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: Macroeconomics and Business Cycles: An Overview

Chapter Author: Victor Zarnowitz

Chapter URL: http://www.nber.org/chapters/c10372

Chapter pages in book: (p. 1 - 19)

1   Macroeconomics and Business Cycles: An Overview

"Theories and Evidence" is a fitting general title but it conceals the variety of subjects covered in part I of this book. The purpose of this chapter is to introduce these themes, point out the connections between them, and provide some explanations and extensions. References and detail will be mostly left to chapters 2–5, for which this is to serve as a guide.

1.1   Early and Recent Theories

Theories of business cycles should presumably help us to understand the salient characteristics of the observed pervasive and persistent nonseasonal fluctuations of the economy. The next chapter first presents a summary of these "stylized facts." The discussion then proceeds from historical theories, which were largely endogenous, deterministic, multicausal, and descriptive, to contemporary theories, which are largely exogenous, stochastic, monocausal or limited to very few types of shocks, and based on small formal models.

The early authors emphasized the instability of investment in fixed capital and inventories as well as of the supply of credit used for the acquisition of real and financial assets. Changes in relative input and output prices, interest rates, and profits constituted another important area of interest. Uncertainty about the profitability of future business ventures and volatility of the associated expectations received much attention even before Keynes elevated these factors to particular importance. The investment accelerator was among the earliest discoveries of the theories of economic growth and instability.

The rise of mathematical analysis in the 1930s and 1940s was marked by a dichotomy between the endogenous accelerator-multiplier models driven by nonlinearities and lags and the exogenous models, often of similar content but driven by shocks. The shocks can be few and large or many and small, random or serially correlated. The macroeconometric models, which first became

prominent at about the same time, embraced the exogenous, stochastic approach, perhaps largely for reasons of technical convenience as nonlinearities are difficult to handle on a large scale. This did not change with the development of ever more sophisticated analytical and computational techniques and ever larger econometric models. The latter came under heavy criticism recently in the mainstream literature, which favors simple and stable but also linear and stochastic models. However, there is no conclusive evidence that nonlinearities are not very important or that identifiable random shocks do in fact drive business cycles.

Money and prices received little or no attention in the accelerator-multiplier interaction and other Keynesian disequilibrium models. In contrast, shocks to the stock of money, notably those due to shifts and errors in monetary policy, represent the principal source of instability in the monetarist theory. Monetary change has real effects, except in the long run where it influences prices only as in the classical Walrasian prototype theory. Any trade-off with unemployment can exist only for unanticipated changes in the price level; in equilibrium with fully anticipated inflation, the "expectations-augmented" Phillips curve is approximately vertical.

In the Keynesian literature of the 1960s and 1970s the role of money and monetary policy in business cycles was increasingly recognized, but other influences retained their importance as well, notably the variability of private investment and shifts in fiscal policy. The natural-rate hypothesis has been generally accepted, too, but the shifts in the Phillips curve due to changes in inflationary expectations are mostly viewed as slow. A common feature of these theories is that they are concerned mainly with changes in aggregate demand that are translated into fluctuations in output and employment because wages and prices are considerably less than fully flexible over the relevant time horizons. This view of the world has long prevailed in the work on business cycles, which thus sharply departed from the competitive market-clearing paradigm that dominated the theory of price and long-term growth.

It has only been in the last two decades that this contradiction has led to an important new-classical resurgence and new-Keynesian response in macroeconomics. Two strong motivations stirred up this ferment in the recent literature. The first is to build dynamic models consistent with what are viewed as the basic empirical characteristics of business cycles. The second is to ensure that these models have the desired microeconomic foundations, that is, that they are consistent with utility- and profit-maximizing behavior of individuals and firms. The task of achieving both of these objectives simultaneously is very ambitious, far from accomplishment, and indeed of uncertain feasibility. One reason for this is the complex and changing nature of business cycles, as discussed briefly below (and in more detail in chapters 2 and 3). Other reasons lie in problems of theory introduced in sections 1.4–1.6 below and examined more thoroughly in chapters 2 and 4.

1.2   The Diversity and Evolution of Business Cycles

The National Bureau of Economic Research, in its comprehensive studies of long time series for the United States, Britain, France, and Germany, documented the timing and amplitudes of recurrent nonseasonal fluctuations in numerous indexes and aggregates covering a variety of economic processes. Based on the strong elements of consensus in these movements, chronologies of peaks and troughs in macroeconomic activity were constructed for the four countries (see chapter 5 for a brief survey of this research). There is much evidence of the pervasiveness and persistence of business cycles. However, although individual cycles share important family characteristics, they are by no means all alike. The historical experience includes some violent swings of boom and bust, some extended depressions, and (especially in recent times) some long periods of relative tranquillity.

In short, the record shows much diversity of the so measured business cycles. Although most contractions were short and mild, some were long and/ or severe. Although most expansions were neither very long nor very vigorous, some had one or both of these attributes.

The long and widely held view that the U.S. economy tended to be less unstable after World War II than before has recently been challenged. A review of data and literature in chapter 3 upholds the earlier view, with some modifications. In particular, business contractions became somewhat shorter, and much milder and less frequent, whereas both real growth and inflation grew much more persistent. When the possible sources of the moderation of business cycles are considered, they are found to form a rather long list. Several of the examined hypotheses are affirmed, and the selection has some implications for the general analysis of business cycles.

