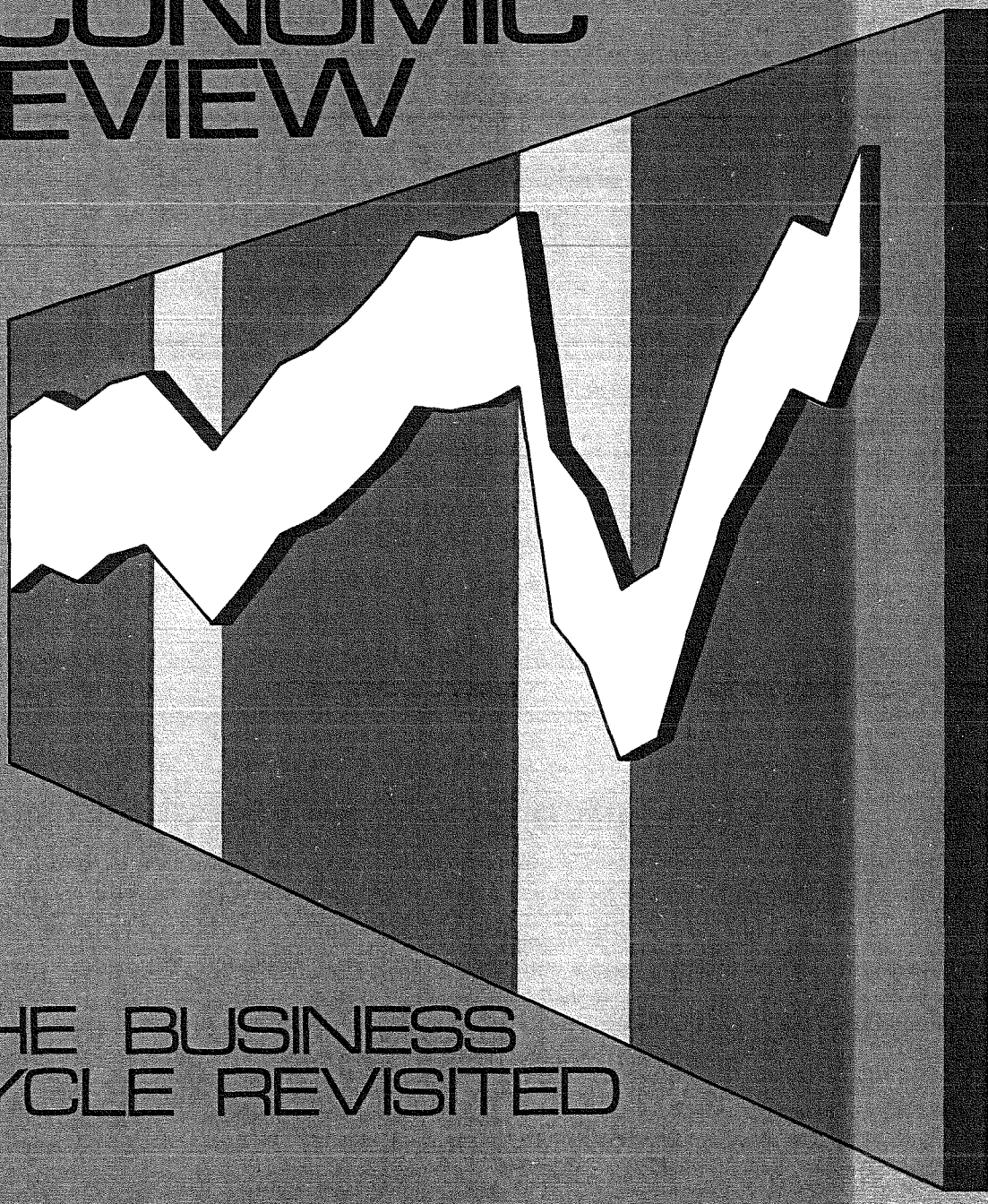
FEDERAL RESERVE BANK OF SAN FRANCISCO

ECONOMIC REVIEW



THE BUSINESS CYCLE REVISITED

SPRING 1977

The "New" Theory of the Business Cycle: Are Recessions Just Random?

