

This PDF is a selection from an out-of-print volume from the National Bureau of Economic Research

Volume Title: Business Cycles: Theory, History, Indicators, and Forecasting

Volume Author/Editor: Victor Zarnowitz

Volume Publisher: University of Chicago Press

Volume ISBN: 0-226-97890-7

Volume URL: <http://www.nber.org/books/zarn92-1>

Conference Date: n/a

Publication Date: January 1992

Chapter Title: Recent Work on Business Cycles in Historical Perspective

Chapter Author: Victor Zarnowitz

Chapter URL: <http://www.nber.org/chapters/c10373>

Chapter pages in book: (p. 20 - 76)

2 Recent Work on Business Cycles in Historical Perspective

2.1 Introduction

Interest in business cycles is itself subject to a wavelike movement, waxing during and after periods of turbulence and depression, waning in periods of substantial stability and continuing growth.¹ At time, confidence in government institutions and actions persuaded many that cyclical instability had ceased to be a serious problem. Thus in the early heyday of the Federal Reserve System, 1922–29, monetary policies were expected to help maintain prosperity. In the 1960s, the late heyday of Keynesian economics, fiscal fine-tuning evoked similar hopes.

The present is another time of disillusionment—now extending to both types of stabilization policy. The sequence of serious worldwide recessions in the last decade soon refuted the perennially attractive idea that business cycles had become obsolete. Beyond that, the credibility of both Keynesian and monetarist explanations has diminished. Once again, the apparent failure of old solutions prompts the profession to pay more attention to the continued existence of business cycles.

Reprinted from the *Journal of Economic Literature* 23 (June 1985): 523–80.

The author thanks Alan Blinder, William Fellner, Stanley Fischer, Milton Friedman, Robert Gordon, Robert Hall, Bert Hickman, Wilhelm Krelle, David Laidler, Ronald McKinnon, John Taylor, and two anonymous referees for many helpful suggestions and criticisms of the first outline and earlier drafts of this paper. Also gratefully acknowledged is the financial and intellectual support received from the National Bureau of Economic Research and the Graduate School of Business of the University of Chicago. Naturally, responsibility for the expressed opinions and any remaining errors is exclusively the author's.

1. This is well illustrated in the early literature, which focused on the episodes of commercial crises, but it is also reflected in the timing of later, classical studies of the nature and causes of business cycles at large. There is little doubt about the impetus provided in this context by the major depressions of the late 1830s, 1870s, 1890s, 1907–8, 1920–21, and most strikingly the 1930s.

The rediscovery of an important subject is always welcome, even if long overdue. However, much of the recent work has neglected the long history of both the phenomena of major economic fluctuations and their interpretations, concentrating instead on contemporary theoretical and policy controversies, mainly in the United States. An overview of selected literature will attempt to demonstrate that this myopia is costly and needs to be corrected.

The study of business cycles is almost coextensive with short-term macrodynamics and it has a large interface with the economics of growth, money, inflation, and expectations. The literature is huge; its level of difficulty is in general high. This survey attempts to provide a historical background and outline the evolution of thought leading to the recent developments in theory and related evidence. The coverage is extensive, yet of necessity much is left out, including theories that are largely concerned with unemployment and inflation, much less with business cycles directly.²

In particular, no attempt can be made here to discuss in any detail the statistical and historical work on the observed regularities and idiosyncrasies of business cycles and their possible long-term changes. This empirical literature is rich and important: it deserves a separate review. After all, it is the “stylized facts” which it provides that ought to be explained by the theory. Those main facts are summarized in section 2.2, but with a minimum of references and commentary.

Section 2.3 discusses the main elements of older theories, before and after the Great Depression, and proceeds to more recent models driven by changes in investment, credit, and price-cost-profit relations. Most of these theories and models are primarily endogenous and deterministic. Exogenous factors and stochastic elements are introduced early in section 2.4.

The sections that follow deal, first, with the monetarist interpretation of business cycles, then with the newer equilibrium models with price misperceptions and intertemporal substitution. The route leads generally from “adaptive” to “rational” expectations. The approach is generally monetarist in the sense of relying on monetary shocks, but the emphasis shifts from nominal demand changes and lagged price adjustments to informational lags and supply reactions. Various problems and complications arise, revealed in large part by intensive testing and criticism. This leads to new attempts to explain the persistence of cyclical movements, the role of uncertainty and financial instability, real shocks, gradual price adjustments, etc. Conclusions are drawn in the last section, which stresses the need for a realistic synthesis.

2. Two sets of writings should be mentioned in this context, namely, the theories of the new radical economists and those embodied in some of the recent “disequilibrium” models. For summaries or surveys, see Sherman 1976; Malinvaud 1977; and Drazen 1980.

Also, the early mathematical models and the more recent theories of the “political business cycle” receive little attention in the present paper; monographs surveying this literature are Rau 1974; Gapinski 1982; and Mullineux 1984.

2.2 Stylized Facts

2.2.1 The Overall Aspects and Varying Dimensions of Business Cycles

The term “business cycle,” is a misnomer insofar as no unique periodicities are involved, but its wide acceptance reflects the recognition of important regularities of long standing. The observed fluctuations vary greatly in amplitude and scope as well as duration, yet they also have much in common. First, they are national, indeed often international, in scope, showing up in a multitude of processes, not just in total output, employment, and unemployment. Second, they are persistent—lasting, as a rule several years, that is, long enough to permit the development of cumulative movements in the downward as well as upward direction. This is well established by the historical chronologies of business cycles in the United States, Great Britain, France, and Germany, a product of a long series of studies by the National Bureau of Economic Research (Burns and Mitchell 1946; Moore 1961, 1983; Moore and Zarnowitz 1986; see also chapter 5). For all their differences, business expansions and contractions consist of patterns of recurrent, serially correlated and cross-correlated movements in many economic (and even other) activities.

Seasonal movements, which are periodic but often variable in size and pattern, may obscure the cyclical developments from an observer of current changes in individual time series. The same applies to short, erratic movements, which are similarly ubiquitous. But, looking back across monthly or quarterly data representing many different variables, business cycles can be clearly distinguished from the other fluctuations in that they are as a rule larger, longer, and more widely diffused. They dominate changes in the economy over spans of several years, in contrast to the seasonal and other variations which spend themselves over spans of a year or less. They reflect, and interact with, long growth trends which dominate developments across decades.

Peacetime expansions in the United States averaged about 3 years in the last half-century, 2 years in the earlier periods containing 10 cycles each (table 2.1). Each of the wartime expansions was much longer. Contractions have averaged close to 1 year since 1933, almost twice as long in the earlier periods. This suggests a strong shift toward longer and more variable expansions and shorter and more uniform contractions since the Great Contraction of the early 1930s. However, the NBER dates for the early business cycles are based on limited information and may overstate the length of some of the recessions (see note to the table). The mean duration of full peacetime cycles has remained approximately stable at 4 years.

The individual phase and cycle durations show considerable variability over time, as shown by the standard deviations in table 2.1. However, when the relatively rare outliers are discounted, fairly clear central tendencies emerge. Thus the ranges of 1.5–3 years, 1–2 years, and 2.5–5 years account for three

Table 2.1 Average Duration of Business Cycles in the United States, 1854–1982

Period (years, T to T)	No. of Business Cycles Covered (1)	Average Measures of Phase and Cycle Durations					
		Expansion		Contraction		Full Cycle (T to T)	
		Mean (2)	S.D. (3)	Mean (4)	S.D. (5)	Mean (6)	S.D. (7)
1854–97	10	27	9	24	17	51	24
1897–1933	10	23	10	20	10	43	10
1933–82	10	49	27	11	3	60	26
1933–82, excl. wars	7	37	15	11	4	48	14
1854–1982	30	33	20	18	12	51	22
1854–1982, excl. wars	25	27	11	19	13	46	16

Note: All means and standard deviations (S.D.) are rounded to full months. Expansions are measured from troughs (T) to peaks (P), contractions from P to T, the full cycles from T to T. Figures in line 4 exclude the expansions during World War II, the Korean War, and the Vietnam War and the immediately following contractions. Figures in line 6 exclude also the expansions during the Civil War and World War I and the immediately following contractions. For references and underlying detail see Moore and Zarnowitz 1986.

It should be noted that the data available for identifying and measuring the historical business cycles are fragmentary and often weak. Much of the evidence relates to cyclically sensitive sectors and processes. Hence some of the early fluctuations may have involved only slowdowns rather than absolute declines in total output and employment. If so, the averages in lines 1 and 2 should be somewhat larger for expansions and smaller for contractions. This would only moderately reduce the contrast with the entries in line 3 and would not significantly alter the overall conclusions drawn in the text (see chapter 7).

fourths or more of the peacetime expansions, contractions, and full cycles in the United States, respectively.

The amplitudes of cyclical expansions vary as much as their durations, with which they tend to be well correlated. The rates of change (velocities) and diffusion show less variability across the cycles. Table 2.2 provides some evidence in support of these generalizations.

In the 20 years between the two world wars three major depressions occurred, including the uniquely deep one of 1929–33. Since then no general declines of comparable magnitude have occurred, notwithstanding the gravity of recent conditions of rising and high unemployment in some countries, such as the United Kingdom. On the whole recessions have become not only much shorter but also shallower and less diffused. Table 2.3, using a sampling of measures for the U.S. business contractions of 1920–82, illustrates the contrasting dimensions of major depressions versus other declines and the much smaller but consistent differences between the “severe” and “mild” recessions.

2.2.2 Main Features of Cyclical Behavior

Most industries and sectors of the economy participate in the general business cycles with substantial regularity (i.e., they exhibit high conformity or

Table 2.2 Selected Characteristics of Seven Expansions, United States, 1949–82

Line	Statistic	Largest Value (1)	Smallest Value (2)	Mean (3)	Standard Deviation (4)
Real GNP:					
1	Duration (months)	106	12	46	30
2	Total increases (%)	49.2	4.4	21.1	14.7
3	Rate of increase (% per year)	6.4	3.5	4.7	1.0
Unemployment rate:					
4	Total decline (% points)	-5.3	-0.6	-2.7	1.5
Nonfarm employment:					
5	Percent of industries expanding	100	73	89	9

Source: Moore and Zarnowitz 1986, table 6.

Note: The entries in col. 1 refer to the expansion of 2/1961–12/1969 (lines 1–3) and 10/1948–7/1953 (lines 4 and 5). The entries in col. 2 refer to the expansion of 7/1980–7/1981. The entries in col. 3 and 4 cover all seven expansions. Line 5 shows the maximum percentage of nonagricultural industries with rising employment, based on changes over 6-month spans.

Table 2.3 Average Duration, Depth, and Diffusion of Thirteen Contractions, United States, 1920–82

Line	Statistic	Great Depression (1)	Two Major Depressions (2)	Six Severe Recessions (3)	Four Mild Recessions (4)
1	Average duration (months)	43	16	12	10
Percentage decline:					
2	Real GNP	-32.6	-13.4	-3.3	-1.7
3	Industrial production	-53.4	-32.4	-13.1	-7.8
4	Nonfarm employment	-31.6	-10.6	-3.8	-1.7
Unemployment rate:					
5	Total increase (% points)	21.7	9.6	3.8	2.3
Nonfarm employment:					
6	Percent of industries contracting	100	97	88	77

Source: Moore and Zarnowitz 1986, table 7.

Note: The contraction of 8/1929–3/1933 is referred to as the Great Depression; the contractions of 1/1920–7/1921 and 5/1937–6/1938 are the major depressions. The dates of the six severe recessions are 5/1923–7/1924, 11/1948–10/1949, 7/1953–5/1954, 8/1957–4/1958, 11/1973–3/1975, and 7/1981–11/1982. The dates of the four mild recessions are 10/1926–11/1927, 4/1960–2/1961, 12/1969–11/1970; and 1/1980–7/1980.

coherence), but some do not (e.g., agriculture, which depends heavily on the weather, and production of naturally scarce resources).³ Durable producer and consumer goods tend to have high conformity and large amplitudes of cyclical movements in production, employment, and inventories. The amplitudes are

3. This section is based primarily on studies of U.S. economic history, but many of the *qualitative* features of cyclical behavior summarized here are found as well in the data for other major industrialized countries with private enterprise and free markets: Mitchell 1913, 1927; Schumpe-

much smaller for nondurable goods, and still smaller for most of the (nonstorable) services. Manufacturers' sales move with greater amplitudes than wholesalers' sales, and the latter with greater amplitudes than retailers' sales. In many industries, particularly manufacturing of durables, production is, in large measure, guided by advance orders, which show large fluctuations followed, with variable lags, by much smaller fluctuations in outputs and shipments. The resulting changes in backlogs of unfilled orders and average delivery lags are themselves procyclical.

Private investment expenditures, although much smaller in the aggregate than consumer spending, have much larger cycles in percentage terms. Aggregate production typically fluctuates more widely than aggregate sales, which implies a procyclical behavior of inventory investment. Business profits show very high conformity and much greater amplitude of cyclical movements than wages and salaries, dividends, net interest, and rental income.

The level of industrial prices tends to have wider fluctuations than the levels of retail prices and wages. Virtually all U.S. business contractions before World War II were associated with declines in wholesale prices.⁴ However, the last recession to be accompanied by a significant deflation was that of 1948–49. Since then the price level never fell cyclically, but each of the seven U.S. recessions of 1953–82 resulted in a temporary reduction of the rate at which prices rose, that is, in some disinflation. But, in contrast to the general price indexes for consumer and producer goods, prices of industrial commodities and raw materials traded in organized auction markets continued to show high sensitivity to business cycles, often turning down early in slowdowns as well as contractions.

Narrowly and broadly defined monetary aggregates usually experience only reduced growth rates, not absolute declines, in connection with ordinary recessions. Only in cycles with severe contractions do substantial downward movements interrupt the pronounced upward trends in these series. The income velocity of money (i.e., ratio of income to the stock of currency and commercial bank deposits held by the public) tends to move procyclically (up in expansions and down in contractions), allowing for its long trends (downward before World War II, then upward for some time).

Short-term interest rates display high positive conformity and generally large amplitudes of movements relative to their average level in each cycle. However, when measured in basis points, cyclical changes in these series are typically small when the interest-rate levels are low. Long-term rates usually

ter 1939; Frickey 1942; Burns and Mitchell 1946; Abramovitz 1950; Mitchell 1951; Gayer, Ros-tow, and Schwartz 1953; Matthews, 1959; Moore 1961, 1983; R. A. Gordon 1961; M. Friedman and Schwartz 1963a; Hultgren 1965; Zarnowitz 1972a, 1973; Zarnowitz and Moore 1984; P. Klein and Moore 1985.

4. This is true both for the periods of long-term inflationary trends (1843–64, 1896–1920) and for those of long-term deflationary trends (1864–96, 1920–32).

lag behind the short-term rates and have much lower conformity and much smaller amplitudes. The relative movements in both short-term market rates and bond yields increased significantly in the recent past compared with their historical averages. Near cyclical peaks, short rates tend to come close to or exceed the long rates; near cyclical troughs, they tend to be much lower.

Along with these conformity and amplitude characteristics, the recurring features of business cycles include an array of timing sequences. Months before total employment, output, and real income turn down, activities marking the early stages of investment process begin to decline. These include the formation of new business enterprises, corporate appropriations for capital expenditure, contracts for commercial and industrial construction, new orders of machinery and equipment, and new bond and equity issues. Investment realizations—construction put in place, deliveries and installations of equipment—keep increasing long after the decline in these investment commitments as work continues on the backlog of orders accumulated during the busiest stages of expansion. Indeed, business expenditures for new plant and equipment often peak when the overall economic contraction is already well under way. At business cycle troughs, with lower levels of capacity utilization, the delivery lags are generally shorter, but investment commitments still tend to lead and expenditures coincide or lag.

Long before the downturn in total sales, profits per unit of sales decline. Total profits (a product of margins times sales) also lead, but by shorter intervals. Stock prices move early as well, reflecting expected changes in corporate earnings. Bond prices tend to turn earlier yet (bond yields are generally lagging).

Labor productivity (output per hour) fluctuates procyclically around a secularly rising trend, generally with leads. Money wages often rise less than prices in recoveries and more than prices in late expansion stages. This combines with the marked and persistent productivity changes to induce a procyclical and lagging movement in labor costs per unit of output.

Net changes in consumer installment credit and in mortgage credit outstanding have similar procyclical, leading behavior patterns. So has the net change in bank loans to business, but here the leads tend to be shorter and less consistent. Compared with the overall credit flows, the rates of growth in monetary aggregates show, in general, lower cyclical conformities and amplitudes and more random variations. They have historically led at business cycle turns by highly variable but mostly long intervals. Indeed, these leads are often so long as to produce strong elements of inverted behavior in the monetary growth rates: that is, extended declines during expansions and rises during short recessions.

Consumers' "sentiment," that is, anticipations concerning their economic and financial fortunes, also has a predominantly leading pattern. Recent recessions in the United States have been preceded, more often than not, by downturns, and recoveries by upturns, in consumer buying plans and actual ex-

penditures on automobiles, housing, and related durable goods. Residential construction commitments, such as new-building permits and housing starts, have particularly long leads at peaks and often also at troughs of the business cycle. Here the gestation periods are fairly short so that the expenditures themselves show sizable leads.