The following factors have probably contributed significantly to the increased stability of the economy:

1. Shifts in employment to less cyclical, largely service-producing industries. This is consistent with the net destabilizing role of changes in business inventories and durable-goods purchases.
2. Increased size of the acyclical government sector.
3. Increased role of automatic (mainly fiscal) stabilizers. Points 2 and 3 are consonant with the idea that growth in normal government activities can, up to a point and given certain rules, offset some of the instability of aggregate private demand.
4. Reduced frequency and intensity of financial crises. Here the implication is simply that such crises can aggravate (if not necessarily cause) business cycles.
5. Some favorable fiscal (mainly tax) policies.
6. Reduced volatility of monetary growth (to the extent that money plays an independent role rather than merely adjusting to changes in income or

business activity). Points 5 and 6 imply that fiscal and monetary policy actions can add to or reduce macroeconomic instability.

7. Greater confidence of private economic agents, both induced by the observed business cycle moderation itself and inducing behavior favorable to more stable economic growth. This suggests a role for endogenous and self-validating expectations in reinforcing positive (or negative) macroeconomic trends.

1.3   Some Implications and Evidence: Problems with Monocausal Explanations

What theories can do justice to such diverse and evolving phenomena? It is doubtful that monocausal hypotheses can, whether the single factor to which they attribute business cycles be real, monetary, or expectational. Theories that rely on a simple interaction (such as that of the accelerator-multiplier) or a single type of shock (e.g., to the money supply, investment demand, or technology) are attractively simple and can be instructive. But they fare on the whole poorly when confronted with data, presumably because they are seriously incomplete.

Models that rely on cumulative propagation effects of small and frequent white-noise shocks from a single source produce fluctuations with common properties that may be fairly regular. They cannot account for any systematic variation or historical evolution of business cycles.

Although a model should be as simple as possible, the wider the range of the relevant facts it covers, the greater is its claim to validity. There is no way to discriminate effectively between the many models that address "the" business cycle on the sole basis of this or that of its selected aspects; one must ask of the models that they be reasonably consistent with all or most of the important stylized facts supplied by history.

Empirically implemented stochastic models suggest that the U.S. economy is exposed to a mixture of shocks, mostly small but occasionally large, with no one source being dominant. Small "structural" models using vector autoregressions (VAR) have aggregate demand disturbances accounting for the largest proportion of the variance of total output; aggregate supply, fiscal, money, and credit shocks exert smaller effects. As the forecast or simulation horizon grows longer, the relative importance of demand shocks tends to decline, that of supply shocks tends to rise. [1]

---

1. In the model of Blanchard and Watson 1986 (quarterly, 1947–82), demand shocks contributed 54% to the variance of the log of real GNP four quarters ahead; shocks to supply, fiscal operations, and the stock of money contributed 15%–16% each. One quarter ahead, the role of aggregate demand was much larger yet (74%); the fiscal and monetary influences were much smaller (3%–4%). For a model with bank loans and M1, Bernanke (1986) reports somewhat larger demand effects four quarters out and smaller supply effects; money and credit account for 11%–12% of GNP variance each.

Different and larger macroeconometric models detect different and more types of shock, which here affect the stochastic equations and the variables treated as exogenous. But the reported simulation results, when properly aggregated, are similar in that they assign the largest effects to demand shocks (mainly to durables consumption, housing, and business fixed and inventory investment). Fiscal, supply, and monetary shocks taken together explain statistically about 30%–40% of the variance of output (Fair 1988; also see Eckstein and Sinai 1986).

The numerical results of any such decomposition are questionable because of differential limitations of the applied econometric and VAR methods, the likely specification errors and identification problems, the reliance on linear relations with constant parameters, and the restriction to short time series covering only the moderate postwar fluctuations. Different models may well be required in application to different conditions, for example, the interwar depressions. But the consensus of a wide range of recent models on at least one point is telling: the theory of business cycles must recognize more than one critical factor and come to grips with the interaction of shifts in demand and supply, real and nominal variables, and changes originating in private behavior and government actions.

1.4   New Developments

Three trends are manifest in the contemporary literature: (1) the increased emphasis on the role of aggregate supply, its determinants and effects; (2) the new insistence on maintaining the assumption of competitive markets cleared by relative price changes; and (3) the adoption of the hypothesis of rational expectations (RE). All of these converged in the family of new-classical equilibrium theories of the business cycle.

The first generation of these models relied on imperfect information: prices outside one's own market are known only with a lag. Random monetary shocks induce changes in the price level that are temporarily mistaken for changes in relative prices, which gives rise to wrong production and investment decisions. The second generation, currently favored by equilibrium theorists, is the theory of real business cycle (RBC), which can be viewed as a stochastic version of the Walrasian prototype. It assumes full information, replaces perfect foresight by RE, and links fluctuations in output and employment to real disturbances only, not nominal disturbances misperceived as real ones.

The existing RBC models use shocks to technology and hence to the representative worker's productivity (other conceivable real shocks are to preferences, weather, government expenditures, and tax rules). Favorable impulses of this type raise the current return relative to the expected future return and so stimulate the present productive effort at the expense of the future effort; negative impulses have the opposite effect of inducing workers to substitute

more leisure now for less later. A high degree of intertemporal substitution of labor is assumed in both the imperfect-information and the RBC equilibrium theories. Productivity of capital varies with that of labor, so positive shocks cause investment to rise. A long gestation lag spreads the effects of investment on output over the time needed to build the finished capital goods.