Larry Butler*

Business cycles are features of all market-oriented economies. In the United States, there have been six recessions since the end of World War II, separated by generally long-lived periods of expansion. Measured from trough to trough, these cycles have varied in length from just under three years to ten full years. The associated downturns have varied greatly in severity. Until the most recent recession, whose trough was reached in early 1975, it was possible to argue that government stabilization efforts had become increasingly successful, judging by the reduction in observed movements in income. But the last recession, the most severe of the postwar period, destroyed any thoughts that we had in fact learned to control the cycle.

Despite the varying depth and duration of these business cycles, they have displayed striking similarities both in the U.S. and in other market-oriented economies. In each cycle, for example,

1. the major components of output have moved together;
2. the output of producer goods and consumer durable goods have fluctuated much more than the output of non-durable goods and services; and
3. both wages and profits have moved with output, although with a greater variability in the profits share of income. Thus income and its components have displayed a highly consistent relationship to each other.[1]

The principal features of the expansions we have experienced include the consistency of income shares and the highly irregular timing of cyclical turning points. In this article we attempt to explain the feature of timing—why recessions occur when they do—which is probably the least understood feature of the cycle. In fact, both regularities and erratic timing have been so pronounced as to require an explanation of observed cycles, that is, a *theory* of the cycle. According to the "new" theory of the cycle analyzed here, cyclical events can be seen as arising from random shocks to the economy. In this paper, we will discuss *how* such shocks can generate cycles, and more importantly, *why* we should expect them to do so in market economies.

Our analysis shows, first, that the renewal of interest in "shock" theories of the cycle stems from the recent development of the "rational expectations" literature in economics. According to that view, the public forms expectations, particularly of prices, which incorporate knowledge of both the economic structure and of the behavior of policymakers, and may offset the actions of policymakers. In this context, the business cycle can only be explained as the economy's response to "outside" shocks. The rational explanations approach is closely related to much of the pre-Keynesian theoretical tradition. As this development has proceeded, however, the new expectational models have become difficult to distinguish from older Keynesian models, which attempted to explain cycles in terms of the failure of certain prices to adjust quickly enough to clear markets (especially the labor market). The new cycle models provide important insights, the most important being the view of the cycle as a sequence of random shocks to the economy. We use a simplified version of such a model to generate business-cycle fluctuations similar to the ones experienced in the postwar period.

Despite the challenge of finding a common explanation for observed cycles and price move-

*Senior Economist, Federal Reserve Bank of San Francisco

ments, little work on such a theory was done from the mid-1930's until quite recently. The reasons for this hiatus are outlined in Section 1 below. Section 2 describes the recent development of the rational expectations literature, which has been the source of the recent renewal of interest in "shock" cycle theory. Section 3 provides a discussion of the principles governing the new "random shocks" cycle model. Finally, Section 4 provides a description of a very simple "new" cycle model.

I. From Classical to Keynesian Theory

Classical economic theory is based on the assumption that all prices can move to levels which equate supply and demand in each market. In such a world, people offer labor and capital as long as they find it to be profitable, and wages and interest adjust automatically to clear the labor and capital markets. There are no unused resources in this world, and in particular no involuntary unemployment, for the real wage adjusts to equate the supply of and the demand for labor. Though this classical approach provides an elegant way of showing how relative prices are determined, it essentially assumes away the business cycle and thus does not further our understanding of the rather large observed short-term movements in output and employment.

During the early 1930's and even before, theorists were aware of the need for some device which would allow the integration of classical value theory with the harsh facts about income and employment fluctuations which characterized business cycles. A large business-cycle literature existed, much of it focusing on the role of monetary factors in the cycle. The literature often emphasized the role of institutional rigidities in keeping the economic system away from classical equilibrium, and thus tended to favor removing such obstacles in order to dampen the cycle. Much of this work sounds quite modern, especially in its description of how external shocks initiate cycles. As a statement of what the "new" cycle theory is about, it would be hard to improve on this passage from Gottfried Haberler (1937):