Change in business inventories not only conforms positively to cycles in general economic activity but is highly sensitive and volatile, often leading, albeit by variable and, on the average, short intervals. Total manufacturing and trade inventories, on the other hand, are dominated by long trends and tend to lag. Inventory investment plays a very important role in short and mild cycles, whereas fluctuations in fixed investment acquire a greater weight in the longer and larger cycles.

Table 2.4 provides a conspectus of the timing relationships found to be typical of business cycles.

2.2.3 Some International Aspects and Recent Developments

Business cycles have tended to be shorter in the United States than in Europe (e.g., the 1854–1938 period witnessed 21 U.S. cycles averaging 4 years and only 16 British cycles averaging $5\frac{1}{3}$ years). However, before World War II, more than 60% of the cyclical turning points can be matched for all four countries covered by the NBER chronologies, and only 10% cannot be matched at all (Moore and Zarnowitz 1986). After World War II, an era of great reconstruction in Western Europe and Japan set in, which witnessed, first, a restoration of sound currencies and free markets, then rapid growth. For some time cyclical setbacks in these countries assumed the form of retardations of growth rather than absolute declines. However, these slowdowns and the intervening speedup phases continued to show a high degree of international diffusion. Then growth slackened and the “classical” business cycles (with absolute declines in total output and employment) reappeared everywhere in the 1970s. The tendency for these cycles to be roughly synchronized across the major trading countries became visible again, even without allowances for discrepancies in the longer growth trends.⁵

In a large economy dominated by production for domestic markets, business cycles are likely to be induced primarily by internal mechanisms (e.g., fluctuations in spending on durable goods endogenously and elastically financed) but they are then transmitted abroad through the movements in imports that are a positive function of production and income. For small and, particularly, less developed countries, fluctuations in exports usually call the tune. Of course, foreign influences can be critical, at times, for even the largest and relatively least open economy. This is well illustrated by the adverse effects on the United States of the OPEC oil price boosts in 1973–74 and

5. These events rekindled interest in business cycles of larger dimensions and essentially endogenous nature; see Volcker 1978.

Table 2.4 Typical Leads and Lags among Major Economic Indicators

Leading	Roughly Coincident	Lagging
<i>I. Investment in Fixed Capital and Inventories</i>		
New building permits; housing starts; residential fixed investment	Production of business equipment Machinery and equipment sales*	Backlog of capital appropriations, mfg.* Business expenditures for new plant and equipment* Mfg. and trade inventories
New business formation		
New capital appropriations, mfg.*; contracts and orders for plant and equipment		
Change in business inventories		
<i>II. Consumption, Trade, Orders, and Deliveries</i>		
New orders for consumer goods and materials	Production of consumer goods	
Change in unfilled orders, durable goods*	Mfg. and trade sales	
Vendor performance (speed of deliveries)		
Index of consumer sentiment		
<i>III. Employment, Production, and Income</i>		
Average workweek; overtime hours (mfg.)	Nonagricultural employment Unemployment rate	Average duration of unemployment
Accession rate; layoff rate (mfg.)	GNP; personal income Industrial production, total	Long-term unemployment
New unemployment insurance claims		
Productivity (output per hour)		
Rate of capacity utilization (mfg. mtls.)		
<i>IV. Prices, Costs, and Profits</i>		
Bond prices*		Unit labor costs
Stock prices*		Labor share in national income
Sensitive materials prices*		
Profit margins		
Total corporate profits; net cash flows		
<i>V. Money, Credit, and Interest</i>		
Monetary growth rates*	Velocity of money	Short-term interest rates*
Change in liquid assets*		Bond yields*
Change in consumer credit*		Consumer credit outstanding*
Total private borrowing*		Commercial and industrial loans outstanding
Real money supply		

Note: Series marked * are in nominal terms (some have the same average timing properties when deflated). All other series are in real terms (constant dollars, physical units) or related indexes and ratio numbers. The selection is based on U.S. indicators published in *Business Conditions Digest*, a monthly report by the Bureau of Economic Analysis, U.S. Department of Commerce. The timing relations among corresponding series for other countries are in many respects similar (Moore 1983, ch. 6).

Abbreviations: mfg. = manufacturing; mtls. = materials.

1979–80 through increased costs and prices (leftward shifts in the aggregate supply schedule) and reduced real disposable income (hence, presumably also, some leftward shifts in the aggregate demand schedule). Such worldwide supply shocks, although clearly of major importance in the context of contemporary problems of productivity, growth, and development, are new and rare phenomena whose role in business cycles generally is modest but in danger of being overemphasized.⁶ The more persistent effects come from changes on the demand side. Thus the volume, prices, and value of U.S. exports show fluctuations that correspond well to cycles in the dollar value of imports by the outside world (Mintz 1967). The demand changes are powerfully reinforced when the links among the major countries convert their independent cyclical tendencies into fluctuations that are roughly synchronized.

These links result not only from international trade (current-account transactions in goods and services) but also from international lending and investment (capital-account transactions in assets). The latter factor became particularly important in recent times when asset markets became highly integrated worldwide. Interest rates (adjusted for the anticipated exchange-rate movements) are now linked across the open economies, and capital flows are extremely sensitive to the risk-adjusted differentials in expected rates of return.

Partly because of the increased capital mobility, the shift from fixed to flexible exchange rates in the early 1970s provided much less insulation against foreign disturbances than was hoped for, and also much less autonomy for effective national macroeconomic policies. The price levels adjusted for exchange rates (i.e., the “real exchange rates”) show large and persistent movements over time: the purchasing power parity does not hold over time spans relevant for the study of business cycles. The recent fluctuations in real economic activity show a very considerable degree of international convergence, which presumably reflects not only the exposure to common disturbances but also the increased interdependence among (openness of) nations (Whitman 1976; P. A. Klein 1976; Moore 1983, ch. 6).

In the 1960s, when it appeared that business contractions in Europe and Japan were being replaced by mere retardations, there was a revival of interest in cycles defined in terms of deviations from long trends rather than levels of economic aggregates. For lack of a better term, the alternations of above-trend and below-trend growth phases came to be called growth cycles. These short fluctuations are defined by the consensus of detrended indicators just as business cycles are defined by the consensus of the same time series with no allowance for their long-term trends. The trends are estimated only to be eliminated from each series separately. Growth cycles in this sense are thus sharply dif-

6. James Hamilton (1983), writes about the role of changes in crude oil prices in the U.S. recessions after World War II. It should be noted that many recessions are preceded by upward cost pressures and supply restrictions associated with the boom-and-slowdown sequence of middle and late expansion stages. On the developments during 1973–76, see Zarnowitz and Moore, 1977.

ferent from, and should not be confused with, any fluctuations in the long-term growth rates themselves.

They also need to be clearly distinguished from business cycles. Most persistent and pervasive economic slowdowns begin with much reduced but still positive growth rates and then develop into actual declines—recessions. Thus the high-growth phase typically coincides with the business cycle recovery and middle-expansion, and the low-growth phase with late expansion and contraction. But some slowdowns stay in the range of positive growth rates and result in renewed expansion, not recession. Thus growth cycles are more numerous than business cycles and more symmetrical, being measured from rising trends.

One can imagine a lengthy period of low, positive growth that would be associated with as much deterioration in business conditions and rise in unemployment as a short and moderate recession—and more. But it is also possible for a slowdown mainly to reduce inflationary excess demand created in the preceding boom, without causing much surplus capacity and real hardship. The policy implications of such a deceleration in economic growth are entirely different from those of a recession, which always depresses real incomes and spending, outputs, and employment.

In actual experience, those decelerations in growth that have *not* led to absolute declines in aggregate economic activity (in recent U.S. history: 1951–52, 1962–64, and 1966–67) occupy an intermediate position between the two hypothetical cases considered above. Their adverse effects were felt primarily in areas of particular sensitivity, notably as declines in housing activity and stock prices. Unemployment would cease falling rather than rise substantially, and profits would weaken rather than tumble. Thus the overall impact of each of these slowdowns on economic activity was definitely less than even the mildest of recent recessions.

Some economists focus on the nature and sources of expansions and contractions, that is, on the business cycles. Others, by abstracting from the long-run trend, actually address growth cycle phenomena while aiming at an analysis of business cycles; that is, they fail to differentiate between the two categories. The latter treatment, frequently implicit in the theoretical literature of recent years, may not be a good practice. General business contractions need to be distinguished from periods of low but positive growth. However, mild recessions and severe depressions are also quite different. Also, many important regularities described in the previous section are observed, to a large extent, in the context of growth cycles as well. Thus, when the series that tend to lead at business cycle turns are adjusted for their own long trends, the resulting detrended series are generally found to be leading at growth cycle turns. An analogous statement can be made for the roughly coincident and lagging indicators. Systematic differences among the series with respect to their conformity and amplitude characteristics are likewise largely retained after the necessary transformations.

2.3 Theories of Self-Sustaining Cycles

2.3.1 Disparities and Common Elements in Some Early Theories

The classics of business cycle literature made lasting contributions to the description and analysis of the motion of industrialized market economies. They addressed the cumulative processes of inflationary expansions and deflationary contractions induced by bank credit fluctuations constrained by the availability of reserves under the gold standard (Hawtrey 1913). The role of discrepancies between the market and the “natural” interest rates in this process was much explored following Knut Wicksell ([1898] 1936). At below-equilibrium market rates, excessive bank credit creation produces overinvestment in capital-goods industries and imposes “forced saving” on those whose incomes lag behind inflation (Hayek 1933). But banks will have to curtail the supply of credit, and individuals will tend to restore their old consumption standards. As the demand and resources shift back to consumer-goods industries, undersaving—real capital shortage and losses to the producers of capital goods—results, causing a decline in these industries which cannot be compensated elsewhere in the short run. A deflationary downturn cannot be avoided. Here the monetary changes are linked to real “vertical maladjustments”: that is, imbalances between production of capital and consumer goods or between aggregates of investment plans and savings decisions (Tugan-Baranovskii [1894] 1913; Spiethoff [1925] 1953).

Other writers worked out the importance of long gestation and life periods of capital goods and developed some cyclical aspects of the acceleration principle (Aftalion 1913; J. M. Clark 1917, 1934). Schumpeter (1939) saw economic growth itself as a cyclical process, reflecting technological progress and spurts of innovations—opening up and temporary exhaustion of opportunities for new, profitable investment. Related factors include the failure of foresight, intersectoral shifts, and changes in relative prices. Thus, under uncertainty, interdependent expectations of business people generate widespread errors of optimism in expansions and pessimism in contractions (Pigou 1927). Unpredictable shifts in demand or supply lead to “horizontal maladjustments”—say, overinvestment in a particular sector, which involves indivisible and durable fixed capital, high costs of adjustments, and temporary but cumulative depressant effects (Robertson [1915] 1948). Unit costs of labor and production tend to rise relative to output prices before and after the downturn, and they tend to fall before and after the upturn, reflecting changes in capacity utilization and productivity; as a result, business profits show large fluctuations, which help explain the cyclical movements in investment and output (Mitchell 1913, 1927).

This capsule summary can merely illustrate the broad range of views held by these early students of business cycles. It is clear that there are important disagreements among their theories, particularly with respect to the relative

importance of monetary and real factors, long a major point of contention. But the dominant tone is one of awareness that what matters most is the *interaction* of changes in money and credit with changes in economic activity, particularly those connected with business investment. Most of the writers considered business cycles to be caused and conditioned by a number of factors and circumstances, and so their theories typically overlap and vary mainly in the emphasis accorded the different elements (Haberler [1937] 1964).

Not surprisingly, there is much in these individual theories that is unsatisfactory, unduly restrictive, or out-of-date. Here we must abstract from the detail and note that it is the high degree of consensus and achievement that is much more remarkable from the present point of view.

The first aspect of essential agreement is that the theories are mainly endogenous. That is, they purposely concentrate on internal dynamics of the system (interrelations and lagged reactions among its components). The authors generally held that contemporary industrial economies are, as a result of such dynamics, subject to recurrent fluctuations with major regularities that can be explained economically. They believed that “the cyclical movement has a strong tendency to persist, even where there are no outstanding extraneous influences at work which can plausibly be held responsible.” Hence they viewed the role of the exogenous forces as secondary, even though acknowledging that the latter continually act “as the originators or disturbers of endogenous processes, with power to accelerate, retard, interrupt, or reverse the endogenous movement of the economic system.”⁷

Second, these economists all basically adhered to the standard economic theory of their times, which is what Keynes later labeled the “classical school”; indeed, the latter is well personified by some of them. At the same time, they generally appreciated the seriousness of the problem of economic instability. The business cycles of their principal concern were major fluctuations measured in years of cumulative expansions and contractions. The recurrent phases of widespread unemployment and underutilization of productive capacities did (and still do) present a deep puzzle to the classical doctrine, according to which the economy is always in, or at least tending closely to, the general equilibrium. Thus, for a long time, business cycles were simply ignored by most economic theorists or, at best, were viewed as merely temporary “frictional” interference with, and departure from, equilibrium. But students of the subject, including those who were themselves committed to the equilibrium theory, have done much to counteract this evasive and untenable position.

7. Both quotations are from Haberler (1937) 1964, p. 10. This characterization is strongly confirmed by numerous passages in works by Robertson (1915, in particular pt. II, ch. 4); Mitchell (1927, esp. chs. 1 and 5), Hayek (1933, esp. ch. 4), and Pigou (1927, chs. 6–8, 21). For Schumpeter, the basic mechanism of credit-financed innovations is of much greater intrinsic interest than the multitude of diverse “external factors,” no matter how important the latter may be on any particular occasion (1935–1939, vol. 1, chs. 1–4).

Third, in the historical periods addressed by these studies the level of prices tended to move up during the general business expansions and down during contractions. The positive correlation between cyclical movements in broad price indexes and real activity measures implied that fluctuations in total nominal expenditures parallel the fluctuations in the aggregates of real income, employment, and output. This was generally accepted as a central characteristic of business cycles by the early theories, in which the fluctuations in aggregate money flows of income and spending play a large, proximately “causal” role. Of course, for these fluctuations to produce cyclical movements in real variables, it is necessary that wages and prices adjust with some sufficient lags rather than being highly flexible. Sometimes this condition was assumed explicitly but it was not much discussed and often only implied (Haberler [1937] 1964, pp. 459–61).⁸

2.3.2 Uncertain Expectations, Unstable Investment, and Long Depression Cycles

Keynes (1936, ch. 22, esp. pp. 314–15) attributed to the trade cycle a sudden, sharp downturn, a protracted decline, and a gradual, sluggish upturn. These are all characteristics of the 1929 peaks and the depressions and recoveries of the 1930s in Great Britain and the United States; also, in part, of the British experience in the depressed early 1920s. They are not typical of most cycles in these countries and elsewhere.

The sharp downturn, or “crisis,” is explained mainly by “a sudden collapse in the marginal efficiency of capital.” During a boom, the supply of new capital goods and the cost of their production and financing rise, with growing adverse effects on the current returns on investment. The inducement to invest is further weakened if the current costs come to be viewed as higher than the probable future costs. Optimism about the always uncertain future returns lingers for some time, but sooner or later doubts arise about the reliability of the hopeful expectations engendered by the boom.

Investment expectations are highly volatile because even those forecasts of long-term profitability of specific business projects which are viewed as most probable inspire little confidence. Observable frequencies of past outcomes are not generally a source of predictive knowledge in these matters. Keynes’s world is thus one of pervasive uncertainty which is sharply distinguished from calculable and insurable risk (as in Knight 1921).

It is easier and more rewarding to predict the short-term movements in the stock market, which are strongly affected by “mass psychology,” than to divine the long-term prospects for individual business concerns. The market reacts promptly to news of fluctuating business profits with revaluations which

8. In terms of the present-day conventional macroeconomic model, let us add, the positive correlation of fluctuations in prices and real variables would indicate that shifts in aggregate demand dominate the shifts in aggregate supply over the business cycle. If this sounds rather alien to the early theories, it is because the latter are typically more disaggregate.

inevitably exert a decisive influence on the rate of current investment. A new business will not attract investors if a similar existing one can be acquired on the exchange at lower costs (Keynes 1936, p. 151).

This last insight gave rise to an influential theory of James Tobin (1969), which makes investment in new plant and equipment an increasing function of q , the ratio of the value placed by the security markets on the existing firm to the replacement cost of its capital. This approach has several advantages: it is relatively simple, uses observable variables, and provides an analytically attractive linkage between investment and the expectations of the financial asset markets. Implicitly, it also relates the expected profit rate to the required rate of return on capital in the stock market and, hence, to the interest rate. However, the hypothesis has not fared well in empirical tests (Von Furstenberg 1977; P. K. Clark 1979; Blanchard and Wyplosz 1981; Abel and Blanchard 1983; R. J. Gordon and Veitch 1984). This is perhaps partly because of the use of average q instead of the theoretically more appropriate marginal q^9 but more likely because of various simplifying restrictions used in this work: homogeneous capital and perfect financial markets with no liquidity constraints on firms. These idealizations are poorly suited for an analysis of cyclical movements in investment and they certainly clash with the Keynesian views on the instability of financial markets (sect. 2.4.6).¹⁰

Once aroused, the doubts about profitability of planned and current investment projects spread rapidly, bringing down in "disillusion" the stock market, which is revealed to have been overoptimistic and overbought. As the pendulum swings to overpessimism, the demand for broadly defined money will increase, raising the rate of interest and hence seriously aggravating the crisis. The revival of investment will require a "return of confidence . . . an aspect of the slump which bankers and businessmen have been right in emphasizing" (Keynes 1936, p. 317). But confidence, once severely shaken, takes time to mend. Also, the propensity to consume is adversely affected by the fall in the value of equities.