Although lacking clear support from the data, the hypothesis of RE was widely adopted as a maintained premise, with important implications for the analytical methods used. The hypothesis of price flexibility was not, but many of those economists who continued to treat wages and prices as "sticky" now embarked on a search for a reconciliation of this assumption with economic theory. The resulting work on the "microeconomic foundations" of macroeconomics gave particular impetus to studies of aspects of aggregate supply.

A sharp distinction has been drawn in some recent writings between new-classical (NC) and new-Keynesian (NK) macroeconomics (see especially R. J. Gordon 1990). The assumption that prices continuously clear the markets is common to both generations of NC but is not accepted by NK. The NK literature uses imperfections of competition, information, and capital markets in attempts to explain *why* prices and wages are to a large extent set and sticky. This is to provide the microtheoretic foundation for why a decline in nominal demand should be associated with a fall in aggregate output (recession) rather than a proportional downward adjustment in the price level. The new theories of cost and price rigidities are discussed in chapter 4 from the viewpoint of their contributions to our understanding of business cycles.

1.5   The New-Classical Single-Shock Models

Since monetary and price data increased greatly in availability, quality, and promptness, the theory of recurrent price misperceptions resulting from un-anticipated monetary disturbances should be less applicable now than in earlier times. Indeed, the main recognized defect of this theory is the implausibility of its informational assumptions. At the present, aggregate data on nominal variables are not so much stale as of uncertain dependability and subject to revisions over time. Moreover, output does not appear to be independent of that component of monetary growth which could have been observed or anticipated (and which should have been therefore met exclusively by price adjustments).

The hypothesis that has people confuse absolute with relative price changes also requires prices to be flexible in the short run. It is impaired by the evidence that the responsiveness of the price level to fluctuations in monetary growth and income has diminished. As discussed in chapters 3 and 4, industrial prices used to decline in cyclical slowdowns and contractions, but they no longer do so in the postwar era of continuing inflation.

Prices are flexible also in the RBC theory, now the more active class of equilibrium models. Here inside money and credit provide inputs into the pro-

duction process in the form of transactions and financial services; they are driven by the economic activity and this explains their positive correlation with output. In reality, passive responses of money and credit are common, though they take different forms under different monetary regimes (chapter 3). But this is not the whole story: on occasions active money changes occur that are of paramount importance. Thus, monetary authorities can take actions that result in curtailment of monetary growth or even the stock of money. Under other conditions, banks may restrain the availability and raise the real cost of credit.

The aggregate RBC model postulates overall productivity shifts that have strong effects on the total demand for labor. This requires that the diffusion of the shocks to technology across the economy be wide and prompt, but actually technical changes are likely to be localized and gradual. Thus, most of the interest in the dynamics of Schumpeterian innovations—new products, new technologies, new markets—centers on their role in "long waves," not business cycles (cf. van Duijn 1983, ch. 6; Tichy 1983, esp. pp. 80–81). For five major innovations in the United States, 1953–74, the estimated average half-life of the diffusion process was about eight years (Mansfield 1980, pp. 578–80).

The shocks to productivity must be sufficiently large in either direction to result in general business expansions and contractions in the absence of multiplier effects. (The RBC model makes no distinction between the actual and the potential output and so has no multipliers.) It is difficult to think of what would cause such recurrent disturbances and particularly of a technical regress that could account for a major depression. The only supply shocks with demonstrably large power that have been identified so far are the oil cartel's explosive price increases in the middle and late 1970s, which were associated with unique inflationary recessions. These are important cases that the RBC model can help explain but no other similar examples come to mind. Historically, the predominant relationship between the cyclical movements in output and the price level (or its deviations from long trends) was positive, not inverse.

1.6   Technical Change, Productivity, and Investment

The rate of neutral technical change has been estimated residually from an aggregate production function under the assumptions of perfect competition and constant returns to scale (Solow 1957). Updating such calculations confirms that the "Solow residual," an indirect measure of the rate of change in total factor productivity, is positively correlated with the rate of growth in output (see, e.g., Mankiw 1989, p. 84). When interpreted as a proxy for exogenous shifts in the technology of production, these estimates show that such shocks fall to negative values (technical regress) in the years of recessions and rise and are positive in the years of business recoveries and expansions.

Whether the shocks derived from this model are viewed as highly persistent (Prescott 1986, pp. 25–27) or as random (Plosser 1989, p. 63), this reading supports the RBC theory.

It is very unlikely, however, that any nationwide technological decline occurred in the recessions of 1949, 1954, 1958, 1960, 1979, or 1980–82 without anyone having noted it at the time (the OPEC shocks in 1974 and 1979, of course, aroused great public concern). It is far more plausible that the productivities of both labor and capital fall in business slowdowns and contractions because workers and machines have less work to do as demand declines. Skilled and specialized labor is "hoarded" since dismissing and reacquiring it is costly, and speedy capacity adjustments are costly as well. It is important to make investment in human capital pay and difficult to predict the occurrence and duration of demand fluctuations; hence uncertainty impedes the proper timing of input changes.

Labor hoarding can be added to a model of RBC with indivisible labor (i.e., a fixed shift length in hours, as in Hansen 1985) by making firms vary the effort required from their workers in response to exogenous shocks. The work effort is then procyclical and positively correlated with the Solow residual and government consumption or military expenditures used as proxy for demand shocks. In this model, which has perfectly competitive and complete markets, the ability of technology shocks to account for fluctuations in aggregate output is drastically reduced (Burnside, Eichenbaum, and Rebelo 1990). The model shows some improvement on the standard RBC analysis in that it approximates better the observed leading cyclical behavior of the average productivity of labor (which is here positively correlated with future, and negatively correlated with lagged, total hours worked).

