Economics / criticisms of models by other authors, 1935-1982

ok

Jay
This is Mayer's Presidential address given to the Western Economics Association. I gave you copies of my correspondence with him earlier. You might want to keep this along with the Mayer correspondence.

P.S. Mayer's letter to me is in reference to Nat's my paper, "Alternative Tests for the Selection of Model Variables"

Requiring such tests of robustness would substantially toughen the standards of acceptability, and should also work towards reducing the frequency of contradictory results in the literature. Authors would, so to speak, be their own critics by mining the data *against* themselves, and bring out objections that are now frequently raised by others in comments. This raising of standards is justified because those who present new results should have the burden of the proof thrust upon them, and the burden of the proof amounts to more than just showing that there happens to be one particular form of the basic regression equation that yields the desired result.[20] As both Keynes (1973, pp. 287 and 294) and Friedman (1951, p. 108) for once in agreement, have pointed out, one can get a good statistical fit merely by repeated testing, but such a fit proves very little.[21]

Fifth, it is important that authors not use up all their data in *fitting* their regressions, but leave some as a hold-out sample against which to *test* the regressions. This would not really waste data points because, once the hypothesis has been successfully tested, the regression can be rerun using all the data points to refine the coefficient estimates.

Sixth, the journals should publish papers that find statistically insignificant results. This would not only remove the great pressure to stomp on the data until they give in and yield a t value of 2 or more (as a saying has it: "if you just torture the data long enough, they will confess"), but it would also prevent others from wasting time replicating an unsuccessful project (see Feige, 1975). Journals should also encourage economists to replicate previously published results.[22] In other sciences replication is done continually. Presumably this is because replication is a good way of training students in laboratory techniques. But in economics we do this much less frequently probably, in part, because, in the absence of lab experiments that produce new data, all the "replication" we can do is rerunning the old regressions. Hence, erroneous results are allowed to remain in the literature. This lack of replications is particularly serious because, as just discussed, economists pay insufficient attention to avoiding calculating errors. Some foundation should finance a program wherein graduate students would rerun each year, say 10 percent of the empirical work published the previous year in the journals. This would not only catch many mistakes, but the potential embarrassment of being caught in a calculating error should make researchers more careful.

Seventh, authors using unpublished data should be required to make

20. Would this requirement result in journals having to close down because they have an insufficient number of acceptable manuscripts? No it would not, because a paper that shows that a particular hypothesis cannot at present be tested because some forms of the underlying regression equation support it, while others reject it, is a contribution that warrants publication.

21. For a rigorous demonstration, see Bacon (1977).

22. Currently at least two journals, the *Journal of Political Economy* and *Journal of Consumer Research* have offered to publish replications.

Massachusetts Institute of Technology
Alfred P. Sloan School of Management
50 Memorial Drive
Cambridge, Massachusetts, 02139

System Dynamics Group
N51-321
November 10, 1980

Professor Thomas Mayer
Professor of Economics
University of California, Davis
Davis, California 95616

Dear Professor Mayer:

Thank you for your letter of September 28, including your comments on the paper by Professor Mass and myself. I very much appreciate your taking the time to read the paper as carefully as you obviously have, and I am pleased by the similarity in our views regarding statistical testing.

A letter is hardly an adequate way to continue this conversation and I hope that we will be able to continue it in person sometime in the future. However, let me make a few comments in response to the questions that you raised.

You are absolutely correct that the usefulness of statistical significance testing depends very much on the purpose of the model that you are building. If, for example, the purpose of the model is to forecast a particular variable with a minimum forecast error, and the model consists of only one equation, then statistically insignificant influences on the variable in question are not important variables for that model. In this spirit, I think the example you described of a variable which is statistically insignificant in one period, then becomes statistically significant in a subsequent period is perfectly consistent. One could say the variable in question is insignificant during the initial period and then becomes significant during the subsequent period. I, like yourself, see no paradox here.

On the other hand, if the purpose of your model is to understand and increase ability to influence some particular economic phenomenon, then the usefulness of traditional statistical significance tests becomes much less clear-cut. There are several different ways in which a particular variable may be statistically insignificant yet be very important as a causal factor and as a channel for influencing a particular phenomenon. One way in which this can happen, of course, is the problem of collinearity. Another, pointed out in Professor Mass' and my paper is the problem of measurement error. In thinking about these problems, we felt that the clearest, most reliable interpretation of the statistical significance test for such a model is that it tells you the extent to which a particular variable's hypothesized influence is estimable. That is, given some particular body of data and some particular hypothesis, to what extent does the data allow you to get a "tight" estimate of the parameter associated with the hypothesis? Now, clearly, given this interpretation, a hypothesized influence may be difficult to estimate yet be extremely important. "Estimability" and importance may be related, but there is no necessary reason why they should be related in all cases.

You suggested in your letter, if I understood the argument, that statistical insignificance could allow one to reject a theory that a variable X "dominates" a variable Y. Now, I take this phrase to mean that a variable X has a particularly important influence on Y. In our paper, we showed an example of a hypothesized relationship which was statistically insignificant given data with 10% measurement error. Yet, if that relationship was removed from the model which generated the data, the simulated behavior of the model was altered dramatically. In fact, the statistically insignificant relationship was crucial for the behavior of the model and for most policy insights that might be generated by the model. This seems to me to be a good example of a case where one cannot reject a theory that X "dominates" Y on the basis of the statistically insignificant effect of X on Y.

These sort of problems in the interpretation of statistical significance tests lead us to propose the " model behavior" testing approach as an alternative to statistical significance testing. Again, this comes back to the purpose for which a model is being built, and this may in fact, be a link which is not clearly made in our original paper. If one wants to understand some particular economic phenomenon such as persistent inflation or the causes of the short-term business cycle, then we propose that modelling should focus on the economic phenomena as a pattern of behavior over time. It seems to us that most crucial economic issues can be addressed in this way. The question then becomes, if a model is capable of internally generating the particular pattern of behavior over time, what relationships within the model are most important for replicating the observed behavioral pattern? Such testing is most interesting when the model incorporates competing theories of the phenomenon in question. All of this testing takes place relative to a purpose of understanding specific observed patterns of economic behavior. I even question the applicability of the model behavior testing approach if one's purpose is to generate a minimum forecast error. It seems to me, if such a purpose dominates, the whole question of causality falls into an ambiguous state.

As for your detailed comments, your observation concerning the t-test as the sample size increases illuminates the problem in a rather different way. Clearly, if all parameters have statistically significant t-tests, then there is no discriminating power in the tests. I suspect, in such a situation, many would be tempted to interpret the t-test or partial correlation coefficient still as a measure of relative strength--a higher partial correlation coefficient signifying that a particular variable "explains" more of the movement of the dependent variable. Under such circumstances, one might test a hypothesis by looking at the associated coefficient estimate. If a parameter is statistically significant and has the incorrect sign, then that should be some indication of the correctness of the related hypothesis. All this, of course, depends on satisfying the basic assumptions underlying the estimator's consistency. The sensitivity of standard econometric estimators to reasonable variations in their underlying assumptions represents, I believe, another line of legitimate criticism. Some years earlier, after finding some discouraging results in this area, I was again greeted with the response that there's nothing new here. I enclose an earlier paper of mine on this point.

The beta coefficient allows one to contrast the relative "strength" of different explanatory variables. Consider an ideal case of a regression where all consistency assumptions are satisfied and all parameter estimates are statistically significant. In such a case, the beta coefficients reliably indicate how large a variation in Y one can expect from a certain variation in each explanatory variable X_i, both variations measured as fractions of the respective variables' standard deviations. One might conclude that, if X_1 has a beta coefficient of .8 and X_2 has a beta coefficient of .2, then X_1 is 4 times "more important." That is, a one standard deviation variation in X_1 produces 4 times the effect of a one standard deviation variation in X_2. This gets us very close to a crucial issue I believe. (An issue which is equally important to the significance of the elasticities which are the focus of so much applied econometric work.)

One way to approach the issue is to ask the question, "Over what time period is X_1 more important than X_2?" If all variables in the regression analysis are contemporaneous (i.e., not lagged), the answer implicitly is "during the current period." But, a variable may have no effect during the current period yet have major influences in the longer term. For example, the effect of X_2 on Y may be part of a positive feedback loop which dominates the behavior of Y over the long term. In such a case, the size of the parameter or beta coefficient for X_2 gives <u>absolutely no indication</u> of the importance of the hypothesized dependence of Y on X_2 for the behavior of Y over time. The importance of the dependence can only be determined by examining the larger feedback structure within which the effect of X_2 on Y is embedded. So, in a sense, one can say the regression results in themselves can only tell us how important alternative variables are in determining the immediate (current period) response of Y. This is in the <u>ideal</u> case where all assumptions regarding consistency are satisfied and statistically significant estimates are obtained.

Of course, all variables need not be contemporaneous--some may be lagged and some may enter through distributed lags. But I do not think this alters the basic problem. One can show through distributed-lag analysis that the strongest effect of one variable on another occurs with a two- to four-period lag. But, this does not help in determining the relative importance of different variables. Distributed lag estimation will not allow you to discover an important oscillatory mechanism (which must involve the *interaction* of at least two variables) nor will it even handle the simple case of the positive feedback loop described above. In that case, the effect of X_2 on Y may occur during the current period, may have a small coefficient and a small beta coefficient, yet dominate behavior *due to the way* Y feeds back to reinforce changes in X_2.

Taken together, these arguments concerning the limitations of t-tests, partial correlation coefficients, regular regression coefficients or beta coefficients, can be legitimately generalized, I believe, to understanding the limitations of *all* tests based on standard regression analysis, including the R^2. If dynamic behavior is due to the interactions over time of different economic variables, how can we expect to test the significance of alternative theories except by explicitly considering the interactions they imply? This is the basic idea behind model-behavior testing. Many econometric system model builders are coming to realize these same problems, but few, as far as I can see, have "cut the strings" to their regression training and really begun serious development of appropriate whole-system testing methodologies.

I'm not exactly sure how to respond to your questions regarding the synthetic data experiment. Perhaps the above discussion will clarify matters regarding the regression coefficient and the R_2. It is important to keep in mind that the causal relationships in the model being estimated in the experiment are perfectly specified. One cannot say that the theory is "confused." The theory is in fact perfect. Moreover, we can assume that the modeler correctly expects the coefficient associated with the theory to be negative and even has a good *a priori* estimate of its value. Neverthe-less, the estimated effect of the theoretical relationship is statistically insignificant given error-corrupted data. The point here is simply this: is the modeler justified in rejecting the theory? We argue no. The magnitude of the modeler's mistake if he rejected the theory is illustrated by showing the theory's importance for the behavior of the model which generated the data.

I think there are several similarities between Lucas's critique of econometrics and mine, although the specific points differ. Surely, statistical data generated when a system was operated under one set of policies may be of little use in understanding the possible consequences of an alternative set of policies. When statistical analysis becomes the modeling and policy analysis paradigm, emphasis is placed on predictions within the regime of past behavior rather than broad policy analysis. This can significantly bias policy analysis, by prejudicing us against a new policy which pushes the system into a range of behavior not exhibited in recent history. Does that mean that we should not search for policies which can alter undesirable behavior patterns?

Yes, the focus on investment is an important aspect of the Keynesian vision of the business cycle. Although European cycles are somewhat longer than American cycles, I do not think they are so long as to reasonably admit to investment being a significant determining factor. This is not to say that investment might not fluctuate substantially over the business cycle, as of course it does in the American economy as well. However, again, correlation with business cycle behavior does not necessarily indicate a significant causal role. In virtually all versions of the System Dynamics National Model that we have explored, the delays involved in deciding upon, acquiring, and depreciating capital investment are too long to make it a significant causal mechanism in generating the business cycle.

I agree that, when all is said and done, the primary reasons statistical significance tests are relied upon so heavily in econometrics <u>is</u> habit and cost. Ultimately, no matter how much one understands theoretical limitations, decisions regarding equation specification must be made, and they will be made on the basis of the information one has available. If the only testing information available comes from standard regression tests than it will strongly influence the decision. Moreover, of course, the problem is much subtler because of the demands of journal publication. Regression tests have the particular advantage that the entire testing process can be summarized in a few lines. This is consistent with the demands of journal editors for short articles that can be read in 15 minutes or less. One can scan the statistical significance results in a matter of seconds and find out the sum total of an individual's research. I do think this is a considerable problem. Effectively communicating the results of model behavior testing demands much of the reader. For, not only are the results of the test per se important, but one must understand <u>why</u> those results were generated. Model behavior testing is illuminating <u>only</u> insofar as it clarifies underlying causal mechanisms, distinguishing those mechanisms capable of generating the economic phenomena of interest from those that are not.

Thank you once again for your encouraging letter. Please forgive the lengthiness of my reply. Your letter raised many interesting questions, several of which I had never had the opportunity to think through carefully. I hope that we have a chance to interact further. Our work on the System Dynamics National Model focuses on developing the model-behavior testing approach and applying it to test alternative theories of persistent inflation, economic cycles, and related economic issues. We also hope to continue experimental evaluations of statistical methods, such as the "Granger causality tests." Please let me know if this research is of interest to you.

Sincerely,

Peter M. Senge
Assistant Professor of Management

UNIVERSITY OF CALIFORNIA, DAVIS

BERKELEY • DAVIS • IRVINE • LOS ANGELES • RIVERSIDE • SAN DIEGO • SAN FRANCISCO — SANTA BARBARA • SANTA CRUZ

DEPARTMENT OF ECONOMICS — DAVIS, CALIFORNIA 95616

Professor Peter Senge
Department of Management
MIT, Cambridge, Mass.

September 28, 1980

Dear Professor Senge:

First, I would like to apologize for being so tardy in replying to your letter of July 17 and commenting on your manuscript. I read it with great interest soon after receiving it and intened to write to you right away. Somehow, I never got the letter written, and only recently, when cleaning up my desk did I find it buried in a pile of unfinsihed correspondence.

I think your paper is a very important one, and I hope it will have a powerful influence on the way econometricians proceed. However, I wonder if one couldn't argue along the following lines: suppose that x has a low t coefficient because the variance of the dependent variable ,y, is high, that is, it is influenced by many other variables besides x. On the usual criteria an econometrician would then say that since x is insignificant we can treat this variable as " unimportant". Isn't there a sense in which this is right because it tells us that x does not really account for the behavior of y? Now, of course, it is true also that if in the next period the other variables that affect y are fairly stable then, all of a sudden, as the variance of y declines, x will turn out to be significant. But why wouldn't it be fair then to say that now x is " important"? Admittedly, this implies that x can be unimportant in one period and important in the next even though the way in which , and the amount by which, x affects y has not changed. But this does not strike me as so apardoxical.

Second, why do we want to know whether x is significant ? If it is because we want to influence y by changing x, then an insignificant coefficient tells us that (assuming the sign is right) we will move y in the right direction we will not really do much good. And this seems legitimate. On the other hand, assume that we want to test a theory that tells us that x <u>affects</u> y, as distinct from a theory that tells us that x dominates y. If the coefficient of x is not significant we can reject the latter ,but not the former theory. But if the coefficient is significant with the wrong sign we can reject both theories. And in in-between cases we can learn something about the plausibility of the theory. Hence, while I agree with you that much of the significance testing that goes on is mindless, it still seems to me that there are some carefully circumscribed situations in which it is useful.

Here are some detailed comments:

p.8. You might want to discuss here also the extreme case in which the " sample" is really the entire relevant universe so that t tests are meaningless. And also as N approaches infinity everything should become significant, so that there is an obvious warning here about treating t= 2 as the way to test a theory.

p.9. Actually, for some reason, very few people give the partials.

p.10. Does the criticism of the partial R^2 also apply to the full R^2 ?

p.11. Is the beta coefficient any help in this sort of situation ?

p.21. Isn't the point you discuss here a confusion of the regression coefficient and R^2 ?

p.25. I am confused here and may well be missing the point, but isn't the problem you discuss really something else: the theory that is being brought to the empirical test is confused. The thing that validates the theory is a low coefficient, but the econometrician thinks it is a high coefficient that validates his theory. This is not really the fault of regression analysis,but of bad theorizing. Suppose you had taken another example in which there is less confusion about what coefficient size validates the theory, for example a regression of GNP on the size of the federal deficit and money. Would your argument still apply ?

p.28. The point you are making here seems very similar to Lucas' point that if one changes policy, then the old data may no longer contain useful information.

p.41. The focus on investment as the cause of the cycle is really a major part of the " Keynesian vision."

p.47. I am not sure about this, but arn't European cycles longer, so that the fixed capital explanation may apply better to them ?

p.49. Isn't another reason for using single equations tests that it is much cheaper?

But, all in all, thank you for sending me your very important paper. If econometricians tell you that you are kicking a dead horse, just tell ~~than~~ that it is better to kick a dead horse than to try to ride it as they do. them

With best wishes,

Sincerely,

Thomas Mayer

Thomas Mayer
Professor of
Economics

ECONOMICS AS A HARD SCIENCE: REALISTIC GOAL OR WISHFUL THINKING?

THOMAS MAYER*

Many economists believe that we should strive to turn economics into a hard science. Indeed, if this goal is feasible it is hard to see how anyone could possibly object to it. But, I will argue that it is overly ambitious, premature, and more likely to do harm than good. I do not deny that economics may perhaps ultimately become a hard science. To be sure, at present it is hard to see how we can ever achieve such a high degree of certainty, given the limited potential for meaningful controlled experiments, and also the fact that our conclusions affect the data (human behavior) they apply to. But the fact that one cannot see at present how something could be invented in the future is hardly evidence that it will never be invented (Cf. Popper, 1961). My argument that economics is currently very different from a hard science, therefore, does not rest on any fundamental dichotomy between the natural and the social sciences, but is based on much more mundane considerations.

I.

Those who claim that economics is becoming a hard science point to the greatly increased use of mathematics in economics. And, indeed, economists no longer are differentiated from the hard scientists by their ignorance of advanced mathematics. If knowledge of mathematics is all that would be required to make a hard science, then many, though certainly not all economists could indeed claim to be hard scientists. But mathematics itself is not an empirical science. To make economics into a rigorous *empirical* science requires that we have reliable methods of testing hypotheses.[1] With all due respect to the great contribution that mathematical economics has made, it is the ability to test hypotheses rather than the use of advanced mathematics to formulate hypotheses that is the distinguishing mark of a hard empirical science.[2] Hence, to

*University of California, Davis. Presidential address delivered at the 54th Annual Conference of the Western Economic Association, Las Vegas, 1979. I am indebted for helpful comments to Cliff Attfield, George Benston, Martin Bronfenbrenner, Robert Ferber, Jay Helms, David Laidler, Alan Olmstead, Boris Pesek, Robert Renshaw, Steven Sheffrin, Paul Strassmann, and R.M. Sundrum, who are not responsible for any remaining errors.

1. Another possibility is, of course, to turn economics into a deductive science; i.e., praxeology as advocated by the Austrian school. But most economists, rightly I believe, reject this course.

2. Norman Storer (1967) has argued that the hard sciences are distinguished from the soft sciences by their greater use of mathematics. The hard sciences are those in which "error, irrelevance or sloppy thinking" can be detected relatively easily, and mathematics facilitates such detection. However, Storer includes statistical testing in his concept of mathematics, and as I will try to show below, in economics such testing has not eliminated sloppy thinking. Moreover, it would be hard to argue that mathematics has helped to purge economics of irrelevance.

see whether economics is within hailing distance of being a hard empirical science, we have to see how reliable our techniques for testing hypotheses are.

II.

The answer, unfortunately, is that they are not at all good. This judgment is based, not on any sophisticated and subtle criticism of *theoretical* econometrics, but on the actual procedures followed by the workaday econometrician in turning out the "applied econometrics" papers that appear so frequently in our journals.[3] Techniques that are not subject to many of the following criticisms may exist, but they are not the ones being used in most of the applied work.

Suppose a new question arises and econometric studies are undertaken to answer it. Will we be much closer to an answer after these studies are completed than before? This is surely the acid test of whether economics is a hard science. And the answer is disheartening. After the first study is completed one can easily feel optimistic; arguments based on arm-waving can now be replaced with precise coefficients calculated to several decimal places, *t* statistics, Durbin-Watson statistics, etc. So far so good. But sooner or later some of the other studies that have been undertaken will also be completed. Then optimism is likely to vanish. It is highly probable that some of these studies will reach conflicting conclusions. And what is just as bad, we do not really have an effective way of deciding which ones are correct. Hence, everyone can continue to adhere to the position he or she held prior to the appearance of the empirical tests, and justify this position by citing the supportive econometric results.[4] All that has happened is that arm-waving has been replaced by *t* coefficient waving. Perhaps this is a bit of an overstatement — a few false hypotheses are rejected, but many hypotheses — however contradictory — cannot be rejected, and hence coexist.

And what is just as bad, even if all or most of the econometric evidence points in the same direction, those who hold the contrary view need not be intimidated by this since there is now some empirical evidence that many econometric results are not reliable. The prime piece of evidence is a study by Michael Lovell (1975) in which he used a Monte-Carlo technique to test a standard applied econometrics procedure. His first step was to take consumption functions and use them to generate consump-

3. For criticisms of theoretical econometrics, see Brunner (1972) as well as Streissler (1970) who also discusses some problems in actually applying econometrics. In the subsequent discussion my focus will be primarily on the use of applied econometrics in macroeconomics. In other fields of economics, such as agricultural economics (see Leontief, 1971, p. 4) the problem may be less severe. For a criticism of the literature in finance along the lines of this paper see Friend (1973).

4. As Don Patinkin (1972, p. 142) has remarked, "I will begin to believe in economics as a science when out of Yale there comes an empirical Ph.D. thesis demonstrating the supremacy of monetary policy in some historical episode and out of Chicago, one demonstrating the supremacy of fiscal policy."

tion data. He then added random error terms of reasonable size, so that he got a set of consumption data such as those that would be generated by stochastic consumption functions. His next step was to act as though these were data generated by the real world, and to try to see, by the usual econometric procedure, which of several possible consumption functions generated each of them. And what happened? It turned out that someone who did not know ahead of time which independent variables generated the data would usually not have been able to discover this by looking at the diagnostic statistics of the various consumption functions. This is not necessarily the fault of theoretical econometrics; it simply means that, given the limited powers of these tests, the available data are insufficient to distinguish sharply between valid and invalid hypotheses. The theoretically correct thing is, therefore, not to try to do so, and instead to devote more effort to data collection. But our journals are full of papers that use the techniques which Lovell's paper shows to be unreliable.

A similar conclusion is indicated by Robert Ferber's (1953 and 1956) comparisons of consumption functions. He showed that the consumption function that predicts best during the sample period frequently does not do so during the post-sample period. Similarly, Martin Schupak (1962) in looking at the demand for various types of food products, beverages, tobacco and fuels found that there was only a weak relation between the fit over a ten-year sample period, and predictive accuracy. Comparing five different regressions, it turned out that in less than a quarter of the cases did the regression that gave the best sample period fit also predict best. When I extended this approach by looking at the use of regression equations for a wide variety of problems, investment functions, money supply and demand functions, econometric models, etc. it again turned out that goodness of fit during the sample period is an unreliable guide to performance in the post-sample period (Mayer, 1975). Even if one compares only three hypotheses, the probability turns out to be less than two thirds that the one that performed best in the sample period also performed best in the postsample period. And if one compares four hypotheses, the probability drops to about one third. While this is better than one would expect on a completely random basis, it is a far distance from our claim to verify hypotheses at the 5 percent level.[5]

To be sure, one might well argue that the results of just a few studies like this are not sufficient to invalidate our standard testing procedures.

5. For an interesting discussion of how to evaluate the forecasting accuracy of regressions, see Armstrong (1978a, Chapter 13). He ranks goodness of fit in the sample period, which we use so much in econometrics as the least reliable method. Since I am dealing with hypothesis testing rather than with forecasting, it is not necessary to consider the question whether econometric models forecast better than naive models. For evidence on this, see Armstrong (1978b) and Zarnowitz (1979). Any superiority of econometric models need not result from their being an accurate representation of the economy, but could result from their picking up various autoregressive features of the economy, as well as from the ad hoc adjustments that the managers of models usually make to the raw output of their models.

But where is the opposing empirical evidence that these techniques actually work in the sense of furnishing reliable guides to policy?

It would, therefore, be interesting to see the results of some further tests of the validity of applied econometrics. One possible test is to rerun a sample of regressions several times, each time omitting a year, to see if this changes the results substantially. It shouldn't, but the only time I did rerun someone's regression, leaving out a single year in which circumstances were unusual, it changed the results entirely (Mayer 1978). Similarly, when Howrey and Hymans (1978, p. 666) dropped a single year from a consumption function that had been used to estimate the effect of interest rates on saving, the *t* value of the coefficient of the yield on savings fell from −3.24 to −1.62.[6] Another test is to take a series of empirical studies that rely heavily on the National Income Accounts data and rerun them, substituting the recently revised data for the ones used in the original study. In principle, the use of more accurate data should lead to a better fit, but if the fit was raised artificially by data mining, then it is likely to deteriorate when revised data are substituted for the data originally used. A third test is to take a sample of empirical studies and rerun them to see what proportion of them contain careless errors large enough to change the conclusions significantly. Until we have the results of these or other studies, and can say that they support our standard econometric results, we should conclude, on the admittedly limited empirical evidence that *is* currently available, that our procedures are just not adequate.

III.

Lovell's results suggest that the procedures actually followed in much applied econometrics work do not enable us to distinguish between true and false hypotheses, presumably because of high multicollinearity and the paucity of data points. This is an inherent problem faced by economics, as is the familiar and more general problem that behavior parameters are not as stable as those in the natural sciences. But to these unavoidable problems we have added avoidable ones, and it is these that I mainly want to discuss. In doing so I do not want to sound like a hellfire preacher. Not only do we find examples of "unscientific" behavior also in the hard sciences,[7] but the fact that despite the high pay-off that publication has, there is a limit to the gilding of weeds that economists do, suggests that the moral standards of our profession are fairly high.

But even so, econometric work includes far too many examples of

6. This problem could be ameliorated by the use of some techniques suggested by Belsley, Kuh and Welch (forthcoming).

7. Thus the great physicist Max Plank lamented that "a new scientific truth does not triumph by convincing its opponents and making them see the light, but rather because its opponents eventually die, and a new generation grows up that is familiar with it. (Cited in Kuhn, 1962, p. 150.)

game playing, or what Frisch (1970) in his criticism of certain types of mathematical economics has called "playometrics." Admittedly, game playing has an obvious role in science since it provides a stimulus to work, and economics is not the only field in which game playing occurs in its undesirable as well as its desirable aspects (see Mahoeny 1976). But I suspect that it is a particularly serious problem in economics because of the very fact that so few studies of the validity of econometric methods of testing hypotheses have been undertaken. In a field in which there is, on the one hand, so much methodological contention, and on the other hand, so much emphasis on testing, one would surely expect our standard testing techniques to be themselves subject to many more tests than have actually been undertaken. In this connection, the experience I had upon completing the previously discussed comparison of the forecasting performance of hypotheses within and beyond their sample periods was rather strange. The response of econometricians was usually: yes, you are right, but there is nothing new here, we have known this all along. But then how can one explain the continuation of these regression studies, unless it is a matter of "playing the game"?

Moreover, the rewards for publishing a paper are usually high, while both the chance — and the cost — of being caught in game playing, as long as it is not too egregious, are both low.