Only as the downswing develops will it bring the level of interest rates down. This decline will not be as prompt and large as would be necessary to counter the "collapse" of investment. The demand for money is interest elastic, highly so at low levels of the rates, because of bearish speculation in the face of basic uncertainty as to future changes in the rate of interest. The con-

9. This is the ratio of the increase in the value of the firm from acquiring an additional unit of capital to the marginal cost of that unit (which, in contrast to the measured average q , is an *ex ante* and not directly observable quantity).

10. The same observations apply *a fortiori* to the "neoclassical" investment theory dating from Jorgenson 1963, which concentrates on the average long-term behavior as determined by the requirement that the expected returns over the life of a project exceed its costs. The short-run deterrent effect on investment of the rising flow supply price of capital goods (stressed in Keynes 1936, ch. 11) is not well captured in this approach, and the expectational lags are not distinguishable from the gestation periods or delivery lags (Abel 1980). For recent tests of this and other investment theories, see also Bischoff 1971; and Kopcke 1977.

clusion here is that a recovery from a severe slump is possible only after the capital stock of business has been reduced sufficiently to restore its profitability. This may take several years, through use, wear and tear, or obsolescence (Keynes 1936, p. 151).

It might appear that overbuilding is the cause of the downturn and long slump but Keynes insists that it is not. Rather, the effective private demand fails to sustain full employment because investment is too unstable and the propensity to consume is stable but too low. It is only relative to the deficient demand that “overinvestment” can occur; there is no saturation of profitable investment opportunities at full employment.

Keynes’s concern was with long and severe depressions characterized by very large declines to low levels of both real investment and stock market values. Such major depressions have occurred at intervals of a few decades through the 1930s but most business contractions were much milder and shorter. Even in long contractions the stock of capital usually continues to increase, although at much reduced rates; also, an *abrupt* collapse of investment is rare.¹¹ Consumer spending is much less stable in the short run than Keynes assumed (but also much more stable and supportive of growth in the long run). As for the demand for money, there is a mass of evidence that its interest-elasticity tends to be relatively low (Laidler 1969). These observations raise serious questions about some elements of Keynes’s theory.

2.3.3 Wage and Price Dynamics in Business Cycles

Despite the great rise and persistence of unemployment, real wages *increased* throughout the 1930s, thereby failing to provide one classical cure for the apparent disequilibrium in the labor market.¹² Keynes (1936, chs. 19 and 21) did not argue that money wages are entirely rigid downward but rather that they adjust but sluggishly to excess supplies of labor. Such slow wage declines are apt to reduce incomes, consumption, and prices before they begin to improve profitability and stimulate investment. When the resulting gradual deflation becomes widely anticipated, people repeatedly postpone purchases, mainly of durable goods, while waiting for prices to fall further. The demand for money increases at the expense of the demand for goods and equities.

Moreover, unexpected deflation increases the burden of accumulated debt, which falls most heavily on businesses and individuals with high propensities to borrow, invest, and spend. The activities of these units are severely curtailed as their bank credit lines are cut. Business failures and personal bankruptcies rise in numbers and size. Irving Fisher (1932, 1933) ascribed the

11. These points were made early by Burns (in his collected essays, 1954, pp. 3–25, 207–35). On the dispersion of peaks in various categories of investment commitments and expenditures, see Zarnowitz 1973, ch. 4.

12. In the United States, average hourly earnings in manufacturing divided either by the consumer price index or by the wholesale price index rose approximately 20% between 1929 and 1934, for example. Money wage rates declined less than prices. Hours of work fell along with the number of the employed workers (Temin 1976, pp. 138–41).

depth of the Depression to the confluence and mutual reinforcement of deflation and “overindebtedness” inherited from the boom. (This suggests overinvestment as the cause of the downturn.) His policy prescription was monetary reflation, a reversal of the price decline.

A one-time large drop in the general wage level is a theoretical but hardly a practical alternative in a large decentralized economy with numerous, strongly differentiated labor markets and a complicated structure of relative wages. A spreading depression in an open system is more likely to result in competitive devaluations and various protectionist “beggar my neighbor” measures, with deeply damaging overall consequences.

The actual and expected changes in the *rates of change* in wages and prices can certainly be of great importance in business cycle dynamics. The effects on aggregate demand of changes in the *levels* of wages and prices, on the other hand, are believed to have their main roles in comparative statics and the long run.¹³ The static and dynamic elements were never clearly distinguished by Keynes and his immediate critics; the debate proceeded for a long time in the framework of comparative statics, which obscured the essentially dynamic disequilibrium nature of Keynes’s theory (Leijonhufvud 1968).

After World War II inflation became, for the first time, a chronic condition in peacetime, drawing attention away from the concurrent, relatively mild recessions and at the same time making their understanding apparently much more difficult. The old problem of depression-cum-deflation was replaced by the new problem of unemployment-cum-inflation. The famous Phillips curve emerged first as a nonlinear and inverse dependence of the rate of change in nominal wages (w) on the rate of unemployment (U) and was rationalized by relating U to the excess supply of labor (Phillips 1958; Lipsey 1960). Soon the rate of inflation (p) was similarly related to U , on the ground that p and w normally differ by a steady rate of growth in labor productivity (Samuelson and Solow 1960).¹⁴

The classical view of an aggregated labor market posits the existence at any time of a unique equilibrium, or “natural,” unemployment rate (U^N) as a function of *real* wages (M. Friedman 1968). An inflation that has lasted for some significant time will be expected to persist at some positive average rate (p^*). Changes in the price level that are generally and correctly anticipated are

13. The reference here is, first, to the “Keynes effect” (at lower prices, a given quantity of money represents a larger real quantity) and the “Pigou effect” (at lower prices, a given quantity of net nominal private wealth represents a larger real quantity). The former would raise investment through lower interest rates; the latter would raise consumption. To stimulate the economy in the short run both require downward flexibility of prices generally. The Keynes effect depends inversely on the interest-elasticity of the demand for money. The Pigou effect depends positively on the magnitude of *net* private wealth, which is probably small in relative terms. Deflationary expectations and distributional shifts may also greatly weaken this process (Pigou 1947; Patinkin 1948).

14. It should be noted that from the cyclical perspective it is far from innocuous to substitute price for wage inflation. As will be shown in section 2.3.6, labor productivity and the related price and cost variables undergo partly systematic changes over the cycle.

matched by wage changes and hence cannot cause deviations of U from U^N . Only unanticipated inflation, that is, forecast errors ($p^e - p$), can cause such deviations. Equilibrium requires that $p^e = p$, hence a stable long-run tradeoff between p and U cannot exist (the natural-rate hypothesis—NRH: Friedman 1966, 1968; Phelps 1967). A short-run Phillips curve, associated with unanticipated or disequilibrium inflation, stays in place only as long as p^e remains unchanged.¹⁵

The initial reaction of most economists to the Friedman-Phelps critique was to embrace the NRH without questioning the existence of an inverse relationship between inflation and unemployment *in the short run*. This was because they assumed expectations to be “adaptive,” that is, backward looking and involving only partial and lagging corrections of past errors.¹⁶ Indeed, it was frequently assumed that the errors of inflation forecasts are fully eliminated only on the average over the business cycle. On this permissive interpretation of a “long run,” the NRH is entirely consistent with continuing parallel fluctuations in inflation and real economic activity around their (uncorrelated) long-trend movements.

The expectations-augmented but only slowly shifting Phillips curve appears in Tobin’s 1975 analysis of Keynesian models of cyclical contractions, with the qualification that this does not imply a full acceptance of the NRH. In this dynamic model, output (Q) moves in reaction to changes in aggregate real demand (E). In the short run the two variables can differ: say, $E < Q$ when Q is rising because of lags in consumption and unintended inventory changes. E depends positively on Q and the expected price change p^e and negatively on the price level P . Actual inflation p (= rate of change in P) adjusts to changes in p^e and in Q relative to Q^N (the full-employment output). Finally p^e reacts to the divergencies of p from p^e . The equilibrium conditions are $E = Q$, $Q = Q^N$, and $p = p^e$. The main inference from the model is that “a strong negative price-level effect on aggregate demand, a weak price-expectations effect, and a slow response of price expectations to experience are conducive to stability” (Tobin 1975, pp. 199–200). Large adverse shocks to E can push the economy into a depression, and market price adjustments will provide no reasonably prompt and effective remedy under conditions where the price-change effects on E are stronger than the price-level effects.¹⁷

2.3.4 Disequilibrium Models with Capital Accumulation

Keynes’s analysis is only implicitly and partially dynamic. Since net investment varies, so does total capital, which influences output, investment,

15. The general form for the original equation is $w_t = f(U_t)$; for the “expectations-augmented” equation satisfying the NRH it is $w_t = f(U_t) + p_t^e$.

16. A simple model of this type if $p_t^e - p_{t-1}^e = k(p_{t-1} - p_{t-1}^e)$, where $0 < k < 1$.

The early *locus classicus* for adaptive expectations is Cagan 1956; also Muth 1960; Nerlove and Wage 1964; and Mincer 1969a.

17. Tobin, 1980, ch. 1, has a more recent reassessment, with similar conclusions concerning the cyclical effects of deflation.

and savings. But these effects are ignored and the stocks of production factors and technology are treated as constant. The older acceleration principle has no role in the *General Theory*, and indeed is not even mentioned. But the 1930s and 1940s saw a proliferation of formal models of essentially endogenous cycles in aggregate output, which use various versions of the investment accelerator and the consumption multiplier and let the two interact (Harrod 1936; Kalecki 1937; Samuelson 1939; Metzler 1941; Hicks 1950).

In most of these formulations, net investment is a function of *changes* in output, which implies that fluctuations in consumption are transmitted with increasing amplitudes to the higher stages of production—the derived demand for intermediate and producer goods. This contrast in amplitudes is broadly consistent with the evidence (sec. 2.2.2), which has long made the acceleration principle attractive to business cycle theorists. Yet it is clear from the data that investment series have much higher persistence or serial correlations than the series of first differences in output or sales. The simultaneous relationship between these variables over short unit periods is weak, not surprisingly, since investment depends on expectational and financial factors as well as on changes in technology, cost of labor, etc.¹⁸

With consumption lagging behind income and induced investment behind change in output, the multiplier-accelerator interaction can produce a model of fluctuating output in the form of a second-order difference equation. Small values of the coefficient of acceleration result in damped cycles, large values in explosive cycles. When a trend is added reflecting continuous technological change embodied in “autonomous” investment, the cycles are slanted upward around a line whose slope represents the equilibrium rate of the economy’s real growth. In the potentially explosive case, fluctuations in actual output are constrained between a “ceiling” along which the (growing) resources are fully utilized and a “floor” set by the non-negativeness of gross investment (Hicks 1950). This model uses distributed lags in consumption and investment functions and suspends the accelerator during a steep downswing, with the result that the slump is both cushioned and prolonged as the excess stocks of capital depreciate but slowly down to the levels required by the low production at the floor.

A closely related but more general class of models is based on the capital stock adjustment (or “flexible accelerator”) principle: current investment equals some fraction of the gap between the desired and the actual capital. The desired stock varies directly with output (taken, questionably, as a proxy for the expected demand for output that the capital is to help produce). Net

18. The theory of the production function of the firm includes the acceleration effect but also, in the general case of variable factor proportions and prices, a substitution effect (e.g., Gapinski 1982, chs. 2 and 4). Net investment cannot decline in any period by more than the capital stock can be worn out, which suggests asymmetrical behavior between upswings and downswings. In principle, it is in the long run, along the rising trend of capacity output, that the simple accelerator making net investment proportional to the change in output should work best, not in the short run of cyclical analysis.

investment therefore depends positively on output and inversely on the initially available stock of capital. This formulation (used early in Kalecki 1935; Kaldor 1940; Goodwin 1951) is capable of being improved, with some significant advantages. Since profitability should depend positively on the output-capital ratio, the role of profits in the investment function (stressed by many authors, particularly after Tinbergen 1938–39) is given at least an indirect consideration. The fraction of the capital gap closed in the short unit period (the speed of adjustment) may be made a function of the interest rate and, more generally, the costs of capital. The costlier the speed, the more gradual will be the optimal adjustment; this opens the way to a potentially useful dynamic analysis. On the other hand, the cyclical changes in capital stock, being relatively small (and rarely negative), have probably rather weak short-term effects on investment and output.

The dynamics in the models under review come from lags, nonlinearities, or both. Nonlinearities are likely to be very common in economic relationships, in part because the numerical values of certain important parameters should vary with the phases of business cycles, which cover a wide variety of macroeconomic conditions. Yet, few theoretical models of the cycle make important use of nonlinearities. In Kaldor 1940, investment (I) is a sigmoid function of output (Q), with much lower positive slopes at both extremes than in the broad middle range of the Q scale. I is deterred by both surplus capacity in slumps and rising construction and financial costs in booms. Saving (S) has a converse shape, with higher positive slopes at both extremes than in the middle range of Q : people stabilize consumption over time by temporarily reducing (raising) the average rate of savings when their incomes are unusually low (high).¹⁹ Given Q , I depends inversely and S depends positively on the capital stock K . There are three possible equilibria; two of which are stable. The result is a self-sustained cycle in the real aggregates, from stable to unstable to another stable equilibrium.²⁰

The substantive uses of nonlinearities in the theory of business cycles are yet to be systematically explored. There are, to be sure, various nonlinearities in mathematical and econometric models but they are, on the whole, scattered and treated mainly as technical detail. The early models by Kaldor, Hicks, and Goodwin remain influential in this literature.²¹

19. The argument seems to anticipate the more recent theories of consumption if one adds the assumption that people perceive larger proportions of their incomes as being transitory in booms and slumps than in the more "normal" times closer to the trend. Transitory income is taken to be largely saved; permanent income consumed.

20. The path of output is $dQ/dt = \alpha[I(Q, K) - S(Q, K)]$, with $\alpha > 0$ denoting the speed of adjustment. The equilibria are defined by $I(\cdot) = S(\cdot)$; they are stable when $\delta S/\delta Q > \delta I/\delta Q$, which occurs at both low and high levels of Q , and unstable when $\delta S/\delta Q < \delta I/\delta Q$, which occurs in an intermediate position. The responses of Q , that is, movements along the I or S curves, are speedier than the shifts of these curves caused by the changes in K .

21. A list of some references must suffice: Ichimura 1954; Rose 1967; Bober 1968; Chang and Smyth 1971; Kosobud and O'Neil 1972; Torre 1977; Scarfe 1977, ch. 4; Gapinski 1982, chs. 4–6; Schinasi 1981, 1982.

Recently, some methods of the “singularity” and “catastrophe” theories, new branches of applied mathematics, began to be applied to the analysis of large economic fluctuations involving crises, depressions, and rapid recoveries. The theory is concerned with the interaction of “fast” and “slow” variables in dynamic systems described by differential equations, where the short-run equilibrium may jump from one region of the state space to another. For example, the rate of change in output may depend on the level of output, a “fast” variable, and physical capital and financial assets, “slow” variables (expectations or other parameters may also be treated as slow variables). An illustration using a modified Kaldor model is offered by Varian (1979). This is an interesting approach, though still in its infancy.

Although innovative in their dynamics, the models discussed here are severely limited in their contents. They slight or ignore the monetary, financial, and expectational factors which, theory and evidence suggest, are particularly important in major cycles, crises, and depressions. Most of these models, too, neglect the role of new technology as a determinant of investment (Goodwin 1951 is a notable exception). These criticisms were made early in the heyday of the endogenous investment theories of the cycle (Burns 1952a; Haberler 1956; Lundberg 1958). A. Hansen (1964) viewed the accelerator as relatively weak and stressed the driving force of autonomous investment related to innovations which require more capital per worker.

However, it is also clear that the models in question contain important elements and are capable of being expanded and improved, in particular by incorporating monetary factors. Indeed, Hicks (1950, chs. 11 and 12), added to his main real model a monetary subsystem which could aggravate some downturns.²² Tobin’s 1955 model has similar cycles but also explicit roles for the supply of monetary assets and the inflexibility of money wages. More recently, Laidler (1973b) makes the desired capital stock depend on the lagged real interest rate as well as the lagged output, thereby modifying Hicks’s equation for induced investment so that it includes monetary effects. Inflation expectations, formed adaptively, appear in a Phillips-type relationship which complies with the NRH but permits gradual price adjustments. This model retains an accelerator-multiplier mechanism and can generate fluctuations in output and prices. However, unlike in Hicks, the cycles are here damped, and exogenous changes in either the nominal interest rate or autonomous expenditures may be required to move the economy off either the floor or the ceiling.²³

22. The “monetary complication” takes the form of a cobweb cycle in the IS-LM space, with rather complex lags and nonlinearities. A quarter-century later, Hicks (1974) revised his views, stressing the need for greater reliance on monetary factors and autonomous investment.

23. Laidler considers, alternatively, the nominal rate of interest and the money stock as exogenous. Output is systematically related not to inflation or deflation but to the rate of change in either.