Still another interpretation of the procyclical movement of the Solow residual is that it reflects market power and excess capacity (Hall 1986). The markup ratio of price over marginal cost (P/MC) exceeds 1 (its value under competition) and firms often operate in the region of decreasing average cost or increasing returns. This would impart a procyclical bias to the residual estimate even if the true factor productivity were constant (neither influenced by nor causing business cycles). This approach represents the opposite extreme to the aggregate RBC model; it also conflicts with the alternative explanation of procyclical productivity based on labor hoarding, constant returns, and sticky prices (Rotemberg and Summers 1988).[2]

"Autonomous" investment in fixed capital driven by innovations that require more capital per worker has long been recognized as an important force

2. Finally, one should mention a strong statistical argument against the reliability of the method of residual estimation of total factor productivity with the available data. In this approach, technical change accounts for more than 50% of real GNP growth in the United States between 1909 and 1949. Alternative calculations suggest that this makes "technical change" a catchall for omitted factors and measurement errors, much of it a "measure of our ignorance" (Denison 1974, 1980; Jorgenson, Griliches, and Denison 1972).

in both economic development and business cycles (e.g., in such diverse theories as those of Schumpeter and Hansen). In Keynesian models the autonomous investment is volatile, and its unpredicted shifts have important multiplier effects on consumption and income. But the desired stock of capital increases with output (more strictly, the expected demand for output), so investment is also in part induced by real growth through a "flexible accelerator" relation. Accelerator-multiplier interaction processes play a central role in various dynamic models. Depending on the parameters involved, they can give rise to explosive, maintained, or damped fluctuations, that is, either to endogenous nonlinear theories or to exogenous stochastic theories of business cycles.

The NC equilibrium models have no multiplier mechanisms, and compensatory combinations of large shocks to technology and a strong instantaneous accelerator lack credibility. The models also fail to produce the sizable positive autocorrelation of investment that is observed.

1.7   Cyclical Movements in Prices, Cost, and Profits

Indexes of U.S. wholesale (and to a lesser extent consumer) prices tended to fluctuate procyclically around alternating upward and downward trends between 1789 and 1932 (Zarnowitz and Moore 1986, pp. 525–31). These movements canceled each other to a large extent over the long stretch covered by historical statistics, even when the inflationary war periods are included. The last half-century, however, witnessed the longest, largest, and most continuous inflation on record. The recent recessions and major slowdowns were no longer associated with any general deflation, only and at most with some disinflation. (Indeed, inflation accelerated during the contractions in 1974 and 1980, in part as a result of the novel supply shocks—the huge increases in oil prices imposed by the OPEC cartel.) The variability of quarterly log differences in prices was much smaller after World War II than in the earlier periods.

A review of the evidence suggests that in recent times money wages as well as prices have become less sensitive to retardations and declines in aggregate demand, even while the cyclical instability of the economy has decreased (chapter 4). This seems inconsistent with the principle of the market-clearing role of prices but it really is not. Depending on the underlying conditions and sources of the shifts in the economy, departures from price flexibility may or may not be destabilizing. The main contrast is between the stabilizing potential of flexible relative prices and the destabilizing potential of large movements in the general price level.

Major deflations of the past had strong adverse expectational and distributional effects, and their nonoccurrence after the Great Contraction of 1929–33 was salutary. But the long persistent inflation of the postwar era could not stay moderate or unanticipated and became itself a major source of instability. It

accelcrated in the 1970s, grew increasingly volatile, and gave rise to varying inflationary expectations, much uncertainty, and popular discontent. It led to oscillating, tardy, and excessive policy interventions that had disturbing consequences of their own.

The effects on the economy of changes in expected inflation depend on the adjustments of nominal interest rates. The latter have become prompter and stronger in the last two decades than they had been before. They can be stabilizing, up to a point, if permitted by the monetary authorities.

Relative price adjustments work best when the general price level is reasonably stable. Moderatc fluctuations in the price level or the rate of inflation are not necessarily detrimental to real growth. In contrast, protracted and anticipated declines in wages and prices have aggravated demand contractions and accompanied the worst depressions in the history of U.S. business cycles. Long and rapid inflations can also be associated with very depressed economic conditions, as illustrated by recent developments in several Latin American countries. (The United States has avoided such inflations in peacetime.)

The structural changes that have contributed to the relative stabilization of the economy in recent times are probably also responsible in part for the reduction in the cyclicality of wages and prices. Not only quantities but prices as well tend to be cyclically less sensitive for services than for goods. Hence, the large rise in the overall weight of services should have made the comprehensive cost and price indexes more sticky.

Another factor working in the same direction was the spread of long and staggered wage contracts prompted by the growth of labor unions (which peaked in the first half of the postwar period). More generally, technical progress and competition require large investments in human capital that pay off best when employment relations are long term, maintaining stably rising levels of productivity and real compensation. Implicit contracts designed to produce such conditions would favor, and might be inferred from, the observed growth of career employment in the corporate sector, professions, and government.

A rise in confidence and the perceived probability that business contractions and slowdowns will remain brief and mild would help to keep up spending and hence firm up prices in general. So expectations of prosperity may also counteract price declines or cuts.

Procyclical movements in delivery lags help absorb changes in demand. The frequency of price changes is reduced by nonprice market clearing. An increase in the role of production to order and adjustments in delivery periods would have made the measured price level more sticky.