One aspect of game playing is that certain mundane aspects of research that are critical in obtaining correct conclusions are deemphasized relative to another aspect, the use of the latest complex techniques. These techniques are used not so much because they are likely to lead to the right answer, but probably as much, or more, because their very complexity creates a fascinating challenge, and also generates the applause of one's peers.[8] The deemphasis of that most mundane and unglamorous of tasks, getting the arithmetic right, also suggests that econometrics is in large part, "game playing." While obviously no data on this are available, it is likely that most econometricians do not bother to check their arithmetic, that is the copying of the data from the source onto the sheets from which cards are punched, and then the copying from computer print-out to the tables. The fact that few textbooks warn students about the need to check data suggests that they are infrequently checked. Yet, as anyone who does check his or her data has probably observed, data errors are quite common. This raises a rather nasty question. Suppose there are two conflicting econometric studies, one

8. The use of complex procedures is, of course, not the only possible type of game playing. For example, assume that mathematical sophistication would be held in low esteem by economists, but that wide-ranging scholarship would be highly regarded. In this case, articles would contain many totally irrelevant footnotes to esoteric sources. Whatever the standards that are used as proxies for the hard to discern true worth of research, there will be a tendency to meet these standards in ways that add little to the true worth of the project. I certainly do *not* mean to imply that econometricians are more "careerist" than are other economists; nor do I mean to imply that I am better than others in this respect; I certainly do not claim to have always met the standards recommended below.

using sophisticated state of the art techniques but unchecked data, and the other using much less sophisticated techniques but (you happen to know) checked data. Which one should you believe? The most advanced methods do little good, if in transcribing the results a decimal point is allowed to slip one digit.[9]

A related point is the apparently frequent nonreplicability of results, i.e., the inability to determine what data were used in a published paper, and to use these same data to reproduce the results. Yet it is one of the most basic rules of scientific practice that one's methods must be reproducible, so that one's results can be checked.[10] A study of the criteria used by scientists in evaluating scientific publications found that 62 percent of natural scientists considered "replicability of research techniques" to be "essential", 18 percent considered it to be "very important but not essential", 12 percent to be "somewhat important", and only 7 percent "not very, or not at all important". Neither originality, logical rigor, or any other criterion was ranked as "essential" by so many natural scientists as was replicability.[11] In this respect, economists differ very sharply from natural scientists. Again, there are no data on what proportion of the published econometric literature can be replicated, but word of mouth folklore at least is that much of it cannot be. Many economists who have tried to reproduce results of others have probably experienced an author being unable to provide either his data or sources. In general, we do not teach our students, and to not practice ourselves, one of the standard techniques of the sciences, that is, to keep adequate lab books. Yet I suspect that the maintenance of adequate research records is at least as much a basic requirement for real scientific status of a subject as is the use of advance mathematics.[12]

Another example of the game elements in applied econometrics

9. Failure to check one's data is not the only example of carelessness in economic research. My colleague, Alan Olmstead, informs me that in refereeing articles in economic history (in which quotations play an important role) he has his research assistant check them against their source. He finds that most papers have mistakes in quotations, sometimes serious ones, such as omitting the word "not". While no data are available on the extent to which this occurs in other fields too, carelessness is probably more common in economics than in the majority of other fields. One reason is that scholarship is not held in high esteem in economics. The story is told about a Princeton historian who failed a graduate student's paper because in a footnote a page number was wrong. This is hardly likely to occur in economics! And while the natural sciences too do not place much weight on scholarship, their students are trained to be careful and painstaking in laboratory work.

10. The alternative criterion, "exportability" preferred by some scientists (see Agnew and Pyke, 1978, pp. 162-63) to take account of the fact that no laboratory experiment can be *precisely* reproduced, amounts to the same thing here.

11. (Chase, 1970). For social scientists, the ratios are 41.9 percent, 24.4 percent, 29.1 percent and 4.7 percent. However, in psychology, the reproducibility criterion seems to be largely ignored. (See Wolin, 1962 and Mahoney, 1976, pp. 53 and 97.) Mahoney states (p. 97) "the average physical scientist would probably shudder at the number of social science" facts "which rest on unreplicated research."

12. Moreover, adherence to rules of replicability should help to inhibit cases of outright fraud such as the recent Cyril Burts scandal in psychology.

(though certainly not in theoretical econometrics) is the automatic pardon for the crime of using upside-down significance tests. Although undergraduates are warned about this in statistics courses, one can find many examples in professional journals where authors grossly misuse significance tests. They treat the fact that a certain coefficient has a *t* value that is not significant at the 5 percent level as *evidence* that variable A has no effect on variable B. Even though it may be significant at, say the 20 percent level, the authors then conclude that they have shown that A has *no* effect on B. In this way, by using a small enough sample, it would be easy to prove that price has no effect on the quantity bought, income has no effect on consumption, and so on. Now these authors know better. But the rules of the game permit them to make this elementary mistake, while they prohibit the sin of using a simple technique, when there exists a more complex and reliable technique.

Another peculiarity with respect to significance tests in applied econometrics is the blanket exemption from having to use them that is extended to maximum likelihood methods. An economist may draw important conclusions from the fact that the regression selected as best by the computer has, say a positive gamma coefficient, without bothering to test whether the difference in the goodness of fit between this regression and one with a negative gamma is actually significant at any meaningful level. It is not clear why the "rules of the game" permit this.

Moreover, as William White (1967) has pointed out, econometricians often seem satisfied with having demonstrated that a certain variable is significant. But the policy-maker must usually be concerned, not just with the question of whether variable A has an effect on variable B, but he needs to know also how large this effect is. A 50 percent error in the point estimate can make a big difference to the success of a policy! But given a normal distribution, a variable that is just significant at the 5 percent level on a two-tailed test will, 16 percent of the time, have a point estimate that is at least 50 percent too low. And if one makes allowance for the usual data mining, this 16 percent figure should be raised substantially (White, 1967, pp. 28-34).

A further example of our doing something which all of us know to be wrong is the common tendency to interpret an insignificant regression coefficient or low correlation as evidence that a certain independent variable has no effect on the dependent variable. But surely all it means is that there is no *stable* relationship between the two variables. The independent variable could still have a high, though variable, effect on the dependent variable.

Another game aspect of economics is the way in which we ignore the quality of the data. As Wassily Leontief (1971, p. 3) has complained: "in all too many instances sophisticated statistical analysis is performed on a set of data whose exact meaning and validity are unknown to the author or rather so well known to him that at the very end he warns the reader not to take the material conclusions of the entire 'exercise' seriously." Moreover, it is probably quite common for economists to go to the library (or send their research assistants) to copy some data with which they are unfamiliar without reading the description of the data given in the source. This could lead to some peculiar results.[13]

To cite a specific instance of how questions about the quality of the data are ignored, Paul Taubman (1968) pointed out that the data used in time series studies of postwar saving are just one of three different sets of saving estimates that can be derived from government data, all of which purport to measure the same thing. He then used all three sets of data to see if they yield similar results when plugged into regressions. It turned out that, "judging both by the significance and size of the coefficients, the character of the saving function depends crucially on the choice of the savings series" (Taubman 1968, p. 128). And what is just as bad, there is no way of determining which one of the series is the best. This is a result that should surely have shaken up everyone working on savings functions and consumption functions. Yet this paper has had only little impact.

Moreover, the computer revolution, by allowing us to work with masses of data has an unfortunate by-product. We tend to dump the data into a computer without ever looking at them. Yet if the data are subject to large errors, (e.g., household budget data) a sample of, say, 300 observations for which the researcher has actually read each questionnaire with some care may be more reliable than a sample of 3000 observations that were dumped into the computer with very little editing for inappropriate responses.

It would be interesting to discover what would happen if econometricians were offered the following gamble: "You will get $1000 if you are right, but have to pay me $3000 if you are wrong." How many (even among those with little risk aversion) would take this gamble, and what does this tell us about our claim to have established something at the 5 percent significance level?[14]

IV.

Given all of these problems, what can be done? One possible answer is that we should abandon applied econometrics altogether. But this is an unappetizing alternative because it would usually leave us with no way

13. Thus, many years ago, a then well-known economist used some Census Bureau data and complained that he could not carry his analysis up to date because, for some reason, the Census Bureau no longer published the data. But the description of the data said that they are untrustworthy and were published only because Congress insisted on it. And eventually Congress no longer insisted. Another potentially serious problem, that research assistants may have low morale and do slipshod work, is discussed by Roth (1966) who also suggests a way of ameliorating it.

14. Admittedly, this test may not be quite fair because it is said that "no one believes an hypothesis except its originator, but everyone believes an experiment except the experimenter." (Beveridge, 1957, p. 47).

of selecting, from among a plethora of possible explanations, the one that actually *does* explain events. For example, there used to be a large number of business cycle theories. Each author pointed to some factor that, in principle, could generate cycles, without offering any evidence that it was *this* factor, and not some other, that actually did cause the observed cycle (see Mitchell, 1927, Chapter 4). Perhaps the particular factor singled out by his theory, while capable of generating a cycle, is much too small to account for the observed fluctuations. But without econometrics we will never know. Hence, my own attitude towards econometrics is like that of the person who upon being told that the craps game he was about to participate in is crooked; replied, "Sure, I know that, but it is the only game in town." Admittedly, this is a bit of an overstatement since at least some hypotheses could be tested by economic history, or by a simple cross-country comparison without the use of econometrics. But not many could be tested in these ways.

Hence, instead of abandoning applied econometrics altogether, we should do two things. One is to be more skeptical of our results. Instead of treating an econometric result as evidence from a "crucial experiment," we should think of it more as circumstantial evidence.[15] Only if a whole number of separate studies using different data, e.g., cross section data as well as time series data, and perhaps data for various countries, tell us the same thing, should we take it seriously. Such evidence from repetitive "experiments" is much more convincing than is the fact that the critical coefficient in a single study is highly significant (see Tukey, 1969, pp. 84-85). Until we have such massive evidence, we should not pretend, particularly to policy-makers, that we have the answer merely because we have some significant regression coefficients. If we are thus to rely on the weight of many pieces of evidence rather than on a single crucial experiment, then it would be most useful to have, from time to time, surveys that pull the evidence together, in the sense, not just of describing what various studies show, but by resolving any contradictions between them.[16] At present on too many questions we are buried in an inchoate mass of seemingly contradictory evidence.

Second, it would surely help if we can raise the standards of what is acceptable work by evaluating research more on the basis of the likely validity of its results, and less on the technical sophistication of its techniques. For example, it is likely that despite their well-known problems survey methods would be used more in economics if it were not for the fact that since they seem so simple, there is little kudos to be gained by using them.[17] Third, much more emphasis should be placed on collecting

15. For a general argument that economics should model itself on law rather than on the exact sciences, see Benjamin Ward (1972).

16. For an attempt to do this with respect to some of the consumption function literature, see Mayer (1972, Part 2).

17. Actually, survey techniques are not at all so simple to use. There exists an extensive literature in sociology on questionnaire construction. For a plea to use survey techniques more see Friend (1973).

relevant data (Leontief, 1971).

Fourth, we should guard against data mining. Admittedly, not all data mining is necessarily bad, and it is often the fault of the theory for not specifying such critical "details" as the appropriate lags. As Griliches (1967) has pointed out, one should not expect the data both to specify the lag and to test the hypothesis embodying it. Moreover, Leamer (1978) has recently set out rules and safeguards for valid data mining. But he agrees that "without judgment and purpose [i.e., the way data mining is usually done] a specification search is merely a fishing expedition and the product of the search will have a value that is difficult or impossible to assess" (p. 2). It is particularly important to guard against mere fishing expeditions at present when there is such a strong tendency in economics to carry the rationality assumption to extremes. It is usual to justify reliance on an implausible rationality assumption by saying that this assumption is validated by the fact that the hypothesis based on it is consistent with the data. But there is then strong temptation to run regressions until one of them finally "confirms" the hypothesis. Hence, there is a case for requiring researchers to list all the regressions they ran, and not just to present the particular regression, perhaps the only one of a large number, that happens to support their hypothesis.[18]

But this may not be sufficiently stringent. Suppose a researcher who planned to run, if necessary, say fifty regressions is lucky, and the very first regression happens to support the hypothesis. This regression may have been chosen quite arbitrarily, and data mining that just happens to hit a vein of gold on the first strike of the pick-axe is still, in a way, data mining (Bronfenbrenner, 1972, p. 57). Hence, before taking the results seriously, it would be reasonable to require that authors run their regressions in all or many of the numerous and varied forms that are consistent with their hypotheses, and are both plausible and econometrically valid. One should then accept only those results that are robust with respect to a wide variety of reasonable techniques. They could be robust in the sense that all the variants that are run generate similar results, or in the sense that the one variant that does support the maintained hypothesis gives a substantially better prediction. If they are not robust in either of these ways, then any support they provide to the maintained hypothesis should be considered as highly tentative.[19]

18. There is, of course, no way this requirement could be enforced. But if authors would know that not reporting regression results that contradict their hypothesis is considered flagrant intellectual dishonesty, they would be reluctant to do this.

19. Unfortunately, there are many possible tests of robustness aside from the standard ones, such as using in alternative regressions logs instead of natural numbers, yearly instead of quarterly data, and first differences instead of levels. For example, Evans (1967) showed that the Ball-Drake model, while confirmed by a time series regression in which the data are deflated by the GNP deflator, is rejected by the same data if they are deflated by the CPI instead. Obviously, it is not always feasible to run all reasonable variants of a regression, but enough should be run to give at least a presumption of robustness.

Requiring such tests of robustness would substantially toughen the standards of acceptability, and should also work towards reducing the frequency of contradictory results in the literature. Authors would, so to speak, be their own critics by mining the data *against* themselves, and bring out objections that are now frequently raised by others in comments. This raising of standards is justified because those who present new results should have the burden of the proof thrust upon them, and the burden of the proof amounts to more than just showing that there happens to be one particular form of the basic regression equation that yields the desired result.[20] As both Keynes (1973, pp. 287 and 294) and Friedman (1951, p. 108) for once in agreement, have pointed out, one can get a good statistical fit merely by repeated testing, but such a fit proves very little.[21]

Fifth, it is important that authors not use up all their data in *fitting* their regressions, but leave some as a hold-out sample against which to *test* the regressions. This would not really waste data points because, once the hypothesis has been successfully tested, the regression can be rerun using all the data points to refine the coefficient estimates.

Sixth, the journals should publish papers that find statistically insignificant results. This would not only remove the great pressure to stomp on the data until they give in and yield a t value of 2 or more (as a saying has it: "if you just torture the data long enough, they will confess"), but it would also prevent others from wasting time replicating an unsuccessful project (see Feige, 1975). Journals should also encourage economists to replicate previously published results.[22] In other sciences replication is done continually. Presumably this is because replication is a good way of training students in laboratory techniques. But in economics we do this much less frequently probably, in part, because, in the absence of lab experiments that produce new data, all the "replication" we can do is rerunning the old regressions. Hence, erroneous results are allowed to remain in the literature. This lack of replications is particularly serious because, as just discussed, economists pay insufficient attention to avoiding calculating errors. Some foundation should finance a program wherein graduate students would rerun each year, say 10 percent of the empirical work published the previous year in the journals. This would not only catch many mistakes, but the potential embarrassment of being caught in a calculating error should make researchers more careful.

Seventh, authors using unpublished data should be required to make

20. Would this requirement result in journals having to close down because they have an insufficient number of acceptable manuscripts? No it would not, because a paper that shows that a particular hypothesis cannot at present be tested because some forms of the underlying regression equation support it, while others reject it, is a contribution that warrants publication.

21. For a rigorous demonstration, see Bacon (1977).

22. Currently at least two journals, the *Journal of Political Economy* and *Journal of Consumer Research* have offered to publish replications.

them available so that their work can be verified by others.[23] Some may object to this since they want to enjoy a monopoly while they exploit their data source for additional papers. This is something that occurs in the hard sciences too (Barnes and Dolby, 1970, p. 15). And indeed, like patent protection, such a temporary monopoly may provide a desirable incentive to look for new data. Hence, there is a case for allowing authors perhaps a one or two year leeway on such a rule. One or two years, combined with the usual publication lag for the initial paper should give them sufficient time to exploit their data mine ahead of others. Another requirement that journals should impose is that the author state whether, and to what extent, the reported computations were checked for transcription errors, etc.[24]

Finally, given all the weaknesses of econometric techniques, we should be open-minded enough to accept that truth does not always wear the garb of equations, and is not always born inside a computer. Other ways of testing, such as appeals to qualitative economic history, should not be treated as archaic.

V.

To return to the earlier theme, I have so far given reasons why economics appears to be still far from a hard science. But is anything lost by taking this as a goal? I believe the answer is yes, because we then tend to act as though we are already close to that goal, which causes us to oversell the validity of our results both to policy-makers and the general public. Moreover, it induces us to reject the plausible in favor of the seemingly "proven." This can easily result in the choice of the wrong hypothesis, since we may then reject an hypothesis that accords with common sense in favor of one that appears to fit the data much better only because of one of the weaknesses of our methods just discussed. It also distracts us from looking for merely plausible evidence because we believe a paper based on plausible evidence rather than on seeming "scientific" evidence will not be published, or if published, will receive little recognition. As a result, we lose not only some interesting evidence, but in addition, some important problems get much too little discussion. Furthermore, the gathering of much needed data is underemphasized. In addition, the stress on using advanced mathematical tools allows us to have a good conscience while ignoring some very elementary rules of

23. The way to enforce this would be for journals to require that authors send, perhaps with the final revised draft of the paper, a copy of their data. The journal, and not the author, would then be responsible for making the data available.

24. There is, of course, no way this could be verified, but the mere fact of having to make a statement about it should induce economists to be more careful in their treatment of data.

good research procedure. Much of the published research consists of taking a new technique out for a walk rather than of really trying to solve a problem. And also, economics has become much too isolated from other social sciences, since being hard scientists we do not want to use either the results or tools of those who cannot claim our exalted status (Cf. Leontief, 1971).

REFERENCES

Agnew, Neil and Sandra Pyke, *The Science Game* (1978) Englewood Cliffs, NJ: Prentice-Hall.

Armstrong, J. Scott, (1978a) *Long Range Economic Forecasting* New York: John Wiley & Sons.

____________, (1978b) "Forecasting with Econometric Methods: Folklore versus Fact," *Journal of Business*, 51 (October), pp. 549-64.

Bacon, Robert (1977), "Some Evidence on the Largest Squared Correlation Coefficient from Several Samples," *Econometrica*, 45 (November), pp. 1997-2001.

Barnes, S.B., and Dolby, R.G.A. (1970), "Scientific Ethos: A Deviant Viewpoint," *European Journal of Sociology*, No. 1, pp. 3-25.

Belsley, David, Kuh, Edwin, and Welsch, Roy (forthcoming), *Regression Diagnostics: Identifying Influential Data and Sources of Collinearity*, New York: John Wiley & Sons.

Beveridge, W.I.B. (1957), *The Art of Scientific Investigation*, New York: W.W. Norton.

Bronfenbrenner, Martin (1972), "Sensitivity Analysis for Econometricians," *Nebraska Journal of Economics*, II (Winter), pp. 57-66.

Brunner, Karl (ed.) (1972), *Problems and Issues in Current Econometric Practice*, Columbus: Ohio State University Press.

Chase, Janet (1970), "Normative Criteria for Scientific Publication," *American Sociologist*, 5 (August), pp. 262-64.

Evans, Michael (1967), "The Importance of Wealth in the Consumption Function," *Journal of Political Economy*, 75-Part I (August), pp. 335-351.

Feige, Edgar (1975), "The Consequences of Journal Editorial Policies and a Suggestion for Revision," *Journal of Political Economy*, 83 (December), pp. 1291-96.

Ferber, Robert (1953), *A Study of Aggregate Savings Functions*, NBER Technical Paper No. 8, New York: National Bureau of Economic Research.

____________, (1956), "Are Correlations any Guide to Predictive Value," *Applied Statistics*, 5 (June), pp. 113-31.

Friedman, Milton (1951), "Comment," in Universities-National Bureau Committee for Economic Research, *Conference on Business Cycles*, New York, NBER.

Friend, Irwin (1973), "Mythodology in Finance," *Journal of Finance*, 28 (May), pp. 257-72.

Frisch, Ragnar (1970), "Econometrics in the World Today," in W.E. Eltis, M. Scott and J. Wolfe, *Induction, Trade and Growth*, Oxford: Oxford University Press.

Griliches, Zvi (1967), "Distributed Lags: A Survey," *Econometrica*, 35 (January), pp. 16-49.

Howrey, E. Phillip and Saul Hymans (1978), "The Measurement and Determination of Loanable Funds Savings," *Brookings Papers on Economic Activity*, 1978: 3, pp. 655-685.

Keynes, J.M. (1973), *Collected Writings*, 14, London: Royal Economic Society, Vol. 14.

Kuhn, Thomas (1962), *The Structure of Scientific Revolutions*, Chicago: University of Chicago Press.

Leamer, Edward (1978), *Specification Searches*, New York: John Wiley & Sons.

What We Know and Don't Know About Inflation

Robert M. Solow

Why is our money ever less valuable? Perhaps it is simply that we have inflation because we expect inflation, and we expect inflation because we've had it.

Inflation is going on when you need more and more money to buy some representative bundle of goods and services. It is a sustained fall in the purchasing power of money, or — what comes to exactly the same thing — a sustained rise in the general level of prices.

Those definitions contain a useful lesson. You often read in the newspapers or hear in everyday conversation that some particular price has gone up, and the event is described as inflationary. That description might be correct, if the price is really a typical one and if most prices have been rising in terms of money. But the description needn't be true: even if there were no inflation at all — if the price of a representative bundle of goods and services were absolutely flat — it could still be that the price of coal, meat, or houses is rising while some equally significant price — of, say, medical care, wheat, or motor vehicles — is falling. In any live economy, you have to expect the prices of goods to change in terms of each other, because relative costs of production change, or fashions change, or a cartel forms. Though those changes in relative prices may have something to do with the generation of inflation, they are not the same thing as inflation. We're entitled to talk about inflation only when some representative average of prices is rising in terms of money.

This distinction between relative prices and the general price level is important. We define a *pure inflation* as an

---

The general price level, 1865-1975, measured in terms of what is called the gross national product (G.N.P.) deflator which is, for the author's argument, sufficiently correlated with the consumer price index — the cost of a representative bundle of goods and services. Depressions followed the Civil War (decreasing prices for 30 years from 1865 to 1895) and World War I (the downturn from 1920 to 1934); but since 1940, despite World War II and military involvements in Korea and Vietnam, the price level has gone only one way: up. "Something has happened since 1940," writes Professor Solow, "that's worth talking about."

1940 1950 1960 1970

inflation in which all prices are rising at exactly the same rate, the ratio of any one price to any other particular price remaining the same. If actual inflations were pure inflations like that, then no confusion would ever arise. But in fact both things always happen at once: prices rise on the average, and some prices rise faster than others. Between December, 1976, and December, 1977, the consumer price index (which is the cost of one of those representative bundles of goods and services) rose by 6.8 per cent. But in the same 12 months the price of clothing went up by only 4.2 per cent while the price of medical care went up by 8.8 per cent. How fortunate for the clothes horse, and how unfortunate for the invalid. Or, to put it the other way, how fortunate for the doctor, and how unfortunate for the haberdasher. What is so often forgotten in everyday discussion is that the doctor would have been just as lucky, and the haberdasher just as worried, if there had been no inflation at all (the consumer price index remaining unchanged) while the price of medical care went up by 2 per cent and the price of clothing down by 2.6 per cent. The distinction between relative prices and the general price level is important; the two things may be connected, as I happen to think they are, but they're not the same.

Forty Years in One Direction: Up

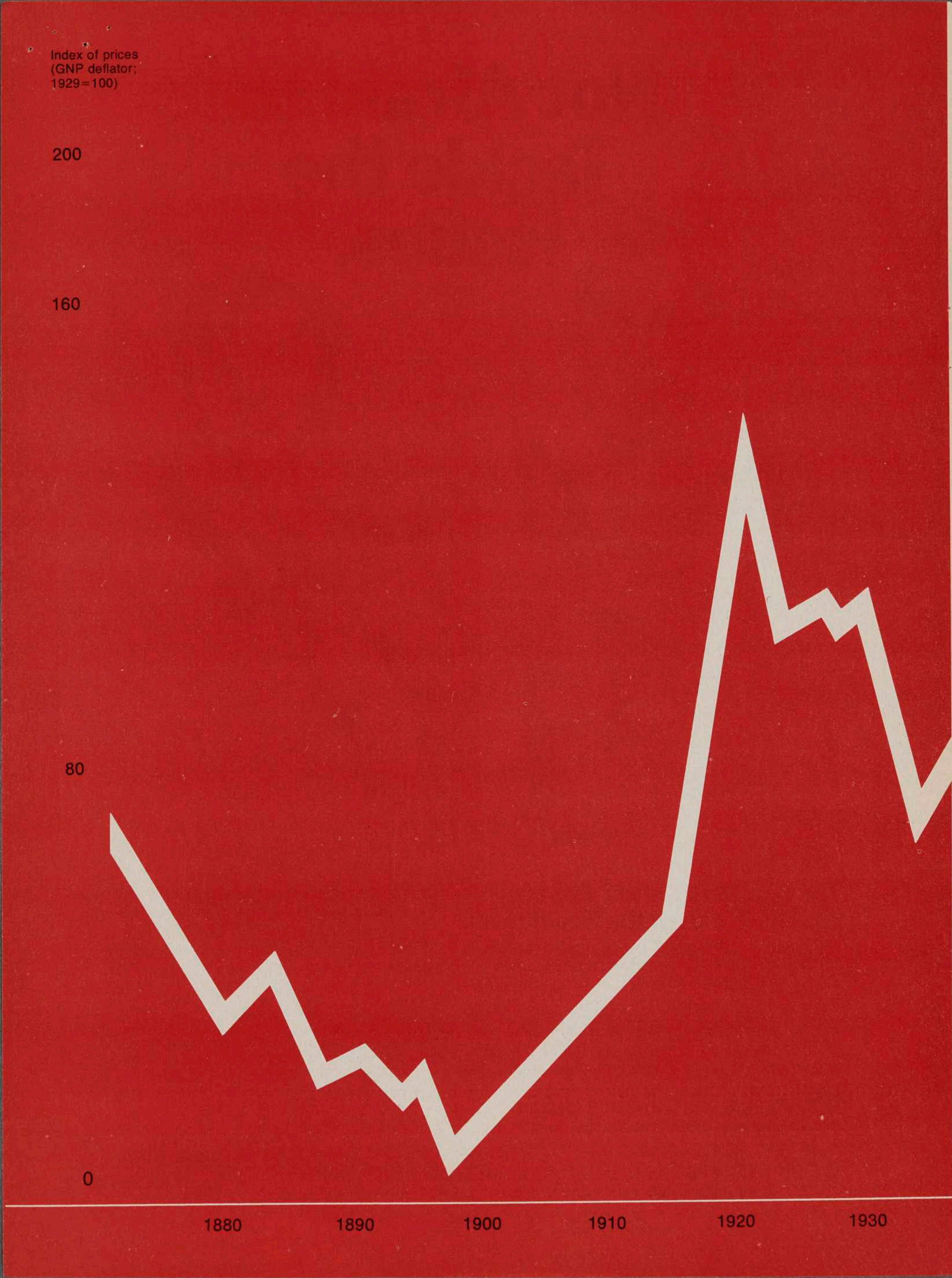
History has something to tell us about this matter. From about 1867 to 1897, for the first thirty years after the Civil War, the trend of the general price level was down; in fact, the price level deflated by over 40 per cent during that time *(chart, previous page)*. We're all so accustomed to deploring our current inflation that three decades of deflation may seem to have been a golden age. Not at all. There were two deep depressions during that period, one in the 1870s and one in the 1890s. Farmers found themselves paying off their mortgages in dollars of rising purchasing power — a situation that only bankers appreciated. In those days, unlike today, farmers outnumbered bankers, and the cruel eviction of Little Nell by Mr. Moneybags became the standard playhouse theme, the soap opera of the time. The price level finally turned upward about 1897, but by the time the First World War broke out in 1914 it had recovered only to the level of 1873. Then came the big war-time inflations: between 1914 and 1920 the price level almost doubled, and it happened again between 1940 and 1948. The wars in Korea and Vietnam also added sharply to the price level, but of course much less spectacularly than the two World Wars.