2.3.5 On Causes and Consequences of Fluctuations in Inventory Investment

Inventories have a much shorter expected life span than fixed capital and can be adjusted much more quickly to the desired levels. Nevertheless, the success of attempts at such adjustment is by no means assured as it depends partly on accurate sales forecasts and partly on how promptly production reacts to unanticipated sales changes. Some of the observed inventory investment is planned, and some is unintended.

It is difficult to use stocks as a buffer protecting output from the variation in sales, except when the changes in demand are small and short, calling for no alteration in production. In the more persistent business cycle movements, inventory investment tends to be, on the contrary, *destabilizing*, as shown by the already noted fact that aggregate output fluctuates more widely than final sales.

An early and influential theory of inventory cycles was developed in Metzler's (1941, 1947) multiplier-accelerator models. Here the desired level of stocks of consumer goods varies with anticipated sales to consumers, which reflect sales observed in the recent past. Output lags behind consumption, which is proportional to contemporaneous income. An initial rise in the level of noninventory investment, which is treated as autonomous, depletes inventories; hence business people attempt to increase them. But inventory investment has the feedback effect of raising income and consumption, which reduces the stocks still more. This causes further rounds of increase in inventory investment, output, and sales. However, at some point during the expansion, the rates of increase in sales, and hence also in inventory investment, will start falling. Declines in income and consumption will ultimately result, reducing the desired stock levels. Again, the very efforts to get rid of the unwanted stocks depress income and consumption further and are thus temporarily self-defeating, but eventually the rate of decline in sales and the disinvestment in stocks will begin to diminish, and an upturn in income and consumption will initiate a new cycle.

Several papers report generally favorable tests of Metzler's model based on annual and quarterly data, mainly for the United States in the first two decades after World War II (Coppock 1959, 1962, 1965; Hillinger 1966, 1979). The periodicities are heavily concentrated in the 3–4 year range and are not seriously disturbed by inclusion of random error terms, but there is substantial damping. Anticipated values are derived from distributed lags (adaptive expectations). The techniques include cyclical (NBER) measures, regression, and spectral analyses. The hypothesis that fluctuations in nonfarm business inventory investment and nonresidential fixed investment are essentially periodic, with cycles of about 3–5 and 7–10 years, respectively, has been recently revived by Hillinger (1982, 1983). However, the unconventional methods

used in this work are subject to serious doubts and the results are yet to be fully described and evaluated.²⁴

Metzler's hypothesis, like other cyclical investment models of the multiplier-accelerator type, runs entirely in real terms and pays no attention to price adjustments and monetary and financial factors. Also, it fits best the finished-goods inventories subject to changes reflecting errors in sales forecasts. But studies which disaggregate inventories by stage of fabrication and type of production (Abramovitz 1950; Stanback 1962; Lovell 1964; Mack 1967; Zarnowitz 1973; Popkin 1984) show that for good reasons the behavior of stocks varies systematically between these categories. Thus finished-goods inventories are important primarily in production to stock; in production to order, which plays a very large role in durable-goods and particularly capital-goods industries, inventories consist mainly of goods-in-process, which depend positively on the rate of output, and materials, which are strongly influenced by cyclical changes in supply conditions (delivery periods, availability). The stocks of materials can be promptly adjusted in the ordering stage. Of course, it is difficult to allow for such details in aggregate models but an important lesson here is that desired inventories depend importantly on other variables in addition to sales.

2.3.6 The Role of Changes in Prices, Costs, and Profits

In his 1913 volume, Mitchell linked the major changes in business activity to the outlook for profits or (in time of crisis) the quest for solvency. Prospective profits depend on sales experience and expectations and on the price-cost relation, which is itself changing with the rate of employment and capacity utilization. Business costs tend to rise faster than product prices in the late stages of expansion, which depresses profit margins and expectations. Accordingly, new investment commitments are curtailed well before sales flatten. Income receipts and consumption expenditures weaken, inventories pile up, and production cuts multiply, particularly in durable-goods industries. Pessimistic expectations spread and are confirmed and worsened when output and employment turn down. In the contraction that follows, similarly, price-cost margins and profits first deteriorate and then improve, excess stocks and other imbalances are gradually liquidated, and new investment orders, sales, and output eventually revive.

When Mitchell first proposed a theoretical account of these developments, he had little empirical knowledge of them from the very inadequate data then in existence. By now, however, there is much evidence that the relations he stressed are generally consistent with the stylized facts discussed in section 2 above. (On their validity, see Fabricant 1959; Hultgren 1950, 1965; Kuh

24. Cycles are estimated by cosine functions applied to residuals from polynomial trends. The latter lack a theoretical rationale and are limited to the period of estimation. The paucity of annual observations presents another grave problem of statistical nature.

1960; Moore 1962, 1975a; Zarnowitz, 1973; Moore and Cullity 1983; Boehm 1990.) What is particularly well established and important is the typically procyclical but lagging pattern in labor costs per unit of output, which reflects primarily the positive conformity and lead times of labor productivity (output per hour of work). Real wages, on the other hand, normally do not show large deviations from trend that are consistently associated with business cycles.²⁵

In a 1967 model of Rose, employment and labor supply fluctuate relative to the stock of capital, which grows with net investment. The rates of change in wages (w) and prices (p) are equal in the long-run equilibrium but differ in the short run, reflecting these fluctuations. During a recession, the ratio of labor supply to capital increases, the rate of employment falls, and prices start rising faster than wages. The improvement in profitability leads to an upturn in the rate of employment. During the recovery, the condition $p > w$, persists, but gradually investment revives and growth of capital accelerates. First employment and then capital start growing faster than the labor supply. The upswing eventually causes wages to overtake prices. In the new phase, where $p < w$, the profit rate and the employment-capital ratio turn down. Investment is reduced, the rate of employment declines, and a new recession begins.

The model has debatable implications for the real-wage movements and its shortcomings are apparent, given the lessons of the recent inflationary era.²⁶ But all formal models are heavily restricted and the aspect covered here, namely, the cyclical role of changes in the relative input/output prices, is important enough to make the attempt interesting. Earlier theories of this type, although not worked out mathematically, are in some respects broader and more satisfactory.²⁷

The price-cost-profit nexus can and should be combined with monetary elements, since, as stressed by Mitchell, business cycles arise only in a "money economy" in its late stage of development and are incompatible with pure barter. (For one interesting attempt in this direction, see Rose 1969.) In mod-

25. Countercyclical movements in real wages are implied by the classical marginal productivity theory of the demand for labor, which was accepted by Keynes, and they are also suggested by the view that prices generally are more flexible and procyclical than money wages. However, the evidence is mixed and not conclusive. It varies with the choice of the deflator, the characteristics of the period covered, methods and data (e.g., the intracycle changes may not show up well in regressions with annual series). A few studies favor the countercyclical hypothesis (Neftegi 1978; Sargent 1978), but there is more support for both procyclical behavior (Dunlop 1938; Tarshis 1939; Modigliani 1977; Stockman 1983; Bils 1985) and no significant relationship between real wages and employment (Kuh 1960; Bodkin 1969; Geary and Kennan 1982).

26. Under perfect competition, Rose's hypothesis has real wages increasing in the boom and early contraction, decreasing in late contraction and recovery. Under imperfect competition, however, real wages could be either invariant or procyclical, depending on the elasticities of demand and marginal costs. A single nonlinear Phillips wage-employment curve is involved; shifts in it would have to be introduced, lest the model be applicable at best to a short period before any endogenous expectations of inflation changes develop and take effect.

27. Krelle (1981) demonstrates the similarity of Rose's theory to that of Erich Preiser (1933). The main difference is that Preiser had a two-sector (consumer-goods, producer-goods) model, whereas Rose has a simpler, one-sector model.

erate cycles, the effective limit on the volume of transactions is set by demand, whereas in “intense booms” a higher limit of monetary nature is reached.²⁸

2.4 Theories of Cyclical Response to Monetary and Real Disturbances

2.4.1 Exogenous Factors, Stochastic Elements, and Types of Theory

Section 2.3 discussed mainly the work of economists who attribute business cycles to the modus operandi of industrialized private-enterprise economies. Here the cycle itself is the principal source of the stresses and imbalances that keep it going. A nonlinear model that requires only a single initial disturbance to produce self-sustaining cycles has maximum endogeneity.²⁹

In reality, of course, the economy is always influenced by outside factors (e.g., weather) so that a comprehensive explanation of its motion cannot be purely endogenous. But no outside influences can by themselves produce the recurrent sequences of expansions and contractions; in the first place, this presumably requires the particular dynamics of an interdependent economic system. A really satisfactory theory, therefore, should explain how business cycles are generated by the internal mechanism of an economy exposed to the impact of a great many potentially relevant external events. What matters, then, is the relative role of the inside and outside factors, not the extreme cases. Nevertheless, a *mainly* endogenous model of business cycles differs in principle sharply from a *mainly* exogenous one.

The specific events and variables that are usually treated as exogenous include wars and changes in population, technology, weather, government spending, tax laws, etc. They clearly can and often do have major economic consequences that affect cyclical behavior. In addition to these factors, which usually show considerable persistence over time (are serially correlated), there are also the random shocks—uncorrelated disturbances of various kinds which impinge upon the structure of economic relationships. Both the white-noise and the identifiable exogenous factors play important roles in the linear and dynamically stable model of Frisch (1933). Here a low value for the accelerator is assumed but also a sufficiently close succession of erratic impulses which keep the system fluctuating. That is, the response of the economy to

28. For example, an expansion can be halted by the constraint on a further rise in bank credit imposed by the gold standard, as in Hawtrey (Mitchell 1927, ch. 2). On Mitchell's efforts to synthesize real, monetary, and expectational factors in viewing both the causes and effects of business cycles, see M. Friedman 1952; Zarnowitz 1968a, and ch. 5 this volume.

29. Fluctuations that are neither damped nor explosive but self-sustaining are simply not credible in linear models. It would be extremely unlikely, for example, for the accelerator always to assume that precise middle value which is needed to keep the system in a constant cycle. With random shocks imposed upon such a model, the cycles would increase over time (Samuelson 1947, pp. 268–69). This does not apply to nonlinear models, which can produce recurrent cycles with bounded variances, as shown for a Kaldor-type model by L. Klein and R. S. Preston (1969).

random but continual disturbances is such that what would otherwise be damped (fading) oscillations are converted into the recurrent business cycles. This hypothesis (which was suggested earlier by Wicksell and Slutsky) gained much recognition in recent theoretical writings and it particularly influenced the work of macroeconomic model builders.

Several econometric models of the U.S. economy in the post-world War II period have been found generally noncyclical in the absence of outside disturbances, as shown by simulation studies (I. Adelman and F. Adelman 1959; Hickman 1972). But random shocks applied to the more recent quarterly models proved to be insufficient to generate movements with the observed cyclical properties; to induce fluctuations in these models, it was necessary to use serially correlated disturbances (Zarnowitz, Boschan, and Moore 1972; Howrey 1972). Moreover, even the best simulations show only residual cyclical elements, much weaker than those found in the historical series used in the estimation of the models. This could be due to errors in either the structure of the models or the estimates of the disturbances or both.³⁰

It should be noted that the large macroeconomic models used primarily for short-term forecasting are frequently and extensively revised, and the more recent versions of them may well be substantially more cyclical. Simulation of one commercially successful model suggests that random noise from equation errors accounts for only 7% of an overall measure of "cyclicality" and that some two thirds of the latter would remain even with stable monetary policy, no financial disintermediation "crunches," and no oil price shocks (Eckstein and Sinai 1986).

The validity of the evidence from macroeconomic models which appear to refute the endogenous cycle and favor the random shock theory is, for these and other reasons, open to doubts. Thus Blatt (1978) constructed artificial time series on income, consumption, and investment based on Hicks's model with a high accelerator, implying unstable behavior. An econometric analysis of these data shows that they are seemingly well explained by a linear model with random shocks, which has a low accelerator yielding stable conditions. The problem is attributed to the limitations of linear models (Hicks's theory is essentially nonlinear).³¹

Business cycles interact with long-term trends in varied and subtle ways, so the separation of the two is difficult conceptually and empirically (see chapter

30. Parameters that may well vary with changes in policy regimes (Lucas 1976), the structure of the economy, and major departures from average cyclical patterns are estimated as constant by the models. Models estimated with data from periods with mild business cycles such as those of the 1950s and 1960s are known to be unable to reproduce much more violent fluctuations such as those of the 1930s (Zellner and Peck 1973).

31. Further, Blatt (1980) argues that the random shock theory is inconsistent with the evidence that deviations of many economic time series from smooth long-term trends show a pronounced asymmetry: the rises tend to be longer than the declines and also smaller per unit period. The measures are based on long monthly time series examined in trend-adjusted form by Burns and Mitchell (1946, ch. 7).

7). Decompositions using purely deterministic (say, log-linear) trends ignore such interactions, for example, the imprint that major cycles leave on the growth rates for some considerable time. Cyclical analyses based on deviations from such trends are suspect on statistical grounds (Nelson and Plosser 1982). The proposed alternative is to use stochastic trends approximated by random walks, but this means in practice that most of the contractions as well as expansions are included in such trend constructs, whereas the residual components labeled the “cycle” are largely pure noise (Beveridge and Nelson 1981). There is no good *economic* theory to justify this way of looking at the world.

The regularities reviewed in section 2.2 of this paper cannot be reconciled with the suggestion made by Fisher in 1925 and reviewed by McCulloch (1975, p. 303) that business cycles resemble “the cycles superstitious gamblers believe they can discern in their luck at casinos like the one at Monte Carlo.” As shown by McCulloch, once an expansion or contraction has exceeded its minimum historical duration, the probability of its being reversed in a given month is independent of its age. But that is merely a proof of non-periodicity, not of randomness, of the fluctuations called “business cycles.” Endogenous processes can still bring about the downturns and upturns, but they interact with all types of random and serially correlated outside events, which makes the timing of the reversals unpredictable (cf. Matthews 1959, pp. 199–205).³²

Certainly, the processes and relations economists study are in general stochastic, and purely deterministic explanations of macroeconomic movements cannot be sufficient. But purely stochastic explanations have no theoretical content, and it is the factors which can be integrated in an economic theory that are naturally of primary interest to any economist who attempts to understand the nature and causes of business cycles.

In the generally prosperous times after World War II, however, business cycles slipped way down in the public and professional interest. The weight of the public sector increased greatly throughout the industrialized world, and government actions and policies attracted growing attention as a likely source of large macroeconomic effects, notably the rising inflation. The idea that business contractions are also policy induced and episodic rather than a part of self-sustaining cycles seemed increasingly plausible.

Even while the style of macroeconomic analysis and policy remained predominantly Keynesian, the theory soon veered sharply away from the unstable accelerator-multiplier models to concepts that imputed much more stability to the private sector. This evolution shows up strongly in Duesenberry

32. Such processes include those arising from particular historical and institutional developments, for example, the “misintermediation” practice of banks and thrift institutions which McCulloch elsewhere (1977, 1981) asserts are a cause of financial instability and recurrent, if non-periodic, business fluctuations. Similarly, Irving Fisher’s debt-deflation hypothesis of the 1930s treats the business cycle problem much more seriously than his writings in the 1920s.

1958.³³ The notion of strong cyclical effects from high values of the accelerator and multiplier was further deflated by new theoretical developments: the permanent income and life-cycle hypotheses of consumption and the “neo-classical” models of investment. Used along with adaptive expectations reacting only gradually to past events, these formulations suggested relatively stable trends in private demands. The rise of monetarism (see next section) worked in the same direction.

Historically, the main substantive differences among the theories centered on the relative importance of real versus monetary factors. This can be linked to the distinction between the impulses and the propagation mechanism, introduced formally by Frisch in 1933. There can be models with monetary shocks and real propagation, models with real shocks and monetary propagation, and various mixtures of the two. Even in the largely endogenous theories it is sometimes possible to differentiate in a somewhat similar manner between the mainly originating factors and the mainly conditioning or response factors.³⁴

In sum, the several dichotomies encountered in modern business cycle dynamics intersect in various ways, and the extant theories actually represent many of the possible resulting combinations. They generally resist being neatly characterized by these categories, but even attempts at approximations can be instructive. A few illustrations for some theories, lightly treated above, are given in table 2.5. (See the first four cases; the others refer to materials to be discussed later.)

2.4.2 The Monetarist Interpretation of Business Fluctuations

In the 1960s, the rise of monetarism mounted a frontal challenge to Keynesian economics, starting from the simple quantity-theoretic proposition: changes in the stock of money are the main determinant of changes in nominal income. The demand for money is a relatively stable function linking real balances to wealth or permanent income and to expected rates of return on money and alternative assets. Given significant lags in wage and price adjustments, sequences of alternating phases of high and low growth rates in the quantity of money lead to corresponding fluctuations in aggregate demand and real economic activity relative to the secular trends. Sufficiently long periods of low but predominantly positive monetary growth rates are likely to produce business slowdowns or recessions; sufficiently long periods of negative monetary growth rates lead to depressions (Friedman and Schwartz 1963a, 1963b).³⁵

33. In this relatively disaggregate and complex theory, growth is explained by the interaction of a capital-adjustment process with autonomous investment, downturns by the operation of various exogenous factors, and upturns by the corrective forces inherent in the basically stable system.