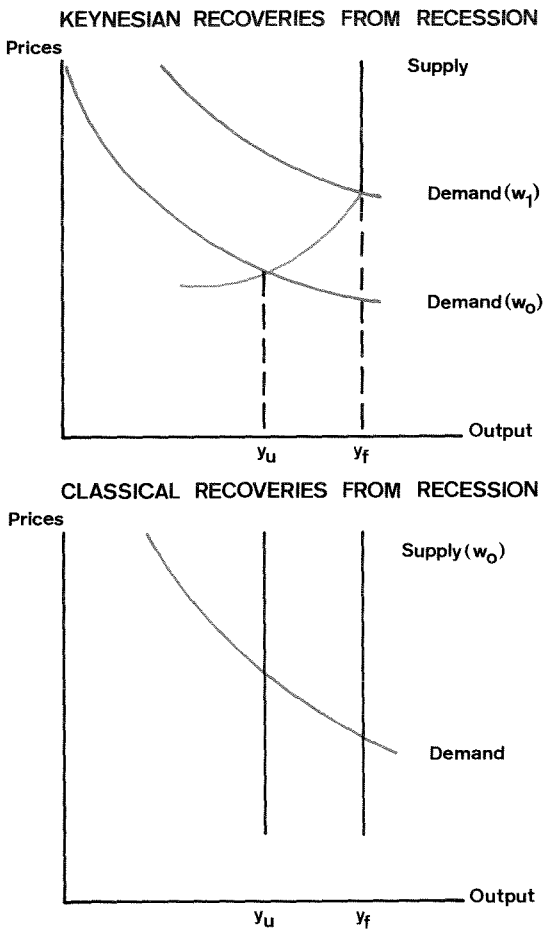
> We can compare the economic system with a pendulum or with a rocking-chair. A rocking-chair may be made to perform fairly regular swings by quite irregular impulses (shocks) from outside. (Besides, it may have a mechanism installed which makes it swing without outside forces operating on it.) In the explanation of the movement of the chair we must now distinguish two factors: the structure of the chair and the impulses from the outside—endogenous and exogenous factors. The structure of the chair is responsible for the fact that irregular shocks are transformed into fairly regular swings. An ordinary chair would ordinarily respond quite differently, although some kinds of impulse are thinkable (regular pushes and pulls) which would make it move in regular swings.[2]

Classical business cycles thus consisted of a sequence of shocks to an economy which, in most respects, was able to produce a fairly quick return to full relative-price equilibrium and thus full employment.

The Keynesian alternative to this analysis was developed in the middle and late 1930's, with the main tools of Keynesian theory in place in J.R. Hicks' *Value and Capital* (1939). This disequilibrium approach, which drops the classical assumption that all markets clear simultaneously, has come to characterize almost all macroeconomic work since Keynes. Specifically, Keynes assumed that wages are inflexible downward in the short run when output is below its full-employment level, so that a fall in prices leads to a rise in real wages and a fall in the demand for labor. This produces an underemployment equilibrium, which can be eliminated only by aggregate stimulus, in the form of expansive fiscal and monetary policy.

The distinction between the Keynesian and classical cycle models is illustrated in Chart 1. The curves describe aggregate supply and demand for output as functions of the price level. The principal difference between the two models lies in the supply curves. The vertical classical supply curve (lower panel) embodies the assumption that prices can always adjust to produce full-employment output. In contrast, the Keynesian aggregate supply curves (upper panel) assume the presence of a rigid wage rate W_0, which may yield a less-than-full-employment level of output Y_u. Expansive policy will shift the demand schedule to the right and eventually produce full employment at the level Y_f. In the bottom panel,

Chart 1



recession comes instead from a classical shock to supply, which reduces output to Y_u. Given the vertical aggregate supply curve, which reflects the assumption of flexible prices, the price effects of the shock work through the economy, and the supply curve shifts back to full employment at the output level Y_f.

The Keynesian revolution replaced the quite sophisticated relative-price mechanism of the classical model, where wages adjust to clear the labor market, with the simple assumption that nominal wages are determined "outside of the model." There was an advantage to such a shift—real income is no longer always at the full employment level—but this advantage was purchased at some cost. The relative-price mechanism, with flexible wages playing a major adjustment role, is the heart of the classical model, serving to allocate scarce labor and capital and also to determine the mix of output. Yet the old question of integrating such a price mechanism with a cycle-generating mechanism failed to surface until the late 1960's, thirty years after the Keynesian revolution.

There is a cogent theoretical reason for this anomaly. Once one accepts the key role of underemployment disequilibrium in the Keynesian short-term apparatus, it becomes clear that there is no necessary contradiction between a Keynesian short run and a classical long run. The former is characterized by disequilibrium in at least some markets, the latter by full equilibrium. In particular, it is easy to devise models in which an increase in, say, money supply increases real incomes in the short run but affects only prices in the long. Out of equilibrium, both price and output respond to a shock; on return to equilibrium, only prices are affected by the shock.[3] Keynesian theorists, in developing a way of describing the behavior of economic units which are not in equilibrium, did not see a clear need for a separate cycle theory. Their cycle theory was one of aggregate demand disequilibrium, with only a limited role for and no explanation of price movements.