But there's a subtler point to notice in the graph: there was a sharp deflation after the First World War, when the depression of the 1930s, like that of the 1870s to 1890s, pushed the price level down. The low point was reached in 1933, and in 1940 the price index was still 20 per cent below its 1929 level. But that was the last time. For the past forty years the price level has gone only one way, and that's up. Some years it has risen very slowly, some years rapidly, but at no time has it gone down.

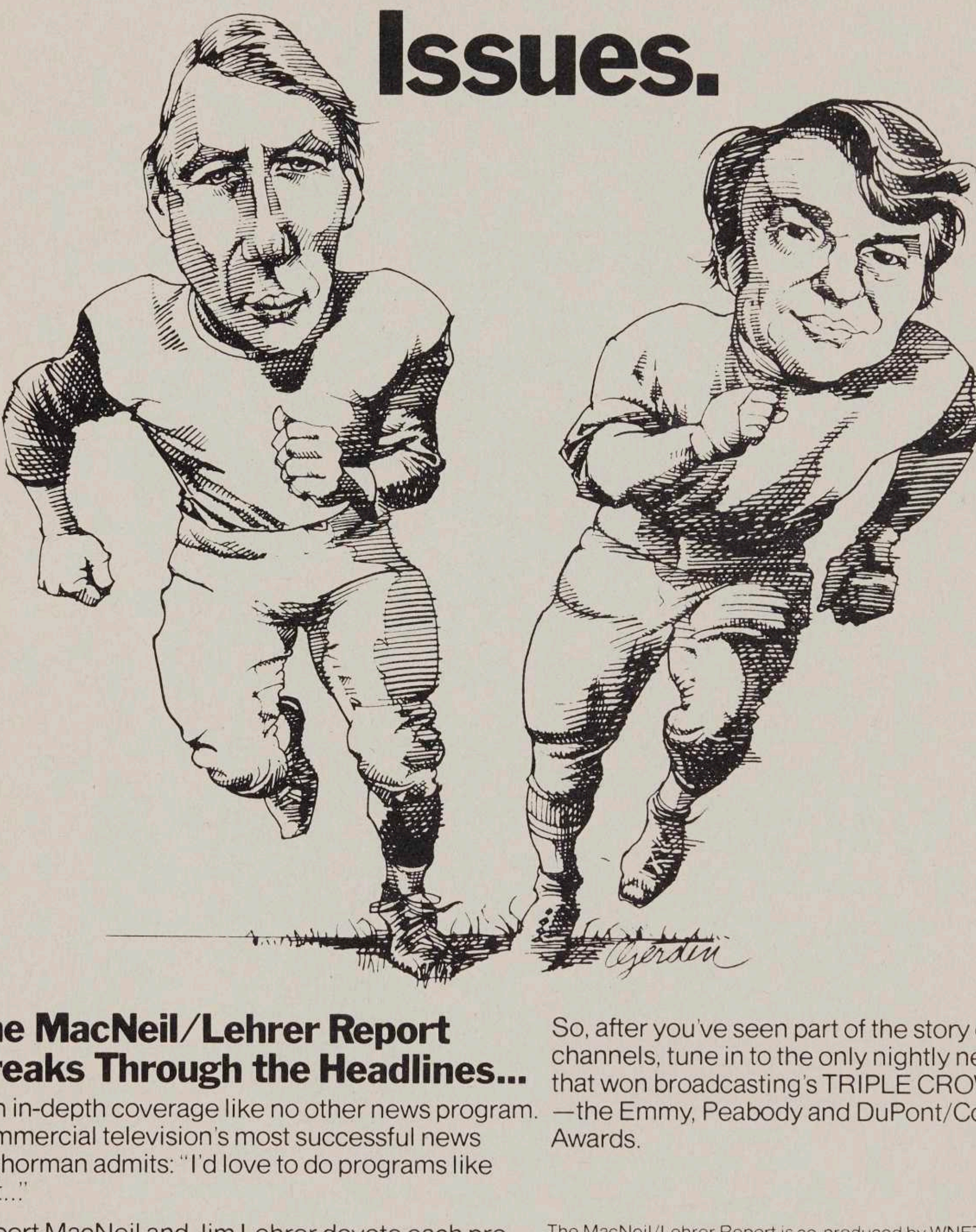
I'm pretty sure I wouldn't want the 1870s to come around again, and I know I wouldn't want the 1930s to come around again. Indeed, I know of no one who expects to live to see deflation again. So something has happened that's worth talking about.

Now consider these same facts from another angle. This time, in the interest of drama, I switch to the *rate of change* of the wholesale price index *(chart, pages 34-35)*. This measure of the price level gives too much weight to raw materials, which are notoriously volatile in price, but that's exactly why I want to use it now: it tells the story very dramatically. You can see the two big periods of sharp deflation in the 1890s. You can see the inflation of the First World War, that period of many months when the rate of change of the wholesale price index runs at an annual rate of between 20 and 40 per cent. You can see that prices collapsed sharply in 1920 and 1921, when for a couple of months the wholesale price index was falling at a rate of 90 per cent a year (which is a lot!) and you can see the period of falling prices from 1929 to 1933. You can see the initial inflation at the time of the outbreak of the Second World War in 1941 and 1942, the period of price controls during which the wholesale price index was essentially stable, and the loud "boing!" when the price controls were lifted in 1946. You can see the run-up in prices, especially in those of raw materials, at the beginning of the Korean war. And then you can see the rather curious episode when prices actually fell during the Korean war in 1951 because everybody knew price controls were coming and had jacked prices up higher than the traffic would bear; so when the price controls went on some prices had to be brought down.

The astounding thing that I want you to notice about this picture is the difference between the first sixty years and the last twenty, ending in 1971 — before the oil embargo. In the early part of the period, even in peacetime, wholesale prices went up and down continuously; in some months the index rose or fell at an annual rate of 10 or 20 per cent or even more. After 1951 there is a remarkable change. The trend of the index is clearly up — there are more plusses than minuses — but the main thing is the stability: the volatility in the rate of change of the price level has suddenly disappeared. It looks uncannily like one of those television ads for a headache remedy: from 1890 to 1950 it's pain, pain, pain. But then the se-



Index of prices
(GNP deflator;
1929=100)
200
160
80
0
1880
1890
1900
1910
1920
1930



I know of no one who expects to live to see deflation again. So something has happened that's worth talking about.

cret ingredient reaches the bloodstream and suddenly you can sleep again.

Inflation: The Abstract Has Reality

Now a very deep question begins to arise.

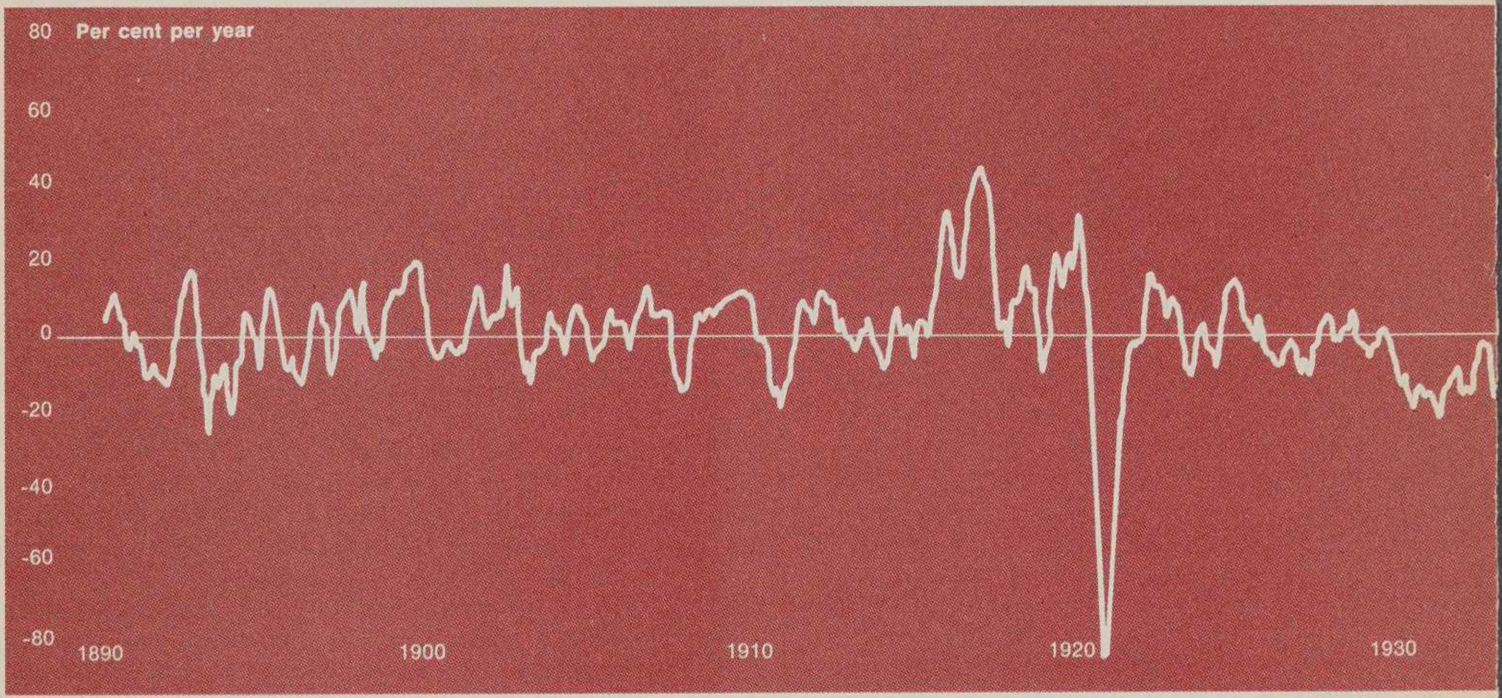
In explaining what we are supposed to mean by inflation, I distinguished carefully between changes in relative prices and changes in the value of money, the absolute price level. Inflation has to do with absolute prices, in money terms. That distinction between absolute prices and relative prices is an example of a broader distinction — that between nominal magnitudes and real magnitudes. A real magnitude is a physical quantity entirely independent of the monetary unit. By contrast, a nominal magnitude is a quantity that's scaled in the monetary unit. Halve the size of the monetary unit and you double all nominal magnitudes.

The price level, which is the price of a representative bundle of goods in money, is clearly a nominal magnitude. Things like the annual output of coal, the number of hours worked, or the number of people unemployed are clearly real magnitudes. And relative prices are real magnitudes, too. The price of *a* in terms of *b* is the number of units of *b* you have to give up in order to obtain one unit of *a*. And that's clearly independent of nickels, dimes, or dollars.

The deep question to which I referred is: How are the price level and inflation connected to what happens in the real economy of production and employment? How are nominal and real magnitudes related — if they are related at all?

This is not a question of semantics. We want to know how our economy actually works — whether events in the real economy have effects on the price level, whether (and how) changes in the price level have effects on the real economy. In a way, I should point out, we've already prejudged part of the answer to that question. If we seriously thought that the price level and the real economy were unconnected, we would hardly be interested in inflation at all. You eat, you wear, you drive, you combust, and you enjoy real quantities, not nominal ones. If your twin sister's life consisted of exactly the same real events as yours, but just in a place with a different price level or a different rate of inflation, your curiosity might be aroused, but you would hardly think that one of you was better off than the other. So we act as if we believe that there were a connection, though of course we could be wrong about it. But the nature of the connection is not obvious, as you will see.

Now look at one more time series — the year-to-year percentage rate of change of real gross national product



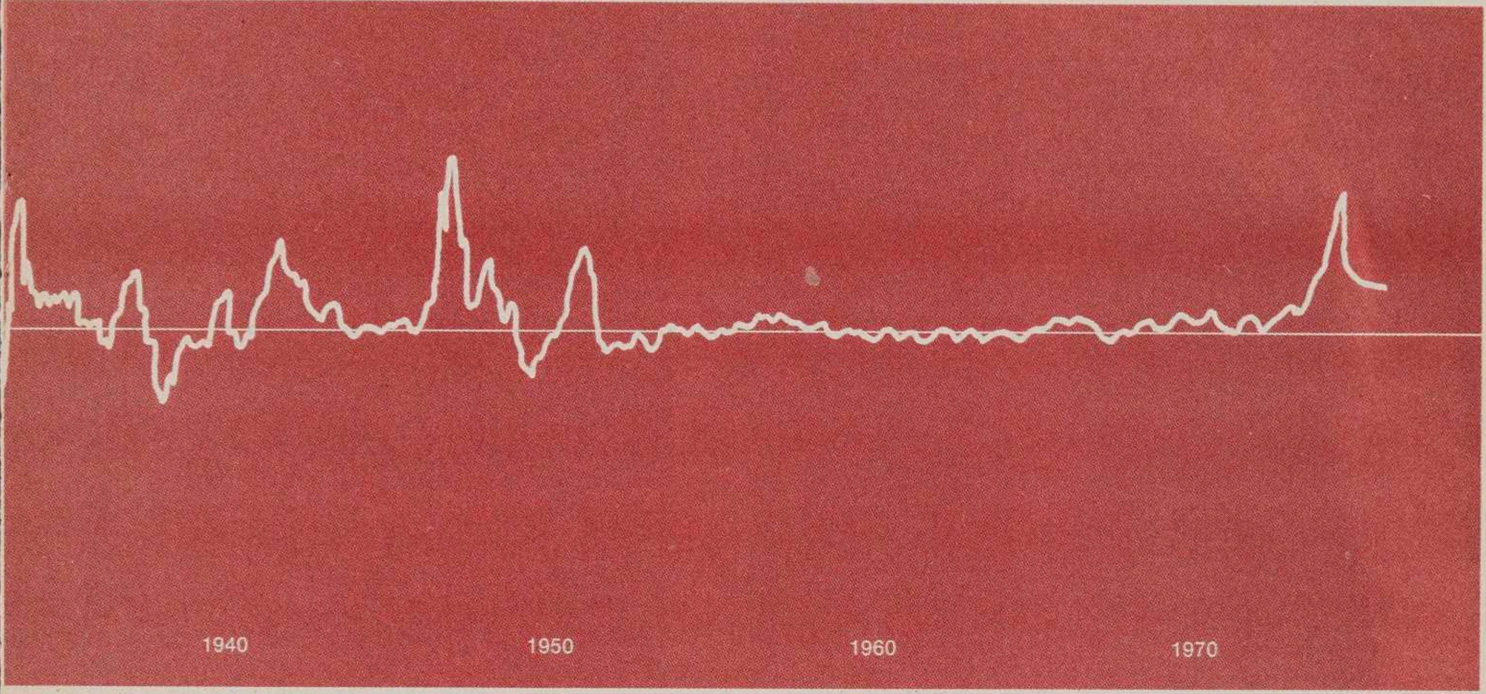
(G.N.P.) from 1901 to the present, at the top of this page. Real G.N.P. is an admittedly imperfect measure of the flow of new production available to our economy to consume, or to add to the stock of plant and equipment, or to shoot off into space, or whatever. It is used here as a summary measure of our production of useful goods or services; it is therefore a real magnitude that does not change if you change the monetary units. You can pick out the boom of the First World War, the very sharp depression that occurred immediately after it, the prosperous 1920s interrupted by two minor recessions during which the rate of growth of real G.N.P. barely reached zero, the traumatic depression of the 1930s, and the very large burst of output during the Second World War — really quite remarkable, although not so strange if you realize that it came after all those years of depression so that there was a lot of idle capacity and unemployment to use up in increasing output. But what I want you to notice is the rather striking improvement in the stability of the real economy after about 1950, after the Korean war. Qualitatively, it's very much like the picture of the wholesale price index. Since about 1950, the peaks are very much less pronounced than the peaks of the previous fifty years, and the valleys don't come very often, last only a very short time, and never go very deep. Whatever else you may say about it, we run a much smoother economy than we used to. Sometime around 1950, after at least half a century of sharply fluctuating price and production levels, we seem to have entered a period in which the trend of these is almost always up, and in which the rate of inflation is unusually steady by any historical comparison.

That description requires some qualification in view of the events of 1973 to 1975. But I want to preface that with a deeper analysis of the connection between the real world of production and employment and the nominal world of price and inflation.

Wages, Prices, and the Phillips Curve

It's tempting, and likely correct, to conclude that there must be some systematic relationship between the greater stability of real output in the post-war years and the greater persistence and smoothness of inflation during the same period. The kind of casual observation and interpretation that I've given you thus far can hardly establish that connection, and my main business now is to do so.

To set the stage, I'm going to begin with a sort of quick, crude caricature of the ideas with which economics, or macroeconomics, entered the 1950s. The mainstream economists of the 19th and early 20th centuries liked to believe that there was no long-run connection between the real world and the nominal world. The real economy, the economy of goods and services and employment and output and relative prices, was thought to be a rather stable mechanism which operated according to the general rules of supply and demand. If undisturbed by bad harvests, wars, or other such catastrophes, that mechanism would fairly quickly gravitate toward a natural equilibrium in which the supply and demand sides



The rate of change of wholesale prices, 1890 to 1974.
Note that because this chart shows a *rate of change* the index is rising when the line on the chart is above zero and falling when it is below zero. The wholesale price index is more volatile than the consumer price index because it gives extra weight to raw materials prices, but this chart, because it shows the change in a six-month moving average, has the effect of smoothing the actual fluctuations. Despite this effect, volatility is the predominant characteristic before 1950 and stability that of the years since then.

of every market were near balance. The main determinants of the absolute price level were the size of the existing stock of money in the economy, the technology of making payments, and the amount of purchasing power that people liked to keep in liquid form. In the simplest version of that theory, a change in the existing stock of money would be reflected (as soon as the economy returned to its natural equilibrium) in a proportional change in the price level and in nothing else. Add 20 per cent to the supply of money — according to this theory — and, after some jitters up and down, the general price level would settle at 20 per cent higher, leaving the real economy essentially where it had been before. Inflation — a rise in the price level — was in that view the natural concomitant of a rising money supply.

The economists of those decades were neither blind nor fools. They knew that there were sometimes sharp fluctuations in the real economy, and they developed a number of more-or-less *ad hoc* theories about what was called the business cycle to account for those; mainstream economists were definitely predisposed to regard such fluctuations as transient departures from the equilibrium of a generally stable mechanism. Even beginning with David Hume in the 16th century — a very long time ago — economists realized that purely monetary or nominal events such as a big gold strike could have real effects in the short run; but in the long run, such monetary disturbances were also expected to damp away, leaving the real position unchanged.

Then came the Keynesian revolution, the other strain in post-war macroeconomic theory originating in the work of J. M. Keynes. In the 1930s, and even earlier in Great Britain, it became painfully obvious to everybody that those transient deviations from equilibrium could last a very long time and cause a heap of misery. Keynes succeeded in focusing macroeconomic theory on such situations in which markets — especially the labor market — are not in balance. He showed that under those circumstances some of the propositions of mainstream economics were not merely irrelevant, they were sometimes downright wrong. In particular, during those long periods of slow adaptation, nominal events could have important connections with real economic quantities.

Keynes and his immediate successors were not primarily concerned with the theory of inflation. They held that whenever there was unemployment and excess capacity in the economy, as in a depression, the price level was more or less rigid and nominal events had their full effect on real quantities. But as soon as the economy reached the state of full employment and full utilization of its productive capacity, as it did during the war, the real production and employment situation was necessarily frozen — you couldn't produce any more. And therefore the old divorce between nominal events and real quantities came back in force again. In such a situation of full employment and full utilization of capacity, added nominal expenditure from any source would merely lift the price level. As long as monetary accommodation was forthcoming, as long as the central bank would go along, the inflationary gap would recreate itself and the inflation could continue. If the central bank was not forthcoming, then the expenditure impulse and the accompanying inflation would burn itself out.

I suppose that if you had put the question explicitly to any serious economist, he or she would have granted easily that any expanding economy would reach balance in some markets sooner than others, so that the sharp dichotomy — no inflation short of full employment and nothing but inflation at full employment — could not be taken literally. But that dichotomy was an oversimplification that permitted you to focus on the right problem in the right circumstances, and it was definitely the general habit of thought.

And that brings me to 1958 and William Phillips, an engineer turned economist and econometrician, a New Zealander resident at the London School of Economics. In that year, he published an article that must rank as one of the great public works enterprises of all time. In the past twenty years it has provided more employment than any project since the construction of the Erie Canal. Phillips went through an exercise in history, matching annual unemployment rates in Britain between 1862 and 1957

. . . the sharp dichotomy — no inflation short of full employment and nothing but inflation at full employment — was an oversimplification that permitted you to focus on the right problem in the right circumstances . . .

Economics as a Non-Science

The study of economics is an attempt to understand a very complicated mechanism without any possibility of controlled experimentation. You don't have the option of applying a step voltage and seeing what happens; you don't have the option of taking genetically identical mice and depriving half of them of potassium in their diet while treating them otherwise alike and seeing what happens. All you get to do is to watch the single experimental run that history performs.

When any kind of complicated system is exposed to an erratic environment, there is certain to be more than one plausible explanation of the observed outcome. So two observers, studying the same stretch of history, may each be able to claim that the outcome is consistent with his or her favorite theoretical model, even though the two theoretical models are by no means equivalent. You can imagine an experiment that would consist in holding some of the things in such a stretch of history constant artificially, so that changes in the behavior of the system could be attributed with some confidence to the factors that have in fact changed; then if two competing theories made different predictions about the outcome, at least one of them would turn out to be wrong. The only trouble is that you can't perform that experiment. An economist is entitled to hope that occasionally the next stretch of history's single experiment would actually disqualify one of the alternative theoretical models by producing some events that are simply incompatible with it; and that does sometimes happen, and progress occurs.

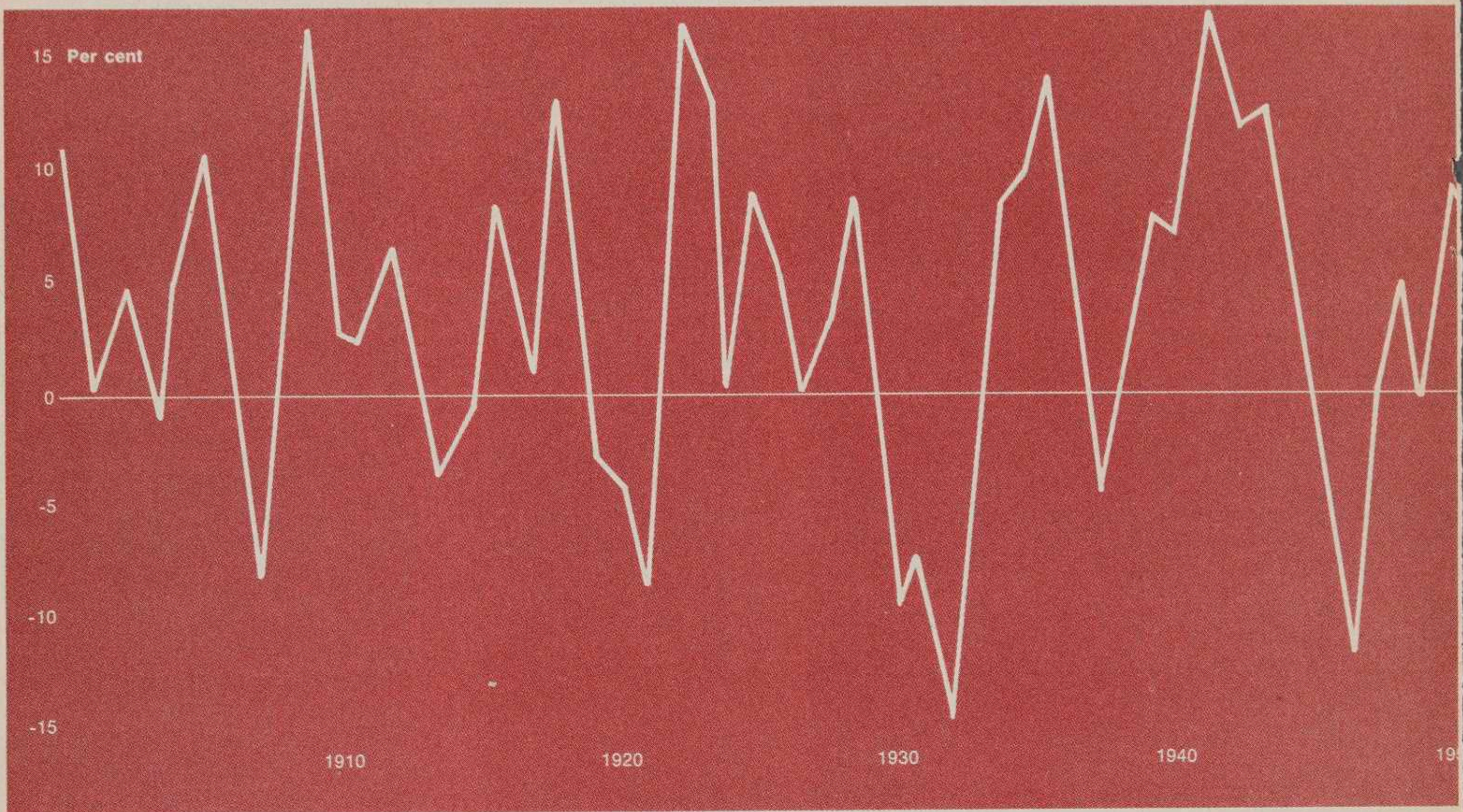
The Inconstancy of Constants

But even that process is a little slower than you might think, for a reason that I want to explain. Every theory about a complicated economic system contains a lot of coefficients, constants whose values are not specified by the theory but have to be chosen to fit the observations and be consistent with the theory. So when a surprise occurs to someone with such a theoretical model or picture of the world, the observer can often claim that the observation disqualifies not the theoretical model but the values which happen to have been assigned to the constants. "If that constant which I used to suppose was nearly zero is actually positive," he or she says, "then I can absorb the new observations and account reasonably well for the old observations as well." It is a very complicated economy, and it's very hard to begrudge a few parameters to someone who is trying to understand it.

There is another closely-related difficulty of which readers should be aware. Economic phenomena such as inflation are acted out by real people living in real societies according to some more-or-less well defined institutional rules of the game. What actually happens depends on the particular social institutions that have evolved and on the attitudes, expectations, and motivations of the various groups and individuals who play roles in the story. Now all these things — institutions, expectations, motivations — can change with time. The Federal Republic of Germany in the 1960s is not the same thing as the Weimar Republic of the 1920s. And one of the differences is that some people in the 1960s remember or have been told about the 1920s. So a theoretical model that's right for the 1920s may be wrong for the 1960s, and you have to expect that to happen from time to time. When a theoretical model that seems to have been working adequately for a decade or so starts to go wrong, there are two possibilites: either it was wrong all along and history just waited until now to perform the crucial experiment, or it was actually giving the right answer while it lasted, but the rules of the game — institutions, expectations, and motivations — are no longer quite what they used to be. The correct theory of inflation may well be just such a moving target.