34. The idea is found in Pigou 1927, p. 8, A. M. Hansen (1964, chs. 17 and 18) applies it to the work by Wicksell, Aftalion, Pigou, and J. M. Clark.

35. Of course, monetarism (like Keynesianism) has come to denote a broad assortment of theoretical concepts and empirical propositions attributed to economists who agree in some respects and disagree in others. Here there is need only for a selective treatment of these character-

The lags of output behind the monetary changes are seen as variable but, on the average, lengthy. This is so because, say, an accelerated increase in the quantity of money must first alter the relative prices or yields on a broad range of assets. The resulting discrepancies between the actual and the desired portfolios prompt the banks and the public to take corrective actions. The stimulus would eventually spread to product markets, causing rises in payments for services of, and the investments in, nonfinancial assets. As spending, income, and prices rise, interest rates will snap back from their earlier decline. If price expectations are adaptive, that is, subject to lagged error corrections, people will tend to underestimate the rising prices; hence they will overestimate their real money balances for some time. This will induce more transitory spending to liquidate the extra amounts of “redundant” money, and in the process, the rates of rise in prices and nominal income will overshoot the new equilibrium paths for these variables.³⁶ Thus, even in the case of a single shock, some cyclical (presumably damped) reaction may well occur.

The rate of growth of output or real income corresponding to full employment (or the “natural unemployment rate”) is exogenous in this model, being determined by real factors. The monetarist theory of macroeconomic fluctuations deals with deviations of output from this trend, that is, with “growth cycles.”³⁷

The early and still influential versions of this theory, due largely to Friedman, treat monetary changes—growth rates or deviations from trend of the quantity of money—as if they were predominantly autonomous, that is, having strong one-way effects on movements in total spending, income, and output. The reverse effects, from business activity to money, are recognized to exist but only as secondary “feedbacks.” The main source of critical monetary disturbances is thus located outside the private economy, in policy actions and institutional changes. Private expenditures, including business investment, are viewed as essentially stable, except when affected by the money shocks: when undisturbed, they tend to be consistent, in real terms, with the natural rates of employment and output.

This is a new emphasis. Earlier monetary theories generally gave most attention to private-sector instabilities, particularly credit fluctuations. The differences are highlighted in table 2.5 above (see also, section 2.3.1).

This brings us to the subject of the determinants of money supply. Under

istics inasmuch as they bear on the evolution of the work on business cycles. (For comprehensive surveys, see R. J. Gordon 1974; Stein 1976; Mayer et al. 1978.)

36. These paths would run parallel but higher than the new equilibrium path for the money stock, because at a higher rate of price rise less money in real terms would be demanded relative to wealth and income (Friedman and Schwartz 1963b; Friedman 1970b).

37. The equilibrium or trend level of output in this sense is associated with less than full utilization of resources; hence, it can be either higher or lower than the actual level of output at any time. This is unlike the concept of a “ceiling” imposed on output by full employment (as in Hicks 1950).

the present fiduciary standard, money consists of currency and, mainly, deposits in private banks. Monetary authorities can affect the quantity of money only indirectly and over time, by trying to control the monetary base (bank reserves plus currency). Subject to legal or regulatory constraints, banks vary their reserve-to-deposits ratios and the public their currency-to-deposits ratios in response to actual and expected changes in interest rates, real income, and probably some other indicators, (e.g., those of business and consumer confidence). In principle, then, a supply function for money in nominal units can be derived, involving the policy-related base and a few endogenous determinants of the money multiplier (ratio of the money stock to the base). The stability of this function is an empirical question on which there is significant disagreement.³⁸ However, there is a substantial consensus, not only among monetarists, that any effects that interest rates, wealth, or income may have on the nominal money supply are much weaker and less consistent than the effects of these variables on the demand for real balances. Moreover, monetarist studies argue that central banks have the power and tools to exercise the dominant influence upon the money stock, except in the very short run.

The potential for an autonomous monetary policy is generally overestimated by analysts who concentrate on the United States, the only large market economy in which balance-of-payments considerations could long be treated as secondary even under fixed exchange rates. For any small economy under this regime, the nominal money supply depends on changes in the available foreign exchange reserves; it must tend to be consistent with prices and incomes that will balance the country's international payments and hence cannot be determined independently by domestic authorities. Under the gold standard, a business expansion would come to a halt when the drain of cash and shortage of gold reserves forced the banking system to curtail credit on a sufficiently large scale. This was the ultimate source of the relatively large and regular British cycles in the nineteenth century for Marshall, Hawtrey, and Lavington.

An early expectation that the monetarist approach may ultimately produce a theory of "a partly self-generating cyclical mechanism . . . including a feedback in the rate of change in money itself" (Friedman and Schwartz 1963b, p. 64) has remained unfulfilled. The feedback, or "reflex influence," running from business activity to monetary growth has not been analytically developed and integrated into a theory of how money, prices, and real factors interact in the short run. The approach produced instead an essentially

38. Cagan (1965) shows that high-powered money fluctuated more erratically than, and often opposite to, the currency and reserve ratios. One source of the interrelation is indirect—common response to business cycles; another is direct—central-bank operations designed to stabilize the economy. He finds that "the dependence of the money stock on prices and business activity, as well as on other variables, is strong but is neither rigid, uniform, nor immediate" (p. 16). Friedman and Schwartz stress that "neither interest rates nor real income have a consistent and sizable influence on the nominal quantity of money supplied" (1982, p. 35). On the other hand, Brunner and Meltzer argue that an empirically stable money supply function exists (1968, 1972).

Table 2.5 A Synopsis of Selected Business Cycle Theories

Type of Theory (1)	Main Factors		Most Sensitive Processes (4)	Are Cycles Linked to Growth? (5)	Special Features (6)	Author and Dates (7)
	Originating (2)	Responsive (3)				
<i>I. Some Largely Endogenous Theories</i>						
Monetary disequilibrium	Unstable flow of money (bank credit)	Interest rate changes; cycles of inflation and deflation	Investment in traders' inventories	No	Cycles tend to be periodic under the gold standard	Hawtrey 1913-37
Monetary overinvestment	Unstable supply of bank credit	Discrepancy between the natural and money interest rates	Capital investment, lengthening and shortening of production processes	No or weakly	Real vertical maladjustments result from monetary disequilibria	Hayek 1931-39
Cyclical real growth	Bursts of innovation (new products, markets, etc.) contested by imitators	Credit financing; excesses of speculation and misjudgment	Business capital investment booms and readjustments in contractions	Yes	Simultaneous interacting long, intermediate, and short cycles	Schumpeter 1912-39
<i>II. Some Theories with Major Exogenous and Stochastic Elements</i>						
Impulse and propagation in a real model	Undefined erratic shocks and discontinuous Schumpeterian innovations	Investment accelerator, lags in output of capital goods, money demand, and imperfectly elastic supply	Capital-goods production, but the system as a whole is damped (dynamically stable)	Yes (through innovations)	Random shocks or innovations bunched in expansions needed to maintain oscillations	Frisch 1933

The original monetarist theory	Sequential shocks: high monetary growth rates followed by low rates, etc.	Relative prices and asset yields, then spending flows	Both consumption and investment react to monetary changes	No	Monetary policies destabilize the private sector	Friedman and Schwartz 1963a, 1963b
Market clearing with rational expectations and incomplete information	Random monetary shocks causing price-level variations	General price changes misperceived for relative price changes; intertemporal substitution of labor and leisure	Prompt and strong reactions to perceived changes in relative prices or real rates of return on the supply side	No	Flexible prices and wages clear markets continuously; money and price sur-prises cause fluctuations in output and investment	Lucas 1977
A disequilibrium theory of investment and financial instability (largely endogenous)	Unstable expected profits drive business investment, which generates fluctuations in realized profits	Money created by bank lending to business; short-term financing of long-term investment	Relative prices of capital assets set in financial markets under uncertainty about future re- turns, costs of capital, and cash flows	Yes	Long expansions produce over-confidence, unsound financing practices, a growing debt burden and illiquidity . . . sources of contractions and crises	Minsky 1982

exogenous and monetary model of the business cycle. Yet the same studies that show an important, independent role of money, particularly in major inflationary booms and severe deflationary slumps, also find much evidence that fluctuations in the monetary variables reflect those in real aggregates, particularly in mild cycles (summarized by Cagan 1965, p. 294). The primacy of the monetary effects cannot be established by the less-than-compelling argument that since money plays a key role in the major cycles, it should also be important in the minor cycles, which are just "less virulent members of the same species" (Friedman and Schwartz 1963b, p. 55).

The idea that cyclical instability is mainly policy induced is not an integral part of, or a necessary inference from, the monetarist theory. It is an empirical judgment held by some monetarists as well as other economists.³⁹ The popularity of this view rose in the late 1960s and 1970s, when both inflation and unemployment drifted upward, the Phillips curve moved north-to-northeast through a series of clockwise cycle-related loops, and failures of attempted stabilization policies were evident. Rather paradoxically, the 1970s also witnessed the culmination of belief in the power of macroeconomic policies that took the conceptual form of a "political business cycle," which is not of monetarist origin.⁴⁰

Monetarist models can acquire some cyclical dynamics by combining an expectations-augmented Phillips curve with the quantity-theoretic demand for money function. Laidler (1973a) presents an intentionally simple model of this type, with exogenous monetary growth, the full-employment output rate, and adaptive expectations of inflation. The causation runs from lagged and current money growth rates and current inflation to the change in output, then back to inflation, and so on, recursively. The model generates a cyclical movement in output and inflation (relative to their natural and expected values, respectively) in response to a single shift in the rate of monetary expansion. Laidler notes that its simulations leave much to be desired, probably in large part because of the missing variables, particularly the interest rate.⁴¹

39. Consider the following statement by authors who contributed much to the development of the money supply function: "Our version of monetarism does not deny that if exogenous shocks take the form of government policies, including fiscal policies, the system may oscillate or even explode into inflation or cumulative deflation. Our proposition asserts that cyclical instability is mainly the product of government policies that are imposed on a stabilizing private sector" (Brunner and Meltzer in J. Stein 1976, p. 180).

40. The political models of business cycles maintain several strong and questionable assumptions. (1) Policymakers know the structure of the economy; the public does not. (2) The inflation-unemployment short-run trade-off can be exploited effectively by the party in power. (3) Voters are myopic and concerned with the current electoral period only. Thus the government is capable of fine-tuning the economy periodically. It is not surprising that these models, though ingenious, find little support in facts. The related evidence is mixed and generally inconclusive: Nordhaus 1975; Lindbeck 1976; MacRae 1977; also Fels 1977; Moore 1977a; Tufte 1978. A very useful, balanced overview is given in Alt and Chrystal 1983, esp., ch. 5.

41. For a corresponding theoretical analysis of a model for a small open economy under fixed exchange rates, see Laidler 1975, ch. 9.

2.4.3 Price Misperceptions and Intertemporal Substitution of Labor

Adaptive expectations often imply long lags of adjustment and persistence of apparently systematic errors. Critics regard models that yield such results as ad hoc and inconsistent with optimizing behavior. They accept instead the view that "expectations are essentially the same as the predictions of the relevant economic theory" (Muth 1961, p. 316). Application of this rational expectations hypothesis (REH) to macroeconomics was part of a new, ambitious program of work undertaken in the 1970s by several economists following the initiative of Lucas (1972). The objective was to develop a general business cycle theory in strict adherence to the basic principles of the analysis of economic equilibrium: consistent pursuit of self-interest by individuals and continuous clearing of all markets by relative prices.⁴²

Under REH, the route of the older monetarist theory which used adaptive expectations to help explain the duration of cyclical movements is closed. Expectations are now taken to be free of any bias and subject to random errors only. All persistent monetary changes, inasmuch as they are predictable, will be correctly anticipated and met directly by proportional changes in prices and related nominal variables. Only random monetary impulses can lead to price surprises and miscalculations, which, in this view, are necessary to explain any cyclical movements in real variables.

The short-run aggregate supply function for labor (Lucas and Rapping 1969) is upward sloping relative to the deviations of the current level of real wages from their expected (discounted future) level. It is seen as typically elastic, reflecting strong competitive incentives to take advantage of temporarily higher rates of real return. Since the substitutability of leisure over time is high, a small change in the return on the current work effort can induce a large change, in the same direction, in the amount of work done. This intertemporal substitution hypothesis (ISH) plays a central role in the recent attempts to explain employment fluctuations as an aggregate result of individual choices on the supply side of the labor market.

By an analogous argument, firms are expected to vary their output positively in response to transitory changes in their selling prices, provided that these are seen as relative or real price changes and not equated with general inflation. But they would likewise vary their output inversely in response to transitory changes in the relative prices of their inputs, particularly labor. The basic equilibrium model of business cycles disregards this complication by

42. See Lucas 1973, 1975, 1977; Sargent and Wallace 1975, 1976; Sargent 1976a; Barro 1976, 1980; Lucas and Sargent 1978. Lucas (1977, p. 7) cites Hayek (1933) as an "intellectual ancestor" who posed the problem of explaining the business cycle as part, not a contradiction, of the equilibrium theory. This was indeed Hayek's intent, but it is also correct to characterize his solution as a theory of monetary disequilibrium and an unstable cumulative process, with excessive credit creation causing distortions of relative prices and the structure of production (as Hayek 1933 and 1939 are commonly interpreted).

simply combining workers and firms into a single group. The representative worker-entrepreneur then generally supplies more (less) of both labor and output when faced with an unanticipated rise (fall) in selling prices.

This approach, by eliminating other prime suspects of earlier business cycle theories (including real disturbances, which are viewed as dispersed and localized), places a heavy explanatory burden on a single causal chain: random monetary shocks induce price misperceptions, which induce wrong production decisions. By assumption, prices other than those in one's own market are known to anyone only with a lag of one unit period of unspecified length. Agents have complete and timely local information but only incomplete and lagging information about other "island" markets (Phelps 1970) and about economy-wide aggregates such as the money stock and the overall price level. Suppose now that an unanticipated acceleration in monetary growth occurs, raising prices in general; then the representative worker-entrepreneur first observes a higher selling price than he expected, takes it to be in some part a temporary increase in his relative price or real rate of return, and raises his output in accordance with the ISH. These reactions prevail whenever the observed prices turn out to be higher than the level most producers had expected on the basis of previous information. In the opposite situation of prices having proved lower than expected, output is, on the average, reduced. The random forecasting errors are unavoidable and can be recognized only after the outside price data become available. However, by then many erroneous decisions will already have been made, and the necessary revisions and corrections will also involve time and costs.

The model of the "representative producer," with its fusion of worker and employer, assumes that labor and business recognize that their interests generally coincide, or at least act as if they did. This is not consistent with the strong evidence of significant patterns of cyclical behavior in cost-price-profit relationships (secs. 2.2 and 2.3.5). These observations can best be interpreted with the commonsense assumption that firms and workers pursue their own interests in reacting rationally to changing business and labor market conditions.⁴³

The issue of the relative timing of output prices versus input costs can be reduced to an informational problem in a number of ways. It might be assumed that firms have more immediate knowledge of price changes than workers or, more generally, that the representative producer unit (firm or worker) knows the prices of things it sells better than the prices of things it buys (Friedman 1968). If so, then inflation will be stimulative because it is largely unanticipated by workers or buyers (or, which is much the same, because it is recognized sooner in output prices than in input costs). Some critics

43. To be sure, the interests of workers and employers may coincide in some respects and are reconciled in negotiated or implicit contracts (sec. 2.4.7).

view these assumed informational asymmetries as arbitrary specifications (B. Friedman, comment on Lucas and Sargent 1978, p. 76; Tobin 1980, p. 42).

The equilibrium approach to business cycles can be explained well in general terms (Lucas 1977) and restated simply. (See the text above and the capsule description in table 2.5.) The individual models based on the REH and the ISH, however, are much too diverse, experimental, and complex to lend themselves to such verbal summarization. Some of them use changes in current prices relative to the next period's expected price level to induce intertemporal substitution on the supply side (Lucas 1972, 1973; Sargent and Wallace 1975). Others allow for the existence of assets that earn a nominal interest rate and add that rate to the above price surprise term to obtain a measure of anticipated one-period real rate of return. This relative price variable then appears with a positive sign as a determinant of supply and with a negative sign as a determinant of demand (Barro 1980).⁴⁴

The general criticism of the price misperception hypothesis is that it requires long informational lags, which are even less likely under rational than under adaptive expectations. Ample, frequent, and low-cost monetary and price statistics are now available, so informational confusion of the type here hypothesized can only be short lived and associated with random changes, not persistent cyclical fluctuations, in output and employment (Hall 1975, 1980b; Tobin 1977; Modigliani 1977). True, this argument is partly countered by the observation that there are indeed serious deficiencies in these and other important data on the economy, which in many cases are reduced only through a time-consuming sequence of revisions. This can distort initial expectations and delay successful signal detection for several months (Zarnowitz 1982c). Still, informational lags are surely much shorter than the average cyclical movements, so they alone cannot account for the duration of the observed fluctuations (Okun 1980) or, one may add, for the large size of procyclical fluctuations in corporate profits and stock prices. Beginning with an economy at full employment,⁴⁵ most errors caused by temporary misperceptions of monetary and price changes would be detected and corrected before they could give rise to large cumulative income movements in either direction. Finally, the knowledge of the nominal interest rates, a set of timely and global variables, may convey information about the unobserved part of money growth (King 1983; Barro 1980).⁴⁶

44. The net effect of a rise in this composite variable, then, will be to stimulate output if on the aggregate across the markets, the induced increase in supply is larger than the induced decrease in demand. This formulation takes into account the debate about the direction of informational asymmetries noted in the preceding paragraph of the text.