Prices of finished industrial products generally move less than average production costs, which tend to rise *relative* to their trends in slowdowns and downswings and decline late in recessions and particularly in early and high expansions. (These lagging cost changes reflect mainly the sensitive behavior

of labor productivity—output per hour—and the sluggish behavior of wages.) As a result, the markups on unit labor costs and profit margins have strong cyclical movements that typically lead at peaks and troughs by long intervals, responding directly to changes in productivity and inversely to changes in cost. Total profits have similar large fluctuations with somewhat shorter leads. These patterns of relative movements in prices and costs, productivity and profitability, are well documented in recent data, but they may have a long history and do not imply that prices in general are necessarily very sticky.

Certain NK hypotheses on why wages and prices are rigid assume monopolistic competition or oligopoly. Labor productivity is taken to be a rising function of real or relative wages, which are therefore kept high ("efficiency wages"). Small deviations from the equilibrium price affect profits only slightly and are therefore not worth correcting, as the individual firm sees it ("menu cost"). These are interesting ideas but their applicability is still to be established. They would gain support from the data if it could be shown that market power has increased in recent times, but it is not clear that it has.

1.8   Elements of Business Cycle Dynamics: Some Illustrations

Lags and nonlinearities provide the essential ingredients of endogenous business cycle models, which are valued for not being dependent on the theoretically unexplained and empirically unobserved outside shocks. There is a strong presumption that these elements are indeed important in the immanent dynamics of the economy. Some simple examples are presented in figure 1.1 and the text below; they draw on theories discussed in more detail in the chapters that follow. But it should be noted at the outset that the available formal and fully articulated models of self-sustaining cycles are generally small, abstract, and of uncertain value in interpreting the world. The main probable reason lies in the limitations imposed by the mathematical complexity of constructing larger models of this type. In any event, what we have is a number of ingenious but fragmentary constructs and a larger number of interesting ideas that may serve as building blocks for a more general theory. My purpose here is merely to introduce some such notions that seem to fit the facts at least some of the time.

One old idea is that near, at, and above full employment the demand for labor and materials tends to outrun supplies, driving up wages and other input costs. At the same time, growth of capital and output may deter comparably large rises in prices of final goods and services. At low employment and capacity utilization, however, prices rise faster than wages and other costs. In other words, the rate of change in costs (c) leads the rate of change in prices (p), as shown by the upper pair of curves in figure 1.1. When the increase in p exceeds the increase in c, profits per unit of sales rise; when the reverse obtains, they fall (see curve Π/S).

As depicted, p and c would rise in booms, decline in slumps, and remain
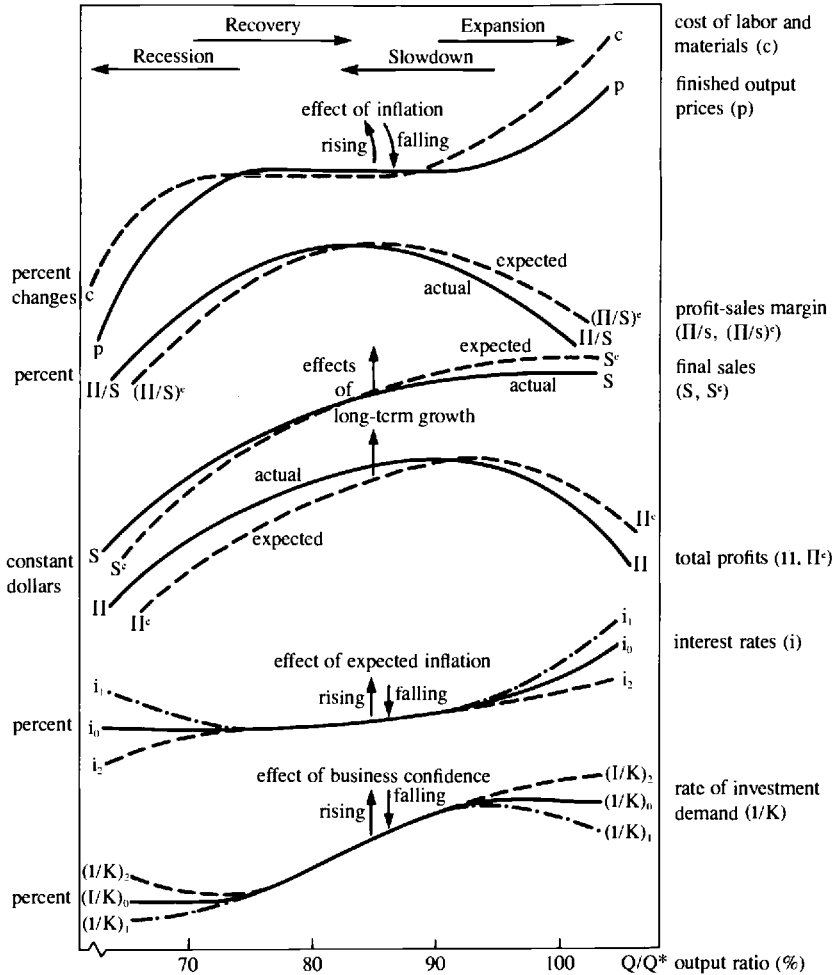
Fig. 1.1   Selected nonlinear relationships used in business cycle models

stable during the intervening periods of moderate growth. The long-term average rate of inflation is assumed to be low (perhaps close to 0) and highly stable. History offers examples of a prevalence of such conditions of long-period monetary and price stability in peacetime, notably during the pre–World War I gold standard era (which presumably explains the good overall fit of the original Phillips curve). But in an age of secular inflation, the expected rates of price and wage changes will not indefinitely lag behind the actual rates, unless the latter continually accelerate. Hence, the simple Phillips relationship between inflation and unemployment (or the ratio of actual to natural output, Q/Q^*) will be destabilized as it was in the late 1960s and

1970s. In such times, then, the c and p curves would steepen (i.e., tilt counterclockwise) and shift upward with anticipations of rising inflation, as indicated by arrows in figure 1.1).[3] It seems plausible that the hypothesized cyclical effects on, and differential behavior of, the relative prices (c. vs. p) could well persist as long as the inflation remained moderate, but one can certainly think of disabling conditions in this context as well (e.g., rapid inflation with c lagging behind p; the two curves tending to common verticality).