New Answers to Old Questions

There's an old joke about the young professor of economics who asks an old professor of economics, "How can you possibly make up new examination questions year after year?" And the answer is, "Oh, we don't change the questions, we just change the answers." As it turns out, that may not be a joke after all. It may not be we who change the answers. And that makes it hard. — *R.M.S.*



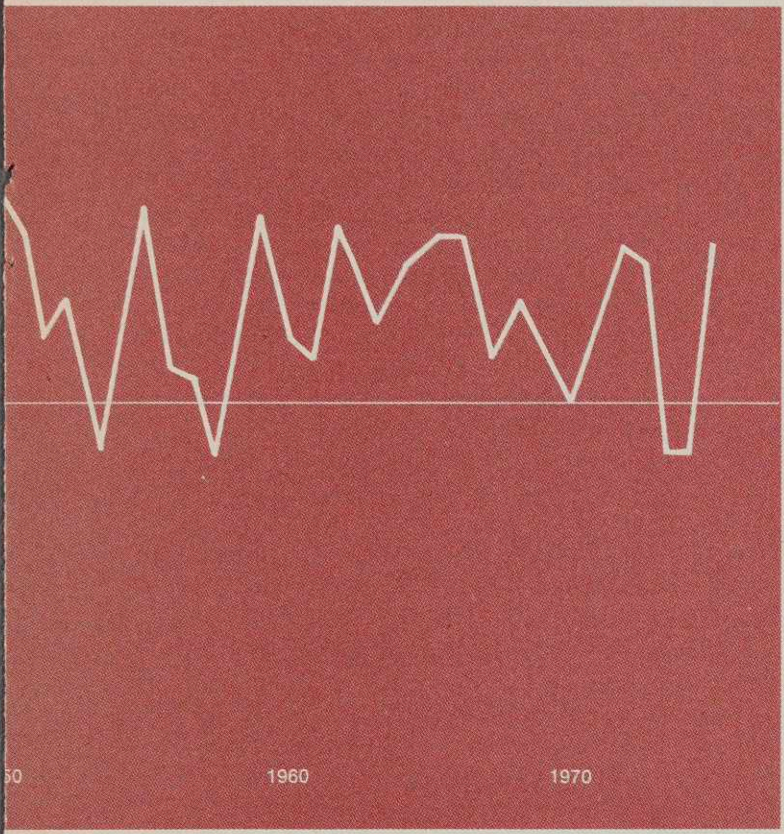
with the change in average hourly wage rates year-by-year during the same period. Notice that he was comparing the rate of change of wages, a nominal quantity, with the percentage of the labor force out of work, a real quantity. If there were no long-run connection between real events and nominal events, then there ought to be no relation between those two time series. If the crude dichotomy in the Keynesian picture were a good description of the world, then the rate of wage inflation ought to be near zero for anything but full employment. And in times of full employment, if there were any to be observed, there ought to be substantial inflation.

What Phillips found was really pretty astonishing. The simple bivariate relation, relating only one real and one nominal variable, held up very well over a very long time during which the nature of British industry and labor changed very drastically. Here was evidence for a strong, and apparently reliable, relation between the nominal world and the real world. It did not appear to be a short-run transient affair, as the mainstream macroeconomics of the 19th and early 20th centuries would have suggested. It seemed not to be a simple dichotomy between less-than-full employment and full employment, as the casual picture of the early 1950s might have suggested. It seemed to say quite clearly that the rate of wage inflation — and probably, therefore, the rate of price inflation — was a smooth function of the tightness of the aggregate economy.

Manipulation Along the Phillips Curve

I remember that Paul Samuelson asked me when we were looking at those diagrams for the first time, "Does that look like a reversible relation to you?" What he meant was, "Do you really think the economy can move back and forth along a curve like that?" And I answered, "Yeah, I'm inclined to believe it," and Paul said, "Me too." And thereby hangs a tale.

A year later, he and I decided to assemble the longest time series analogous to Phillips' data we could find for the United States. We produced a scatter diagram for the United States which didn't look at all like Phillips' for the United Kingdom. There were some very striking anomalies. For example, nominal wage rates were steady or rising in the United States after 1933 although unemployment was still extraordinarily high. But the major difference was that the data points did not cluster closely around a smooth curve, or around any curve, as they did for Phillips; there were points all over the place. That didn't surprise us, because we had expected that changing institutions and changing industrial structure would have a big effect on the responsiveness of wages to unemploy-



The rate of change in U.S. real gross national product (G.N.P.), 1900-1975. This is analogous to a portion of the chart on pages 34 and 35, which shows the similar rate of change in wholesale prices. Beginning in about 1950, the author notes, we seem to begin running "a much smoother economy than we used to. . . . After a half-century of sharply fluctuating price and production levels, we seem to have entered a period in which the trend of these is almost always up and in which the rate of inflation in unusually steady." (Data: U.S. Bureau of the Census)

ment; it was Phillips' result that was surprising, not ours. Then we continued analysis into the post-war period, and our points for 1946 to 1958 *did* cluster around a smooth curve — not the same curve as Phillips', but those thirteen years looked like any representative thirteen years from Phillips' data; their curve was of the same qualitative shape. Then, using no more than a couple of rules of thumb and educated guessing, we converted those post-war observations into a hypothetical relation between the rate of price inflation and the unemployment rate. "This shows the menu of choice between different levels of unemployment and price stability as roughly estimated from the last twenty-five years of American data," we wrote.

Trouble in the Phillips Curve

Now I have continued this series, using the rate of unemployment and the consumer price index for goods other than food, for the years from 1958 through 1977. The first ten years of observations — 1960 through 1969 — are not bad, given that we only intended this as a schematic thing, not the result of formal statistical work. But from even the most casual inspection of the observations for 1970 through 1977 *(see the chart on page 45)*, you would have to say that those were not exactly vintage years for the Phillips curve.

Let me just remind you what was actually happening. The year 1965 was one of negligible inflation. The price of the representative consumer bundle, not counting food, rose by less than 1 per cent, and the unemployment rate, falling throughout the year, reached 4 per cent in December. But 1966 through 1969 were the years of escalation in Vietnam; stepped-up military spending and conscription pushed the economy into the zone where a number of markets became tight. Against the advice of his house economists, President Lyndon Johnson refused to go for a tax increase to siphon off private purchasing power, and by 1969 the unemployment rate averaged 3.5 per cent and the consumer price index for goods other than food was up 4.5 per cent above its 1968 level. By that time, in 1969, the war began to wind down a bit, and the new Republican administration began a perfectly orthodox attempt to reduce the inflation it had inherited by gradually moving the economy back down the Phillips curve. But it didn't quite work out that way; the points for 1970, 1971, and 1972 *(see page 45)* lie further to the right than anyone had a right to expect: inflation was reduced to about 2.5 per cent a year, but at the cost of higher unemployment than the old Phillips curve would have suggested. An observer might reasonably have decided that the economy was still moving along a Phillips curve, only it was moving along a curve that for some reason had slipped about 2 percentage points to the right. But then followed 1974, 1975, 1976, and 1977, producing a relationship between prices and employment hardly compatible with any simple picture of an economy sliding up and down a two-variable Phillips curve. What's more, at this stage of the game, we have no right to pass off the 1970 through 1973 observations as representing just a rightward shift of the curve for which we only need to find a plausible explanation; the whole period from 1970 has to be taken as a major problem for this view of the economy. So we return to the fundamental question again: How is the price level related to the real economy, if at all? And to three other analytical questions:

☐ What happened in the 1970s to blow up the older picture?
☐ Why is today's inflation so persistent and unyielding?
☐ Why don't we have answers to these questions?

To avoid non-essential complications, I have been pretending that there is nothing more to the Phillips curve than the bivariate relationship between the tightness of the economy and the rate of price increase. Of course, that oversimplifies. You should think of this as already having built into it the effects of the other main variables that might be expected to influence levels of wages and prices.

The Search for the New Inflators

The standard reaction to the surprises of the 1970s has been to try introducing some new variables into a wage/price equation. And since the intellectual problem is to reconcile the observations of the 1970s with those of prior years, the most promising variables are those which can be shown to have been strong recently but to have had only a small role, or no role at all, before the 1970s. The trouble is that this way of doing business can make it altogether too easy to explain away a rash of wild observations. All sorts of things are always happening in a complicated economy, and a clever and practiced observer can usually think of one whose timing and direction will correlate with the embarrassing facts that have to be explained. You can find some additional flexibility in the fact — or at least we think it is a fact — that economic effects can often lag behind their causes by a substantial interval of time, maybe even years.

One of the great events of the 1970s was the oil embargo and the spectacular rise in the price of oil that followed it. You don't need to be a deep thinker to figure that this must have had something to do with the inflation since 1974. And indeed there is now a body of work suggesting that a significant part, though only part, of that inflation can be attributed to the special effects of rising oil and food prices.

Inflationary expectations is the other causal factor most commonly relied on to explain the inflation of the late 1970s. It's a very plausible one. In the labor market, for instance, workers who expect prices to rise will certainly demand larger increases in nominal wages than they would under stable conditions. And employers who expect the general price level to be rising will certainly be prepared to offer larger wage increases than under stable conditions, partly because they will find it easier to pay higher wages when prices are rising and partly because they will expect other employers to compete more actively for skilled workers when other employers' selling prices are rising. So it is reasonable to guess that widespread expectations of inflation in fact lead to faster rates of wage increase, and therefore to faster rates of price increase through the influence of rising costs.

In fact, many economists are prepared to go a step further. Both parties to the labor market are presumably interested fundamentally in real wages, the purchasing power of their earnings. So if something in the air leads workers to anticipate another percentage point on the inflation rate, they ought to try to translate that anticipation into another percentage point of wage increase. And they ought to succeed, because employers who share the anticipation of the extra point of inflation ought to be willing to go along. Whatever wage increase the average employer was prepared to offer in the first instance ought to be revised upward by one percentage point for each percentage point of expected inflation. That would leave everything that matters to anybody unchanged — both the relation of costs to prices and the relation of my price to other prices. This means, first of all, that if the market should become convinced that prices will be rising one per cent a year faster than it used to think, then that will in fact make prices rise faster than they would otherwise have risen. Secondly, the stronger assumption that I just described says that an upward revision of one per cent in the expected annual rate of inflation will make prices rise exactly one per cent a year faster than they would otherwise have risen.

I have to say that I regard the propositions just stated with very mixed feelings. I'm always a little dubious about an appeal to expectations as a causal factor; expectations are by definition a force that you intuitively feel must be everpresent and very important but which somehow you are never allowed to observe directly.

But this hypothesis about inflationary expectations has two things going for it. The first is that it really does have a lot of plausibility — it makes a lot of sense in the abstract. The second is that, especially since the traumatic inflation of 1973 and 1974, it does seem to be happening; in both wage-bargaining and industrial pricing people talk about the way their intentions and decisions fit into

> . . . if the market should become convinced that prices will be rising one per cent a year faster than it used to think, then that will in fact make prices rise faster than they would otherwise have risen.

an ongoing inflationary context. Look back to page 41 and see what happens when the Phillips curve is shifted year by year so that it passes through each of the points for the 1970s. A story about the heating-up and cooling-down of inflationary expectations can make plausible sense of the required pattern. But that is hardly a proof.

The Real Power of Expectations

This expectations factor does make sense in the 1970s, and the simultaneously dubious and convincing device of a Phillips curve that shifts with the state of expectations about the price level can explain the recent history of the price level rather well. Here is the history, briefly stated: The perverse political decision to finance the Vietnam war by allowing excess demand to develop created a perfectly orthodox situation. By igniting inflationary expectations, that initial mistake made it much harder to return to the conditions of the 1950s and 1960s. The difficulty was monstrously compounded by the inflationary shock of 1974, which is best understood as an unusual collection of other-factor changes that produced a one-time adverse shift in the Phillips curve. Most of those other-factor changes were on the supply side. The O.P.E.C. oil price increase and the various crop failures that led to a very sharp run-up in food prices are the main ones; but there were others, such as the synchronized prosperity of industrial countries around the world that led to a rise in mineral raw materials prices and set off the speculative boom that ran them up even further. The worst part of that explosion in 1974 was its effect in confirming inflationary expectations and running them up by several notches. What we've been observing in the last three years, according to this hypothesis, is mostly a painfully slow unwinding of those inflationary expectations.

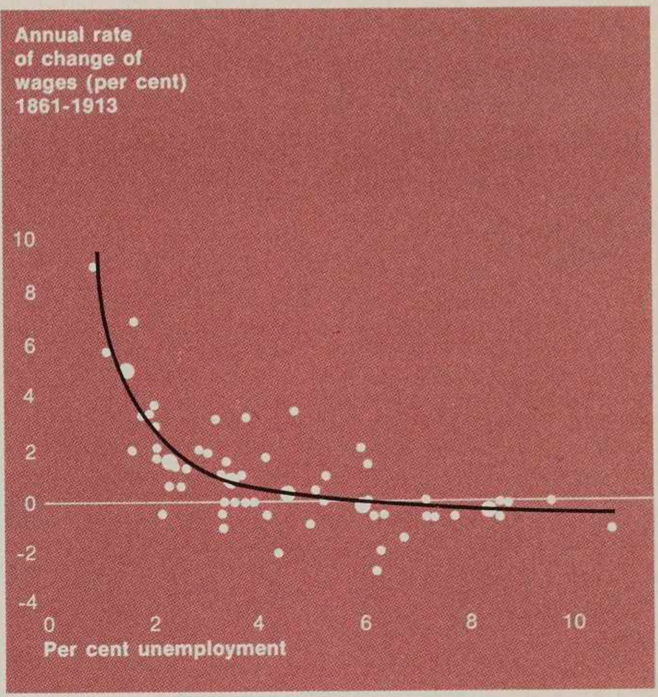
Gum on the Downward Side

Why, then, is the current inflation so persistent and unyielding — so hard to reverse without doing damage to the economy? In terms of the theory I've just been paraphrasing, that question can be rephrased: Why does it take so long for the economy to forget 1970-1974? Why can't we get back to the relatively optimistic Phillips curve of the 1960s?

My first response is almost implicit in one of the earlier diagrams. If you look closely at the behavior of nominal wages and the price level in the roughly 20 business cycles that have occurred since the turn of the century *(pages 34-35)*, a simple generalization asserts itself: on the upswings the price level rises relatively rapidly, but on the downswings the curves flatten.

The upswing behavior should strike you as entirely natural. As the economy prospers, more and more markets become tight. Their individual prices start to rise, and they carry the average of the price level with them. Business cycles are not all alike, of course, and price behavior is not exactly uniform. But the array of figures for the upswings in the 20th century reveals a similar behavior of prices each time; there is nothing that looks like an evolution in the responsiveness of the price level to upswings in the real economy.

But when you turn to the downswings, the picture is rather different: there definitely appears to be an evolution. There were some twelve contractions between 1890 and 1933 in the United States, and the wholesale price index fell during ten of the twelve. In contrast, there have been perhaps six contractions since the end of the Second World War, and only in the first of them did the wholesale price index actually fall. In two it hardly budged at all, and in the remaining three the general price level actually rose at least as fast as in the last year of the preceding upswing. On this crude reckoning, then, the price level is about as flexible in the upward direction as it ever was, but on the down side the price level has become quite sticky. Somehow, the dollar prices of goods and services resist being pushed down.

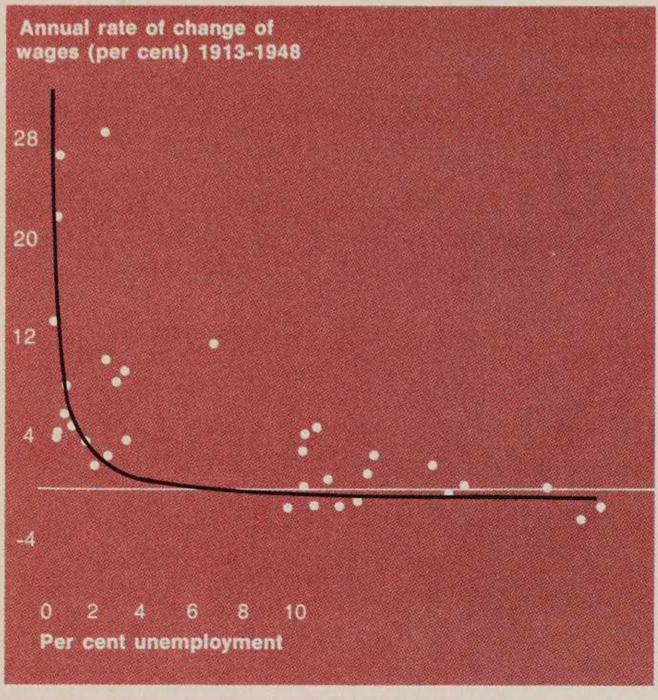


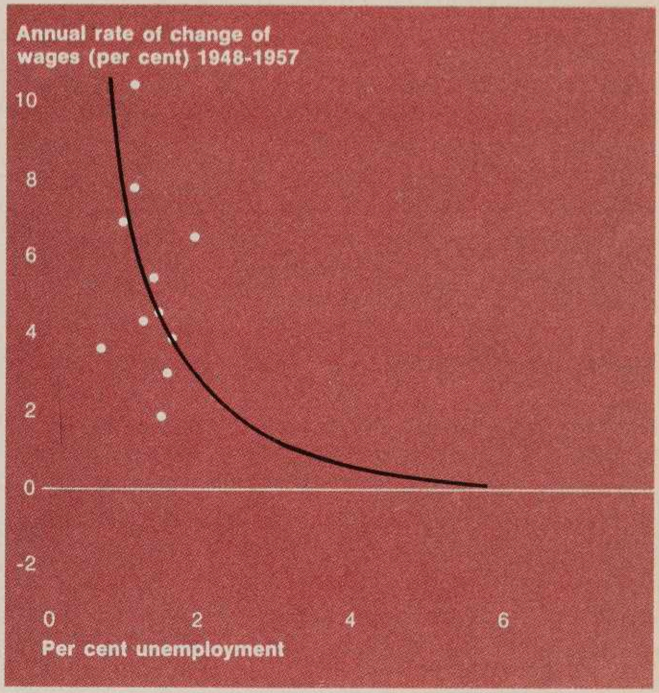
How the famous "Phillips curve," relating rates of wages and of unemployment, entered the mainstream of economics. New Zealander William Phillips (working at the London School of Economics in the 1950s) began by plotting annual rates of change of wages and rates of unemployment in Britain between 1861 and 1913, the small dots in the chart above; then he plotted the decade averages, the large dots on the chart. And finally by statistical methods he found the now-famous curve shown here.

The Upward Bias of Sticky Down

It is easy to see how this asymmetry between upswings and downswings can go a long way toward explaining what has been happening to us. In the first place, it obviously creates an inflationary bias: if the price level rises in the upswings and doesn't fall in the downswings, then the trend of the price level is bound to be upward. A second, more subtle inference to be drawn from this asymmetry may account for the much-reduced volatility of the price level which I discussed earlier. Imagine a recession, and suppose that some commodity prices fall very little despite the weakness in their markets and in the general economy. The rate of change of the general price index will dip little below the zero line — if it dips at all. As the recession comes to an end and economic conditions begin to improve, these sticky downward prices tend to stay put — at least for a while — while their markets recover. If those prices began to rise right away, they would soon enough be way out of line with the prices of those goods whose markets did not weaken so much in the recession.

But then the rate of inflation will not pick up very much in the early stages of expansion, just as it did not fall very much in the recession. Thus, so long as we stay away from extreme booms and extreme depressions, the rate of inflation will not fluctuate very much. That does correspond to the facts of postwar economic life.





William Phillips also plotted wage and unemployment data for 1913 through 1948 (left) and 1948 through 1957 (right); the curve he drew for 1861-1913 (page 42) obviously fits both sets of later data, on which it is superimposed. It was immediately obvious, says the author, that anyone using the 1861-1913 curve could have made good predictions of the wage increase accompanying any unemployment rate in Britain for at least the next 45 years. It was a surprising relationship because it linked a nominal variable (wages) with a real one (unemployment).

Fads and Foibles as an Upward Lever

What can this tell us about our predicament after 1974? Assuming that we accept the generalization about downward inflexibility, how can relative price changes come about? Suppose commodity *a* is trying to become more expensive compared with commodity *b* because costs have risen in the *a* industry or fallen in the *b* industry, or because everybody is mad for *a*'s while *b*'s have gone out of style. If the dollar price of *b* refuses to fall, then the dollar price of *a* has to rise — and rise quite a lot. Such a process pushes the average of the prices of *a* and *b* upward and so raises the whole price level.

Now suppose that after a while the process wants to reverse itself, because an innovation reduces costs in the *a* industry or the in-crowd discovers that *b*'s are beautiful after all. Now it's the turn of *a*'s price to be sticky downward. The only way the relative price change can come about is for the price of *b* to rise. When the relative prices of *a* and *b* return to "normal," the dollar price of both commodities is higher than before. So you see that if prices are sticky downward, the mere churning of relative prices will generate an upward drift of the price level. And the bigger the change in relative prices that the market wants to bring about, the more the price level will have to rise to accommodate them.

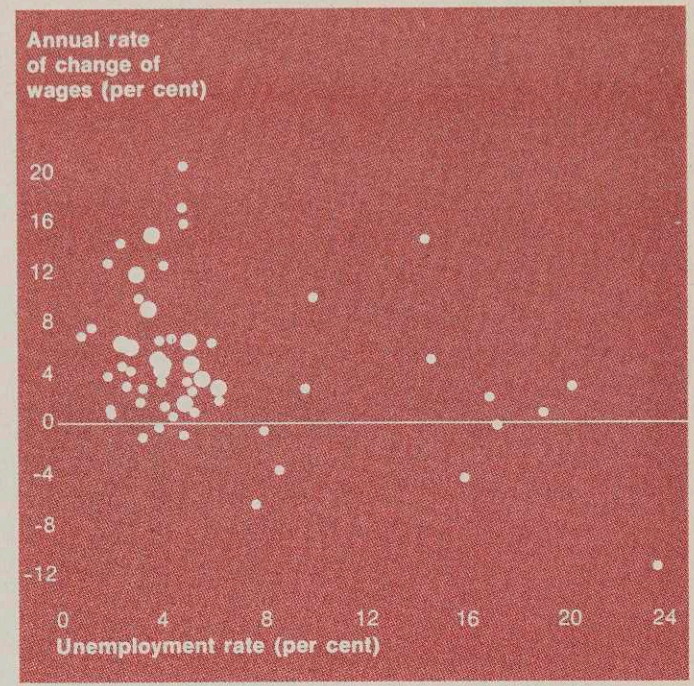
This seems to me especially relevant to the 1970s. Some of the most important special factors that came together in 1973 and 1974 were just the sort of things that call for sharp relative price change — the O.P.E.C. oil embargo, the world shortage of grain and animal feeds, and even the speculative boom in minerals prices. Those were all trying to make fuel, foods, and mineral materials more expensive relative to manufactured goods and various services. Since the prices of manufactured goods and various services were sticky downward, the whole price level floated upward. And because the initiating impulse that started it off was unusually large, the resulting inflation was unusually fast and substantial.

Pressing the Rise Out of Prices

Now we can begin to see why squeezing the inflation of the 1970s out of the system in the conventional way is such a long and painful process. In the past, you could hope to push the economy back down the appropriate Phillips curve; a softer economy would mean slower inflation as more and more prices came under the pressure of weaker markets. The price level might stabilize, with some prices still rising and others falling. But if the weaker prices refuse to fall, then the extra point on the unemployment rate resulting from reduced production for weaker markets will have a smaller effect in reducing inflation. Any given reduction in the rate of inflation requires a bigger and probably a longer-lasting dose of recession and unemployment.

This somewhat superficial explanation comes pretty close to saying that we have inflation because we're able to avoid deflation, and only a little is added by the observation that we have inflation because we expect inflation, and we expect inflation because we've had it. Why have so many prices become sticky downward, and why are inflationary expections so hard to dissipate?

We know that in the inflation of the 1970s each of the Phillips curves in the family is relatively flat; you have to accept a lot of unemployment to push the economy down any one of those curves. Most of the serious estimates suggest that an extra 1 per cent of unemployment maintained for one year would reduce the rate of inflation by something between 0.16 and 0.5 per cent. That trade-off is not very favorable. We also know that the inflationary process involves a great deal of inertia; that is, it takes a long time for the economy to pass from one member of the family of Phillips curves to a lower one, at least under normal circumstances. For instance, an extra 1 per cent of unemployment maintained for three years would reduce the inflation rate by something between 0.5 and 1.75 per cent. (An extra point of unemployment for three years costs the economy about $180 billion of production, which makes this a very expensive way to reduce the inflation rate.)



The relationship of wage changes and unemployment in the United States, 1933 to 1960. Intrigued by William Phillips' results analyzing unemployment and wages in Britain (see pages 42 and 43), the author and Professor Paul A. Samuelson in 1959 sought and plotted the longest time series of similar data they could find for the United States. The data points for the years before 1946 (small) were "all over the place," as Professor Solow puts it; but those for 1946 to 1958 (large) cluster much as Phillips' do. In general, the author felt justified in his skepticism about a continuous relationship of wage changes and unemployment through countless changes in industrial institutions and structure.

We know those two things, albeit in a tentative and gingerly way. What we don't know — which is the last thing I'm supposed to tell you — is why the inertia is so great, why those Phillips Curves are so flat. That is, we do not know what bits of our social and economic structure would have to be changed in order to change those relationships. The last few paragraphs of this article represent some speculations on this matter, and readers should be aware that they are just that and nothing more.

Inflation as the Price of Stabilization

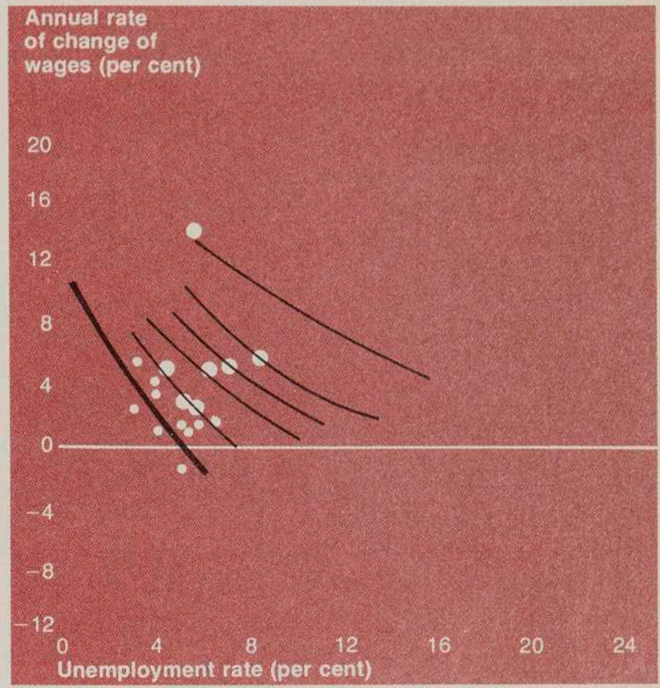
It's not hard to think up reasons why our society might exhibit just those characteristics that I've described. But I have no confidence at all that I know which are the important factors and which are the minor ones, which we should try to change and which would hardly be worth the trouble.

For example, why are wages so sticky downwards, and when would wages *not* be sticky downwards? When might a little bit of unemployment be expected to erode the nominal wage rapidly? Presumably this would be the case when there's active competition for a limited number of jobs. Has anything happened in the modern world to make unemployed workers less willing to compete for jobs held by employed workers?

I can suggest two things.

Since the 1930s we've had a system of unemployment insurance. An unemployed worker who has built up eligibility through previous employment is entitled to a cash benefit for a specified number of weeks. Any such cushion is bound to slow down the process of competition when there are not enough jobs to go around. But this unemployment insurance system has been in place for forty years, so it can hardly account for a change in the character of the labor market which we first observe after 1950. But in the 1950s about 60 per cent of all employment was covered by the unemployment insurance system, whereas now something over 80 per cent is covered. That could make a difference. In the 1950s the average unemployment compensation benefits replaced about 40 per cent of the spendable earnings of the average worker. That replacement rate rose to about 46 per cent in the late 1960s and is now 47 per cent. That seems like a minor difference to me. But the average figure may understate the change because the replacement rate could be larger for the second worker in a two-worker family.