45. It is against the spirit of the new equilibrium theory (and some older scholars such as Hayek and other Austrians) to start an attempted explanation of business cycles by postulating an initial state of recession.

46. This prompts Barro (1981a, p. 51) to observe that the "stress on confusions between temporary and permanent monetary shocks has been overdone. The real effects of temporary, but

The criticism of the ISH centers on observations said to be inconsistent with continuous equilibrium in the labor market (e.g., Okun 1980). Thus during recessions and depressions, indefinite layoffs and involuntary separations account for most of the rise in unemployment. More people are looking for work at current (or even lower) wages over longer average time periods. Fewer people quit their jobs as vacancies drop.

General critiques, however, have limited power of persuasion, particularly against strong priors of the economic equilibrium theory. It is therefore particularly important that the hypotheses under consideration have been subjected to various tests, in large part by their proponents.

In one set of tests, the reaction function of monetary authorities was estimated by regressing the rate of growth in money on its own past values and selected lagged variables, and identifying the *residuals* from this regression with the “unanticipated” component of monetary change (Barro 1977b, 1978; Barro and Rush 1980). These tests could not reject the joint hypotheses of rationality and neutrality of money, but many doubts were raised about the specification and identifiability of Barro’s reaction function as well as its consistency with private *and* public rational behavior.⁴⁷

The neutrality hypothesis that anticipated money growth has no real effects is strongly rejected by tests reported by Mishkin (1982). Here, anticipated as well unanticipated money growth influences output and unemployment, with lags of up to 20 quarters.⁴⁸

Data on monetary aggregates are available promptly and often (now weekly) but they are also repeatedly revised. The revisions are frequently large relative to the average rates of money growth, but on the whole, they appear to be random. They are interpreted as “components of unperceived monetary movements” and found to have no significant real effects (Barro and Hercovitz 1980). Economic agents should not be assumed to be ignorant of the current monetary values for which they do have usable approximations. And, contrary to the RE-IS models, in which prices are fully flexible, output is in fact positively associated with these measured and knowable values of the money stock (King 1981; Boschen and Grossman 1982; McCallum 1982).

The early tests by Lucas and Rapping (1969) favor the ISH but they are based on adaptive expectations. When reestimated by Altonji (1982) under

perceived, money shocks would be eliminated by the appropriate adjustment in the nominal rate of return.”

47. See the comments by Blinder, R. J. Gordon, Weintraub, and Fischer in S. Fischer, ed. 1980, pp. 49–72 and 219–21. On the related basic problems of “observational equivalence” and testability, see also Sargent 1976b; and Sims 1980b.

48. It is these longer lags that are primarily responsible for Mishkin’s conclusion being the opposite of that reached in the Barro papers, where lags of 2 years or 10 quarters are used. No attempts to rationalize the persistence of such long distributed lags under the REH are made in any of these reports. See also Nelson 1981 on the dependence of unemployment on lagged values of nominal GNP.

rational expectations and a variety of alternative assumptions about agents' information sets, the results generally fail to support the ISH.⁴⁹ Weak negative effects of a price surprise term on the unemployment rate are reported by Sargent (1976a) but disputed by Fair (1979); there are simultaneity problems with these estimates, as shown by Barro (1980). These and other tests (Hall 1979) are admitted to be rudimentary and on the whole inconclusive; all participants in this work stress that it presents great difficulties. Still, for the most part, other recent tests addressed to certain manageable aspects of the problem have produced negative results (K. Clark and Summers 1982; Mankiw, Rotemberg and Summers 1982).

To sum up, the evidence can be fairly described, on balance, as unfavorable to the theories here considered. This has led to some reassessments on the part of their authors. Thus, Barro (1981a, p. 74) expresses "doubts about the explanatory value for business cycles of currently available equilibrium theories."⁵⁰ McCallum (1982, p. 4.) argues that the evidence requires abandonment of "flexible-price equilibrium models" but not of the "*equilibrium approach* to macroeconomic analysis," which can rationalize sticky or slowly adjusting wages and prices.

2.4.4 Cyclical Persistence and Extensions of the Equilibrium Models

Can random monetary shocks produce *persistent* fluctuations of real aggregates in an economy with market clearing and incomplete information? Those who deny it attack mainly the ISH as leading to misguided attempts to represent business cycles as a "moving equilibrium" (Modigliani 1977; Tobin 1977, 1980; Solow 1980). Some expect that the postulate of continuous market clearing will have to be abandoned in the RE models (Okun 1980; R. J. Gordon 1981).

Equilibrium theorists recognize that the basic ingredients of their models are not sufficient to produce the persistent movements in output and employment which occur during the business cycle. However, they point out that random shocks to aggregate demand can be converted into persistent movements by suitable propagation mechanisms, as in Frisch 1933. Further, Lucas and Sargent (1978, pp. 65–67) argue that some devices of this type have already been incorporated in the equilibrium models, and others are likely to be developed.

The first of these propagation mechanisms relates to the familiar observation that rapid adjustments of employment and production rates are costly.

49. "For most specifications, the current real wage, the expected future real wage, and the expected real rate of interest are either insignificantly related to unemployment and labour supply or have the wrong sign" (Altonji 1982, p. 784; Altonji and Ashenfelter 1980).

50. He hastens to add that these doubts "do not constitute support for Keynesian disequilibrium analysis," which is incomplete and even more questionable. His argument implies that there are only two sides to the debate, the "new classical macroeconomics" and "Keynesian macroeconomics." This is a widespread but mistaken and, in my opinion, much too restrictive point of view.

Accordingly, firms respond with lags to the relative price signals they perceive.⁵¹ In practice, modeling the cost-of-adjustment effects takes the form of making the demands for factors of production depend partly on their own lagged values. It is important to note that this mechanism is entirely different from and extraneous to that of the random price misperception effects. The latter should not be spread over time by the old device of distributed lags. The unit period here is defined by the lag of data on the nominal aggregates which was already shown to be relatively short. The logic of this approach seems to leave no good reasons for extending the lag to more than one period and so opening up the possibility of autocorrelated forecast errors. Cost-of-adjustment models may provide a rationale for more complex and longer lags. However, in a world without uncertainty about the probability distributions governing the future (see next section), where markets clear continuously, leaving no unexploited opportunities for gain, there should be little ground for any sizable distributed lags in economic decision making (Poole 1976).

Another propagation device incorporates a form of an investment accelerator (Lucas 1975). Positive price surprises induce not only increases in current employment and output but also acquisition of additional capital. Capacity is supposed to increase promptly during a period when the nature of the shock is not yet recognized. This increment to the stock of capacity raises labor productivity and temporarily increases the demand for labor and the supply of commodities; it also retards the general price increase, thereby delaying the recognition of, or adjustment to, the initial shock. Thus a persistence effect is created.

This formulation also poses some major problems. Many capital investment projects involve indivisibilities, high costs and risks, and gestation periods measured in years and quarters, not months. Decisions to initiate them are unlikely to depend on isolated signals that could well prove false after a short information lag (R. J. Gordon 1981, p. 510). Not surprisingly, random monetary disturbances and price misperceptions have not attracted much attention in the literature as potentially important determinants of investment; instead, expectations have—based on systematic changes in demand, profits, credit, and cost of capital. In this class of models neglect of these factors deprives the treatment of business capital investment of much interest.

Furthermore, because of the relatively long ordering and construction lags, investment will tend to add to demand before it adds to capacity. The effects of increased capacity on supply, prices, and the anticipated real rate of return on capital are delayed. It must be recalled, too, that capacity itself normally continues to increase during recessions; it is investment and the rate of capacity utilization that are highly sensitive, cyclically. In models that concentrate

51. It should be noted that such lags could also rationalize a dynamic model of fluctuations in employment and output that is purely "real", that is, independent of the behavior of money and prices (Sargent 1979, pp. 370–79).

on monetary and price effects on the stock of capital, which is presumably optimally utilized throughout (as is the stock of labor), the role of the reformulated "accelerator" would seem to be quite limited.

In models with lagging information, the effects of purely random monetary-price shocks do not cumulate: the responses peak in the first period and decline gradually thereafter (Lucas 1975, p. 202). This feature is not changed by the introduction of investment as long as new capital is installed with a lag of one period only. But these timing specifications are very unrealistic. There is strong evidence that investment in plant and equipment requires, on the average, about seven quarters to complete, with few projects taking less than 1 year (Mayer 1960; Zarnowitz 1973, pp. 505-19; Hall, 1977).

Kydland and Prescott (1980, 1982) use the "time to build" new capital stocks as a feature of technology which dictates the number of periods needed to produce durable producer and consumer goods. These lags are treated as policy invariant and constant for a given type of capital. (This is unlike the observed delivery lags on these goods, which vary procyclically with backlogs of orders: Popkin 1965; Zarnowitz 1973, pt. 3; Carlton 1979.) Time to build contributes to the persistence of output movements over the multiple periods required to produce the finished capital goods (unfinished goods are not part of the productive stocks). The models used in this work rely on an intertemporal labor supply function and are driven basically by real shocks that affect technology and the productivity of the representative worker.⁵²

Investment realizations are a distributed-lag function of investment plans. But the series of new capital appropriations and contracts and orders for plant and equipment that proxy for investment plans are also serially correlated like the expenditures that follow them, only less so (e.g., Zarnowitz 1973). Hence they lack the random shock property which the equilibrium theorists look for in "an essential propagator of business cycle fluctuations" (John Taylor in Fischer 1980, p. 192). Nevertheless they are an important link in the cyclical process. The lag of investment expenditures behind investment decisions is an essential element in several otherwise quite different theories of business cycles (Mitchell 1913; Kalecki 1935; Hicks 1950).

Inventory investment provides still another potential channel for persistence effects in equilibrium models. In Blinder and Fischer (1981), an unanticipated rise in money and prices leads firms to sell out of inventories at the same time as they increase output. In subsequent periods, production is gradually raised to restock the depleted inventories. More specifically, inventory investment depends positively on the excess of the desired stock ($N_i^* - N_i$) and inversely on the price surprise term ($p_i - p_{i-1}^e$). The aggregate supply function

52. The real shocks have permanent (autocorrelated) and transitory components; the former also contribute to the cumulative movements or "momentum" in employment and output. The 1980 paper by Kydland and Prescott includes random shocks to nominal wages as well, but these monetary disturbances are secondary. The 1982 model contains real shocks only.

has Q_t vary positively with Q_t^N and the same price surprise (as in Lucas 1972, 1973) but now also with $(N_t^* - N_t)$. The desired inventory N_t^* will stimulate activity, but a rise in N_t relative to N_t^* will discourage it. Even with N^* a constant, this model can account for some serial correlation of output. With interest-sensitive N_t^* , larger fluctuations would result since even fully anticipated changes in money would have some real effects (Blinder and Fischer 1981, sec. 5).

This approach draws on some old ideas about the aggregate sales-inventory-income nexus in business cycles (sec. 2.3.5) and combines them with elements of the new equilibrium models for a rather uneasy match. It needs to be recalled, too, that the role of inventory adjustments is large during mild recessions and slowdowns, such as prevailed in the post-World War II period, but is otherwise supporting rather than central (Blinder and Holtz-Eakin 1984).

Summing up, the cost-motivated adjustments lags, durability and long gestation periods of capital goods, and desired inventory effects have all been long recognized as important in studies of business cycles. These elements, however, do not exactly mesh with the basic core of the equilibrium model. The random monetary shocks and price surprises have smaller and less intelligible parts to play, while the real factors in the "propagation" processes move to the center of the stage. These extensions, indeed, "may undermine the quantitative role of the underlying intertemporal substitution mechanism as the basis for fluctuations in output and employment" (Barro 1981a, p. 49).

2.4.5 Rationality, Knowledge, and Uncertainty

To be "rational" in a technical sense, expectations must be consistent with the structure of the given model. Unless they are self-fulfilling, on the whole, the model of behavior assumed to be ruled by them is vitiated. Given the relevant information set, it may be possible to solve a RE model for its equilibrium path over all futures. As new information becomes available, the forecast-solution is updated. Used as a principle of modeling dynamic stochastic equilibria, RE gained wide popularity and produced important new insights.⁵³

However, even in relatively simple linear models, the computational problems posed by this radically logical approach are often formidable. Using RE as a model of actual behavior cannot mean imputing to economic agents generally the ability to solve such problems. Rather it is the markets that are supposed to work as if they somehow approximated this capacity. This could

53. These include the criticism of some aspects of macroeconomic models, the analysis of competitive markets with imperfect information, and work on the consequences of endogenous expectations for the effectiveness of economic policies (Lucas 1976; Poole 1976). These matters lie outside the scope of this chapter. Assessments of the RE models abound (Shiller 1978; Kantor 1979; Fischer 1980; McCallum 1980b). For surveys of the literature and references, see Begg 1982; Sheffrin 1983; and Frydman and Phelps 1983.

be interpreted along the lines of what may be called a weak version of the REH: market incentives and penalties favor the dominance of optimal or cost-efficient predictions. Firms and individuals whose forecasts are consistently poor will not be able to survive economically. The anticipations of those who do survive will tend to come true. This implies that predictively valuable information, on which such anticipations are based, is scarce so that collecting it is a profitable activity (Grossman and Stiglitz 1980).

Such propositions, when applied to well-functioning individual markets, seem simply good, standard economics. In the aggregate, they need not imply more than a long-run tendency toward equilibrium, promoted by learning from experience but inhibited by limited opportunities for controlled experiments. Thus no firm link is established between the type of model and the expectational hypothesis used. In particular, RE models have been built without the property of continuous market clearing. In principle, various types of disequilibria could be expected by people and modeled with the aid of RE methodology; or, if expectational confusions prevail, no unique solutions should be found.

The REH of the macroeconomic literature in the 1970s, however, is a strong version which adds to the reasonable premise of rational use of costly information another assumption—namely, that the available data and models provide sufficient knowledge about the future such that the prevailing expectations are free of any systematic errors and consistent with *continuous* aggregate equilibrium. Economic behavior is guided by subjective probabilities which agree, on the average, with the true frequencies of the events in question. Unlike in Knight 1921 and Keynes 1936 there is no uncertainty as to what these objective probabilities are.⁵⁴

Now there seem to be no good a priori reasons why this should generally be so; indeed, the belief that it is not accounts for most objections to the REH. It is evident that there is no agreement on what is the “objective” probability distribution of future outcomes for the economy at any time, since different theories and models coexist. Another reason why forecasts differ across people, firms, etc., is that skills to acquire and use information are not evenly distributed.

Attempts to form rational predictions of any macrovariable that depends on anticipatory actions of many or all agents in the economy involve adjustments through a learning process in which not only the individual forecaster’s own beliefs but also those of others are continually evaluated. Each agent, then, tries to predict the average forecast, or what others are likely to predict that average will be, and so on. This is the difficult “infinite regress” problem, well known from the “beauty contest” example of Keynes (1936, p. 156). Al-

54. Indeed, Lucas (1977, p. 15), argues that “in cases of uncertainty, economic reasoning will be of no value.” For the REH to apply, business cycles must represent “repeated instances of essentially similar events.”

though increasingly and ingeniously attacked, this problem is still far from being fully tractable or understood. However, an important result that is strongly suggested by this work is that a unique, stable RE equilibrium path, along which prices continually clear all markets, entails the collective consistency of individual plans. But individual rationality (the cornerstone of modern microeconomic analysis) does not necessarily imply such a consistency.⁵⁵

In a stochastically stationary environment that has persisted long enough to be familiar, agents are assumed to have learned all they can about the probability distributions they face.⁵⁶ In the economy as it is, however, change is to a large extent unanticipated and learning is perpetual. In models with learning and disparate expectations, convergence to the RE equilibrium requires that agents know no less than the laws governing the change in the key parameters of the economy and the effects of exogenous shocks. Learning itself can act as a cyclical propagator mechanism by including serial and cross-correlations in forecasts (Townsend 1978, 1983a, 1983b).

The strong assumptions of prior knowledge are implausible but there seems to be no alternative to them that would be satisfactory from the RE point of view.⁵⁷ Some simple but flexible rules of adaptive expectations (AE) may be consistent with optimal learning (Taylor 1975; B. Friedman 1979) and some models with a common, simple rule converge to an equilibrium solution (e.g., Bray 1983). But the collective adherence to a rule which, if individually followed, would yield biased forecasts is in conflict with the idea of the optimizing representative agent in a decentralized market economy.

It is important to distinguish between the critique of the RE and the critique of the particular market-clearing models with RE (e.g., Fair 1978). Many critics agree that the pre-RE treatments of expectations are generally arbitrary and, at least in principle, inferior (Tobin 1980, pp. 28–29; Meltzer 1982, p. 3). Further advances in the intensely used and studied RE methodology are to be expected (as argued, e.g., in Lucas and Sargent 1978; Lucas 1981; Grossman 1981). Still, the claim to general validity of the strong form of the REH as applied to market-clearing macromodels is now rejected by a wide range of economists (Arrow 1978; Tobin 1980, 1981; Laidler 1981, pp. 11–15; Friedman and Schwartz 1982, p. 630).