The disequilibrium-equilibrium dichotomy is best exemplified in the natural rate hypothesis (NRH), first presented by Milton Friedman in 1968.[4] Suppose the economy is in equilibrium at some unemployment rate, level of income, and inflation rate. The NRH says that if there is no difference between the actual and expected rate of inflation, unemployment will be at some fixed level, which we define as its natural rate. If the economy is shocked by, let us say, a permanent increase in the growth of money, the unemployment rate will be at its old NRH level when the economy returns to equilibrium, and all of the increase in money growth will be translated into an increase in the rate of inflation. Friedman's proposition follows *entirely* from the properties of the classical model. In the absence of changes in taste or technology, the new equilibrium must be at the same level of real income, and thus at the same level of unemployment, as the old, and all of the increased money growth must appear as an increase in inflation. It is only in the "short run" that increased money supply will increase

output, and thus employment.

The NRH makes no direct statement about the way people form expectations; it just assumes that people do form them, and are correct in the long run. The NRH can thus be considered a direct application of Keynesian disequilibrium theory, early versions of which date from the late 1930's. The NRH, or something very like it, should thus have long been part of the Keynesian macroeconomic tradition. But until the late 1960's none of the main macro-models used any version of the NRH. Most instead contained a Phillips curve, which traces a relation between the rate of inflation and the rate of unemployment. The principle here differs from the NRH, which traces a relation between the difference between the *actual* and *expected* rates of inflation and the rate of unemployment. The NRH thus allows for an accelerating inflation, while the Phillips curve does not.

The importance of the distinction between what people *expect* to occur and what does occur cannot be overemphasized. In a pure classical model, the distinction does not matter, because people have perfect foresight. But if they do not have perfect foresight, they must have some means of forming expectations about their future incomes and prices. The major Keynesian macromodels assume that these expectations are formed as weighted sums of past values of the variables themselves. This device has the virtue of greatly limiting the amount of information which is relevant to the explanation of any one variable, and therefore makes the specification and estimation of particular equations relatively easy.

II. Rational Expectations

That Keynesian approach has a drawback, however, in that it is not based on any notion of how rational people form expectations. But the problem can be dealt with by assuming that people have the ability, based on all currently available information, to form unbiased estimates of future quantities and prices. Most of the economic theory based on this "rational expectations" model is close in spirit to the classical model.

Suppose someone believes that a certain set of prices will prevail, and sets his demands accordingly. Then in terms of *expected* prices, he will be in a classical world. He can be induced to move away from his equilibrium set of demands for goods only when actual prices turn out to be different from his expected price set. If actual prices are different, he immediately incorporates this new information in his expectations and moves to a new set of equilibrium demands. Except for random shocks to his demands caused by unexpected price movements, he is always in equilibrium. Moreover, the random shocks must be unrelated to earlier shocks in order for them to affect individual behavior, for if they were not, the information would be built into the next set of expectations. Because these shocks are random, there can be no possibility that a shortfall of demand in the current period will increase the probability of a further shortfall next quarter. In this world, the mere process of forming expectations prevents business cycles.

The essence of the cycle is a close relation between successive movements in output, and a model whose response to a shock is an immediate return to equilibrium might not seem to be the best vehicle for analyzing such cyclical movements. However, that would ignore a key assumption in the analysis, which is that information is costless. It is possible to devise models where all individuals have rational expectations, but do not adjust fully to new information because the cost of acquiring that information is too high to be worthwhile. This approach could lead to an integrated value and cycle theory, where everyone responds rationally to available price and output data, and yet where short-term output movements are not necessarily random.

III. Random Shocks Model

A basic way of introducing non-random errors is to place some limitation on the amount of information people have at their disposal. Suppose, for instance, that my information set does not include the price of natural gas in New York. If a shortage of gas develops in New York and the price goes up there, I should in principle respond to the increase immediately. But if I do not know of the shortage, or if I do not know how it will affect California prices, I will have no response until the New York price increase spills over to the California market. The aggregate response will be a relatively slow adjustment in both price and quantity, as information about a shock in one segment of the economy slowly becomes reflected in prices in all segments. Shocks will affect output over a span of time, and movements in output will be a moving sum of a number of successive shocks and will be related. That is, a cycle will be possible. Placing arbitrary limits on the information sets available to transactors is not elegant theoretically, but it does yield the real world's highly correlated errors.