Now don't leap from this to the conclusion that the strength of the unemployment insurance system, being a source of downward stickiness which makes the economy inflation-prone, must be a bad thing. Unemployment insurance is a social invention that serves an important and



The last era of the Phillips curve for the United States, 1960-1977. Points relating changes in the price level (represented by the consumer price index for goods other than food) to rates of unemployment from 1960 to 1969 — the small dots on the chart — fall along a line that can easily be connected by a single curve like that of William Phillips for Britain in an earlier era. But similar points for the years 1970 to 1977 (the large dots) seem to require separate curves. Does this failure of a singular Phillips curve on recent U.S. economic data shed light on the special characteristics of present-day U.S. inflation?

useful purpose: a worker who loses his or her job — not for incompetence or goofing off but because the real economy has weakened, therefore for reasons entirely outside the individual worker's control — should not be forced by instant poverty to take *any* kind of a job, to undercut other workers who are still lucky enough to be employed, or to abandon every shred of personal dignity. The problem is precisely that a good social institution *may* have some undesirable side-effects. We don't really know how important that factor is, and it would be a gross error to believe that an easy solution is to be found by tinkering with the unemployment insurance system.

Confidence as a Breeder of Inflation

Another source of downward inflexibility in wages and prices, even harder to prove, may be the widespread belief that we have learned to stabilize the real economy. We have seen that the sharp damping of fluctuations in the rate of inflation after 1950 more or less coincided with a pronounced damping of fluctuations in the real economy. Now you can see how those two things might be related. If workers and employers both believe that recessions will be short and mild, then wages and prices are unlikely to crumble in recessions. Unemployed workers will not compete for jobs because they have confidence that their old jobs will reopen soon or at least that new jobs will soon appear as business recovers. Business firms are unlikely to compete for sales by cutting prices because their costs haven't been reduced and because they expect market conditions to improve soon. These beliefs of both workers and business were reinforced between 1950 and 1974 by the fact that they happened to be true: recessions were in general mild and short. We cannot yet know whether the unusually deep recession of 1974 and 1975 has made any impression on that belief. But it is clear that, even if one source of inflation-proneness in the modern economy is its success in stabilizing itself around a fairly high level of production and employment, you would hardly want to go back to instability and unemployment just so that inflation and deflation could alternate, as they did in the good old days. That would be no bargain at all.

Here is one more hypothesis about the downwards-stickiness of wages and prices. A highly differentiated labor market is a characteristic of this country and of all advanced industrial countries. Even within a single industry, we observe a broad hierarchy of occupations which differ one from the next by the kind and amount of skill they require, the deadliness or variety of the work itself, the comfort or discomfort of the work environment, and in many other ways. A structure of wage differentials accompanies this hierarchy of jobs. This pattern of wage differentials turns out to be very persistent; if disturbed, it tends to reestablish itself. Anomalies in the economy — a construction boom, a bad automobile year, a successful or unsuccessful strike — can disturb it. The traditional wage structure tends to reassert itself when that happens, but that process can take a very long time, depending on accidents of the two-or three-year collective bargaining cycle, uneven prosperity of different regions and different industries, and other factors. During this long period while a discontinuity is working itself out in a segmented labor market we are likely to experience what is sometimes called a wage-wage spiral. I think that this is at least as important in our recent history as the wage-price spiral that is usually held responsible for the inertia of the inflationary process. For example, the outcome of the coal strike of 1978 will be visible in the wage-and-price record in this country for many years as other workers catch up with the coal miners. I do not mean to begrudge the miners any single nickel they got; for my money, one thing worse than not having tenure in a university would be having tenure in a coal mine; I mean only to point out that the wage structure is another example of a social institution that serves a real function — keeping order and equity in the workplace — and that may also make more difficult the shift to a lower Phillips curve.

Please be warned: these last observations are purely guesswork. They make some sense to me, so I hope they make some sense to you; but I can't even imagine how one would go about testing those hypotheses and evaluating their relative significance. There may even be equally or more important effects that I've missed.

What I am fairly sure of, however, is this: there are no easy answers or quick cure-alls for the inflation-proneness of the modern industrial economy, and I suspect that nibbling away at the problem is the best that we can do.

This article is an adaptation of two lectures delivered by **Robert M. Solow**, Institute Professor and professor of economics at M.I.T., and the Institute's distinguished Killian Lecturer in 1977-78. Professor Solow has been a member of the faculty since 1950, a year before he completed his Ph.D. at Harvard. Since then he has been widely honored for studies in economic theory, including the theory of capital and growth, macroeconomics, and the economics of natural resources. Readers will quickly understand and applaud the citation which accompanied Professor Solow's Killian Award from his faculty colleagues: his teaching and research have been accomplished "with such wit, style, and commitment as to give him a special place in our community."

THE REVIEW

OF

ECONOMIC STATISTICS

Quoted in
ChC, long wave.
8/29/85

Kondratieff
1935

✓

1935

TABLE OF CONTENTS

THE REVIEW OF ECONOMIC STATISTICS, VOL. XVII

The pages indicated in roman type refer to the issues of January, February, and March; the pages indicated in bold face type refer to the issues of May, August, and November.

HARVARD ECONOMIC SOCIETY, INC.

(JANUARY–MARCH)

PRINTED AT
CAMBRIDGE, MASS., U. S. A.

THE REVIEW OF ECONOMIC STATISTICS

(MAY–NOVEMBER)

PUBLISHED BY HARVARD UNIVERSITY

PRINTED AT
CAMBRIDGE, MASS., U. S. A.

CONTENTS

The Review *of* Economic Statistics

Volume XVII | November, 1935 | Number 6

THE LONG WAVES IN ECONOMIC LIFE

N. D. KONDRATIEFF

FOREWORD

The editors of the Review of Economic Statistics are happy to be able to present in translation the peculiarly important article by Professor Kondratieff, which, under the title "Die langen Wellen der Konjunktur," appeared in the *Archiv für Sozialwissenschaft und Sozialpolitik* in 1926 (vol. 56, no. 3, pp. 573–609). The combining circumstances of an increasing interest in "long waves" and the difficulty of securing access to the original article would alone justify translation and publication of Kondratieff's contribution to the theory of the trade cycle. In addition, the editors would take this means of indicating their intention from time to time of rendering available to the English-using world outstanding articles in foreign periodicals.

This translation of Professor Kondratieff's article was made by Mr. W. F. Stolper of Harvard University. Due to the limitations of space, the editors have taken the liberty to summarize certain sections of this translation. With the exception of a ten-page appendix of tabular material, however, all tables and charts have been included.

I. INTRODUCTION

The idea that the dynamics of economic life in the capitalistic social order is not of a simple and linear but rather of a complex and cyclical character is nowadays generally recognized. Science, however, has fallen far short of clarifying the nature and the types of these cyclical, wave-like movements.

When in economics we speak of cycles, we generally mean seven to eleven year business cycles. But these seven to eleven year movements are obviously not the only type of economic cycles. The dynamics of economic life is in reality more complicated. In addition to the above-mentioned cycles, which we shall agree to call "intermediate," the existence of still shorter waves of about three and one-half years' length has recently been shown to be probable.[1]

But that is not all. There is, indeed, reason to assume the existence of long waves of an average length of about 50 years in the capitalistic economy, a fact which still further complicates the problem of economic dynamics.

II–III. METHOD

[Sections II and III of Kondratieff's exposition may be summarized as follows:

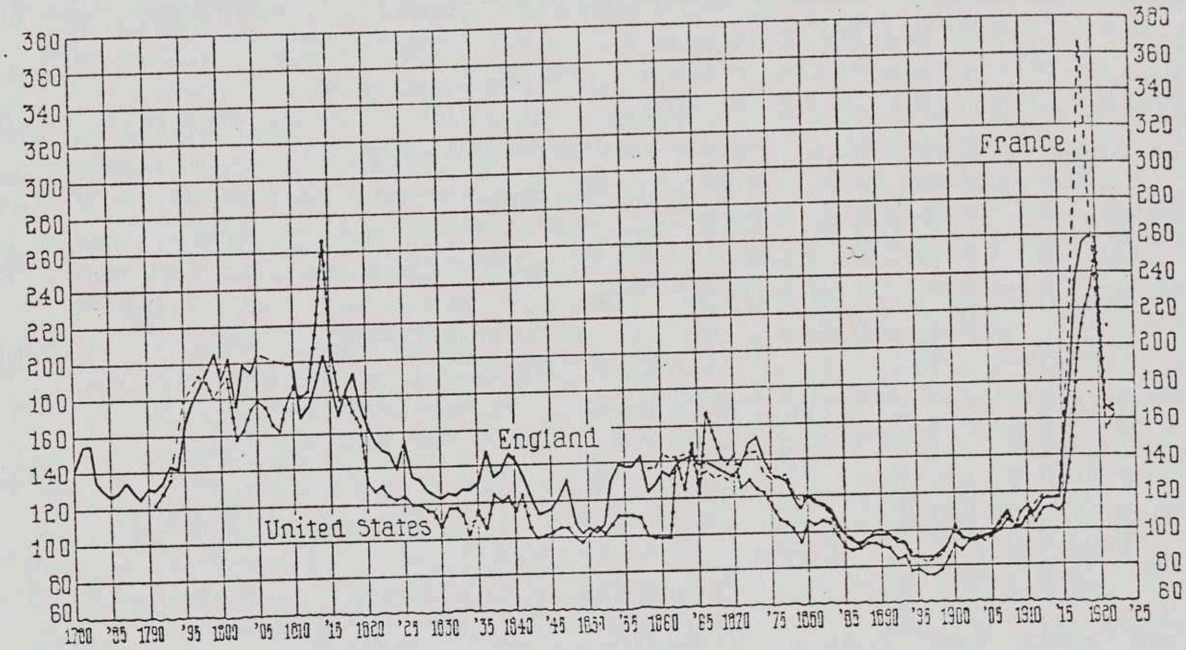
The succeeding study is to be confined solely to an inquiry into various problems connected with these long waves. Investigation here is made difficult by the fact that a very long period of observation is presupposed. We have, however, no data before the end of the eighteenth century and even the data that we do have are too scanty and not entirely reliable. Since the material relating to England and France is the most complete, it has formed the chief basis of this inquiry. The statistical methods used were simple when no secular trend was present in the series. If the series displayed a secular trend, as was the case among physical series, the first step was to divide the annual figures by the population, whenever this was logically possible, in order to allow for changes in territory. Then the secular trend was eliminated by the usual statistical methods applied to each series as a whole; and Kondratieff refers specifically to the methods presented by Dr. Warren M. Persons in this Review in 1919 and 1920. The deviations from the secular trend were then smoothed by a nine-year moving average, in order to eliminate the seven to eleven year business cycles, the short cycles, and random fluctuations possibly present.]

IV. THE WHOLESALE PRICE LEVEL

While the index of French prices goes back only to the end of the 1850's, the English and American indices date back to the close of the eighteenth century. In order not to overburden

[1] Cf. J. Kitchin, "Cycles and Trends in Economic Factors," Review of Economic Statistics [hereafter referred to as "this Review"], v (1923), pp. 10–16.

CHART 1. — INDEX NUMBERS OF COMMODITY PRICES*
(*1901-10 = 100*)



* The *French* data are taken from the *Annuaire Statistique* [Statistique Générale de la France], 1922, p. 341; the index number has been recalculated on a gold basis through use of dollar-franc exchange rates.

For *England*, there is for 1782–1865 the index of Jevons; for 1779–1850, a new index number, computed by Silberling and published in this REVIEW, v (1923); for the period after 1846, we have Sauerbeck's index, which at present is carried on by the *Statist*. Since Silberling's index is based upon more complete data of the prices of individual commodities than that of Jevons, we have used the former for the period 1780–1846. From 1846 on we use Sauerbeck's index number. Both indices have been tied together on the basis of their relation during 1846–50, for which period they are both available; after this procedure, we have shifted the series to a new base, 1901–10. For the period 1801–20 and since 1914, in which periods England was on a paper standard, the index numbers have been recalculated on a gold basis.

For the *United States*, we use the following series, which have been tied together: for 1791–1801, H. V. Roelse (*Quarterly Publications of the American Statistical Association*, December, 1917); 1801–25, A. H. Hansen (*ibid.*, December, 1915); 1825–39, C. H. Juergens (*ibid.*, June, 1911); 1840–90, Falkner (Report from the Committee on Finance of the United States Senate on *Wholesale Prices, Wages, and Transportation*, 52d Congress, 2d session, Report No. 1394, Part 1 [Washington: Government Printing Office, March 3, 1893]); since 1890, the B. L. S. index. All index numbers are on the base 1901–10. For the Greenback period (1862–78), they have been recalculated on a gold basis. All data [except Silberling's index] are taken from the *Annuaire Statistique*, 1922 [which utilizes the sources above cited].

this study with figures, the statistical data are presented exclusively in the form of charts.[1]

The index numbers of prices plotted on Chart 1 have been neither smoothed nor treated in any other way. Nevertheless, a mere glance at the chart shows that the price level, despite all deviations and irregularities, exhibits a succession of long waves.

The upswing of the first long wave embraces the period from 1789 to 1814, i.e., 25 years; its decline begins in 1814 and ends in 1849, a period of 35 years. The cycle is, therefore, completed in 60 years.[2]

The rise of the second wave begins in 1849 and ends in 1873, lasting 24 years. The turning point, however, is not the same in the United States as in England and France; in the United States the peak occurs in the year 1866, but this is to be explained by the Civil War and casts no doubt on the unity of the picture which the course of the wave exhibits in the two continents. The decline of the second wave begins in 1873 and ends in 1896, a period of 23 years. The length of the second wave is 47 years.

[1] [Ten pages of tabular material were given by Kondratieff at the end of his article. The charts presented in this translation are not merely reproductions of those in the original article but have been drawn anew from the data given in his tabular appendix. A few slight discrepancies between the new charts and those of Kondratieff were discovered, but in no case were the discrepancies significant.—Editors.]

[2] In the upswing, the English index exhibits several peaks, which fall in the years 1799, 1805, 1810, and 1814; but since after the year 1814 a distinctly downward tendency can be observed, we regard this year as the turning point.

The upward movement of the third wave begins in 1896 and ends in 1920, its duration being 24 years. The decline of the wave, according to all data, begins in 1920.

It is easily seen that the French prices after the close of the 1850's move generally parallel to the English and American prices. It is, therefore, very probable that this parallelism existed in the preceding period as well.

We conclude, therefore, that three great cycles are present in the movement of the price level during the period since the end of the 1780's, the last of which is only half completed. The waves are not of exactly the same length, their duration varying between 47 and 60 years. The first wave is the longest.

V. THE RATE OF INTEREST

The course of the interest rate can be seen most conveniently from the movement of the discount rate and the quotations of interest-bearing securities. Because the latter depend less on random fluctuations and reflect more accurately the influence of long-run factors, we use here only the quotations of state bonds.

CHART 2

QUOTATIONS OF INTEREST-BEARING SECURITIES

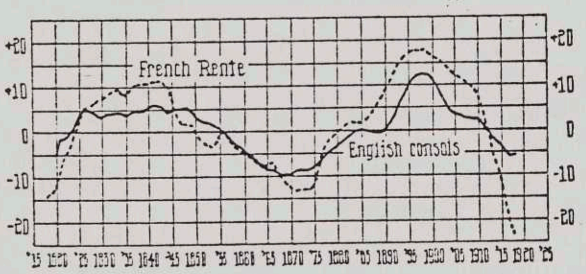


Chart 2 shows the quotations of the French Rente[1] and of English consols.[2] Both have a secular trend during the period of observation. The chart shows the deviations from the secular trend smoothed by means of a nine-year moving average.

The quotations of interest-bearing securities manifest, as is well known, a movement opposite to that of general business activity and of the interest rate. Therefore, if long waves are operative in the fluctuations of the interest rate, the movement of bond quotations must run in a direction counter to that of commodity prices. Just this is shown in our chart, which exhibits clearly the long waves in the movement of the quotations and consequently of the interest rate.

The chart starts only after the Napoleonic Wars, i.e., about the time that the first long wave of commodity prices had reached its peak; it does not cover the period of the upswing of the latter. Considering the data at hand, however, we may suppose that the quotations of state bonds took part in this movement also.

English consols actually manifest a decidedly downward tendency between 1792 and 1813. Their quotation in 1792 is 90.04; in 1813, on the other hand, it is 58.81. Although they drop most rapidly in the years 1797 and 1798, yet this steep decline is only an episode, and the general downward tendency from 1792 to 1813 stands out quite clearly.[3]

Accordingly, the period from the beginning of the 1790's up to 1813 appears to be the phase of rising interest rates. This period agrees perfectly with that of the rising wave of commodity prices.

The wave of bond quotations rises after 1813[4] — or the wave of the interest rate declines — even till the middle of the forties. (See the chart.) According to the unsmoothed data, consols reached their peak in 1844; the Rente, in 1845. With this, the first great cycle in the movement of the interest rate is completed.

The downward movement of bond quotations (or the rise of the interest rate) during the second cycle lasts from 1844–45 to 1870–74.[5] From this time onward until 1897, the market price of interest-bearing securities rises again, and consequently the interest rate goes down. With this, the second great cycle is completed.

The new decline of the quotations (rise in the

[1] Until 1825 the quotations of the five-per-cent Rente, after this the quotations of the three-per-cent Rente. In order to connect both series, we have first computed relatives with the base 1825–30 for both series. Then we shifted the base of the combined series to 1901–10, in order to make them comparable with the price curve. The original data are taken from the *Annuaire Statistique* [Statistique Générale de la France], 1922.

[2] According to the data in William Page, ed., *Commerce and Industry*, Vol. 2 (London, 1919), statistical tables, pp. 224–25. Relatives have been calculated from the figures, with the base 1901–10.

[3] Cf. N. J. Silberling, "British Financial Experience, 1790–1830," this REVIEW, I (1919), p. 289.

[4] The first years have disappeared from our chart because of the use of the nine-year moving average.

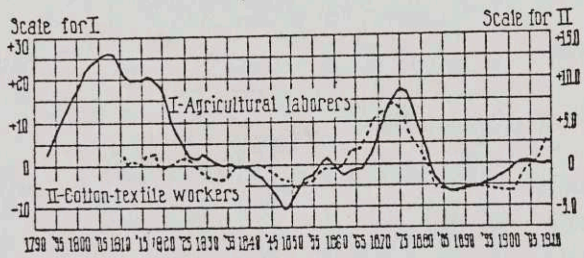
[5] According to the original data, consols actually reach their lowest point in 1866, but the general tendency continues to be one of decline until 1874. The slump of quotations in 1866 is connected with the increase in the interest rate just preceding the money-market crisis of that year, and with the Austro-Prussian War.

rate of interest) lasts from 1897 to 1921. Thus the existence of great cycles in the movement of the interest rate appears very clearly.[1] The periods of these cycles agree rather closely with the corresponding periods in the movement of wholesale commodity prices.

VI–VII. WAGES AND FOREIGN TRADE

[In Section VI, Kondratieff examines the course of weekly wages of workers in the English cotton-textile industry since 1806 and of English agricultural laborers since 1789.[2] The original wage data are reduced to a gold basis and then expressed in the form of index numbers with 1892 as the base year. Chart 3 presents these

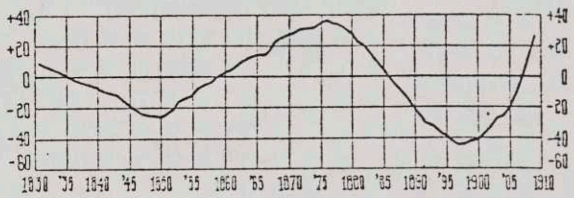
Chart 3.— Wages in England



wage figures as deviations from trend, smoothed by use of a nine-year moving average. Kondratieff devotes the remainder of this section to a description of the series presented in Chart 3, from which analysis he concludes that, despite the scantiness of the available data, "long waves are undoubtedly present in the movement of wages, the periods of which correspond fairly well with those in commodity prices and the interest rate."

For his foreign-trade series presented in Section VII, Kondratieff takes the sum of French exports and imports. The figures were first corrected for population changes, and thereafter the secular trend (in the form of a second-degree parabola) was eliminated. The resulting deviations, smoothed by use of a nine-year moving average, are presented in Chart 4. After an

Chart 4.— French Foreign Trade



examination of the chart, the author concludes that the data on foreign trade also show the existence of two great cycles, the periods of which coincide with those observed in the other data.]

VIII. THE PRODUCTION AND CONSUMPTION OF COAL AND PIG IRON, AND THE PRODUCTION OF LEAD

So far we have examined the movements only of such magnitudes, sensitive to changes in business conditions, as possess either a purely value character, e.g., commodity prices, interest rates, and wages, or at least a mixed character such as the data on foreign trade. Our study, however, would lose much of its force if we did not also analyze the behavior of purely physical series.

For this purpose we choose English coal production,[3] and French consumption of coal,[4] as well as the English production of pig iron and of lead.[5] We divided the original figures by the population, and eliminated from the resulting series the secular trends. The deviations from the lines of trend, after being smoothed by use of a nine-year moving average, were then analyzed. The results are shown in Chart 5.

Continuous data are available, unfortunately, only for the period after the 1830's, in part even only after the 1850's. Consequently, only one and one-half to two great cycles can be shown, but these appear with striking clarity in both charts.

[1] The existence of these cycles is also confirmed by several other studies: P. Wallich, "Beiträge zur Geschichte des Zinsfusses von 1800 bis zur Gegenwart," *Jahrbücher für Nationalökonomie und Statistik*, III. Folge, Vol. 42, pp. 289–312; J. Lescure, "Hausses et Baisses Générales des Prix," *Revue d'Economie Politique*, Nr. 4 (1912); R. A. Macdonald, "The Rate of Interest Since 1844," *Journal of the Royal Statistical Society*, LXXV (1912), pp. 361–79; T. T. Williams, "The Rate of Discount and the Price of Consols," *ibid.*, pp. 380–400. Also *ibid.*, pp. 401–11, the discussion of the last-mentioned studies, especially the speech of E. L. Hartley, pp. 404–06.

[2] [Earnings of cotton-textile workers for 1806–1906 are taken from G. H. Wood, *The History of Wages in the Cotton Trade* (London, 1910), p. 127; beginning with 1906, they are from the *Abstract of Labour Statistics*.

For agricultural laborers, wage data for 1789–1896 are from A. L. Bowley, "The Statistics of Wages in the United Kingdom During the Last Hundred Years: Part IV, Agricultural Wages," *Journal of the Royal Statistical Society*, LXII (1899), pp. 555 ff. Thereafter, the figures are from Page, *op. cit.* The data refer to England and Wales.]

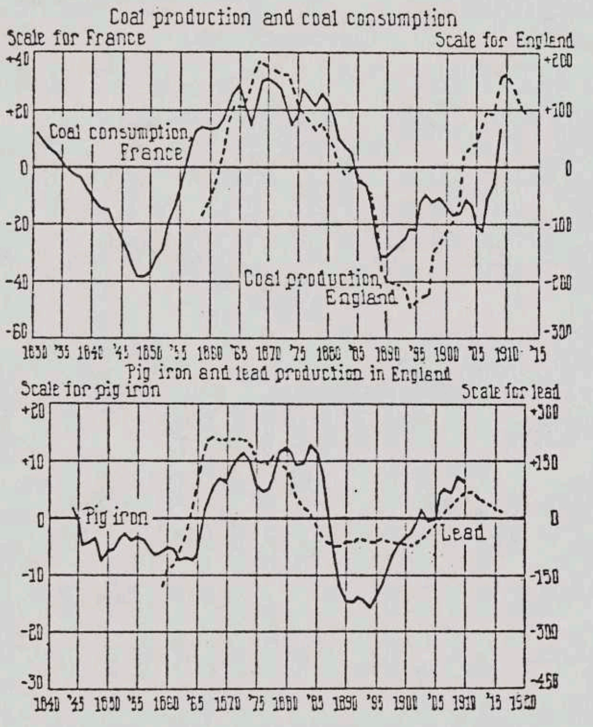
[3] According to the data of W. Page, *op. cit.*

[4] *Annuaire Statistique*, 1908 and 1922.

[5] According to *British and Foreign Trade and Industry*, and the *Statistical Abstract* [*for the United Kingdom*].

There is a retardation in the increase of coal consumption [in France] until the end of the 1840's, then the advance becomes more rapid and reaches its peak in 1865, according to the smoothed curve (on the chart), and in 1873, according to the unsmoothed curve. In the latter year, English coal production also reaches a maximum, according to the unsmoothed curve. Then follows the decline, which comes to an end in 1890–94, giving way to a new long upswing. So we observe in the data relative to the rapidity in the increase of coal production and coal consumption nearly the whole of two large cycles, the periods of which correspond closely to the periods we have already found when considering other series.

Chart 5.— Consumption of Coal in France and Production of Coal, Pig Iron, and Lead in England



Similarly, English production of pig iron and lead indicates sufficiently clearly the existence of one and one-half large cycles.

IX. OTHER SERIES

For the sake of brevity, we break off here the systematic analysis of the long waves in the behavior of individual series. We have also examined other data, some of which likewise showed the same periods as those mentioned above, although several other series did not show the cycles with the same clarity. Value series which show long waves are the deposits and the portfolio of the Bank of France, and deposits at the French savings banks; series of a mixed (quantity x price) character are French imports and English imports, and total English foreign trade. As regards the movement of indices of a physical character, the existence of long waves has been established in the coal production of the United States, of Germany, and of the whole world; in the pig-iron production of the United States and of Germany and of the whole world; in the lead and coal production of the United States; in the number of spindles of the cotton industry in the United States; in the cotton acreage in the United States and the oat acreage in France, etc.

It was absolutely impossible, on the other hand, to establish long waves in French cotton consumption; in the wool and sugar production of the United States; and in the movement of several other series.

X. STATISTICAL FINDINGS

The evidence we have presented thus far permits some conclusions.

(1) The movements of the series which we have examined running from the end of the eighteenth century to the present time show long cycles. Although the statistical-mathematical treatment of the series selected is rather complicated, the cycles discovered cannot be regarded as the accidental result of the methods employed. Against such an interpretation is to be set the fact that these waves have been shown with about the same timing in all the more important of the series examined.

(2) In those series which do not exhibit any marked secular trend — e.g., prices — the long cycles appear as a wave-like movement about the average level. In the series, on the other hand, the movement of which shows such a trend, the cycles accelerate or retard the rate of growth.

(3) In the several series examined, the turning points of the long waves correspond more or less accurately. This is shown clearly by Table 1, which combines the results of the investigation not only of the data considered above but also of several other series.[1]

[1] Table 1 enumerates the maxima and minima according to the original data. The problem of the most accurate method

TABLE 1

Country and series	First cycle		Second cycle		Third cycle	
	Beginning of rise	Beginning of decline	Beginning of rise	Beginning of decline	Beginning of rise	Probable beginning of decline
France						
1. Prices	....	....	....	1873	1896	1920
2. Interest rate	....	1816*	1844	1872	1894	1921
3. Portfolio of the Bank of France	....	1810*	1851	1873	1902	1914
4. Deposits at the savings banks	....	....	1844	1874	1892	....
5. Wages of coal miners	....	....	1849	1874	1895	....
6. Imports	....	....	1848	1880	1896	1914
7. Exports	....	....	1848	1872	1894	1914
8. Total foreign trade	....	....	1848	1872	1896	1914
9. Coal consumption	....	....	1849	1873	1896	1914
10. Oat acreage[1]	....	....	1850*	1875	1892	1915
England						
1. Prices	1789	1814	1849	1873	1896	1920
2. Interest rate	1790	1816	1844	1874	1897	1921
3. Wages of agricultural laborers	1790	1812–17	1844	1875	1889	....
4. Wages of textile workers	....	1810*	1850†	1874	1890	....
5. Foreign trade	....	1810*	1842‡	1873	1894	1914
6. Coal production	....	....	1850*	1873	1893	1914
7. Pig iron production	....	....	....	1871§	1891	1914
8. Lead production	....	....	....	1870	1892	1914
United States						
1. Prices	1790	1814	1849	1866	1896	1920
2. Pig iron production	....	....	....	1875–80	1900	1920
3. Coal production	....	....	....	1893	1896	1918
4. Cotton acreage	....	....	....	1874–81	1892–95	1915
Germany						
Coal production	....	....	....	1873‖	1895	1915
Whole world[2]						
1. Pig iron production	....	....	....	1872¶	1894	1914
2. Coal production	....	....	....	1873	1896	1914

[1] Reversed cycles.
[2] The data which refer to the whole world have not been corrected for population changes.
* Approximate dates.
† Another minimum falls in the year 1835.
‡ Other minima lie in the years 1837 and 1855.
§ Another maximum falls in the year 1881.
‖ Another maximum falls in the year 1883.
¶ Another maximum falls in the year 1882.

