-Etienne comes to a similar conclusion.74 Through its sheer size, the American economy transmitted its business cycles to the world through international markets, and he notes that this had happened in the 1920-21 depression. The cause of the initial downturn in 1929 was tight monetary policy, and the Hawley-Smoot tariff was the main factor in the change from a normal business cycle to an extraordinarily severe one.75

The importance of the Smoot-Hawley tariff is one point upon which many Keynesians, Monetarists, and Austrians agree. The imposition of this wedge between international transactions set off a spiraling contraction of world trade and economic activity. However, recently Charles Kindleberger and Barry Eichengreen have objected to this assessment, pointing out that the influence of the tariff has never been carefully assessed.76 As Figure 7.4 shows, the tariff rates rose more sharply with the 1922 Fordney-McCumber tariff, and the percent of imports that was duty-free dropped as sharply. But there was no depression after this tariff. What made the 1930 Smoot-Hawley tariff seem so much worse after its passage was that dropping prices combined with specific tariff rates to make ad valorem rates rise dramatically and contributed to a sharp decline in the value of imports. (See Figure 7.5.) But these declines in prices and quantities may simply reflect the effect of the worldwide depression on international trade rather than the effect of the tariff in causing the worldwide depression.

The tariff increased the domestic demand for domestic production, and this should have expanded production. The reduction in imports should have caused the prices of imports to decline somewhat.77 Under fixed exchange rates and the gold standard, the reduction in dollars earned by foreign producers will
not necessarily cause a quick reduction in American
exports because foreign buyers can simply buy
American exports with gold rather than the dollars
received from American importers. Therefore, the
tariff may have expanded aggregate demand in the
United States.

As Eichengreen notes, there are two offsets
to this. First, the tariff marginally substitutes less
efficient domestic production for more efficient
foreign production, and this will be contractionary,
though Eichengreen doubts that it was quantitatively
important. The most important effect is tariff and
quota retaliation by foreign nations, a point many
economists have stressed. Though there was some
explicit retaliation, such as the Wais Tariff of Spain,
Eichengreen argues that this has been overstated.
Between 1925 and 1929, 26 European countries
revised their tariffs 33 times, and Latin American
nations revised their tariffs 17 times. He points out
that “In 1927 and 1928 Australia, Canada, and New
Zealand all instituted broad upward revisions in tariff
rates.” Italy’s increase in tariffs on American
automobiles was started several years earlier in order
to aid Fiat, whose sales had slid 30 percent in three
years. In the case of the Eastern European nations’
tariff increases and Britain’s 1932 general tariff
revision, retaliation is never mentioned as the reason.
Eichengreen argues that there is no clear evidence
that Smoot-Hawley was a major factor in creating
and spreading the worldwide contraction.

There is another process by which the tariffs
effect economic activity. Rather than looking at
aggregate effects, consider the individual markets.
The tariff changes the demands for and supplies of
the goods protected and exported, and this requires
reallocations of resources and changes in numerous
relative prices. For example, consider the automobile
industry. The United States exported over 670,000
cars in 1929. With the new protectionist tariffs on
American autos in Italy, Eastern Europe, and other
countries, automobile exports quickly fell,
exacerbating the production declines due to the
contraction. Because resources cannot
instantaneously and costlessly be reallocated and
prices don’t instantaneously adjust, this disruption to
the price system transitionally reduced economic
activity and magnified the contraction already
underway. Thus, Smoot-Hawley’s effects were not
only aggregate ones.