Edmund Phelps' labor-market theory, utilizing the natural rate hypothesis,[5] indicates how the arbitrariness in this problem of information content can be eliminated. Unlike Friedman, Phelps and his followers have emphasized the short-run, rather than the long-run, properties of the NRH. In Phelps' approach, most of the emphasis has been on the role of search and other costs of finding employment, which implies that people bargain about their *incomes* rather than about their *wages*. For example, a construction worker with a high probability of being laid off during bad weather is likely to insist on a higher wage rate than a factory worker with the same skills, to compensate for working fewer hours. Thus, there is a conscious tradeoff between the wage rate and the probability of being laid off.

This result implies that expectations primarily concern quantities rather than price. For what people do is to maximize the value of the stream of their future wages, taking into account any future loss from unemployment. In this environment, despite the rationality of expectations of both prices and quantities, there is no presumption that adjustment to a new classical equilibrium will be instantaneous. It is hard to tell this world from Keynes' (or, more properly, Hicks') on any matter of principle, except that the rational-expectations literature would add one requirement: that the model used should itself generate the expectations of the variables in question. Though such a model need not contain the simple, uncorrelated errors of the pure rational expectations model, we could interpret (as that literature does) the observed errors in the model as a sequence of random shocks to the economy.

As has been known for some time, random events in time series can generate cyclical movements which have a close resemblance to economic cycles. Also, a great portion of the movement in most economic time series can be explained by the series' past history. Because the logic behind the rational-expectations approach involves the ability of transactors to reduce errors in observed price and output forecasts to randomness, the main contribution of this approach may be its ability to explain these correlated error processes and at the same time provide a reasonably good explanation of the business cycle. Yet we cannot be sure that this approach will provide an adequate description of cyclical movements. The difficulty of providing a reasonable expectational interpretation of a model increases enormously with the number of separate errors we must consider, as does also the difficulty of estimating very general lag structures. A general 12-variable model of output with 10 lags on each variable would require the estimation of $12 \times 10 = 120$ parameters, and thus would exhaust the available quarterly postwar data. The basic approach, then, must consist of capturing as much movement as possible in a small number of variables, as we attempt to do in the following model, which contains only one relevant random error.

IV. A Simple Model

Suppose the path of real income through time can be described entirely by its past history, as follows:

(1) $y = .09y^* + 1.4y_{-1} - .49y_{-2} + e$, where

y is real income,
y^* is the trend level of real income at a 3½-percent annual trend growth,
y_{-1} and y_{-2} are past values of this real-income deviation from trend, and
e is random error, uncorrelated with its own past values.[6]

We may ask two questions:

a. Is there a plausible world where this model holds?
b. How well does the model explain observed business cycles?

The answer is yes to the first question. Suppose the world to be a place where the citizenry fixes its real consumption expenditure as a percentage "a" of its expected income.[7] Then rational expectations would indicate that

$$c = ay^e = a(.09y^* + 1.4y_{-1} - .49y_{-2})$$

If we next assume that the rest of income is i, equal to investment plus government expenditure, then

$$i = y\text{-}c = (1\text{-}a)\,(.09y^* + 1.4y_{-1} - .49y_{-2}) + e$$

This simple model is compatible with both classical theory and certain empirical observations on the business cycle. First, real income is independent of nominal magnitudes in the long run, and even in the short run is randomly shocked by those magnitudes only through their impact on the error term. In the long run (say, 20 quarters ahead), the expected value of real income is y^*, the trend level of real income. This fact is compatible with Keynesian and classical theory, and also with the natural rate hypothesis. But the model also says that a rise in nominal magnitudes, such as monetary or fiscal policy variables, will exert a single-period shock effect on the real economy, through its potential effect on the random error term. The model incorporates fiscal or monetary influences into this error term by assuming that the size of these effects is too small to be distinguishable from random noise.