It is easy to see from this table that there is a very close correspondence in the timing of the wave movements of the series in the individual countries, in spite of the difficulties present in the treatment of these data. Deviations from the general rule that prevails in the sequence of the cycles are very rare. It seems to us that the absence of such exceptions is more remarkable than would be their presence.

(4) Although for the time being we consider it to be impossible to fix exactly upon the years that marked the turning points of the long cycles, and although the method according to which the statistical data have been analyzed

for the determination of the maxima and minima would deserve a special analysis; at present we leave this question open. We believe only that the indicated turning points are the most probable ones.

permits an error of 5–7 years in the determination of the years of such turnings, the following limits of these cycles can nevertheless be presented as being those most probable:

First long wave	1. The rise lasted from the end of the 1780's or beginning of the 1790's until 1810–17. 2. The decline lasted from 1810–17 until 1844–51.
Second long wave	1. The rise lasted from 1844–51 until 1870–75. 2. The decline lasted from 1870–75 until 1890–96.
Third long wave	1. The rise lasted from 1890–96 until 1914–20. 2. The decline probably begins in the years 1914–20.

(5) Naturally, the fact that the movement of the series examined runs in long cycles does not yet prove that such cycles also dominate the movement of all other series. A later examination with this point especially in mind will have to be made to show which ones of these share the described wave-like movement. As already pointed out, our investigation has also extended to series in which no such waves were evident. On the other hand, it is by no means essential that the long waves embrace all series.

(6) The long waves that we have established above relative to the series most important in economic life are international; and the timing of these cycles corresponds fairly well for European capitalistic countries. On the basis of the data that we have adduced, we can venture the statement that the same timing holds also for the United States. The dynamics in the development of capitalism, however, and especially the timing of the fluctuations in the latter country may have peculiarities.

XI. EMPIRICAL CHARACTERISTICS

We were led to these conclusions by the study of statistical series characterizing the movement of the capitalist economy. From another point of view, the historical material relating to the development of economic and social life as a whole confirms the hypothesis of long waves. We neither can nor shall undertake here an analysis of this material. Nevertheless, several general propositions which we have arrived at concerning the existence and importance of long waves may be presented.

(1) The long waves belong really to the same complex dynamic process in which the intermediate cycles of the capitalistic economy with their principal phases of upswing and depression run their course. These intermediate cycles, however, secure a certain stamp from the very existence of the long waves. Our investigation demonstrates that during the rise of the long waves, years of prosperity are more numerous, whereas years of depression predominate during the downswing.[1]

(2) During the recession of the long waves, agriculture, as a rule, suffers an especially pronounced and long depression. This was what happened after the Napoleonic Wars; it happened again from the beginning of the 1870's onward; and the same can be observed in the years after the World War.[2]

(3) During the recession of the long waves, an especially large number of important discoveries and inventions in the technique of production and communication are made, which, however, are usually applied on a large scale only at the beginning of the next long upswing.

(4) At the beginning of a long upswing, gold production increases as a rule, and the world market [for goods] is generally enlarged by the assimilation of new and especially of colonial countries.

(5) It is during the period of the rise of the long waves, i.e., during the period of high tension in the expansion of economic forces, that, as a rule, the most disastrous and extensive wars and revolutions occur.

It is to be emphasized that we attribute to these recurring relationships an empirical character only, and that we do not by any means hold that they contain the explanation of the long waves.

XII. THE NATURE OF LONG WAVES

Is it possible to maintain that the existence of long cycles in the dynamics of the capitalist economy is proved on the basis of the preceding statements? The relevant data which we were able to quote cover about 140 years. This period comprises two and one-half cycles only. Although the period embraced by the data is sufficient to

[1] Cf. A. Spiethoff, "Krisen," (*Handwörterbuch der Staatswissenschaften*, 4th edition).

[2] Cf. Ernle, *English Farming Past and Present* (London, 1922), and G. F. Warren and F. A. Pearson, *The Agricultural Situation* (New York, 1924).

decide the question of the existence of long waves, it is not enough to enable us to assert beyond doubt the cyclical character of those waves. Nevertheless we believe that the available data are sufficient to declare this cyclical character to be very probable.

We are led to this conclusion not only by the consideration of the factual material, but also because the objections to the assumption of long cyclical waves are very weak.

It has been objected that long waves lack the regularity which business cycles display. But this is wrong. If one defines "regularity" as repetition in regular time-intervals, then long waves possess this characteristic as much as the intermediate ones. A strict periodicity in social and economic phenomena does not exist at all — neither in the long nor in the intermediate waves. The length of the latter fluctuates at least between 7 and 11 years, i.e., 57 per cent. The length of the long cycles fluctuates between 48 and 60 years, i.e., 25 per cent only.

If regularity is understood to be the similarity and simultaneity of the fluctuations of different series, then it is present to the same degree in the long as in the intermediate waves.

If, finally, regularity is understood to consist in the fact that the intermediate waves are an international phenomenon, then the long waves do not differ from the latter in this respect either.

Consequently, there is no less regularity in the long waves than in the intermediate ones, and if we want to designate the latter as cyclical, we are bound not to deny this characterization to the former.

It has been pointed out [by other critics] that the long waves — as distinct from the intermediate ones which come from causes within the capitalistic system — are conditioned by casual, extra-economic circumstances and events, such as (1) changes in technique, (2) wars and revolutions, (3) the assimilation of new countries into the world economy, and (4) fluctuations in gold production.

These considerations are important. But they, too, are not valid. Their weakness lies in the fact that they reverse the causal connections and take the consequence to be the cause, or see an accident where we have really to deal with a law governing the events. In the preceding paragraphs, we have deliberately, though briefly, considered the establishment of some empirical rules for the movement of long waves. These regularities help us now to evaluate correctly the objections just mentioned.

1. *Changes in technique* have without doubt a very potent influence on the course of capitalistic development. But nobody has proved them to have an accidental and external origin.

Changes in the technique of production presume (1) that the relevant scientific-technical discoveries and inventions have been made, and (2) that it is *economically* possible to use them. It would be an obvious mistake to deny the creative element in scientific-technical discoveries and inventions. But from an objective viewpoint, a still greater error would occur if one believed that the direction and intensity of those discoveries and inventions were entirely accidental; it is much more probable that such direction and intensity are a function of the necessities of real life and of the preceding development of science and technique.[1]

Scientific-technical inventions in themselves, however, are insufficient to bring about a real change in the technique of production. They can remain ineffective so long as economic conditions favorable to their application are absent. This is shown by the example of the scientific-technical inventions of the seventeenth and eighteenth centuries which were used on a large scale only during the industrial revolution at the close of the eighteenth century. If this be true, then the assumption that changes in technique are of a random character and do not in fact spring from economic necessities loses much of its weight. We have seen before that the development of technique itself is part of the rhythm of the long waves.

2. *Wars and revolutions* also influence the course of economic development very strongly. But wars and revolutions do not come out of a clear sky, and they are not caused by arbitrary acts of individual personalities. They originate from real, especially economic, circumstances. The assumption that wars and revolutions acting

[1] One of the best and most compelling arguments for the assumption that scientific and technical inventions and discoveries are not made accidentally but are intimately connected with the needs of practical life is given by the numerous cases in which the same inventions and discoveries are made at the same time at different places and entirely independently of one another. Cf. the long list of such cases in W. F. Ogburn, *Social Change* (New York, 1924), p. 90. Cf. also Dannemann, *Die Naturwissenschaften in ihrer Entwickelung und in ihrem Zusammenhange* (Leipzig, 1923).

from the outside cause long waves evokes the question as to why they themselves follow each other with regularity and solely during the upswing of long waves. Much more probable is the assumption that wars originate in the acceleration of the pace and the increased tension of economic life, in the heightened economic struggle for markets and raw materials, and that social shocks happen most easily under the pressure of new economic forces.

Wars and revolutions, therefore, can also be fitted into the rhythm of the long waves and do not prove to be the forces from which these movements originate, but rather to be one of their symptoms. But once they have occurred, they naturally exercise a potent influence on the pace and direction of economic dynamics.

3. As regards the *opening-up of new countries for the world economy*, it seems to be quite obvious that this cannot be considered an outside factor which will satisfactorily explain the origin of long waves. The United States have been known for a relatively very long time; for some reason or other they begin to be entangled in the world economy on a major scale only from the middle of the nineteenth century. Likewise, the Argentine and Canada, Australia and New Zealand, were discovered long before the end of the nineteenth century, although they begin to be entwined in the world economy to a significant extent only with the coming of the 1890's. It is perfectly clear historically that, in the capitalistic economic system, new regions are opened for commerce during those periods in which the desire of old countries for new markets and new sources of raw materials becomes more urgent than theretofore. It is equally apparent that the limits of this expansion of the world economy are determined by the degree of this urgency. If this be true, then the opening of new countries does not provoke the upswing of a long wave. On the contrary, a new upswing makes the exploitation of new countries, new markets, and new sources of raw materials necessary and possible, in that it accelerates the pace of capitalistic economic development.

4. There remains the question whether the *discovery of new gold mines*, the *increase in gold production*, and a consequent *increase in the gold stock* can be regarded as a casual, outside factor causing the long waves.

An increase in gold production leads ultimately to a rise in prices and to a quickening in the tempo of economic life. But this does not mean that the changes in gold production are of a casual, outside character and that the waves in prices and in economic life are likewise caused by chance. We consider this to be not only unproved but positively wrong. This contention originates from the belief, first, that the discovery of gold mines and the perfection of the technique of gold production are accidental and, secondly, that every discovery of new gold mines and of technical inventions in the sphere of gold production brings about an increase in the latter. However great may be the creative element in these technical inventions and the significance of chance in these discoveries, yet they are not entirely accidental. Still less accidental — and this is the main point — are the fluctuations in gold production itself. These fluctuations are by no means simply a function of the activity of inventors and of the discoveries of new gold mines. On the contrary, the intensity of inventors' and explorers' activity and the application of technical improvement in the sphere of gold production, as well as the resulting increase of the latter, depend upon other, more general causes. The dependence of gold production upon technical inventions and discoveries of new gold mines is only secondary and derived.

Although gold is a generally recognized embodiment of value and, therefore, is generally desired, it is only a commodity. And like every commodity it has a cost of production. But if this be true, then gold production — even in newly discovered mines — can increase significantly only if it becomes more profitable, i.e., if the relation of the value of the gold itself to its cost of production (and this is ultimately the prices of other commodities) becomes more favorable. If this relation is unfavorable, even gold mines the richness of which is by no means yet exhausted may be shut down; if it is favorable, on the other hand, even relatively poor mines will be exploited.

When is the relation of the value of gold to that of other commodities most favorable for gold production? We know that commodity prices reach their lowest level toward the end of a long wave. This means that at this time gold has its highest purchasing power, and gold production becomes most favorable. This can be illustrated by the figures in Table 2.

Gold production, as can be seen from these figures, becomes more profitable as we approach a low point in the price level and a high point in the purchasing power of gold (1895 and the following years).

TABLE 2. — SELECTED STATISTICS OF GOLD MINING IN THE TRANSVAAL, 1890–1913*

Year	Cost of production	Profit
	Per ton of gold ore	
1890........	42 *sh.* 2 *d.*	7 *sh.* 2 *d.*
1895........	33 *sh.* 5 *d.*	11 *sh.* 11 *d.*
1899........	28 *sh.* 0 *d.*	14 *sh.* 3 *d.*
1903........	24 *sh.* 9 *d.*	14 *sh.* 11 *d.*
1906........	22 *sh.* 2 *d.*	11 *sh.* 6 *d.*
1913........	17 *sh.* 11 *d.*	9 *sh.* 10 *d.*

* Cf. W. A. Berridge, "The World's Gold Supply," this REVIEW, II (1920), p. 184.

It is clear, furthermore, that the stimulus to increased gold production necessarily becomes stronger the further a long wave declines. We, therefore, can suppose theoretically that gold production must in general increase most markedly when the wave falls most sharply, and vice versa.

In reality, however, the connection is not as simple as this but becomes more complicated, mainly just because of the effect of the changes in the technique of gold production and the discovery of new mines. It seems to us, indeed, that even improvements in technique and new gold discoveries obey the same fundamental law as does gold production itself, with more or less regularity in timing. Improvements in the technique of gold production and the discovery of new gold mines actually do bring about a lowering in the cost of production of gold; they influence the relation of these costs to the value of gold, and consequently the extent of gold production. But then it is obvious that exactly at the time when the relation of the value of gold to its cost becomes more unfavorable than theretofore, the need for technical improvements in gold mining and for the discovery of new mines necessarily becomes more urgent and thus stimulates research in this field. There is, of course, a time-lag, until this urgent necessity, though already recognized, leads to positive success. In reality, therefore, gold discoveries and technical improvements in gold mining will reach their peak only when the long wave has already passed its peak, i.e., perhaps in the middle of the downswing. The available facts confirm this supposition.[1] In the period after the 1870's, the following gold discoveries were made: 1881 in Alaska, 1884 in the Transvaal, 1887 in West Australia, 1890 in Colorado, 1894 in Mexico, 1896 in the Klondike. The inventions in the field of gold-mining technique, and especially the most important ones of this period (the inventions for the treatment of ore), were also made during the 1880's, as is well known.

Gold discoveries and technical improvements, if they occur, will naturally influence gold production. They can have the effect that the increase in gold production takes place somewhat earlier than at the end of the downswing of the long wave. They also can assist the expansion of gold production, once that limit is reached. This is precisely what happens in reality. Especially after the decline in the 1870's, a persistent, though admittedly slender, increase in gold production begins about the year 1883;[2] whereas, in spite of the disturbing influences of discoveries and inventions, the upswing really begins only after gold has reached its greatest purchasing power; and the increased production is due not only to the newly discovered gold fields but in a considerable degree also to the old ones. This is illustrated by the figures in Table 3.

[1] Berridge, *loc. cit.*, p. 181.
[2] Cf. *Statistical Abstract of the United States*, 1922, pp. 708–09.

TABLE 3. — GOLD PRODUCTION, 1890–1900
(*Unit: thousand ounces*)

	World total	Transvaal	United States	Australia	Russia	Canada	Mexico	India
1890......	5,749	440	1,589	1,588	1,135	65	737	9
1895......	9,615	2,017	2,255	2,356	1,388	101	290	230
1900......	14,838	3,638	3,437	4,461	1,072	1,029	411	412

Source: Berridge, *loc. cit.*, p. 182

From the foregoing one may conclude, it seems to us, that gold production, even though its increase can be a condition for an advance in commodity prices and for a general upswing in economic activity, is yet subordinate to the rhythm of the long waves and consequently cannot be regarded as a causal and random factor that brings about these movements from the outside.

XIII. CONCLUSIONS

The objections to the regular cyclical character of the long waves, therefore, seem to us to be unconvincing.

In view of this circumstance and considering also the positive reasons developed above, we think that, *on the basis of the available data, the existence of long waves of cyclical character is very probable.*

At the same time, we believe ourselves justified in saying that the long waves, if existent at all, are a very important and essential factor in economic development, a factor the effects of which can be found in all the principal fields of social and economic life.

Even granting the existence of long waves, one is, of course, not justified in believing that economic dynamics consists only in fluctuations around a certain level. The course of economic activity represents beyond doubt a process of development, but this development obviously proceeds not only through intermediate waves but also through long ones. The problem of economic development *in toto* cannot be discussed here.

In asserting the existence of long waves and in denying that they arise out of random causes, we are also of the opinion that the long waves arise out of causes which are inherent in the essence of the capitalistic economy. This naturally leads to the question as to the nature of these causes. We are fully aware of the difficulty and great importance of this question; but in the preceding sketch we had no intention of laying the foundations for an appropriate theory of long waves.[1]

[1] I arrived at the hypothesis concerning the existence of long waves in the years 1919–21. Without going into a special analysis, I formulated my general thesis for the first time shortly thereafter in my study, *The World Economy and Economic Fluctuations in the War and Post-War Period* (*Mirovoje chozjajstvo i jego konjunktury vo vremja i posle vojny* [Moscow, 1922]). During the winter and spring of 1925, I wrote a special study on "Long Waves in Economic Life" ("Bol'schije cykly konjunktury"), which was published in the volume of the Institute for Business Cycle Research, *Problems of Economic Fluctuations* (*Voprosy konjunktury*, Vol. 1). Only at the beginning of 1926 did I become acquainted with S. de Wolff's article "Prosperitäts- und Depressionsperioden," *Der lebendige Marxismus, Festgabe zum 70. Geburtstage von Karl Kautsky.* De Wolff in many points reaches the same result as I do. The works of J. van Gelderns, which de Wolff cites and which have evidently been published only in Dutch, are unknown to me.

But in the 1985 translation there are 10 more pages describing a theory and model that seems essentially correct.
JF 8/27/85

Forrester's urban dynamics model has had considerable influence on academicians and policymakers. However, the external validity of the model remains unsettled. Four large American cities which meet the initial assumptions of the model are used to test the model's predictive accuracy. Developments in these cities during a 20-year period are compared with the model's predictions. These comparisons indicate substantial deviations from predicted patterns of change. In general, the predictions of the model appear to be too pessimistic: The model fairly consistently understates developmental change and grossly overstates urban decay. The causes and implications of the model's poor performance are discussed.

THE URBAN DYNAMICS MODEL
A Validation Study

YUNG-MEI TSAI
Texas Tech University
OTOMAR BARTOS
University of Colorado
LEE SIGELMAN
University of Kentucky

It ranks as a gross understatement to say that Jay Forrester's (1969) urban dynamics simulation model has generated considerable interest since it was published over a decade ago. In the scholarly community, where citations to one's work are often used as a measure of influence, the average piece of research is cited only about three times over the course of its lifetime (Cawkell, 1968). By contrast, the *Social Sciences Citation Index* contains several hundred references to *Urban Dynamics*, the book in which Forrester presented his model. But not only has the urban dynamics model had a vast academic influence. By all accounts, it has also been used by high-level planners and decision makers to help

URBAN AFFAIRS QUARTERLY, Vol. 17 No. 2, December 1981 195-218

rethink existing urban policies and formulate new ones (see, e.g., Chen, 1972).

Despite the influence that the urban dynamics model has had both inside and outside of academia, one fundamental question about the model has not yet been answered: How accurately does it simulate actual urban processes? This is a question about the external validity of the urban dynamics model. It is explicitly an empirical question, and as such demands an empirical answer. It is also, for reasons to be articulated below, an extremely important question. In this study, we attempt to answer this question. For purposes of this analysis, we have gathered data pertaining to developmental processes in four large American cities which meet the initial assumptions of the urban dynamics model. These data have been compared with simulation data generated by the urban dynamics model, and the degree of compatibility between the simulation results and what actually happened in the four urban areas has been assessed. Before presenting our empirical results, however, we need to offer a brief description of the urban dynamics model and outline our views of the external validity issue as it applies to simulation models in general.

THE SIMULATION MODEL

The urban dynamics model is one in a series of large-scale social system simulation models that Forrester has developed. Adapting a model that he had originally used to represent growth of industries (Forrester, 1961) and would later use to analyze the dynamics of the world system (Forrester, 1971), Forrester set out to simulate the life cycle of urban areas.

In the model, an urban area is defined as "a system of interacting *industries, housing* and *people*" that operates within fixed physical boundaries (Forrester, 1961: 1). Each of these sectors is composed of three subsectors, corresponding to the growth, maturity, and decay stages of the urban life cycle: Industry is "new," "mature," or "declining"; housing is "premium," "worker," or "underemployed"; and members of the work force are "managerial-professional," "labor," or "underemployed." Growth begins on an "almost empty land area" (Forrester, 1969: 2) and continues in a series of complex interactions among the industrial, housing, and labor sectors. The sheer size and complexity of the urban dynamics model preclude full description of these interactions here, for the model contains scores of equations describing the relationships among a like number of variables. In general terms, the simulation focuses on generic processes associated with urbanism and urbanization rather than developments within any particular urban area. In Forrester's (1969: 14) words, what is modeled is "the general class of system rather than a specific system. Here, this means a model to represent the central processes common to all urban areas rather than to represent those of a specific area." The simulation produces trend lines predicting growth and decay patterns of and within the various urban sectors for periods running up to 250 years. The equilibrium model, a variant of the basic growth model, "is started with the equilibrium conditions that are reached at the end of the growth life cycle. The equilibrium model is used to explore how various changes in policy would cause the condition of the urban area to be altered over the following fifty years" (Forrester, 1969: 2).

In both the growth and equilibrium models, predicted patterns are based on a complex set of feedback processes which operate both internally and externally. As a simplified example of internal feedback, population growth can cause an increase in housing and jobs, which in turn can generate further population growth. However, if the expansions of the job and housing sectors do not meet the demands of the increased population, the effect may be to depress further population growth. External feedback processes relate to the attractiveness the urban area holds for those in the surrounding environment. As another simplified example, if more jobs are available in the urban area than in the surrounding area, people will move in. Population will increase until the employment situation worsens, at which point people will begin moving out. An equilibrium state will be reached when the land in the urban area is fully occupied; from there, further growth will be possible only if old housing or industrial sites are demolished to accommodate new development.

Forrester (1969: 115) designed the urban dynamics model to serve as a learning tool which could enhance the capacity of scholars, planners, and policymakers to envision the often counter-intuitive consequences of complex urban processes. He emphasized the model's potential utility for policy testing, for tracing out the probable outcomes of alternative policies and programs. He did caution that the model probably needed further refinement before policy recommendations emanating from it should be taken seriously, but he himself demonstrated very little reluctance to use the model for recommending policy changes. Many critics consider his policy prescriptions reactionary; in Catanese's (1972: 248) words, Forrester's recommendations reflect the "somewhat Darwinist purist ideology" that underlies the urban dynamics model. To say the very least, Forrester's recommendations for improving the urban situation have proven highly controversial (see, e.g., Ingram, 1970).

This brings us to the rationale for the present study. All modeling begins with abstraction and simplification, which are "the name of the game" in modeling rather than avoidable inconveniences (Raser, 1969: 11). Hence, any model—even one as complex as the urban dynamics model—will not represent a natural system with perfect fidelity. How closely the simulation results need to reflect reality depends upon the purposes of the simulation. If one is using a flight simulator to train airplane pilots, one would hardly be satisfied if the procedures that produced good results in the simulator produced crashes in actual flight. For other purposes, rough correspondence to the natural system may suffice.

What are the purposes of the urban dynamics model? Forrester and others who operate in the system dynamics tradition emphasize the "insight-generating capacity" of their models—the ability of the models to improve one's mental image of the process being modeled and to point out previously unrealized aspects of the system's operation. According to Randers (1980: 18), two common objectives of the approach are to increase understanding of some observed phenomenon and to establish the general consequences of different options available at a decision point. Modelers often draw a sharp distinction between these types of objectives and that of predictive accuracy, which holds sway throughout much of social science. Randers (1980: 19), in fact, ranks nine different objectives of the system dynamics approach, beginning with insight-generating capacity, and concludes that predictive accuracy is the *least* important; ranking prediction so low, he contends, "is rational when the ultimate objective is increased understanding, both of the past and of the likely consequences of future actions" (Randers, 1980: 19).

This is not the place to launch a detailed critique of the epistemological assumptions underlying the urban dynamics model. However, from our perspective, if a model is to be used as a serious tool of understanding or, perhaps even more crucially, if it is to be used as a basis for social engineering, then some acceptable connection must be established between the probable outcomes it predicts and the actual outcomes generated by the natural systems being modeled. Insights based on a model which fails this empirical test are apt to be badly flawed, and policies based on these insights are likely to be fundamentally defective. Very clearly, then, the present study is one that system modelers will not see as very important, because they are not very concerned about the issue of predictive accuracy. Very clearly, too, the present study addresses as issue that we consider to be of the essence: the external validity of the model.

System modelers often speak of the need to "verify" a model, by which they essentially mean establishing that the model is internally consistent and, more subjectively, "makes sense." In this regard, the urban dynamics model is hard to evaluate. Forrester is well-known for his insistence that complex systems behave in counter-intuitive ways, and this assumption plays havoc with the requirement that a model "make sense." Moreover, Forrester is not an urbanist, and although he repeatedly refers to the simulation model as a "theory," it is not grounded in any coherent body of theory as that term is customarily employed. Indeed, Forrester (1969: 113-114) explicitly dismisses existing literature on urban areas as irrelevant to the urban dynamics model. The implicit theory that guides the urban dynamics model strikes some critics as "primitive" (Gray et al., 1972), and it has been demonstrated that when certain propositions upon which urbanists have achieved a fair measure of consensus are substituted into the model, simulation runs produce substantially different results than those reported by Forrester (Stonebraker, 1972).

The possibility of resolving debates concerning verification depends ultimately on the world views of those who are involved in the controversy: What "makes sense" to one person may or may not to another. In contrast to verification, external validity is more of an empirical issue. "Responses emanating from the verified model are compared with available information or data regarding the system being modeled. . . . In general terms, the process of validation concerns the corroboration of the model with the system. If the corroboration is sufficient, the model is said to be validated" (Rausser and Johnson, 1975: 117). Although Forrester (1969: 113) lays no great stress on validation, he does recognize the need to compare the model's operation "with the real systems it should represent." But to this point, the external validity of Forrester's model has received virtually no attention, in spite of the considerable controversy that has surrounded the model.

TESTING THE MODEL'S VALIDITY

Designing a fair test of the model's validity is something of a challenge, because it is no simple matter to decide how the urban dynamics model maps into the real world. In the first place, the model's assumptions are so restrictive that it is difficult to find very many cities to which it legitimately can be compared. Operationalizing the key variables in the model is also difficult, for Forrester is not always clear about what he means by some of the concepts he uses. In choosing measures, we have been guided to the extent possible by Forrester's own words. Where these are not sufficiently specific, we hve consciously tried to make methodological choices that seem consistent with the spirit of his presentation of the model.

SELECTING THE TEST CITIES

The urban dynamics model requires that the land area of a city be fixed at 100,000 acres (approximately 150 square miles) throughout its history. We know of only four large American cities which come close to satisfying this condition: Chicago, Detroit, New Orleans, and Philadelphia. This in itself says something about the generalizability of the model, but since we wanted to test it in a form as close as possible to that given by Forrester, we have restricted our analysis to these four cities.

CHOOSING A TIME FRAME

The equilibrium model runs for a duration of 50 years. However, we have been influenced by Smith and Sage's (1973: 546) argument that the urban dynamics model is most useful for relatively short-range evaluations, running for only 10 or 20 years. Over the long run, technologies can change so fundamentally in ways that could not have been anticipated that model-generated predictions prove to be radically incorrect. Accordingly, it does not seem reasonable to expect the model to produce highly accurate predictions over a 50-year time span, let alone over 250 years. Focusing on a shorter period of time should effectively hold levels of technology constant. Following Smith and Sage, we decided that the most appropriate course of action would be to evaluate the model's predictions over a 20-year span. Forrester provides no indication that the model is intended to simulate urban processes during any particular period of modern history. However, since

Forrester designed the model during the middle and late 1960s, we decided to use the immediately preceding 20-year period (1940-1960) as the temporal framework for the analysis.