Eichengreen believes the most important
effect of Smoot-Hawley was to redistribute “gold
from countries with weak reserve positions to a
country whose position was strong, thereby
undermining the system’s resistance to exceptional
destabilizing events.” And this is part of the
contention that the worldwide depression can be
traced to the gold standard and the First World War, an old position that has gained new followers in recent years. 81

The gold standard was suspended, not abandoned, during the First World War. In none of the developed countries was there ever any serious thought given to not returning to the gold standard after the war, but the reconstitution of the gold standard in the twenties was far from easy. The United States had held 26.5 percent of the world’s gold stocks in 1913, a figure that rose to 40 percent in 1918 and 44 percent in 1923. 82 By the end of the twenties, France and the United States held 60 percent of the world’s gold stocks. 83

Because there was a flood of paper money issues in all of the belligerent countries, price levels everywhere were much higher after the war than before. Returning to the gold standard imparted a deflationary bias everywhere except in the United States, which had excess gold reserves. The resumption of the gold standard in Britain in 1925, at the prewar price of gold, immediately ran into economic difficulties. Peter Temin observes that “The British economy was plunged into social conflict, symbolized by the General Strike of 1926, over the allocation of the burden of a high pound.” 84

The Bank of England had to maintain higher discount rates to keep interest rates higher and attract capital to offset the British trade deficit, and this had adverse effects on production, investment, and unemployment. In December of 1926, France effectively returned to a gold standard and formally did so in June of 1928, but prior to that time the franc had been reduced to one fifth of its prewar value, causing it to be undervalued. All countries adopted contractionary policies to stay on the gold standard. One means of doing so was to withdraw all gold coins from circulation to be used as bullion for central banks’ monetary reserves. Only the United States continued to allow gold coins to circulate. As a result the international market for gold was dominated by central banks.

Germany, worried about stock market speculation and the maintenance of the gold standard, took steps to halt capital inflows in 1926. Between 1919 and 1931 gross capital inflows were over 5 percent of German national income, and net capital inflows were 2 percent of national income. The tax exemption status of foreign holders of German bonds was withdrawn, and the Reichsbank adopted the deflationary policy of reducing discounting by correspondent banks and inducing those banks to reduce their discounting for customers. The German stock market crashed in 1927, and by 1928 the German economy was sliding into a depression. 85

Gene Smiley, The American Economy in the 20th Century, Chapter 7: Page 7-16: REVISED 5-12-93
In the last half of the twenties, to assure the financial community of their commitments to the gold standard, many countries raised their legal gold reserve requirements. This increased the demand for gold and had a deflationary effect across the globe.\textsuperscript{86}

In 1927 the Fed lowered interest rates to reverse the gold flow from Britain to the United States and ease British difficulties in maintaining the gold standard. However, most of the gold actually went to France, not Great Britain. As gold left the United States from May 1927 to June 1928, the Fed became worried and changed policies to raise rates and reverse the gold flow. This caused the credit crunch of late 1928 and 1929 and helped bring on the contraction and American stock market crash.

As the contraction in the United States, Great Britain, France, Germany, and other countries worsened, the governments adopted deflationary policies because that was what the rules of the gold standard required. They desired to see prices and wages fall, and they built up gold reserves. Between 1928 and 1932 the gold reserves of 24 central banks increased by $2.4 billion, $2 billion for the Bank of France alone.\textsuperscript{87} Governments were hamstrung by their commitment to the gold standard and could take no fiscal or monetary policy actions that were contrary to the deflationary rules implicit in the gold standard. Temin describes the situation this way.

The Atlantic economy, in other words, was in the grip of deflationary policies at the end of the 1920s. Each national story differs, but they were all reflections of an underlying theme. The gold standard had been revived, but the conditions that had sustained it before the war no longer existed. The pound was overvalued; the franc, undervalued. Both the Americans and the Germans were trying to stamp out “speculation.” The result was that government policies everywhere were set to discourage economic activity.\textsuperscript{88}

It was the attempt to preserve the gold standard that produced the Great Depression. Great Britain left the gold standard in September of 1931, and its depression ended. France, with its huge reserves of gold, continued on the standard until 1936. The United States effectively began to back off from it with Roosevelt’s new policies and, for all practical purposes, left the gold standard as it had been constituted when the dollar was devalued from $20.67 to $35 an ounce of gold in January of 1934 and private U.S. ownership of gold was prohibited.