The effect on income of any such shock will dissipate only slowly. It will be felt first through its direct impact, then in the following quarter through its effect on the y_{-1} term, in the quarter after that through its effect on both y_{-1} and y_{-2}, and so on, with the equation used as a forecaster of longer and longer periods ahead. The results of such a forecast sequence are given in the table below. This model is compatible with short-term restoration of *price* equilibrium to the economy, as in the pure rational-expectations model, but it is not compatible with short-term *quantity* equilibrium.

Quarter Ahead	Effect of Shock e_0 on Real Income in Quarter K
0	e_0
1	$1.40e_0$
2	$1.47e_0$
3	$1.37e_0$
4	$1.20e_0$
5	$1.01e_0$
6	$.82e_0$
7	$.66e_0$
8	$.52e_0$
.	.
.	.
.	.
12	$.18e_0$
16	$.06e_0$
20	$.02e_0$

The model is also compatible with one of the broader cyclical generalizations—the much greater amplitude of movements in investment than of movements in consumption. In the short run, the impact of any shock to income falls entirely on investment, because consumption is a fixed function of past income. As the model transmits shocks, they appear initially as unanticipated investment, and are then built into consumption over a span of time. Two consecutive large negative shocks to real income—a recession, by the normal definition—will produce a large decline in real investment and only a small movement in consumption.

How well does this simple model describe the cyclical movements of the past several decades? The standard error of the above equation, fitted

to quarterly U.S. data for the 1952-75 period (96 quarters), is 4.0 percent of GNP, with an annual trend growth in income of 3.5 percent of GNP. These figures may be used to indicate how well the model describes actual cycles. Based on the relation between trend growth and standard error, the probability of any one observation showing an actual decline in income is .19,[8] and thus 18 quarters of decline (.19 x 96) should occur in the period of fit. There actually were 18 quarters of decline in the observation period, but this is true almost by definition. The method of fit was designed to produce empirically uncorrelated errors, with high and low errors in roughly the frequency predicted by the bell-shaped curve of the normal statistical distribution.

More interesting is how well the equation predicts a second decline following the first—that is, the actual occurrence of a recession, defined as two quarters of consecutive decline in real GNP. Because the equation's lagged GNP terms make for a very sluggish GNP response to the first decline, the second decline is considerably more likely than the first, with a probability of .38. The probability of two consecutive declines is thus .19x. 38 = .073. The equation thus "predicts" .073 x 96 = 7 recessions in the period, in contrast to the 5 recessions which actually occurred.

Where the equation begins to slip is in predicting longer recessions. Similar, though somewhat more involved, calculations of the type used above yield for the 1952-75 period:

Length of Recession (Quarters)	Predicted Number	Actual Number
2 or more	7	5
3 or more	3	4
4 or more	1	2
5	0	2

Thus the relation tends to slightly understate the frequency of long recessions, and to overstate the frequency of short recessions.

The real problem, though, lies in the prediction of recovery periods. Each of the 5 recessions in the 1952-75 period, including the most recent one, has been followed by about six quarters of extremely high economic growth. The model simply failed to pick up these fluctuations. The model predicts relatively slow turnarounds in real growth rates, so that (for example) a two-quarter recession followed by three quarters of very high real growth would be marginally less probable than a recession of five quarters. And as the table indicates, the model predicts no such lengthy recessions.

The explanation has to do with the nature of simple autoregressive schemes. Whatever their virtues, such schemes tend to say that a variable's level next quarter will be quite similar to its level this quarter. In rate-of-growth terms, our equation says that this quarter's expected growth rate for GNP will equal 60 percent of the trend growth of 3½ percent plus 40 percent of last quarter's actual growth, plus a small weight moving the level of income back toward its trend line.[9] So in a fundamental way, the equation does not have the capacity to produce large quarter-to-quarter swings in the level of income, though the relatively high standard error suggests the occurrence of large *unsystematic* swings in growth rates. Thus the model reproduces the observed short, sharp pattern of recessionary decline with more precision than it does the long, high growth pattern of early recovery.

We have argued that even this simple random-shocks model—a type favored in the "new" cycle theory—can be used to generate behavior which is strongly reminiscent of some of the main characteristics of the observed business cycle. It does so imperfectly, and in particular somewhat understates the duration of the typical downturn and the strength of the ensuing early recovery. But this model assumes a single-source random event, which must thus incorporate every aspect of random influence on the economy from the ordinary monetary and fiscal shocks to world commodity-price booms. Because of the frequent difference in character of these different influences, it should be possible to improve on the single-shock model by providing a better explanation of the sources of shocks.