In various theories of the business cycle, demand rises less than output as the expansion unfolds; that is, final sales flatten or "lag," as shown by the S curve in the chart. One possible source is the short-run consumption function, whose slope is positive but less than 1 and perhaps getting gradually smaller. The accelerator effects on fixed investment and inventory investment of the declining growth in consumption must also be considered. The growth in output itself is thought of as decelerating as full employment is approached and bottlenecks are increasingly encountered. Meanwhile, the accumulation of the real capital stock (K) is picking up steam at high levels of capacity utilization, which according to the flexible accelerator hypothesis has a negative effect on net investment in business plant and equipment.

Given the Π/S and S function, the total profits curve (Π/S) $\times S$ will have an inverted-U shape, like Π/S but lagging and steeper on the downgrade (see fig. 1.1). Since K is a slow-moving stock variable dominated by an upward trend, the profit rate Π/K would reflect mainly the leading cyclical pattern of Π, with an even sharper decline. Expected profits should be influenced by past and present realized profits and have a similar profile of cyclical movement. This supposition is indeed generally consistent with the observed tendencies of both corporate profits and stock price indexes to lead at business cycle turns, particularly peaks.

The decline in profit expectations is likely to dampen the rate of real investment demand, I/K, where I represents investment plans and commitments (appropriations, contracts, and orders) rather than the lagging realizations (outlays and completions of capital projects). The sigmoid function $(I/K)_0$ in figure 1.1 assumes that I responds positively to Q, given K, in the midrange of perceived positive profit opportunities. When expected profits are poor, dI/dQ is low or 0, however, and this is the case at both very low and very high capacity utilization levels. Given Q, net investment should depend inversely on K, but gross investment need not, since depreciation and replacement investment are positively related to the volume of the capital stock. A sufficiently large expansion of K can nevertheless depress the total investment rate temporarily as a result of "overbuilding."

Investment is clearly an inverse function of the level of interest rates (i) to the extent that higher i represents in real terms both higher financing costs and

3. The alternative curves are not shown to avoid cluttering figure 1.1, which tries to reduce complex dynamic situations to their simplest common elements.

higher returns on nominal assets. This relationship too may be nonlinear, as in the case when I is more interest sensitive at high than at low levels of i. In the course of a business expansion, i will often start rising because of any or all of the following: (1) the demand for credit outruns the supply of credit, (2) the real rate of return on physical capital rises, and (3) expectations of inflation increase (note that this last effect relates directly to the nominal, rather than real, interest rates and can occur during business contractions as well). In figure 1.1, this type of behavior is illustrated by the curve i_0. But other developments are just as possible, since interest rates depend not only on the demand for credit and its determinants, notably Q as the measure of aggregate economic activity, but also on the supply of credit and its determinants, notably the role of banking and monetary policy. For example, increasing the money supply growth rate at low levels of Q/Q^* and reducing it at high levels may produce such curves as i_1 and, in response, $(I/K)_1$. The opposite combination of a relatively restrictive (stimulative) policy when Q/Q^* is low (high) can be represented by curves i_2 and $(I/K)_2$.

Movements to the right along the hypothesized curves occur in recoveries and expansions; movements to the left occur in slowdowns and recessions, as indicated by the arrows at the top of figure 1.1. But no simple schematic representation can do justice to business cycle dynamics that include elements of trends, irreversibilities, and shifts over time in some of the depicted functions. As already noted, inflationary trends and expectations tend to shift the c, p, and i curves. Also, long-term real growth will gradually pull upward the aggregate sales and profits (S, Π) curves at any Q/Q^* ratio. The direction of these effects is shown by arrows on the figure.

1.9   Extensions and Qualifications

Business cycles are diversified largely because they are propelled by a number of endogenous forces whose relative intensity varies over time. A major source of differences between the historical explanations of business cycles is that one author places the main stress on factor A, another on B, etc. But the selective emphasis is excessive, and the interaction of A and B may not be given due attention. As is well known, the same observed relationship is frequently predicted by several quite different theories. (Another reason for the diversity of both business cycles and their explanations is that various outside shocks impinge on the system.) Consider the following examples:

1. One of the theories which predict that costs typically rise faster than prices in expansion combines the accelerator principle and differential supply elasticities. The rise of final demand at high levels of capacity utilization stimulates investment in new productive facilities. This drives up further the demand for labor, raw materials, and other producer goods at the very time their supplies are becoming more scarce and less elastic. The unit labor costs increase sharply as the availability of skilled workers is reduced and higher

wages must be paid but growth of average productivity (output per hour) slackens. It takes more time to plan and build capital goods—producers' durable equipment and plant. Many materials are of agricultural origin and have long and rather rigid production periods (e.g., industrial crops, cattle hides).[4] The supply of standardized finished products can be expanded much more readily.