OPERATIONALIZING THE TEST VARIABLES

Our general strategy was to leave the original model as intact as possible, because it is the model, itself, which Forrester and others use as the basis for their diagnosis of urban problems and their prescription of policy changes. Therefore, we decided to treat as test variables only the key factors in the model. According to Forrester (1969: 17), "The changes in housing, population, and industry are the central processes involved in growth and stagnation." This meant that we had to determine where Chicago, Detroit, New Orleans, and Philadelphia stood during the 1940-1960 period with respect to: the percentages of the labor force in the "managerial-professional," "labor," and "underemployed" subsectors; the proportions of "premium," "worker," and "underemployed" housing; the mix of "new," "mature," and "declining" industries; and the total size of each of the three sectors. The primary data sources that we used to make these determinations were the *U.S. Census of Population, U.S. Census of Housing, U.S. City and County Data Book,* and *U.S. Census of Business.* Because the censuses of population and housing are conducted at decennial intervals, our base points for comparing the actual development of these four cities with the simulation results were 1950 and 1960.[1] The 1940 data were used as input for starting up the simulation model. That is, in order to compare the simulation model's predictions with the actual developments in the three sectors of the four cities, we needed to start the comparisons from a common baseline. Thus, for example, the 1940 data on Chicago's job, housing, and industry sectors were substituted into the model. All remaining variables and parameters were left unchanged from the original specifications of the equilibrium model, which was then allowed to run for a 20-year period. Since Chicago and the simulated urban area had begun from a common starting point on the three test sectors, actual developments in Chicago could validly be compared with developments simulated by the model. This process was repeated for Detroit, New Orleans, and Philadelphia. All remaining variables and parameters in the model were left unchanged from the original specifications.

With respect to the *labor force* sector, we took as a point of departure Forrester's (1969: 19) statement that "labor is skilled labor fully participating in the urban economy. Underemployed workers include, in addition to the unemployed and unemployable, people in unskilled jobs, those in marginal economic activity, and those not seeking employment who might work in a period of intense economic activity." We thus used *U.S. Census* classifications to arrive at the following definitions of Forrester's three main labor subsectors. The "managerial-professional" sector included: professional workers and proprietors, managers, and officials, excluding farm. The "labor" sector included: semiprofessional workers: farmers and farm managers; clerical, sales, and kindred workers; craftsmen, foremen, and kindred workers; domestic service workers; protective service workers; and service workers, excluding domestic and protective. The "underemployed" sector included: farm laborers and foremen; laborers except farm and mine; and occupation not reported.

To define the three *housing* subsectors, we again started from Forrester's (1969: 20) discussion: "A corresponding shift in population density defines the principal distinctions between categories of housing units . . . as shown below:

Premium-Housing Population Density	=	3 persons/housing unit
Worker-Housing Population Density	=	6 persons/housing unit
Underemployed-Housing Population Density	=	12 persons/housing unit

These housing densities represent the normal situation if housing and population are in economic balance." We translated this statement into defining a housing unit as "premium" if three or fewer persons live in it, as "worker" if it was occupied by four to six persons, and as "underemployed" if seven or more persons lived in it. It would seem that, if anything, these operational definitions of the three sectors would tend to inflate the underemployed housing category.

To determine the three *industry* subsectors, we used two separate methods which produced very similar results. For Detroit, Philadelphia, and Chicago, we had information about average weekly pay for given types of industry (such as "wholesale") and given occupational categories as of 1940. Thus, for these cities in 1940, we defined the three sectors on the basis of what we call the "personal-mix" method. For New Orleans in 1940 and for all cities in the remaining years, we used what we shall call the "normal distribution" method.

For the "personnel-mix" method, we used the following statement by Forrester (1969: 20) as our point of departure: "The table . . . below defines a productive unit in terms of the employment mix under normal circumstances . . .

TABLE 1
Information Used in "Personnel-Mix" Calculations

	New Enterprise	Mature Business	Declining Industry
	A. Employment Mix		
Managerial-Professional	4	2	1
Labor	20	15	10
Underemployed	10	7.5	5
	B. Employment Mix, In Percentages		
Managerial-Professional	11.8%	8.2%	6.3%
Labor	58.4%	61.2%	62.4%
Underemployed	29.8%	37.6%	31.3%
	100.0%	100.0%	100.0%
	C. Range of Pay		
	$38.43 and above	$36.50-$38.42	$36.41 and below

The manner is which we translated this statement into a decision whether an actual industry is new, mature, or declining is rather involved. In the first place, we noted that the information in Table 1A can be expressed in percentages, as in Table 1B.

Moreover, for some cities in 1940, we had available information about the average pay for each labor sector for a given industry. For example, we found in the *U.S. Census of Business* that the average weekly pay in 1940 Chicago's wholesale industry was $101.23 in the managerial-professional sector, $35.00 in the labor sector, and $24.50 in the underemployed sector. We then used the percentages derived from Forrester's table to arrive at the average weekly pay in each sector of *industry*. For our example of 1940 Chicago's wholesale industry we computed:

New enterprise: Average pay was $39.69 since

$$(\$101.25 \times 11.8\%) + (\$35 \times 58.5\%) + (\$24.5 \times 29.8\%) = \$39.69$$

Mature business: Average pay was $37.15 since

$$(\$101.23 \times 8.16\%) + (\$35 \times 61.2\%) + (\$24.5 \times 30.64\%) = \$37.15$$

Declining industry: Average pay was \$35.85 since

$$(\$101.25 \times 6.25\%) + (\$35 \times 62.5\%) + (\$24.5 \times 31.25\%) = \$35.85.$$

Now it remained to determine the boundaries around these means in order to be able to classify a given industry as "new," "mature," or "declining." We decided to close the midpoint between the means as the boundary, thus obtaining the final result for our example summarized in Table 1C.

As already mentioned, we were unable to obtain average pay by industry for each city and each decade. Where we lacked the necessary information, we adopted the "normal distribution" procedure. We computed the mean pay for a given city at a given time, as well as the standard deviation around the mean. In such cases, we defined the various sectors as follows: "new" = more than one standard deviation above the mean; "mature" = within one standard deviation above and below the mean; and "declining" = more than one standard deviation below the mean.

In order to determine whether the difference in method affected our results, we applied both methods to the wholesale industry in the three cities for which the relevant data were available. We see in Table 2 that the two methods lead to virtually identical percentage distributions among the three industry subsectors.

The labor force classifications are fairly straightforward, because the occupational categories used by the Bureau of the Census are highly compatible with Forrester's subsectors. The housing subsectors employed in the urban dynamics model are also relatively easy to fit, for Forrester provides rough operational definitions of what "premium," "worker," and "underemployed" housing entail. The industrial subsectors, however, are much more challenging to handle empirically. Because any validation test can be no better than the measures used to carry it out, readers are strongly urged to familiarize themselves with our operationalization procedures.

OBSERVING THE PROCESS

Data on the labor, housing, and industry sectors and subsectors in Chicago, Detroit, New Orleans, and Philadelphia in 1940 are presented in Table 3. Once these basic input data were obtained, they were substituted into the equilibrium model. From this starting point, the model was allowed to generate 10- and 20-year predictions of the

TABLE 2
Comparison of Two Measurement Methods

City	Sector	Method of "Personnel-Mix"	"Method of Normal Distribution
Detroit	New Industry	25%	26%
	Mature Business	64%	63%
	Declining Industry	11%	11%
Philadelphia	New Industry	20%	19%
	Mature Business	70%	72%
	Declining Industry	10%	9%
Chicago	New Industry	22%	22%
	Mature Business	71%	71%
	Declining Industry	7%	7%

development of the three sectors and their nine subsectors in each of the four cities, and these predictions were compared with the actual 1950 and 1960 data. In discussing these comparisons, let us first consider the overall growth of the three sectors, and then turn to the more critical question of the growth of the nine subsectors.

Sectoral Change. The rows labeled "Total" in Tables 4, 5, and 6 show the predicted (i.e., simulation-generated) and actual sizes of the labor, housing, and industry sectors in the four cities as of 1950 and 1960. The model's 10-year predictions of labor force size consistently underestimated the actual work force by approximately 15%. The 20-year labor force predictions, however, erred in the opposite direction. For three of the cities, the model's 1960 overestimations were appreciable, and one of these overestimates (Detroit) was quite badly in error. The model's 10-year predictions of the size of the housing market displayed a consistent tendency toward overestimation, although (with the possible exception of Detroit) these errors were not of truly major magnitude. The 20-year housing volume predictions were less consistent: two

TABLE 3

The Starting Point of the Validation: The Four Cities in 1940

Sector	Category	City: Chicago	Detroit	New Orleans	Philadelphia
LABOR FORCE	Managerial-Professional	14%	11%	13%	12%
	Labor	77%	68%	60%	63%
	Underemployed	9%	21%	27%	25%
	TOTAL	100% (1,397,988)	100% (738,646)	100% (220,241)	100% (878,153)
HOUSING	Premium-Housing	58%	52%	56%	53%
	Worker	36%	40%	35%	38%
	Underemployed	6%	8%	9%	8%
	TOTAL	100% (949,744)	100% (425,527)	100% (133,040)	100% (506,980)
INDUSTRY	New Enterprise	7%	9%	8%	3%
	Mature Business	80%	68%	45%	77%
	Declining Industry	13%	23%	47%	19%
	TOTAL	100% (83,469)	100% (22,147)	100% (10,737)	100% (48,628)

TABLE 4
Labor Force in the Selected Cities: Predicted and Actual

City	Year	Type of Labor	Predicted %	Predicted Number	Actual %	Actual Number
Chicago	1950	Mang.-Profess.	15%		17%	
		Labor	67%		75%	
		Underemployed	18%[a]		8%	
		Total	100%	1,421,375	100%	1,694,206
	1960	Manag.-Profess.	8%		15%	
		Labor	60%		80%	
		Underemployed	32%		5%	
		Total	100%	1,634,497	100%	1,635,482
Detroit	1950	Manag.-Profess.	14%		16%	
		Labor	65%		77%	
		Underemployed	21%		7%	
		Total	100%	673,671	100%	816,549
	1960	Manag.-Profess.	10%		14%	
		Labor	57%		75%	
		Underemployed	33%		11%	
		Total	100%	902,371	100%	612,295
New Orleans	1950	Manag.-Profess.	14%		20%	
		Labor	66%		69%	
		Underemployed	19%		11%	
		Total	100%	203,367	100%	229,772
	1960	Manag.-Profess.	11%		17%	
		Labor	65%		67%	
		Underemployed	23%		15%	
		Total	100%	301,006	100%	223,744
Philadelphia	1950	Manag.-Profess.	15%		16%	
		Labor	67%		75%	
		Underemployed	19%		9%	
		Total	100%	808,070	100%	883,034
	1960	Manag.-Profess.	9%		16%	
		Labor	60%		73%	
		Underemployed	22%		11%	
		Total	100%	942,894	100%	816,611

a. Underlined percentages indicate that the predicted percentage is higher than the actual.

(Chicago and Philadelphia) were almost on target, but the other two (Detroit and New Orleans) were badly overestimated, running as much as 60% above the actual figure. By far the model's strongest performance came from its estimates of the number of industrial establishments. There was some tendency to underestimate the size of the industry sector,

TABLE 5
Housing in the Selected Cities: Predicted and Actual

			Predicted		Actual	
City	Year	Type of Housing	%	Number	%	Number
Chicago	1950	Premium	25%		64%	
		Worker	62%[a]		32%	
		Underemployed	13%		4%	
		Total	100%	1,132,051	100%	1,106,008
	1960	Premium	12%		67%	
		Worker	61%		29%	
		Underemployed	28%		4%	
		Total	100%	1,148,074	100%	1,212,229
Detroit	1950	Premium	21%		59%	
		Worker	67%		36%	
		Underemployed	12%		5%	
		Total	100%	678,579	100%	552,319
	1960	Premium	17%		66%	
		Worker	59%		29%	
		Underemployed	25%		5%	
		Total	100%	740,225	100%	439,403
New Orleans	1950	Premium	27%		63%	
		Worker	63%		32%	
		Underemployed	10%		6%	
		Total	100%	191,407	100%	173,590
	1960	Premium	21%		64%	
		Worker	67%		30%	
		Underemployed	12%		6%	
		Total	100%	295,509	100%	201,738
Philadelphia	1950	Premium	26%		59%	
		Worker	60%		36%	
		Underemployed	13%		5%	
		Total	100%	638,348	100%	590,375
	1960	Premium	15%		64%	
		Worker	59%		31%	
		Underemployed	26%		5%	
		Total	100%	663,688	100%	648,968

a. Underlined percentages indicate that the predicted percentage is higher than the actual.

but almost all of the deviations varied from the actual figure by only a few percentage points. In sum, Tables 4, 5, and 6 show that the urban dynamics model tended to:

(1) underestimate the 10-year growth of the labor force, but overestimate its 20-year growth

TABLE 6
Industry in the Selected Cities: Predicted and Actual

City	Year	Type of Industry	Predict %	Predict Number	Actual %	Actual Number
Chicago	1948	New Enterprise	8%		9%	
		Mature Business	51%		67%	
		Declining Indus.	41%[a]		24%	
		Total	100%	84,768	100%	82,944
	1958	New Enterprise	5%		8%	
		Mature Business	28%		84%	
		Declining Indus.	67%		9%	
		Total	100%	77,122	100%	80,673
Detroit	1948	New Enterprise	28%		8%	
		Mature Business	47%		55%	
		Declining Indus.	24%		38%	
		Total	100%	29,320	100%	33,146
	1958	New Enterprise	21%		10%	
		Mature Business	41%		67%	
		Declining Indus.	38%		23%	
		Total	100%	31,923	100%	35,748
New Orleans	1948	New Enterprise	18%		9%	
		Mature Business	37%		67%	
		Declining Indus.	45%		24%	
		Total	100%	11,227	100%	10,763
	1958	New Enterprise	24%		15%	
		Mature Business	32%		51%	
		Declining Indus.	44%		34%	
		Total	100%	11,987	100%	12,047
Philadelphia	1948	New Enterprise	6%		4%	
		Mature Business	48%		71%	
		Declining Indus.	45%		24%	
		Total	100%	49,555	100%	50,637
	1958	New Enterprise	5%		10%	
		Mature Business	28%		67%	
		Declining Indus.	67%		23%	
		Total	100%	44,062	100%	47,775

a. Underlined percentages indicate that the predicted percentage is higher than the actual.

(2) overestimate the growth of the housing sector, especially after 20 years
(3) slightly underestimate the total number of industrial units.

Subsectoral Change. While comparisons of predicted and actual overall sectoral change are interesting, it is critical to determine how well the model performs on a subsector-by-subsector basis. This is so because of the value judgments that underlie the subsectors: "Underemployed" workers or housing units and "declining" industries indicate urban decay, while a city whose "managerial-professional" workforce, "premium" housing stock, and "new" business sector are expanding is seen as developing in a healthy fashion. There is no necessary connection in the model between the growth or shrinkage of a sector, as indicated by its overall size, and the development or decay of the sector, as indicated by the relative preponderance of one or the other of its subsectors. Thus, sectoral growth or decline in and of itself provides little basis for diagnosing urban problems and prospects or for prescribing remedies. Diagnosis and prescription depend fundamentally on the patterns of subsectoral growth and decline that are detected.

Returning to Tables 4, 5, and 6, we see that the model's 10-year predictions concerning the proportion of the labor force comprised of managerial-professional personnel tended to be somewhat underestimated; on the average, the 1950 predictions were only 85% as high as the actual figures for the four cities analyzed here. For 1960, this underestimation of managerial-professional workers became a good deal more severe. According to the simulation model, only about 10% of the workers should have been in this class by 1960, but across the four cities more than 15% of the workers actually were managers or professionals; in proportional terms, then, the model underestimated this subsector by a factor of more than one-third.

The model fared even worse with respect to premium housing. In each of the four cities, well over half of the housing units were premium as of 1950, indicating moderate improvement since 1940 (see Table 5). By contrast, the model predicted that premium housing would have occupied a much less prominent place in the housing sector in 1950 than it did in 1940. As a result, the model's premium housing predictions for 1950 were, on the average, only 40% as high as they should have been. Interestingly enough, the 20-year premium housing forecasts were even farther off. Whereas the model anticipated that the predicted diminution of premium housing would continue down to approximately 15% of the overall housing sector by 1960, in point of fact the percentage of premium housing units continued to grow in all four cities. Accordingly, the 1960 premium housing estimates were badly in error, averaging only 25% of the actual figures for the four cities.

The 1948 new industry estimates were also highly inaccurate, but here we found that the model tended to overestimate rather than underestimate. For example, the proportion of new businesses in Detroit in 1948 was predicted to be 28%, far above the 1940 base point of 9%; in fact, however, the new enterprise proportion in Detroit held virtually constant between 1940 and 1948. Across the four cities, the accuracy of the 1948 new industry predictions was very inconsistent: The Chicago and Philadelphia predictions were reasonable close, but the Detroit and New Orleans predictions were badly in error. Between 1948 and 1958, the model predicted that new businesses would occupy a smaller share of the industrial sector in three of the four cities, but in fact the new industry sector grew in three of the four cities. On the whole, however, the 1958 predictions ran much closer to the actual figures than did the 1948 estimates.

With respect to the three developmental subsectors, then, the model displayed a tendency to:

(1) underestimate both the 10- and particularly the 20-year shares of the labor sector occupied by managerial-professional personnel
(2) badly underestimate the share of the housing sector composed of premium units, particularly after 20 years
(3) overestimate the share of the new industry subsector, especially in 1948 and less so in 1958.

The model's performance at charting the dynamics of the three decay subsectors was somewhat more consistent across sectors, and thus can be readily summarized. First, the size of the underemployed labor sector was greatly overestimated for all four cities. The 1950 predictions of labor underemployment averaged more than twice the actual 1950 rates, and the 1960 predictions were even worse. For example, the model predicted 32% underemployment in Chicago by 1960; when measured, unemployment actually reached only 5%. On the whole, one would do well to divide the model's 1960 estimates of underemployment in the labor sector by a factor of three.

Overestimation of the decay subsector was even more pronounced for housing. The model predicted a substantial 10-year growth in underemployment housing in all four cities. But the measured 1950 figures reveal that the percentage of housing units classified as underemployed declined substantially between 1940 and 1950 in all four cities. The model then went on to predict, for three of the four cities, a vast expansion of the underemployed housing subsector over the next 10 years. In point of fact, however, rates of housing underemployment in all four cities held perfectly constant between 1950 and 1960. On the average, then, the predicted share of the underemployed housing subsector in 1960 was almost five times as high as the actual share.

Finally, the model's predictions concerning the declining industry subsector also tended to be wildly in error, especially by 1958. By our count, for example, 9% of Chicago's business enterprises were classified as declining in 1958, but the model predicted that 67% should have been so classified. This is an extreme example, but with the sole exception of Detroit in 1948, the simulation model yielded highly inflated estimates of business decline in all four cities.

Combining these findings about the decay subsectors, we can see that the model consistently tended to:

(1) badly overestimate the underemployed labor subsector after 10 years and grossly overestimate it after 20 years
(2) grossly overestimate the underemployment housing subsector at both 10- and 20-year intervals
(3) badly overestimate the declining industry subsector after 10 years and grossly overestimate it after 20 years.

DISCUSSION

In sum, our test revealed that the urban dynamics model does *not* accurately represent actual developments over a 20-year period in Chicago, Detroit, New Orleans, or Philadelphia—the four large American cities to which the model is most applicable. The model's predictions of the overall size of the labor, housing, and industry sectors were rather inconsistent: Some predictions, especially those pertaining to the industrial sector, were closely on target, but others contained sizable errors in either direction. More critically, the model performed very poorly with respect to the development and decay predictions it generated for each sector. Again with the exception of the industry sector, developmental change was consistently underestimated. More problematically, the model consistently and grossly overestimated the growth of the decay subsectors in all four cities.

Those who are familiar with the urban dynamics model, including Forrester himself, recognize that it paints a bleak picture of the future of large urban areas. It is no coincidence that later variants of Forrester's model, applied to the world system, have sometimes been labelled "models of doom," for the "life cycle" metaphor that is central to

Forrester's model is hardly what an optimist, convinced of the viability and innate vitality of the process being modeled, would have chosen. Aging in Forrester's frame of reference causes decay, and maturity gives way to stagnation.

The urban dynamics model is pessimistic—of this there can be no question. But what our validation test suggests is that the urban dynamics model is *too* pessimistic, for to a considerable degree, the pessimistic forecasts of urban decay that the model generates are empirically unwarranted. The model purports to represent generic urban processes. But what we left it virtually intact, substituting only starting levels in three key sectors for test purposes, we found that the estimates the model yielded were far removed from what had actually happened in the four test cities. Most important, the labor, housing, and industry sectors in all the four cities consistently exhibited far greater vitality than predicted by the simulation model. *Why* Forrester's model is so overly pessimistic is a question we are frankly unprepared to answer, but it is obvious that Forrester's view of cities and their dynamism is much less positive than seems to be warrented.

At this point, two additional questions must be addressed. First, how do these results compare with those that have been reported elsewhere? This question is easy to answer: We are unaware of any previous validation tests of the urban dynamics model, so no comparison is possible. There are, however, two existing studies that need to be discussed here. The first is a study similiar to the present one: Porter and Henley's (1972) application of the urban dynamics model to the Houston area. The second, though quite different from the present study, is similar in one respect; here we are referring to Alfeld's (1975) application of the urban dynamic model to policy changes in Lowell, Massachusetts.

In the Houston study, the urban dynamics simulation model was used to predict changes in Houston's labor, housing, and industrial sectors, just as in the present study. There, too, the model tended to generate predictions that were too pessimistic. But the predictions generated by Porter and Henley were much closer to actual developments in Houston than proved to be the case in the present study. How can this difference between the two studies be explained? A parsimonious explanation would be that the Forrester model is simply more consistent with the Houston experience than with what happened in Chicago, Detroit, New Orleans, or Philadelphia between 1940 and 1960. But when one juxtaposes the innate pessimism of the urban dynamic model with the "boom" atmosphere that has pervaded postwar Houston, this explanation loses much of its superficial appeal. We believe that the real answer lies in the fact that Porter and Henley introduced several parameter changes into the model to make it more representative of Houston in particular. This was an entirely legitimate strategy in light of their stated purpose—to determine whether the model could be *applied* to Houston. Indeed, Porter and Henley established that the model can predict actual urban developments with a fair degree of accuracy *if* several parameters are adjusted to make it more representative of the special case being investigated. Had we made similar adjustments for each of the four cities investigated here, we would no doubt have obtained a much better fit between predictions and reality.

However, the fact that it is possible to change a model and thereby make its predictions more realistic is not really a point in favor of the model's external validity. As Chen (1973) notes, *any* model can be "tuned" to a particular case; even the most wildly unrealistic model can be made fairly realistic if the change is drastic enough. By far the greatest interest in the urban dynamics model, however, centers on the model *in its original form.* We could have made changes to represent Detroit, others for New Orleans, and still others for Philadelphia. But the model purports to fit "the general class" of urban systems as it stands. It is from the model itself that conclusions concerning the probable future course of large cities have been drawn, and it is the model itself which has been used by Forrester and others as a basis for policy prescriptions. To demonstrate that a specifically tuned revision of the model performs reasonably well in the very system for which it has been specifically tuned is to say nothing at all about the validity of the original model. Nor, to their credit, do Porter and Henley so claim.[2]

Alfeld's (1975) application of the urban dynamics model to Lowell, Massachusetts, is quite different in intent and basic orientation from the present study. The primary purpose of Alfeld's study was to examine whether the introduction of policy changes based on recommendations and insights from the urban dynamics simulation model would indeed bring about favorable results to the city of Lowell, a declining textile industry town. The study was conducted in 1975, also the starting point of the simulation, and generated model projections of the consequences through the year 2000. One could (though, given the date of his study, Alfred obviously did not) compare the predicted results in the first six years, that is, from 1976 to 1981, with actual developments in Lowell, and it is in this sense that Alfeld's study could be considered a test of the urban dynamics model. Needless to say, that was neither the intent nor the spirit of Alfeld's study. However, even if that were the intent, there are still several problems centering around this type of validation study.

First, even assuming that the predicted results fit the real world, it is still very difficult to ascertain that the model is indeed valid. This is especially true in Alfeld's study, where five policy changes were introduced simultaneously. On the other hand, if the predicted results did not fit the real world, one would be hard pressed to identify where the model went wrong. Second, a validation study of this kind, though it is policy oriented and has behavioral implications, is piecemeal at best. The urban dynamics model implies many different policy changes, and in order to test its external validity, one would have to introduce a large number of policy changes. This is definitely not the case in Alfeld's study. Thus, the study falls far short of being a fundamental test of the model.

Third and finally, even if Alfred's intent were to validate the model, the fact is that Lowell is hardly the sort of city envisioned by the urban dynamics model (e.g., it was 8,704 acres in 1970 in comparison with the model specification of 100,000 acres) and the time period in the study would hardly be adequate to make any assessment meaningful.

In this sense, Alfeld's study really says nothing about the validity of the urban dynamics model. There is little doubt that additional validation studies are needed. For the moment, however, the present study provides the only existing basis for assessing the validity of the urban dynamics model, and that assessment is not encouraging. This leads to a second question: How can the urban dynamics model be improved to make it more valid?

Improvement of the model can be approached from either of two directions. The first approach would involve experimenting with the model's variables, equations, and parameters on a piecemeal basis, hoping by trial and error to make it more realistic. This "patching-up" approach seems to us likely to produce nothing more than a great deal of frustration, brought on by the innate difficulties of approaching a large-scale, complex, and counter-intuitive model in such a piecemeal fashion. A potentially more fruitful approach would be explicitly empirical in nature. Because the model is intended to represent "the general class" of urban systems, it should describe what in a sense is an "average" urbanization process. Of course, it is always to be expected that specific cases will deviate from the average. But we have seen that four different large

American cities—each with its own peculiar history, culture, population mix, political traditions, and so on—not only deviate very substantially from the model but also all deviate in the same direction from what is represented in the urban dynamics model. It also suggests a strategy for revising the model.

Such a revision could begin with a large-scale data collection effort ultimately aimed at constructing a profile of a "typical" urbanization process. In order to construct such a profile, the actual life histories of a number of urban areas would have to be examined, historical data located, and central tendencies isolated. With such a profile in hand, the parameters and generating equations in the urban dynamics model could be tuned. But rather than being tuned to a specific city, with all the peculiarities of its past and present development, the model would be tuned to a composite or profile city. In this fashion, the model could be made much more representative of the experiences of a large number of urban areas. Of course, a single profile might do violence to the diversity that one finds among large urban areas. If that diversity undermined the utility of a composite profile, the strategy of constructing two or more *variant* models should be considered; for example, declining northeastern industrial centers might share enough features that they could be adequately represented in terms of a single composite model, but newly developing sunbelt cities need to be represented in terms of a rather different model. Developing an empirically based composite or set of composites to tune the simulation model would obviously be much more difficult than specifying relationships and parameters in the armchair manner that has characterized modeling to this point. But if followed, the strategy outlined here could be used to create an urban model that would generate predictions based on documented historical tendencies and thus could be used with much greater confidence to suggest and test new approaches to urban policy.

NOTES

1. Actually, data on the industrial sector pertain to 1948 and 1958 rather than 1950 and 1960, because this is when the *U. S. Census of Business* surveys were conducted. Accordingly, our comparisons of simulated and actual industrial developments are for 1948 and 1958.

2. Batty (1976: 308) refers to an application of the urban dynamics model to the Venice, Italy, region by Costa and Piasentin. While we have not seen this article (it was

presented at a UNESCO symposium of trends in mathematical modeling held in Venice in 1971 and is written in Italian), Batty's description makes it clear that, like the Potter-Henley study, this was an application rather than an attempt at validation.

REFERENCES

ALFELD, L. E. (1975) "Urban dynamics applied to an old industrial city," pp. 203-217 in W. W. Schroeder III et al. (eds.) Readings in Urban Dynamics. Cambridge, MA: Wright-Allen Press.

BATTY, M. (1976) Urban Modelling: Algorithms, Calibrations, Predictions. Cambridge, England: Cambridge Univ. Press.