In the present context, the critical questions concern the stationarity and predictability of the processes observed during business cycles. Their recurrent and sequential nature is indeed well established, but so is their lack of periodicity and the large intercycle differences in durations and amplitudes (sec. 2.2). The separability of business cycles from the long trends can by no

55. See the several essays and the introductory chapter by the editors in Frydman and Phelps 1983.

56. Most of the early basic RE models, constructed on this premise, simply contain no learning processes.

57. Without some such assumptions, convergence may not occur even in single-market contexts (Cyert and DeGroot 1974; B. Friedman 1979; DeCanio 1979).

means be taken for granted. These are arguments against the applicability of the RE methods. In a nonstationary world with a mixture of random and serially correlated disturbances, uncertainty in the sense of Knight and Keynes is pervasive, even under the (empirically dubious) premises of no structural change and stable policy regimes.

The monetarist models with AE may be interpreted to contain uncertainty.⁵⁸ Meltzer (1982) distinguishes between uncertainty, which is associated with variations in nonstationary means resulting from “permanent” changes in levels or growth rates, and risk, which is associated with transitory, random deviations around stable trends. He argues that the current RE models err in ignoring uncertainty, which is essential to an explanation of the persistence of cyclical contractions. Models of business cycles should allow permanent changes to occur but not to be identified immediately. Stochastic shocks, whether monetary or real, have permanent and transitory components which cannot be reliably separated (new information reduces but does not eliminate the confusion). The rational response to the shocks is adaptive, taking the form of gradual adjustments of beliefs about the permanent values of the endogenous variables.⁵⁹

What is the evidence on how people actually form their expectations? Recent work using survey data has been preoccupied with tests for the rationality of inflation forecasts. Typically, actual values are regressed on predicted values, and the forecasts fail the tests when (1) the sample intercept and slope estimates are significantly different from 0 and 1, respectively, and/or (2) the residual errors are significantly autocorrelated. Data from the semiannual surveys of economic forecasters conducted since 1947 by Joseph A. Livingston, a syndicated financial columnist, reject the REH according to most of these and related tests (Pesando 1975; Carlson 1977; Wachtel 1977; Moore 1977b; Pearce 1979; Gramlich 1983; more favorable results are reported in Mullineaux 1978). Tests of individual forecasts confirm those of the group average forecasts, and the results from the quarterly NBER-ASA economic outlook surveys are consistent with those obtained for the Livingston surveys (Figlewski and Wachtel 1981; see also chapter 16).⁶⁰

Studies of consumer survey data from the Survey Research Center of the University of Michigan show mixed but, in large measure, negative results

58. See M. Friedman 1972, pp. 923–24, on uncertainty; Muth 1960; Frenkel 1975; Mussa 1975; Brunner, Cukierman, and Meltzer 1980; Friedman and Schwartz 1982, pp. 415 and 447, on the rationality and empirical usefulness of AE.

59. Note that this is a general approach, compatible with Keynesian, monetarist, and other theories. It is used in the 1980 model of Brunner et al., in which monetary shocks affect only the price level and the rate of interest, while unemployment is caused by errors due to the inability of workers to distinguish between permanent and transitory real productivity shocks. This model is subject to all doubts concerning the hypotheses of complete neutrality of money and intertemporal substitution in labor supply.

60. The relevant literature on the properties of expectational data is voluminous and can be only briefly summarized here. For a more comprehensive review of the evidence, see Zarnowitz 1984b.

(Juster 1979; Huizinga 1980; Curtin 1982; Gramlich 1983). The same applies to surveys of business executives (de Leeuw and McKelvey 1981) and European and Japanese surveys (Aiginger 1981; Papadia 1982; Visco 1984).⁶¹

Few authors have tested data on anticipations for variables other than inflation. U.S. manufacturers' sales expectations have been on the whole negatively assessed in a comprehensive study by Hirsch and Lovell (1969), more positively by Pashigian (1964) and Irvine (1981). For professional economic forecasters, including econometric service bureaus, there is strong evidence that predictions of inflation in the 1970s have been generally biased, while those of other important aggregates (growth in nominal and real GNP, the unemployment rate) pass the rationality tests much more frequently and better (see chapter 16; also McNees 1978).⁶²

The REH finds the strongest support in the "efficient markets" literature (Fama 1970; Poole 1976). This is readily understandable since financial assets and commodities are traded in well-organized and informed competitive auction markets. However, some tests of forecasts by active participants in these markets reject the REH, as shown in the survey of Wall Street predictions of interest rates by B. Friedman (1980). This could merely mean that this survey was not adequately representative of the most successful traders and hence of the market.⁶³

Expectations which fail the rationality criteria in *ex post* tests may well prove entirely consistent with optimizing *ex ante* behavior once the consequences of uncertainty, unique events, defective models, and fragmentary or costly information are taken into account (Zarnowitz 1982c). Also, they need not be adaptive or otherwise backward looking only. In fact, it appears that time-series models rarely explain statistically much more than half of the variance of expectational data. (See several papers in Mincer 1969a; Aiginger 1979). The evidence suggests that economic expectations are neither mere projections of the past nor flashes of intuition about the future but combinations of both extrapolative and autonomous components. Predictions of real growth and inflation are usually diversified and uncertain, seldom demonstrably self-fulfilling. Anticipatory action is often inhibited by prior commitments incurred in part to reduce uncertainty. Thus not all past is bygones and expect-

61. All these surveys collect largely qualitative responses (on the direction and in some cases range, rather than the precise size, of the expected price movements). The conversion of these data to the quantitative form required for the tests presents some difficult problems.

62. For many early forecasts of U.S. aggregate series, there is evidence of significant bias (Mincer and Zarnowitz 1969). The most common pattern is underestimation of growth and, less frequently, of cyclical changes (Theil 1958; Zarnowitz 1967). There are indications of adaptive learning behavior (Mincer 1969b).

63. Indeed, Mishkin (1981a) constructs forecasts from the bond market data which fail to reject the rationality tests for interest-rate expectations. But the concurrent bond market predictions of inflation fail the tests. The argument that this reflects the unusual character of the 1960s, a period of rising inflation, is not really persuasive. The inflation forecasts in the perhaps even more "unusual" 1970s seem generally worse yet. Long periods of tranquillity are hard to find.

tations are not all-important. Yet they matter a great deal and are a proper subject for empirical as well as theoretical studies.⁶⁴

2.4.6 Models of Financial Instability

Speculative excesses or “manias” have attracted the attention of contemporaries and economic historians at least since the Dutch Tulip Bubble of 1625–37. Financed by credit expansion and characterized by “overtrading” in real or financial assets (e.g., gold, land, or securities) they have been associated with more than two dozen major booms in business activity, often involving many countries (Kindleberger 1978). Historically, they tended to be followed by “panics,” that is, distress selling of the same assets to reacquire money and repay debt, and crashes in the prices of the now illiquid objects of the speculation. The resulting financial crises accompanied or aggravated downturns in the business cycle.

In asset markets, as a rule, the current price depends positively on its own *expected* rate of change. Expectations are influenced by outside events as well as the “market fundamentals”—economic determinants of the rates of return. The markets may react to the events by adopting such price anticipations as would drive the actual prices away from the path consistent with the fundamentals. In this view, accepted much more readily by practitioners than theorists, anticipations of a strong market trend are occasionally capable of causing speculative “bubbles”—cumulative, even explosive, fluctuations in prices of selected assets.

Very recently, these ideas, long mistrusted by most contemporary economists, began to receive support from writers using rigorous techniques of equilibrium analysis. Shiller (1981a, 1981b) concludes that stock price movements are too large to be explained by an efficient-markets model which incorporates future dividends, capital gains, and inflation. Blanchard and Watson (1982) argue that bubbles can occur in efficient markets with new entry and no unexploited arbitrage opportunities; also, that they have potentially strong real effects on relative asset prices, wealth, and aggregate demand.⁶⁵

Much of the recent work in this area uses the overlapping-generations ap-

64. The evidence from surveys has certain important limitations and must be assessed with caution, but the consensus of careful independent tests based on such data should be taken seriously; indeed, the materials now available are rich and in need of much further exploration (Tobin 1980, p. 29). Lessons from such work can usefully counter and correct the freely speculative analysis that treats expectations as being inevitably “unobservable.”

65. The problems of indeterminacy and instability of RE equilibria are now well recognized and much studied. Conditions under which bubbles may be excluded from the RE paths of the price level have been specified for some monetary models (e.g., Brock 1975). Tests for the German hyperinflation of the early 1920s reported by Flood and Garber (1980) give no support to the hypothesis that a price-level bubble occurred in that extreme episode. But it is easy to see that expectations alone cannot account for a long or rapid inflation in the absence of persistent or very high rates of money creation. Clearly, bubbles are much more likely in speculative markets for financial assets.

proach due to Samuelson (1958; Kareken and Wallace 1980). Models of this type, with markets continuously cleared by price movements, typically have a multiplicity of RE solutions, a large proportion of which may involve fluctuations prompted by arbitrary but self-fulfilling shifts in anticipations (Azariadis 1981). Thus, if the belief that sunspots predict future prices were widely held, many individuals would act on it so as to bear out their expectations. Cass and Shell (1983) argue that RE equilibria generally can be influenced by extrinsic uncertainty, that is, random phenomena such as “animal spirits” or “market psychology,” which do not affect the basic parameters of the economy (endowments, preferences, technologies). The conditions under which this would *not* be the case are so unrealistic as to be devoid of any empirical interest.⁶⁶

These abstract treatments rationalize the role in financial crises of “sunspots,” but they do not tell us how much these interacting and self-fulfilling expectations matter. The crises are seldom isolated phenomena; rather they form a part of some business cycles and are thus related to the monetary, institutional, and real factors involved in these cycles. Market psychology plays a large role in some crises, mainly the major ones.

Mild financial crises of recent U.S. history, called “credit crunches,” are associated by some observers primarily with reduced availability of credit, not high interest rates (Wojnilower 1980). Credit rationing by banks is linked to imperfect information about the borrowers’ default risk (Stiglitz and Weiss 1981; Blinder and Stiglitz 1983). When banks refuse to renew their loans, the high-risk borrowers and many small firms are unable to secure credit elsewhere and are forced to curtail investment and perhaps current operations. The retrenchment may or may not be caused by restrictive open-market policies of the central bank. Under fractional banking, loans and deposits are highly collinear, so it is difficult to use the data to distinguish between this “credit” hypothesis and a “money supply” hypothesis of a downturn. Since the early 1950s, broad credit aggregates such as the total debt of U.S. nonfinancial sectors have been closely and stably related to GNP—better or not worse than the money stock and monetary base series (B. Friedman 1983).

In several collected papers, Minsky (1982) argues that long periods of prosperity interrupted only by mild recessions or slowdowns breed overconfidence, excessive short-term financing by banks of long-term business projects, investment booms interacting with stock market booms, and growing indebtedness and illiquidity. Innovative practices and new instruments are used to increase the availability of investment finance: money should be broadly defined and is endogenously determined. Its supply, however, becomes at some point inelastic as uncertainty grows and banks increasingly

66. They include (1) strong RE—unanimity of beliefs, (2) complete markets—there are markets for all types of contingent claims, and (3) unrestricted access to these markets for all agents at all times—across the “generations.” Conditions (1) and (2) are merely improbable; (3) is altogether impossible.

deem it prudent to retrench (or monetary authorities act to constrain inflation). The demand for credit to finance planned and progressing investment projects nevertheless continues to rise. It takes sharp increases in interest rates and declines in the present value of expected net returns on capital assets to check and reverse the expansion in new investment. Business cash flows and profits decline and eventually so do total sales, output, and employment. The resulting losses force many industrial concerns and financial intermediaries into refinancing of debt and liquidation of assets to raise cash. Many retrench; some fail. When a sufficiently large number of these units experience inadequate cash flows from current operations and declining ability to sustain debt, a financial crisis must occur, unless the central bank intervenes, injecting sufficient money into the system and preventing large bank defaults and business failures. If a crisis is averted and stimulative monetary and fiscal policies continue, a mild recession may ensue but another inflationary and eventually destabilizing investment boom will soon follow.

This is a disequilibrium theory with strong endogenous elements. The originating factors are real, but monetary and credit changes have much to do with the propagation of the cycle and are primarily responsible for its worst manifestations (see also the summary in table 2.5). This is the opposite of some recent theories (such as that of Lucas) which have monetary shocks and real propagation mechanisms. Minsky's hypothesis predicts the recurrence of financial crises and business depressions, which invites controversy. It deserves to be carefully evaluated, but it is difficult to test (in part inevitably so since financial crises and depressions are rare and complex events). In general terms, however, the account provided by Minsky tends to be consistent with the history of speculative investment booms, financial crises, and deep depressions in the United States (Burns and Mitchell 1946, ch. 11, sec. 6; Kindleberger 1978; Hoyt [1933] 1970). It also has interesting connections with several diverse business cycle theories.⁶⁷

2.4.7 Demand and Supply Shocks and Responses: A Search for New Concepts

Dissatisfaction with the results obtained by using random money supply shocks and price misperceptions prompted some economists to experiment with different models (while retaining the basic RE methodology in most cases). Abstracting from numerous detail, one can distinguish the following approaches.

67. Minsky credits primarily Keynes, whom he interprets as having a dynamic analysis that does not rely on price rigidities, but his emphasis on the instability of credit flows and business debt relative to income recalls the Austrians and draws on Fisher. Some complementarity may even exist between elements of the Kindleberger-Minsky hypothesis and the monetarist approach, particularly with respect to accounting for financial crises on an international scale (Bordo 1984). Another linkage is to the "corridor effects" suggested by Leijonhufvud: deflationary shocks large enough to produce financial crises and waves of bankruptcies may throw a normally stable system into a deep depression in which disequilibrating forces prevail for some time (Howitt 1978).

Focus on Interest Rates and Related Factors

These received much attention in the literature from Wicksell and the Austrians to Keynes. Monetary intervention or excesses of credit creation were seen as causing interest rates to deviate from their equilibrium levels so that they fail to coordinate saving and investment decisions. Inconsistencies arise in the aggregate between the expectations of those who make these decisions and the expectations of the financial intermediaries. Monetarists opposed these ideas on the ground that investment and savings depend on the *real* interest rates, which cannot be affected by the banking system, except transitively. In the early RE models, too, interest rates had no important roles to play. However, fluctuations in credit and interest, which have been so pronounced in recent years, are receiving renewed attention in the literature since the late 1970s (Leijonhufvud 1981, ch. 7; Frydman and Phelps 1983, ch. 10).

This includes some new departures along the incomplete-information line of approach. McCulloch (1977, 1981) argues that business fluctuations are associated with unanticipated changes in interest rates that misdirect real investment decisions toward a wrong mix of capital goods and so distort the intertemporal production process (as in the real part of the theories of Hayek and Mises).⁶⁸ In an equilibrium model by Grossman and Weiss (1982), random real shocks affect productivity and the real interest rate, causing investment and output to fluctuate. Random monetary shocks affect inflation and the nominal interest rate, with the effect of amplifying the cycle. Agents make errors in trying to infer the *ex ante* values of the real rate and inflation from the nominal rate (which includes the two) because the latter's movements contain much random noise. The critical gap in information concerns the relative rates of real return (on own compared with other investments). The model involves only supply decisions, with the role of demand being purely passive.⁶⁹

Recent empirical work suggests that output (real GNP or industrial production) interacts with price-level indexes, comprehensive credit aggregates, interest rates, and narrow money aggregates. None of these variables is definitely exogenous relative to all the others; in particular, money adds to the

68. The novel element is the stress on "misintermediation" rather than on the money-creating function of the banks (the two are closely connected). Banks and other intermediaries, by borrowing short and lending long, are said to mismatch asset and liability maturities and to add to the uncertainty about interest rates by creating imbalances in their term structure. This hypothesis and the related estimates deal only with nominal, not with real, rates, besides posing other problems that cannot be discussed here in the available space; but the hypothesis does address an important and long-recognized institutional aspect of financial instability (to which the previous section also briefly referred).

69. Investment here is equated with the process of transforming labor input into next period's output. No durable goods are produced. The nominal shocks take the form of disturbances to money demand. These simplifications are made to allow some difficult problems in the dynamic equilibrium theory to be rigorously analyzed, but they also deprive the model of some highly relevant characteristics of the modern capitalist economies, which experience business cycles. This is, to be sure, not uncommon in the recent theoretical literature.

incremental prediction of the real variables along with credit and interest rates, but it is not predetermined or predominant (Sims 1980a; B. Friedman 1983; see also chapter 12).⁷⁰

These results are interesting and may prove robust, but they should be interpreted with much caution. Nominal interest rates may matter because their changes are read as signals of changes in the stock of money and inflation. But they may also be important in their own right as the major part of the true costs of credit. Historically, their adjustments to inflation (and deflation) were sufficiently sluggish and incomplete to suggest that people treated the nominal rates as if they were adequate proxies for the real ones. Evidence going back to the 1860s shows that this condition continued to prevail at least well into the 1950s; only thereafter did the relationship between interest and inflation strengthen to become rather close in the 1960s and 1970s. More recently, it weakened again considerably.⁷¹

The ex post real interest rates, pre- and post-tax, have small cyclical variations, which makes it difficult to detect how they are correlated with movements in real variables, but this is rightly seen as a problem in measurement rather than evidence of no significant relationships (Mishkin 1981b). Real rates rose sharply to unusually high levels in the early years of the Great Depression and again in 1980–82, a period of severe recession and strong disinflation. It is only reasonable to view these changes as mainly unanticipated and their probable net effects as seriously adverse.⁷² The measured real rates are inversely correlated with lagged inflation. Increased money growth, then, by raising inflation, lowers the real rate; but the evidence for a short-term effect of monetary policy, independent of inflation, is mixed and weak (Shiller 1980; Mishkin 1981b).