2. Nonmonetary overinvestment theory argues that the rise in consumer demand causes larger percentage rises in the demands for plant and equipment and, a fortiori, materials processed by industry. Hence, again, c increases relative to p. Investment, financed by the initially easy bank credit, is overextended relative to consumer saving, and must eventually be curtailed as profits fall, banks retrench, and interest rates rise. Thus early analysts of the business cycle as different in their outlook and procedure as Mitchell and Hayek (see 2.3.1 and 2.3.6) had similar views on the role of the price-cost-profit movement.[5]

3. That p is reduced relative to c in contraction is due mainly to wages being more resistant than prices of finished products to downward recessionary pressures. Prices of crude materials tend to soften more than prices of finished goods and services during general business slowdowns and recessions. The strongly procyclical and leading behavior of the profit margin Π/S is implied by the assumed behavior of p and c, and well documented in the data, and so is the downturn of total profits later in expansion, which results not only from the decrease in Π/S but also from the reduced growth of S. The "profit squeeze" is therefore as much a demand as a supply phenomenon. It too features in various business cycle theories emphasizing changes in cost, overinvestment, and other maladjustments, but most notably in the early work of Mitchell (see Haberler [1937] 1964, ch. 4).

4. The deceleration of final demand in the latter stages of expansion has been given explanations varying from full-employment ceiling and accelerator effects to cyclical changes in functional income distribution and consumption. The share of labor in national income has a mildly countercyclical pattern because real hourly wages increase somewhat less than output per hour (labor productivity) in expansion, whereas the opposite prevails in contraction. The proportion of consumption is larger in labor income than in property income, which is consistent with the finding that the ratio of total consump-

4. It should be noted that a large proportion of raw materials is imported, which at times can strongly influence their relative prices. Thus, a change in the terms of trade against materials and in favor of finished products occurred in the 1950s and 1960s, tending to reduce the rise in c vs. p.

5. Still other early business cycle theorists with widely varying views agreed on the strong procyclical fluctuation of the relative price of capital goods, including Marx and Kalecki, as shown in Sherman 1991, esp. chs. 10 and 11. This new monograph offers a very accessible treatment of many theories, particularly the endogenous historical ones, plus interesting illustrations of the average behavior of many time series during the post–World War II business cycles in the United States.

tion to national income (the average propensity to consume) is also mildly countercyclical (Sherman 1991, chs. 5 and 8). Theories of underconsumption span a long era from some early classics and social reformers (Lauderdale, Malthus, Sismondi, Rodbertus) to Hobson 1922, Neisser 1934, A. H. Hansen 1939, and others. Keynes's idea that in an ex ante sense saving outruns investment at full employment has an underconsumptionist flavor.[6] But, unlike investment in producer and consumer durable goods, consumption properly defined is not very sensitive cyclically. Except perhaps in some past episodes of deflation and stagnation, examples of a systematic tendency for underconsumption seem hard to find.[7]

5. Errors of expectations play a critical role in several otherwise quite different theories. Thus, Lavington, Pigou, and Keynes all see a spread of overoptimism in expansion and disappointment and overpessimism in contraction. Business people form their expectations adaptively by learning gradually from the past, but they also share in pervasive moods and opinions, seeking safety in numbers, especially in times of high uncertainty. The assumption made in figure 1.1 is very simply that the prevailing expectations of sales and profits $(S^*, (\Pi/S)^*, \text{and } \Pi^*)$ are lagged versions of the corresponding actual values $(S, \Pi/S, \text{and } \Pi)$. At high levels of economic activity, as measured by Q/Q^*, expectations exceed realizations; at low levels, the reverse is the case (compare the paired solid and broken curves). An overcorrection of errors of optimism generates errors of pessimism. This depresses business confidence and the demand for investment I/K, contributing to the economy's downturn. Similarly, the shift from pessimistic to optimistic forecasts would raise confidence and investment, aiding the recovery.

6. To make the model of expectations less crude and more convincing, one would have to show what causes many agents to commit similar errors simultaneously, and why such errors are not avoided or at least not discovered and corrected more promptly. This leads to a study of how competition works under conditions of uncertainty, that is, deficient knowledge, fragmentary and partly wrong information, and very limited foresight. Economists have only begun to explore this difficult subject (see 2.4.5). Empirical studies show that economic forecasts tend to have much in common with extrapolations from the recent past and so make large errors around turning points.[8] The above

---

6. Note the long quotation from and sympathetic references to Hobson and Mummery 1889 in Keynes 1936, ch. 23.

7. According to Haberler. "The under-consumption theory is a theory of the crisis and depression rather than a theory of the cycle" ([1937] 1964, ch. 5, pp. 119ff.). Sherman argues against this view, though recognizing the serious incompleteness of the theory (1991, ch. 9). He presents an "underconsumptionist cycle model," which reduces to a second-order difference equation in the national income, but that model treats as constant several parameters that are unlikely to be so (the same applies to his models for some other endogenous theories that are probably essentially nonlinear).

8. For evidence, see part IV; also, Lovell 1986. This argues against the hypothesis that expectations are based immediately on the correct model of how the economy works. Insisting on superrationality in this sense often seems equivalent to the fallacy of *petitio principii*. But this is not to

error model, though primitive, still has some useful implications. Investment in new plant and equipment is often associated with investment in new products, markets, and technologies; as such, it typically requires a considerable degree of confidence in the future. Confidence is an intangible but important and clearly procyclical factor. Its influence on business capital investment can be very strong at times, especially near and after cyclical turning points.