The Great Depression: An Assessment

We have traveled a long and difficult path through a dense thicket of explanations for Great Depression of 1929-33, a path laden with theory and critical points and distinctions. And we have concluded with an explanation contending that the fundamental force or impulse bringing on the Great Depression was the gold standard and changes wrought by the First World War. Because we concluded with this explanation, does it mean that it is the correct one? Not necessarily. The explanation still leaves considerable room for an important role for other institutional changes and longer term growth dynamics. It is not likely that those economists who contend that the Great Depression was started in the United States and spread to the rest of the world, and there are many, will be convinced. Was the gold standard at fault, or was it the manner in which countries adopted the reconstituted gold standard, the gold exchange standard, that was at fault? In other words, how much was due to the institution of the gold standard, and how much was due to the policies nations chose when adopting the gold standard?

Finally, why is it that we cannot determine the primary reason for the Great Depression? Facts alone tell us nothing without some theory to guide their organization and selection. In the case of the Great Depression, facts alone cannot discriminate among alternative explanations. Theories allow us to select among the overwhelming mountains of available evidence, but not all theories select the same facts, and the explanation for the same fact often differs between theories. Perhaps one day we will have a better theory of aggregate economic activity that will allow us to come up with a comprehensive theory that weighs the importance of these factors in the Great Depression, but that day has yet to arrive.

Selected References


George, P. J. and Oksanen, E. H. “Saturation in the Automobile Market in the Late Twenties: Some Further Results.” Explorations in Economic History 11 (Fall 1973): 73-86.


“Internal Funds and Automobile Industry Investment: An Evaluation of the
1. The extent of this disagreement can be easily seen in the papers by Paul A. Samuelson, Charles P. Kindleberger, and Milton and Rose D. Friedman in *The Journal of Portfolio Management* 6 (Fall 1979): 7-21.

2. There is no agreed upon standard to distinguish between a depression and a recession. Prior to the Second World War, economic contractions were generally called depressions. Perhaps to distinguish postwar contractions from these earlier ones, they were called recessions rather than depressions, and depression was reserved for a much more serious downturn. By the standards of contractions in the interwar period, the recession of 1981-1982 would have been called a depression.

3. This is indicated by the existence of frictional unemployment.

4. This discussion draws upon the writings of Friedrich A. Hayek. In particular, see the essays in *Individualism and Economic Order* (Chicago: University of Chicago Press, 1948). A more recent presentation of this can be found in Gerald P. O’Driscoll, Jr., *Economics as a Coordination Problem: The Contributions of Friedrich A. Hayek* (Kansas City: Sheed, Andrews, and McMeel, Inc., 1977).

5. These, of course, do not exhaust the explanations of the Great Depression, but they are the ones that have for a long time been given reasonable consideration. Many early explanations are no longer fashionable. For example, “underconsumption” explanations argued that the growing inequality in the distribution of income during the 1920s meant that consumers had less and less income to buy the increasing amount of goods and services produced, and these decreasing sales brought on a decline in production and investment. The Roosevelt administration attempted to deal with this by changes in the tax laws aimed at redistributing income from those with more to those with less. Another explanation fashionable in the mid-1930s argued that increasing monopoly power allowed firms to “administer” prices. These administered prices did not fall when there were demand decreases; rather, employment declines occurred, leading to falling incomes and further decreasing demand. The Roosevelt...

Austrians term this an inflationary increase, however, they use define inflation as an increase in the fiat money not backed by commodity money (gold or silver). The common meaning of inflation is a falling value of money relative to commodities or services, or, usually the same thing, a rising money price of commodities and services.


These criticisms involve both the specific theoretical model the Austrians have developed as well as its application to the Great Depression.