Focus on Real Factors

Consider two polar models. (1) Exogenous fluctuations in real investment cause the business cycle. Banks meet the borrowers' demand for money and the depositors' demand for money because the monetary authority provides

70. These exogeneity tests ask whether, say, the lagged values of money help explain or predict output in a regression equation containing the lagged values of output itself and of the other variables. This question is answered for each variable in the system in terms of all variables included. This method of variance decomposition or accounting for the interactions among the "innovations" (serially uncorrelated residuals) of the series in question avoids some major shortcomings of simple distributed-lag regressions but encounters some difficult problems of its own (see the following paragraph and chapter 12).

71. These findings refer to the United States; for the United Kingdom the corresponding relation was much weaker throughout (see Friedman and Schwartz 1982, ch. 10). According to Summers (1983) the impact of inflation on interest rates has been much smaller throughout the post-World War II period than a general equilibrium model would predict on classical assumptions, particularly when properly allowing for the effective tax rates (see also Feldstein 1976a).

72. But note that the real rates stayed relatively high in 1983–84 without visibly obstructing a vigorous recovery. Presumably, once the economy adjusted to the large changes, the smaller ones had only weak effects.

them with the necessary reserves at any given level of interest rates. (2) Exogenous fluctuations in money supply cause the business cycle. The demand for money adjusts continuously to the supply through changes in permanent income associated with much larger changes in measured real income or output.

Tobin (1970) sets the two models up so that model (1) implies a consistent pattern of early cyclical timing for monetary growth rates or deviations from trend, whereas (2) does not. Friedman denies that these examples cast doubt on his views. The two debaters agree that evidence on leads and lags alone cannot prove any hypothesis about causation (though it can disprove some); also, that both of the above models are far from adequate. Indeed, (1) is an oversimplification and partial distortion of Keynes, (2) of Friedman. The widely used general IS/LM-AD/AS framework can accommodate some of the ideas of some of the Keynesians and monetarists with respect to the effects and interactions of both autonomous expenditure and monetary changes.⁷³ But this long-ruling paradigm was designed for comparative statics of exogenous interventions with a stationary environment; its adaptation to dynamic and expectational processes confronts a host of difficult and unresolved problems.

Earlier sections contain several diverse illustrations of recent work oriented toward nonmonetary (real, psychological, institutional) explanations of macroeconomic instability. Other such efforts relate to the old idea (preceding Keynes, though often associated with him) that both money and real activity respond to the common factor of fluctuations in aggregate demand and the corresponding changes in interest rates and the value of existing assets. Sims (1980a), referring to Tobin (1970), urges the development of a stochastic version of this theory. Some new historical studies also emphasize the importance of factors other than the independent influence on business conditions of changes in the stock of money.⁷⁴

73. IS and LM relate output and interest rates; AD and AS (aggregate demand and supply functions) relate output and the price level. It is assumed that the IS and AD curves slope downward; the LM and AS curves upward. The Keynesian view is often associated with relatively steep IS and flat LM; the monetarist view with relatively flat IS and steep LM curves. But this interpretation is by no means universally acceptable; even if accepted, it leaves open other important issues in the controversy. The IS/LM-AD/AS model evolved from Hicks 1935, 1937, through a long series of writings including, notably, Patinkin [1954] 1965.

74. Thus Temin (1976) attributes the 1929 downturn to a decline in "autonomous" spending (interestingly, mainly in consumption and exports rather than investment). For monetarist counterarguments, see the essays by Schwartz and Brunner (Brunner 1981); for a view that both money and investment (and other spending) mattered, see R. J. Gordon and J. Wilcox in the same volume.

King and Plosser (1984) examine the possibility that the positive correlation between the rates of growth in output and monetary aggregates reflects, in large part, the causal chain running from business activity to "inside money," that is, bank deposits.

Bernanke (1983) produces evidence linking the sharp drop in output during the early 1930s to the reduced quality and higher real costs of credit intermediation services, a result of the institutional weakness and crises of banking.

It remains to take note of the most radical reactions against monetarist and monetary shock theories by some strong believers in general-equilibrium modeling and the neutrality of money. Here the main idea is that unanticipated changes in tastes and technology cause intersectoral shifts of human and physical capital associated with much friction and temporary idleness. Unemployment generated by a large number of partly independent shocks to different sectors will persist for considerable time because rapid transfers of resources are costly, the more so the greater the specialization of the production factors (Black 1982).⁷⁵ But changes in tastes and technology will tend to penalize some sectors and benefit others, causing numerous shifts in relative prices and outputs; they may occasionally have significant net favorable or adverse effects on growth but can hardly be responsible for *recurrent sequences* of expansions and contractions in aggregate economic activity.

The most restrictive formal model of purely real “business cycles” (Long and Plosser 1983) abstracts not only from money and government but also from technological change, durable goods, and adjustment costs; it assumes RE, complete information, and stable preferences. There are random shocks to outputs of the many commodities, each of which can be either consumed or used as input in the production of any other commodity during each unit period. The optimal allocation rule is such that when the i th output is unexpectedly high (low), the simultaneous inputs of i in all its productive uses are increased (reduced). It is shown that this works to propagate the effects of the output shocks both forward in time and across the sectors of the economy. The model is intended to serve as a benchmark for evaluating the importance in actual business cycles of the many factors it omits, but it admittedly presents serious problems and the approach itself is questionable.⁷⁶

Focus on the Causes and Effects of Contractual Wage and Price Setting

Recent literature is much concerned with the specification of economic reasons for the existence of explicit or implicit multiperiod contracts that limit the flexibility of wages and prices in response to unanticipated shifts in demand and supply.⁷⁷ The “contract-theoretic” models based on this work drop

75. This recalls the “horizontal maladjustments,” which were treated as a potentially important aggravating factor, but not a prime cause, in classical writings (see sec. 2.3.1).

76. In particular, the model has constant employment of labor while the commodity outputs fluctuate. Simulations based on a 1967 input-output table for six major industry divisions result in output series that show considerable comovements and two complete quasi cycles in runs of 100-unit periods of undefined length. It is not clear whether the fluctuations due to the output shocks are separable from those due to other factors.

77. One hypothesis is that fixed-wage, variable employment contracts sell risk-averse workers partial insurance against the variability of their incomes (Baily 1974; D. F. Gordon 1974; Azariadis 1975). Another is that labor contracts or understandings treat incomes as normal returns on the loyalty and efficiency of long-term employees and protect large investments in firm-specific human capital (Hall 1980a; Okun 1981, chs. 2 and 3). Steady pricing policies are used to reduce the costs of shopping and attract steady customers (Okun 1981, ch. 4). In sum, contracts are attempts to deal economically with a variety of “transaction” (information, negotiation, adjust-

the assumption of market clearing and the reliance on price errors and intertemporal substitution effects. They restore the changes in aggregate demand to their old role as a direct determinant of fluctuations in the real aggregates by delaying the adjustments in the contractually predetermined wages and the wage-related prices.

The main models of this type are monetarist in the original sense of relating short-run fluctuations in output to monetary policy operating via changes in demand. This is so even where the policy is based on a fixed feedback rule known to the public, because the money stock can be changed more frequently than the wage is renegotiated (Phelps and Taylor 1977; Fischer 1977a; Taylor 1980a). The authors work with staggered contracts in which wages are linked to expectations that are rational (i.e., model-consistent; sec. 2.4.5). Fischer relates wages to price-level forecasts; his contracts are (frustrated) attempts to keep real wages from falling in the face of inflationary policy. Taylor relates wages to other past and anticipated wage settlements; the degree of inertia in these interrelations helps explain the persistence of unemployment. These models have attracted much attention and criticism.⁷⁸

According to the “credibility” hypothesis (Fellner 1976, 1980), a *consistent* policy of adhering to preannounced moderate money growth targets can deter inflationary wage and price setting and reduce instability in the long run. A correlate is that unions will push for higher wages in labor contract negotiations and business will respond by raising prices if both parties expect the monetary authorities to “ratify” their decisions through accommodative policies. This view differs in some important respects from monetarism and the newer RE and contract theories, but it also incorporates some of their elements. It recognizes that, in practice, systematic policy components are often weak and only belatedly detectable (Sims 1982). It is related to the idea of “coordination failures” (Leijonhufvud 1981) in a game-theoretic form: it is possible for expectational impasses to develop, that is, a recession-cum-inflation may continue while the central bank waits for wage demands to moderate and business and labor wait for a resumption of monetary expansion (Phelps 1981). All this certainly raises issues that are very important in dealing with inflation; how much bearing it has here depends on how strongly policy-induced changes in monetary growth and inflation influence business cycles. The notion that such changes are the principal *cause* of the cycles (which many now accept as if it were self-evident) is not supported by the

ment) costs in the face of price instability and general uncertainty. For surveys, see R. J. Gordon 1976; Barro 1981a.

78. The issues include the consistency of the assumed contract features with rational behavior (Barro 1977a; Fischer 1977b) and the reasons why wages are not fully indexed under anticipated inflation (Brunner and Meltzer 1977). For a detailed critique of Fischer and Taylor and a proposed improved model, see McCallum 1982. For some evidence and further discussion, see Taylor 1980b and 1983.

long record of cyclical instability coexisting with alternating extended periods of inflations and deflations. It is also countered by the argument that business contractions, whatever their causes, have deflationary or disinflationary *effects*.⁷⁹

Most recent RE models, including those of the contract-theoretic type, consider only one determinant of aggregate demand, the real money balances, and concentrate on the effects of monetary policy. But shifts in aggregate demand may be due to real and expectational as well as monetary factors. They may reflect the instability of any of the major expenditure components of real GNP. Neither theory nor evidence supports the focus on nominal demand shocks and real supply shocks alone. In recent times, at least, consumer capital outlays have been no less cyclical than business investment (R. J. Gordon and Veitch 1986). Indeed, new models of consumption which embrace RE have consumers respond promptly and strongly to any new information that revises their forecasts of future income. This in itself would tend to increase the sensitivity of consumption, and hence of income, to unanticipated changes in aggregate demand or "autonomous expenditures" (compared with the earlier models which used adaptive rather than rational expectations in the calculation of permanent income). The increased instability on the demand side is offset only when the REH is combined with speedy clearing of the markets and a highly stable and inelastic aggregate supply curve, that is, in the new-classical equilibrium models (Bilson 1980, pp. 279–83; and comment by Hall, p. 301).⁸⁰

2.5 Conclusion: The Needed Synthesis

Business cycles belong to the history of modern economies with interdependent markets, free enterprise, and private ownership of financial assets and capital goods. They developed in the era of great growth of industry, banking, and credit. They are varied and changing, even while retaining their general characteristics of persistence and pervasiveness as well as specific regularities of amplitude and timing. All this was long recognized by serious scholars and careful observers with abiding interest in the subject.

For a long time, too, there was a substantial consensus among these economists that business cycles have mainly endogenous explanations involving recurrent fluctuations in interrelated monetary and real variables, prices and quantities, expectations and realizations. The theories tended to agree on the

79. None of this, of course, is to deny that the inflationary bias of government actions and the reactions to it have economic consequences of great importance; what is questioned is the view (going far beyond the credibility hypothesis) that these consequences include the business cycle.

80. Many interesting theoretical and empirical studies of the consumption function have appeared in recent years, but the implications of this work for business cycles remain to be worked out (Hall 1978; Davidson and Hendry 1981; Blanchard and Wyplosz 1981; Flavin 1981; Bernanke 1982).

cast and setting, disagreeing principally on which factors should play star and which supporting roles. Chance and outside disturbances were left behind the stage, interfering with and modifying the action but not determining its main course. The cycle had plural causes and effects; it produced and resolved its own stresses and resources, nationally and internationally. Few attributed it to any single factor or defect, and few believed that it can be eliminated by any single, low-cost type of policy intervention or institutional reform.

The 1930s convinced many that the economy is not merely cyclical but depression prone. The instability of profits, investment, and credit (old concerns) attracted renewed attention. In the 1940s and 1950s, however, there was a rise of interest in the dynamics of multiplier-accelerator interaction which yielded highly aggregative and purely endogenous models of potentially unstable output fluctuations. The monetary, financial, and expectational aspects of the cycle were largely neglected. Soon, strong reactions developed against this one-sided conception. One took the form of stochastic, dynamically stable mathematical models. This greatly influenced macroeconomic models which emphasized the destabilizing role of random shocks and exogenous factors.

The other reaction against theories of endogenous instability was monetarism. Its targets included not only the Keynesian models of aggregate demand, with fluctuating but often weak investment and a major role for fiscal policies to combat unemployment, but also the older models with unstable credit-investment interactions. Fluctuations in monetary growth attributed to erratic or misguided policies were made primarily responsible for disturbing the basically stable private economy and creating "business cycles." The real, financial, and expectational factors received little attention.

The natural-rate and rational expectations hypotheses, plus the discouraging record of the 1970s, led to new hopes for a program of study with a grand design: to reconcile business cycles with the postulates of the competitive general-equilibrium theory in its modern dynamic form. The initial approach relied upon random monetary shocks, price misperceptions, and intertemporal substitution of labor. However, critical analyses and tests soon cast serious doubts on this construction and thereby on the underlying premises, mainly that of flexible prices continuously clearing all markets. No convincing remedy could be provided by adaptation of familiar elements of older disequilibrium theories such as the accelerator and lags due to costs of adjustments in the stocks of production factors.

Some equilibrium theorists questioned the importance of the disturbances to money supply and tried to explain business cycles in real terms only. Their RE models are more restrictive yet. They consist of supply reactions of a "representative" producer-consumer to productivity or input-output shocks. The environment is purely competitive or of a Robinson Crusoe type. There is little or no evidence to support these models.

Contract theories, designed to explain why most wages and prices are tem-

porarily “sticky,” formed the basis of very different models in which aggregate demand fluctuations are restored to their usual prominence. Other work, also using the currently favored RE methodology, turned to factors with a long history of service in business cycle models: uncertainty and shifts in “market psychology,” interest-rate misperceptions or maladjustments, and long investment gestation periods.

There is not much empirical validation that random shocks of all kinds play as large a role in business cycles as has been attributed to them in recent literature. The weight of exogenous policy factors, too, seems more often than not overstated. The theoretical interest in self-sustaining cycles, or elements thereof, declined in times when economic events and thought favored a revival of the faith in the private sector’s capacity for stable growth. New work in this area, therefore, has been limited, but it still produced some interesting and significant results.

The most disturbing aspect of what must be viewed as the new mainstream literature (the RE models) is its increasing fragmentation in the face of various theoretical problems and recalcitrant facts. The ruling research strategy in these studies is to demonstrate for each particular model that one or more of the selected factors can contribute to fluctuations in total output or employment, while recognizing that others can do so as well. The authors are mainly concerned with theoretical possibilities rather than with explanations of what actually happens. There is in general little regard for how the pieces fit each other or the “real world.” The variety of the models is only loosely limited by the ingenuity of the theorists, but many of the offered hypotheses are not tested and some are not testable. Small linear models are favored because of their mathematical tractability in the equilibrium RE framework, but this certainly does not mean that larger and/or nonlinear systems are somehow inferior. The criterion of conformity to stylized facts would, in fact, suggest the opposite.

As a matter of simple logic, if each of a number of models is indeed valid, then it should in principle be possible to integrate them and improve the theory. Of course, this is easier said than done, but when each model is treated in isolation, there is little chance for the job to be ever seriously attacked. The conclusion reached here, then, is that a movement toward a synthesis of the tested and nonfalsified hypotheses is urgently needed if real progress is to be made in our understanding of business cycles.

In this view, research in this area will profit most from the confrontation of testable hypotheses with a broad range of evidence on stylized facts and from efforts to combine those hypotheses that survive the tests. This assigns high priorities to the work on (1) the consolidation and updating of findings from the historical and statistical business cycle studies and (2) using the results to eliminate those elements of the extant theories that are definitely inconsistent with the evidence and to improve the modeling of the other elements. This strategy could lead us away from the proliferation of models which give lop-

sided stress to either the monetary or the real factors, either the supply or the demand behavior, either random shocks or purely endogenous movements; instead, we would be working toward a better comprehension of how these forces interact. As was recognized early by the "classics," the sharp dichotomies and monocausal theories tend to be invalidated by experience.

Also, the hypothesis that business cycles are all alike would itself be tested in the course of this research. It is important to know not only what the common core of the cyclical process has been for the U.S. economy in recent times but also how it may have changed historically and how it differs from similar processes elsewhere. The knowledge is necessary for an assessment of the temporal and spatial reference points of business cycle theory. (For the most recent work on these problems, see R. J. Gordon, 1986a).

We have witnessed a period of great intellectual ferment, activity, and controversy in the theory of macroeconomic fluctuations and policies, but the debate seems to be growing less heated lately, and some signs of a future rapprochement can be discerned. The present may be a good time to ponder the needed synthesis.