7. Expected sales are presumably the major determinant of business inventories. If total inventory depends on S^e as assumed in figure 1.1, then it should be increasingly perceived as too large relative to S in late expansion when S^e rises above S. By the same token, inventory should be increasingly perceived as too small relative to S in contraction when S^e falls below S. The data indicate that total stocks of goods held in manufacturing and trade tend to increase less than sales in recoveries (when S typically grows fast) and more than sales in late stages of expansion and in contraction (when growth of S declines and becomes negative). The inventory-sales ratio is mainly countercyclical and lagging; inventory investment is procyclical and leading, but very volatile and irregular. These patterns reflect a mixture of intended and unintended investment.

1.10   Approaches to Diversity

Because business cycles are not all alike and are subject to historical changes along with the structure and institutions of the economy, it is not surprising that the numerous efforts to model them as a uniform by-product of one type of random shock or another have failed. Endogenous nonlinear models can capture important aspects of the economy's motion, but they are very underdeveloped at present and far from the stage of empirical usefulness. Whether they can reach that stage without the aid of various shocks is uncertain.

It is difficult but necessary to explain not only why business cycles occur but also why they are diverse and changing. One possible approach is to look for identifiable major events that might have "caused" the fluctuations with the observed differential features. This view is widely adopted in popular discussions of current developments, where market reactions to economic and political, domestic and foreign news receive much continuing attention. There is certainly a great deal to be learned from keen observation of the contemporary scene aided by good economics and intuition. But the large-shock view in itself is very limited and potentially misleading because it neglects the common core of business cycles and inclines to a barren episodic treatment of the subject that ignores important lessons of both theory and history. Still, major disturbances can help to allow for the diversity of cyclical experience.

---

deny that expectations which drive forward-looking behavior are likely to be rational in the more limited sense of attempting to make the best use of the information deemed worth collecting.

With regard to the moderation of business cycles, it is not at all clear that the last 40 years had fewer major shocks than the periods 1894–1914 and 1919–39. Wars cause strong perturbations, and recent times witnessed the wars in Korea and Vietnam as well as the high tensions of the cold war, all of which produced a rising but very unsteady flow of defense expenditures. In 1989–90 the cold war was effectively won by the United States and its allies, and in 1990–91 the conflict and war with Iraq ended in a quick and low-cost victory, so military spending may be reduced. In contrast, peace prevailed in the earlier periods, though it was increasingly threatened. When the wartime cycles are excluded, recent contractions still turn out to be on average shorter and milder than their historical counterparts, and recent expansions longer, if not more vigorous.

The great advance in the production of comprehensive economic information may have contributed to the reduction in general business fluctuations by enabling both the policymakers and the public to monitor and react to macroeconomic developments in a much more informed and therefore potentially more efficient manner. It is only after World War II that quarterly data for national income and product accounts, monthly data on the money stock, and monthly data on new investment commitments and other leading indicators, among others, became publicly and relatively promptly and freely available. However, this is a neglected factor and little is known about its effects.

Another possible way of dealing with diversity is through disaggregation. In systems with many goods and markets, information is always incomplete and partly private as people must be selective and economical about what data they acquire. The resulting gaps and asymmetries, especially when combined with temporary price rigidities in many markets, can give rise to recurrent maladjustments of relative prices and quantities. Although varying by sectoral origin, such imbalances can, if sufficiently large and widespread, contribute significantly to the instability of total output and employment. They are reflected in the cyclical patterns of inventory-sales ratios, price-cost markups, and profit margins. These imbalances have been noted in early NBER studies but they appear to be important in the present moderate business cycles and mild growth cycles as well (for references, see chapters 2 and 5 and Boehm 1990). There is still much to be learned about the sources and dynamics of these diffusion processes.

In the disaggregate and extended versions of the RBC theory, random local shocks to technology and tastes are propagated across markets. Although prices are flexible, complex lags may arise because sectoral shifts are not recognized promptly and correctly, and specialized labor and capital require time to be reallocated (Black 1982; Lilien 1982). Conceivably the shifts and transfers could be concentrated during recessions because the value of the forgone output is then relatively low, but the idea encounters some serious problems.[9]

---

9. In particular, this hypothesis (Darby, Haltiwanger, and Plant 1985; S. Davis 1987) is difficult to reconcile with the observed strong inverse correlation between cyclical movement in the un-

Diversified market economies appear to be very resilient with respect to even large and persistent shifts in the composition of demand, relative prices, and technological advances. For example, in 1980–85 important parts of U.S. manufacturing suffered strong setbacks from intensive foreign competition aided by a massive appreciation of the dollar. Yet the expansion that began in 1982 endured because of the prevalence of gains elsewhere in the economy. Most sectoral disturbances are much smaller, more varied, and localized, and it is reasonable to think of their effects as canceling each other out across markets within fairly short periods of time. But this argument is not entirely compelling, and there is certainly much more scope for useful research in studying disaggregate systems than in dealing with "representative agent" models.

What generally matters more than industry detail, however, is the coverage of the main variables and specification of their interactions in the cyclical process. A theory relevant to both the major and minor cycles experienced over long time periods by the industrialized market economies will probably have to be eclectic and accommodative in the spirit of Mitchell (see chapter 5, sec. 5.2 below). It will combine tested elements from early and recent business cycle studies and from ongoing work but avoid excessive generalization and spurious rigorism.

---

employment rate and the job vacancy (or help wanted) rate. In contrast, fluctuations in aggregate demand can readily account for this relationship as well as for the clearly procyclical movements in the quit rate (Abraham and Katz 1986).