V. Summary and Conclusions

Interest in a "new" business-cycle model began with the development of rational-expectations models in the late 1960's. In these models, it was found that with complete (or nearly complete) information, rational transactors would act in a way which would reduce observed errors in both prices and quantities to uncorrelated random noise. In the case of non-random errors, transactors would incorporate their information in succeeding price forecasts. No cycle, in the ordinary sense, would be possible. The next step in developing a cyclical model involved the attempt, by now largely successful, to provide limitations on the information available to transactors, which would allow for serially correlated observations in quantities and perhaps prices as well.

We argued initially that, in light of this development, it has become much harder to tell these models apart from the much older (and numerous) Keynesian disequilibrium models. Models which embody both rational expectations and slow adjustment are clearly feasible. In the work of Phelps and others, quantity disequilibrium in the labor market results from discontinuous search and transactions costs of various kinds—factors which tend to limit the information available to transactors in that market. And in the rational-expectations model with correlated errors, quantities at least do not fully adjust to shocks instantly, so that this model fits into the Hicksian dichotomy between short-term disequilibrium and long-term equilibrium. Moreover, Phelps' argument is essentially that people bargain over their incomes and not their wages, trading future layoffs against wage increases. Thus the formation of rational quantity (and price) expectations adds one requirement to the usual disequilibrium model, that the model itself generate expectations. In that event, it will be possible to interpret observed errors as they are interpreted in the "new" cycle theory (and in our simple model), as a sequence of random shocks to the economy.

The principal achievement of the "new" cycle model is an accurate description of cyclical timing. In the context of our very simple model, there is no problem in explaining why recessions are short, sharp, and irregular in timing. The timing factor suggests that the economy is subject to random shocks from a variety of sources, and that these will sometimes be severe enough to generate recessions. Further, if the shocks are in fact random, the recessions we observe will in fact be short and sharp. The major thing missing from our simple model is an adequate description of Haberler's "rocking chair": the perception of the economy embodied in the model is too simple to explain how the economy works itself out of recession.

FOOTNOTES

1. The consistency of these similarities is documented by Herbert Runyon in this issue of the **Review.**

2. Gottfried Haberler, **Prosperity and Depression,** Geneva, League of Nations, 1939. His book is perhaps the culmination of the classical cycle-theory tradition. With its late date, it contains an extensive discussion of Keynesian theory, but little reference to the formal disequilibrium theory which was then emerging.

3. This statement summarizes what Samuelson calls the "neoclassical synthesis" of Keynesian and classical theory.

4. Milton Friedman, "The Role of Monetary Policy," **American Economic Review,** 1968, pp. 1-17.

5. Edmund Phelps, "Money-wage Dynamics and Labor-Market Equilibrium," **Journal of Political Economy,** 1968, pp. 678-711. Friedman and Phelps are given credit for simultaneous authorship of the NRH. It is of course a feature of the older classical model as well.

6. This relation is in fact the best description of real income solely in terms of its past values and a random error, as fitted by Box-Jenkins methods to real GNP data for the 1952-75 period.

7. This formulation is a very simple version of the standard behavioral explanation of movements in consumption, the permanent-income hypothesis. For a more detailed explanation of the relation between permanent income and rational-expectations hypothesis, see Kurt Dew, "Market Response to Economic Policies," this **Review,** Fall 1976, pp. 20-30.

8. This calculation assumes normally distributed errors with a mean of 3.5 percent and a standard error of 4.0 percent. Zero growth in the calculation is .88 standard errors below the mean, and 19 percent of the normal distribution is more than .88 standard errors less than the mean.

9. This small weight is what gives the model its long-run classical properties.

FURTHER REFERENCES

R.J. Barro, "Rational Expectations and the Role of Monetary Policy," **Journal of Monetary Economics** 1976, pp. 1-32.

R.J. Gordon, "Recent Developments in the Theory of Inflation and Unemployment," **Journal of Monetary Economics** 1976, pp. 185-220.

R.E. Lucas, "An Equilibrium Model of the Business Cycle," **Journal of Political Economy** 1975, pp. 1113-44.

R.E. Lucas, "Understanding Business Cycles," prepared for the Kiel Conference on Growth Without Inflation, 1976.

T.J. Sargent, "A Classical Macroeconometric Model for the United States," **Journal of Political Economy** 1976, pp. 207-37.