16. The interaction of the spending multiplier, from the consumption function, and the investment accelerator, from the investment function, leads to cyclical expansion and contractions of economic activity. It is generally called the multiplier-accelerator mechanism.


20. To do this, they formulated and estimated an investment demand function based upon a partial adjustment of the actual stock of capital toward the desired stock of capital. From this they could then estimate, given the capital, the capacity output.

21. They formulated a definition of market saturation and they chose to examine the desired and actual stock of automobiles. State saturation exists when the desired stock of automobiles in a year is less than or equal to the stock in the previous year. Dynamic saturation exists when the rate of growth of the desired stock of automobiles this year is less than or equal to the rate of growth of the actual stock of automobiles in the previous year. Mercer and Morgan believe that the desired versus actual stock of automobiles was the preferred way to measure saturation. The desired stock of automobiles was considered to be a function of the relative price of automobiles, real disposable income, and a time trend to capture other effects not explicitly included such as the growing availability of credit (installment financing), growing stock of household wealth, improved roads and service facilities, and so forth. Mercer and Morgan, “Alternative Interpretations of Market Saturation.”

22. P. J. George and E. H. Oksanen, “Saturation in the Automobile Market in the Late Twenties: Some Further Results,” *Explorations in Economic History* 11 (Fall 1973): 73-86. They argued that the appropriate income variable in “stock adjustment models” is not current real
income but the expected real income, because durable goods yield their services over a period of time. Second, George and Oksanen suggested that Mercer and Morgan should have examined the possibility of a structural break around 1929.


25. The real average investment per unit was assumed to depend upon rents relative to construction costs, permanent real income per family, and the interest rate. Housing starts were assumed to be determined by the relative shortage or abundance of housing as indicated by the ratio of total nonfarm families to total available housing units and by real construction costs.

26. For overbuilding in terms of too many housing units they let the ratio of nonfarm families to housing units available take its mean value for the 1920s and allowed the real price of housing to vary as it actually had.


28. Lloyd J. Mercer and W. Douglas Morgan, “Housing Surplus in the 1920’s? Another Evaluation,” *Explorations in Economic History* 10 (Spring 1973): 295-304. Housing starts were a function of the ratio of nonfarm families to available housing and the real cost of new housing. When they held the ratio of nonfarm families to available housing constant in their simulation, the variation in housing starts could only arise from variation in the real cost of housing. Because this did not vary to any large extent, their simulation produced largely constant gross expenditures on housing during the twenties.

29. Temin, *Did Monetary Forces Cause the Great Depression?*, 45-7.


32. Temin, *Did Monetary Forces Cause the Great Depression?*, 69-83.


35. Temin, *Did Monetary Forces Cause the Great Depression?* 62-8.

36. These comparisons can only be suggestive because they are not accurate measures of the changes. The 1920-21 contraction began quite early in 1920 and sharply led to the trough in the first quarter of 1921, after which a relatively strong expansion began. The average values for all of 1920 underestimate the peak values reached, and the average values for all of 1921 overstate the values at the trough of the contraction. In 1929 there was an unusually strong expansion of real economic activity in the first half of the year. The contraction in the last half of 1929 does not seem to be as severe as the contraction after the peak in 1920. The contraction continued all through 1930, whereas the earlier contraction troughed early in 1921. Because there are more months of strong expansion in 1929 than 1920, the average values for 1929 will be relatively larger. And because the contraction continued all through 1930 whereas the 1920-21 contraction troughed early in 1921, the 1930 values will be relatively smaller than the 1921 values. Similar problems exist for the comparisons to the 1937-38 percentage changes. For these reasons Temin also examined percentage changes from 1919-21, 1928-1930, and 1936-38. These temper, but certainly do not eliminate, the data problems.

38. This should not be too surprising. The rapid inflation of 1918-19 induced firms to speculate and build up large inventories of raw materials and intermediate and finished products to gain from the rising prices. The Federal Reserve System’s rapid turn to a very restrictive monetary policy at the end of 1919 and beginning of 1920 set off the contraction and a rapid price deflation causing firms to undertake a massive liquidation of those accumulated inventories. In 1935-37 the pace of economic recovery quickened after the Supreme Court ruled the NIRA unconstitutional. Wholesale prices, which had fallen 9.2 percent per year from 1929 to 1933, rose 9.7 percent per year from 1933 to 1935, largely because of the NIRA. They rose 1 percent from 1935 to 1936 and 6.6 percent from 1936 to 1937. Firms were rebuilding inventories dramatically depleted by the depression, and the rising prices probably led to more inventory accumulation than otherwise would have occurred. When the Fed doubled member bank reserve requirements between August of 1936 and May of 1937, this initiated the 1937-38 contraction. Wholesale prices fell 9.3 percent from 1937 to 1938. The contraction and change from rising to falling prices led to sharp inventory liquidations. Wholesale prices were falling in the late twenties, and with falling prices firms generally had an incentive to minimize inventory investment. When the contraction began in the middle of 1929, firms did not have excessive inventories to be liquidated. Consequently, less of the declining investment in 1929-30 was in inventories as compared to the 1920-21 and 1937-38 contractions.


43. Friedman and Schwartz, A Monetary History of the United States.


48. White was limited to an examination of national banks that were members of the Federal Reserve System, whereas the majority of the small rural banks which failed were state banks. However, he argued that the small rural national banks that failed were good indicators of the problems besetting the state banks.


50. The key city bank in chain banking held much of the reserves for the smaller banks in the chain.


52. Ibid., 577, 579-80.


54. Temin, Did Monetary Forces Cause the Great Depression? 90-5.


58.  Ibid., 582. There were a number of closings in Mississippi and North Carolina. Forty percent of the Mississippi closings were within a 30-mile radius in the northeast corner of the state. The bulk of the North Carolina closings were again concentrated in western North Carolina, where the November failures were. The 12 banks in Arkansas that suspended did so “when the public learned that a director of the closed American Exchange Bank of Little Rock had an interest in eleven of the banks!”

59.  Temin, Did Monetary Forces Cause the Great Depression? 97-8. See also Milton Friedman and Anna J. Schwartz, “Money and Business Cycles,” Review of Economics and Statistics 45 (February 1963): 32-78. This scenario, which Temin called the “pebble in the pond” theory, indicates that those individuals who experience reduced money balances will sell financial assets—securities—which are closest to money in liquidity. As the prices of those assets fall, asset holders will begin turning to less liquid financial assets and, as these prices begin falling, to other financial assets, then nonfinancial assets, and finally begin reducing the purchases of current services.


64.  Brunner and Meltzer, “What Did We Learn From the Monetary Experience of the Great Depression?”

65.  Elmus Wicker in a study extending his previous one of 1980, found that in the southeast, where the late November and December 1930 banking crisis occurred, the evidence does not suggest that this initiated a sharp change in depositor and banker behavior that would have led into a decline in the supply of money. Elmus Wicker, “Interest Rate and Expenditure Effects of the Banking Panic of 1930,” Explorations in Economic History 19 (October 1982): 435-45.


69.  Ibid., 292-93.

70.  Ibid., 292-94.

71.  Ibid., 148-52.

72.  Ibid., chapter 14.


75.  Ibid., 18-20, 48, and 55.

77. In other words, the United States was too large in international trade for the small country assumption. A decrease in the American demand for imports would significantly effect the foreign producers' demand curves and cause prices to fall because of the demand reduction.

78. If the terms of the gold standard were adhered to, as these countries lost gold to the United States to purchase American exports the money stock would have fallen, causing prices to fall, reducing the demand for American exports. But this does not happen quickly. So temporarily, *without retaliation*, the demand for American exports should have been little affected.


80. Ibid., 30.


83. Temin *Lessons*, 22-23.

84. Ibid., 19.