CATANESE, A. J. (1972) Scientific Methods of Urban Analysis. Urbana: Univ. of Illinois Press.

CAWKELL, A. E. (1968) "Citation practices." J. of Documentation 24: 299-302.

CHEN, K. (1973) "An evaluation of Forrester-type growth models." IEEE Transactions on Systems, Man, and Cybernetics 3: 631-632.

——— (1972) "Preface." IEEE Transactions on Systems, Man, and Cybernetics 2: 122-123.

COSTA P. and V. PIASENTIN (1971) "Un modello di simulazione dello sviluppo urbano di Venezia." Presented at the UNESCO Symposium on Trends in Mathematical Modeling, Venice, Italy.

FORRESTER, J. W. (1971) World Dynamics. Cambridge, MA: Wright-Allen Press.

——— (1969) Urban Dynamics. Cambridge: MIT Press.

——— (1961) Industrial Dynamics. Cambridge: MIT Press.

GRAY, J. N., D. PESSEL, and P. P. VARAIYA (1972) "A critique of Forrester's model of an urban area." IEEE Transactions on Systems, Man, and Cybernetics 2: 139-144.

INGRAM, G. K. (1970) "Urban dynamics." J. of the Amer. Institute of Planners 36: 206-208.

PORTER, H. and E. J. HENLEY (1972) "Application of the Forrester model to Harris County, Texas." IEEE Transactions on Systems, Man, and Cybernetics 2: 139-144.

RANDERS, J. (1980) "Introduction," pp. 17-21 in J. Randers (ed.) Elements of the System Dynamics Method. Cambridge: MIT Press.

RASER, J. R. (1969) Simulation and Society. Boston: Allyn & Bacon.

RAUSSER, G. C. and S. R. JOHNSON (1975) "On the limitations of simulation in model evaluation and decision analysis." Simulation & Games 6: 115-150.

SMITH, N. J. and A. P. SAGE (1973) "Hierarchical system identification of models for urban dynamics." Socio-Economic Planning Sciences 7: 545-569.

STONEBRAKER, M. (1972) "A simplification of Forrester's model of an urban area." IEEE Transactions on Systems, Man, and Cybernetics 2: 468-472.

Yung-Mei Tsai is Associate Professor of Sociology at Texas University. His fields of special interest are urban sociology and mathematical sociology. He received the Ph.D. from the University of Colorado in 1973.

Otomar Bartos is Professor of Sociology at the University of Colorado. His primary interests are in mathematical models and sociological theory. He has written extensively on the mathematical modeling of negotiations.

Lee Sigelman is Professor and Chair of the Department of Political Science at the University of Kentucky. He has worked in a number of social science fields, including comparative national, state, and local analyses.

12/6/82

Texas Tech University

Department of Sociology

Nov. 29, 1982

Professor Jay Forrester
Massachusetts Institute of Technology
Cambridge, Mass.

Dear Professor Forrester:

Enclosed please find a copy of our paper. It is a study of four U.S. cities based on your Urban Dynamics Model. I thought you might be interested in seeing it. I hope it is useful to you.

Sincerely yours,

Yung-mei Tsai

Yung-mei Tsai, Ph.D.
Associate Professor

Box 4590/Lubbock, Texas 79409-4590/(806) 742-2400

A Look at *Urban Dynamics*: The Forrester Model and Public Policy

HARVEY A. GARN AND ROBERT H. WILSON

Abstract—A review of *Urban Dynamics* [1] and additional simulation by the authors is presented. Structural choices of boundary conditions, major subsectors, and multiplier relationships in the Forrester model are examined. Some of their implications and shortcomings are noted. Forrester's policy conclusions and implicit evaluation criteria are discussed and related to our alternative assumptions and simulation.

Introduction

FOR A NUMBER of reasons a recent model developed by Forrester and described in *Urban Dynamics* [1] provides an interesting focal point for discussing models of urban systems. First, Forrester has demonstrated, in his previous work [2] as well as in this work, great skill in developing operational dynamic models of complex systems. Second, the model is one of the relatively few attempts to examine the city as a whole. Third Forrester has described possible applications of his model for policy recommendations. Fourth, because of its ambitiousness, the effort provides insights into both the problems and possibilities of applying quantitative models in the solution of urban problems.

There appears to be an increasing conviction that formal models and other analytical tools can be useful for solving social problems which has partly resulted from our increased ability to handle large-scale models and their utility in such fields as defense and space exploration. However, one characteristic of such uses is that many of the critical variables, their values, and their interrelationships are known. Unfortunately, this is much less true for social systems. In some cases knowledge is so limited that it is true that "one guy's guess is as good as another's."

This conviction is sometimes accompanied by a tendency to imagine that models have not been previously used for viewing social systems. As a result there have been, in some cases, exaggerated claims for more formal quantitative models which they cannot fulfill. Such an attitude contains the danger that the more humble but real possibilities of model development will be rejected by potential users—in this case decision makers. While it is true, as Forrester emphasizes, that many of these models are highly simplified and implicit, this should not be taken to mean that people will necessarily abandon their implicit models or that they should.

Manuscript received August 31, 1971. This paper was presented at the Second Annual Pittsburgh Conference on Modeling and Simulation, Pittsburgh, Pa., March 29–30, 1971.

H. A. Garn is with The Urban Institute, Washington, D.C. 20037.
R. H. Wilson was with Yale University, New Haven, Conn., and The Urban Institute, Washington, D.C. 20037.

It is not startling to say that all models are a representation, and therefore an incomplete description of reality, but it does mean that the model builders have many important choices to make about which parts will be represented, how they will be related, what the outputs will be, and how they are to be evaluated if policy conclusions are desired. Evaluating the claims made for the model involves looking at what these choices were and making an assessment of the representation which results.

The Forrester *Urban Dynamics* Model

Forrester is straightforward about the claims he is making. He claims, with considerable justice, that many of our existing models have serious shortcomings for dealing with complex dynamic systems. Many, if not most, of our intuitive models derive from our common experience in dealing with single-loop negative-feedback systems in which cause-and-effect relationships are closely connected. Complex systems contain positive-feedback loops as well as negative-feedback loops. Often it is difficult to determine causal connections since they are not as closely linked as our intuitive models may lead us to believe. One important result of this is that we may take two variables to be related as cause and effect when they are jointly determined as effects of another cause(s). Another important result is that actions designed to improve a situation may be directed at changing a variable which cannot, in fact, affect the desired outcome. The change also may produce adverse effects or, at least, unintended consequences. Forrester argues that complex systems are counterintuitive in the sense that there is a high probability of such adverse effects. He further argues that many of our analytic models are too static to deal with dynamic properties of interesting systems and that the mathematical problems involved in modeling system nonlinearities analytically have led analysts to assume linearity. There is little in this characterization of the state of the art with which to argue.

Forrester's urban dynamic model discussed here attempts to demonstrate the applicability of the methodology which he has developed to the problems of the cities as a means of overcoming these shortcomings. In particular, his model is designed to do two things: 1) reproduce the development of a city, starting with mostly vacant land through a cycle of growth, maturity, decline, and stagnation, showing the interrelationships between the population and where they live and work; and 2) simulate the effects on model output of introducing various policies designed to solve urban problems. His modeling choices in carrying out this design will now be discussed.

Structural Choices: Boundary Conditions and Major Subsectors

Forrester's first set of structural choices related to determination of a boundary for the model and the major subsectors of the model. He chose to develop a closed boundary model of a fixed land area communicating with but not affecting the external environment. By his boundary choice he explicitly ruled out consideration of problems of central city–suburban relations, and possible effects of actions taken to improve the situation in the city on the larger society. Those who believe that central city–suburban problems and relations of cities to the larger society have a high priority and also set some of the important political and social constraints on what is possible in a city will be disappointed in this model. We share this concern, although it would be quite legitimate for Forrester to reply that what is needed to accomplish this is a different model, rather than a criticism of his model.

Awareness of the meaning of the boundary assumption is important for assessing the policy implications of the model. It implies that the causes of the behaviour of the system modeled can be found within the system boundary rather than in interactions with the environment. If this is true, in fact as well as in the model, one should look for both the cause and cure of city problems within the city itself. Forrester seems to accept, during the course of his book, that the boundary assumption which he made to close the model does accurately reflect the real situation. He concludes his book with these words [1, p. 129]:

> The city has been presented here as a living, self-regulating system which generates its own evolution through time. It is not a victim of outside circumstances but of its own internal practices. As long as present practices continue, infusion of outside money can produce only fleeting benefit, if any. If the city needs outside help, it may be legislative action to force on the city those practices that will lead to long-term revival. Such outside pressure may be necessary if internal short-term considerations make the reversal of present trends politically impossible. The revival of the city depends not on massive programs of external aid but on changed internal administration.

This interpretation is open to serious question because, as indicated in the foregoing, the boundary choices made by Forrester rule out central city–suburban problems and effects on the larger society. In particular, many have noticed that one of the major problems for central city mayors has been to develop a means of coping with the out-migration of many of the higher income groups to the suburbs and the concomitant loss of tax revenues which this move entails. Many of these people continue to hold jobs in the central city—their exodus for residential purposes has not opened up anything like the same number of jobs to population in most central cities relative to their suburbs. In Forrester's model there is no commuting across the boundary—people who work in the city also reside there. One of the most interesting and difficult problems, therefore, is lost in his formulation. We will return to the question of effects of policies recommended by Forrester on the larger society after some of them have been described.

Forrester concluded that the major subsectors of the model should be business, housing, and population, although one may reasonably question whether these "are more fundamental than city government, social culture, or fiscal policy" [1, p. 17]. These subsectors include an aging process for housing and industry, and through employment of positive and negative feedbacks it produces periods of growth, decline, and stagnation. This cycle is central to Forrester's intent to show the inevitability of stagnation and decay of the "normal" city. It is presumed that this is the fate of any city with a fixed land area, and, since declining industry, dilapidated housing, and a high proportion of underemployed people characterize the stagnant city, it is not surprising that his policy recommendations are aimed at clearing land of declining industry and old housing and making the city less attractive to the underemployed.

At a more detailed level, the mechanics of the model involve the interrelationships between the initial parameters, the rate equations, and the auxiliary equations which translate information about the level of variables through various multipliers to produce changes on the rates within the system. Within this framework, we find that the average life of mature and older businesses is assumed to exceed that of new enterprises and that new enterprises are strongly discouraged by the lack of land. It is not surprising then that the composition of industry is heavily weighted with the older industries. Similarly, given the distribution of taxation and expenditures and the distribution of people by employment, there will be a preponderance of labor and and underemployed groups in the stagnant city. Later we will examine the effects of changing some of these assumed relationships.

Multiplier Definitions and Their Role in the Model

However, for the moment we will consider some of the critical definitions as well as their behavioral justification and their consequences in the model. The first is the set of influences specified as a set of multipliers which seek to state the impact of such variables as public expenditures, housing availability, jobs, and mobility on the migration of the underemployed. A similar set of relationships exists for both labor and managers. In comparing the effects of each of these multipliers on each category of the population it will be observed that the underemployed are much more sensitive to housing conditions than either of the other two groups. Just why this is assumed to be so is not explained. One would expect at least equal responsiveness on the part of managers and labor, and even more likely, that managers would be the most sensitive to housing conditions. (For reasons of space these equations are not spelled out: the equations referred to are to be found in [1, pp. 135–146, 161, 169].)

Since it is a contention of Forrester that there are "few parameters that can affect policy recommendations" (which presumably implies that parameter errors are not of great importance), it is interesting that he found that the underemployed/housing multiplier is one such parameter [1, p. 237]. In fact, by assuming a less sensitive relationship, the policies of revival, which he urges in [1, ch. 5], "are in several ways less effective." Presumably a corollary would be that the low-cost housing program, described in [1, ch. 4], would be somewhat less ineffective. A similar query can be made about the low-cost housing program to which the underemployed are also assumed to be highly sensitive.

It should be noted that the specification of the influences on arrivals allows for a greater response by labor and managers than underemployed to job opportunities. Since new industry creates relatively more jobs for labor and managers, it is not surprising that encouraging new enterprises will cause a population shift against the underemployed. Again it would appear that the policy recommendations have been built into the model and further, that they are in no sense counterintuitive.

A further interesting but perplexing feature of the model is the effect of land occupancy on construction of new enterprise, premium housing, and worker housing. In the model new construction is possible only on land unoccupied either because of a "natural tendency for demolition . . . as usage becomes more complete" or "forced demolition" [1, p. 201]. The model would more reasonably reflect shifting land uses if such changing uses could come about as a result of market mechanisms rather than through natural aging processes or forced demolition. In Forrester's formulation lack of land becomes an insuperable barrier to new construction instead of being another factor to be considered along with taxation, labor, and management availability.

The specification of taxes is also rather curious. Even leaving aside the question as to whether industry really pays taxes but receives no benefits, it is surely untrue, especially in an historical context, that the poorer members of the city have been the recipients of greater expenditures per capita than other members. As in many of the specifications it would appear that Forrester has accepted the "conventional wisdom" on this matter without having investigated the plausibility of his construction relative to existent social science research.[1]

Although it is possible to examine other multiplier relations and raise similar questions, the included examples make the points we are trying to bring out. These are as follows.

1) The multipliers hypothesized are the driving forces in the model.

2) The assumptions which have been made about the shape and range of the functions which determine the multiplier values are arbitrary.

3) It would take much more empirical work to determine if the multipliers selected are the appropriate ones and even more to determine the shape and ranges of the functions.

4) While Forrester is correct in saying that complex dynamic systems may be insensitive to parameter changes, it is not correct to imply that the policy recommendations in this model are insensitive to simultaneous changes in several of the multiplier functions; Forrester acknowledges this briefly in discussing sensitivity as quoted in the preceding. He even says that "combinations of several parameter changes might react to defeat the desirability of a policy proposal" [1, p. 236] but does not indicate which, if any, such combinations were tested.

5) For all these reasons as well as the earlier points about selection of variables, the particular policy recommendations which derive from the model should be viewed with considerable reservation.

Forrester's Policy Conclusions

Of course Forrester has indicated in several places that he views this work as a preliminary effort which is "not presented as a set of final answers to guide urban policy makers" [1, p. 11], but it is difficult to reconcile this with the strength with which he states the policy conclusions in chapters 4, 5, and 7. It is undoubtedly both those policies which he accepts and rejects which have caused much of the controversy surrounding *Urban Dynamics*. Among those policies rejected (separately rather than in combination with one or other policy proposals) are specially created jobs for the underemployed, a job training program to raise 5 percent of the underemployed into the labor sector, an external tax subsidy, and a low-cost housing program. Forrester concludes that these programs have either neutral or negative effects with the low-cost housing program being the worst. For, although creating short-run improvements, they generally lead in the long run to more underemployed, a worse tax situation, and relatively more declining industry and underemployed housing. Furthermore, this change is at the expense of new enterprise and the attractiveness of the city to labor and managers, and it is in this sense that Forrester counts the effects as negative.

The programs which he considers favorable include construction of worker housing, construction of premium housing, construction of new enterprise, demolition of declining industry, demolition of underemployed housing, a combination of restraining new worker housing with underemployed housing demolition, and, finally, encouraging new enterprise plus the demolition of underemployed housing. The last two are deemed the most successful for, by clearing land, they increase the attractiveness of the city to new enterprise, thus encouraging managers and labor to enter the city. At the same time, demolishing underemployed housing makes the city less attractive to the underemployed.

[1] In his book [1] Forrester cites a total of six references—none of which is a part of the rapidly growing literature in the field of urban studies. In the preface Forrester justifies his approach because he expected "the most valuable source of information to be, not documents, but people with practical experience in urban affairs [1, p. ix]. Without wishing to disparage their contribution, the gap between experience gained in practice and the knowledge that research produces is often large and in favor of research. One cannot escape the feeling that reference to research done in this area would have led to significant changes in the specification of many of the functional relationships in *Urban Dynamics*.

Experimentation with Alternative Program Mixes and Parameter Changes

In order to examine further the importance of some of the assumptions described in the foregoing, we present the results of two simulations of the model which we have performed. The first starts with Forrester's model at equilibrium after 250 years. (The values describing this state are given in [1, p. 217].) Two development programs are run simultaneously for 50 years, the first being the underemployed job training program which is designed to shift 5 percent of the underemployed into the labor category (which, operated alone, is considered a failure by Forrester), and the second is an incentive scheme to attract new enterprise into the city at the rate of approximately 1.2 percent per year. The combination of these programs should provide more laborers as well as the employment opportunities to keep them in the city. Further, such a combination could and might be carried out simultaneously in a comprehensive program.

The results of this experiment can be seen in Table I. The column, "Time −5 years," gives the equilibrium values at the start of the test, and the programs are put into effect at time zero. Generally speaking, these programs produce an overall growth of industry housing and population. However, there has also been a change in their composition. Premium and worker housing have increased by 21 and 35 percent, respectively, while underemployed housing has decreased by 1 percent. In the various categories of the work force the manager–professional group has increased by 43 percent, labor by 46 percent, but underemployed by only 15 percent. Further, the underemployed/job ratio has fallen by 22 percent, and the number of underemployed shifting to the labor category has risen from 5500 to 29 900 per annum, an increase of approximately 550 percent. Finally, the tax ratio needed has declined by 8 percent.

Although it is true that the improvement is not as dramatic as that given by a combination of slum-housing demolition and the encouragement of new enterprise construction, which appears to be the most highly favored combination in *Urban Dynamics*, the essential point is that policies traditionally followed do have favorable influences on the city if properly combined. This is an important point because one of the claims of *Urban Dynamics* is that many of these kinds of programs have failed and could be expected to fail in the future. (See, for example, the comments in [1, p. 70].)

The second simulation we performed involves the effects of changing some of the structural parameters so that they are more consistent with other research. In this case the model was run from time zero with the initial values taken from *Urban Dynamics*. The changes introduced are as follows. The sensitivity of underemployed arrivals to housing conditions has been lessened while that of manager arrivals has been increased. In addition, premium housing, worker housing, and new enterprise are assumed to be less sensitive to the fraction of the city's land area occupied. Finally, it is assumed that the public expenditures needed by managers has increased, while that of the underemployed has decreased. (The actual changes, with the equation numbers used in *Urban Dynamics* may be found in the Appendix.) As it turned out, equilibrium was not reached after 250 years, and the model was run for 350 years. This should not be surprising as the effect of the land area, which plays a vital part in completing the system is given a less prominant role in our simulation. Despite the lengthening of time to equilibrium, the values reached after 250 an 350 years are very similar. The results for both periods a well as the equilibrium values generated in *Urban Dynamic* are presented in Table II.

It can be readily seen that a more "healthy" city has bee generated in the sense that the relative proportions declining industry. underemployed, and underemploye housing have fallen substantially. Most interesting of a perhaps, is that the tax ratio needed has fallen from 2 to 1.01. There has, however, been a change in the net u ward mobility of the underemployed which, although ve small, appears to have been reversed. The overall impressi resulting from these changes is unmistakable, however. T changes made in the structure have produced an equilibri state which can scarcely be regarded as stagnation. important conclusion is that changing the structural p meters can make a considerable difference. Hence i important that they be correctly specified. It is simply good enough for Forrester to claim, as he does repeate that the system will not respond to selective changes examples, see his comments in [1, ch. 6]). Furthermore. system does not seem to be so complex that the effec making such changes, at least the direction of these ef are not predictable in advance. This is an important

TABLE I

The Effect of the Simultaneous Application of an Underemployed Job Training Program (UTR = 0.05) and the Encouragement of New Enterprise (NECR = 0.02)

Variable	Symbol	Time (years) −5	Time (years) 50	Change (percent)
a. New enterprise construction	NEC	462	820	+ 78%
b. Underemployed training program	UTP	0	21,700	-
c. Underemployed to labor	UTL	16,800	46,500	+177
d. Underemployed arrivals	UA	17,300	36,200	+109
e. Underemployed departures	UD	17,300	12,400	- 28
f. Labor arrivals	LA	7,400	6,300	- 15
g. Labor departures	LD	13,200	35,900	+172
h. Labor/underemployed	L/U	1.04	1.32	+ 27
1. New enterprise	NE	4,900	7,200	+ 47
2. Mature businesse	MB	7,800	12,100	+ 55
3. Declining industry	DI	16,500	23,500	+ 43
4. Premium housing	PH	110,900	134,200	+ 21
5. Worker housing	WH	335,700	346,500	+ 3
6. Underemployed housing	UH	310,100	308,700	- 1
7. Managerial-professional	MP	71,100	101,600	+ 43
8. Labor	L	392,600	574,100	+ 46
9. Underemployed	U	377,300	433,500	+ 15
10. Manager/housing ratio	MHR	1.07	1.26	+ 18
11. Labor/housing ratio	LHR	1.17	1.66	+ 42
12. Underemployed/housing ratio	UHR	.81	.94	+ 16
13. Manager/job ratio	MR	1.38	1.33	- 4
14. Labor/job ratio	LR	.97	.97	0
15. Underemployed/job ratio	UR	1.81	1.41	- 22
16. Tax ratio needed	TRN	2.25	2.08	- 8
17. Underemployed to labor	UTLN	5,500	29,900	+543

TABLE II
CHANGES IN EQUILIBRIUM CAUSED BY MODIFICATIONS TO *URBAN DYNAMICS*

Variable	Symbol	Forrester	Garn-Wilson		Changes	
		1	2	3	2/1	3/1
		250 years	250 years	350 years	Percent	
1. New enterprise	NE	4,900	7,300	8,400	+49	+71
2. Mature business	MB	7,800	12,800	13,000	+64	+67
3. Declining industry	DI	16,500	5,100	5,000	-69	-70
4. Premium housing	PH	111,100	178,000	183,000	+60	+65
5. Worker housing	WH	335,800	566,500	561,000	+69	+67
6. Underemployed housing	UH	309,900	184,300	185,000	-40	-40
7. Managerial-professional	MP	71,200	89,200	93,700	+25	+32
8. Labor	L	392,900	448,500	471,400	+14	+20
9. Underemployed	U	377,200	303,900	307,700	-19	-18
10. Manager/housing ratio	MHR	1.07	.84	.85	-22	-21
11. Labor/housing ratio	LHR	1.17	.79	.84	-32	-28
12. Underemployed/housing ratio	UHR	.81	1.10	1.11	+36	+37
13. Manager/job ratio	MR	1.38	1.49	1.45	+ 8	+ 5
14. Labor/job ratio	LR	.97	1.06	1.04	+ 9	+ 7
15. Underemployed/job ratio	UR	1.81	1.53	1.41	-15	-22
16. Tax ratio needed	TRN	2.25	1.01	1.00	-55	-56
17. Underemployed to labor net	UTLN	5,500	-4,200	-1,500	-	-

because it takes the edge off Forrester's comments that past attempts at modeling social effects and causes have failed and will continue to fail because of their simplicity. It would still seem that correctness, not complexity or simplicity, is the issue on which attention should be focused.

EVALUATING THE RESULTS OF THE MODEL

It should be emphasized that the changes made are merely some of many possible changes that can be made. Although we believe that in the particular examples chosen we have derived a more accurate specification, this should not be taken to imply that we believe the *results* generated are necessarily better than those of Forrester. In order to do this we need evaluative criteria. Such criteria cannot be treated as *a priori* truths. They represent assessments, vary with different points of view, and are not subject to scientific determination. Forrester does not specify the criteria which he uses to determine whether one policy outcome is better than another. It is possible, however, to infer something about the evaluative criteria he is using from the outcomes he most strongly supports. He prefers outcomes which have a higher proportion of managers and laborers to underemployed in the city population than is the case at his system equilibrium. He is prepared to significantly reduce the attractiveness of the city to the underemployed in order to increase the attractiveness to managers, labor, and new enterprises. He prefers to accomplish this in both cases by a program which eliminates substantial amounts of low-cost housing. In short, a richer city can be obtained by having fewer poor people in it.

Aside from the questions of whether or not an individual city could remain an island of wealth with the poor outside or the feasibility of all major cities adopting his policies simultaneously, it is possible to suggest alternative evaluative criteria. We suspect that the underemployed, for instance, would evaluate the outcomes differently from Forrester. Evaluating the outcomes from a national perspective, as opposed to an individual city's perspective, would affect the appraisal as well. What happens in the cities cannot be divorced from the surrounding environment, and our evaluation of policies for cities should take this into account. Hence we wish to stress that in making the comparisons of the results we obtained with those reported in *Urban Dynamics* we have been doing so according to what appears to be Forrester's criteria. Within the framework of this paper this was the only reasonable way of making comparisons, but it does not mean an acceptance on our part of Forrester's implied criteria. Even on his criteria, however, the changes made produce a "better" or, at least, less stagnant city.

SUMMARY AND CONCLUSIONS

The model developed in *Urban Dynamics* represents an attempt to model the city as a system in such a way as to overcome some of the deficiencies of existing intuitive and analytical models and to simulate the effects of policy alternatives in model outputs. In this model Forrester has demonstrated considerable methodological skill in developing an operational dynamic model. He has shown that a city and many of the interrelationships within it can be modeled. As indicated in the foregoing, however, we have reservations about the applicability of the model and the policy recommendations derived from it. The particular boundary assumptions made in the model rule out the treatment of some of the critical problems—those related to central city–suburban relations and the effects of city-oriented policies on the rest of society. The assumption that the housing, business, and population variables in the model represent the fundamental structure and are more important than government, fiscal policy, and social structure in understanding city problems is open to question. In any case, little empirical or theoretical justification for the assumption is presented. The assumptions about parameters and multiplier functions, also, are open to question. It is one thing to make the methodological point that the latter functions should be nonlinear. It is quite another to know what the critical functions are, as well as their range and shape. Again, there is little empirical support provided for the assumptions that have been made about the ranges and shapes of these functions.

As Forrester indicates, the applicability of the policy conclusions depends upon the prior determination that the model reflects the central structure and critical relationships satisfactorily. We have shown how certain changes in the

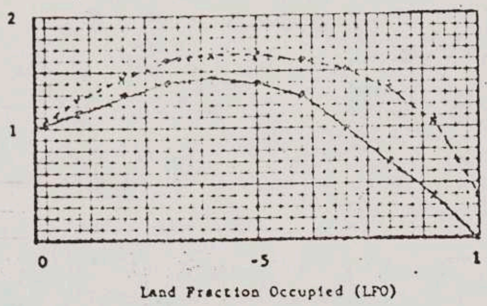


Fig. 1. Enterprise land multiplier (ELM) [1, fig. A-51]. -----Garn-Wilson; ——Forrester.

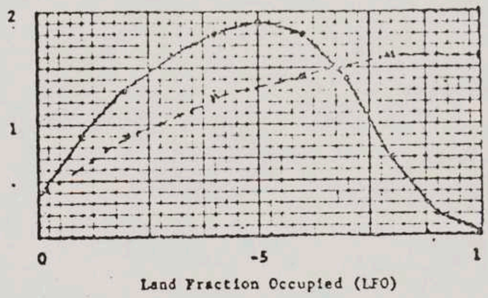


Fig. 2. Premium housing land multiplier (PHLM) [1, fig. A-32]. -----Garn-Wilson; ——Forrester.

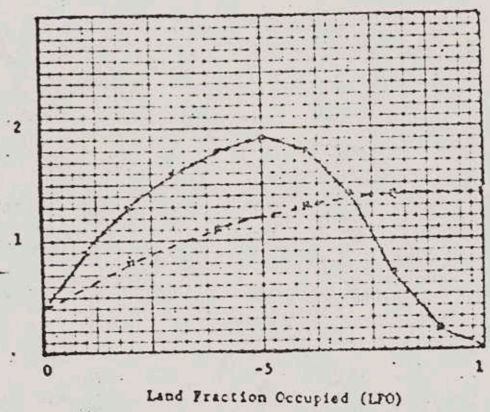


Fig. 3. Worker housing land multiplier (WHLM) [1, fig. A-40]. -----Garn-Wilson; ——Forrester.

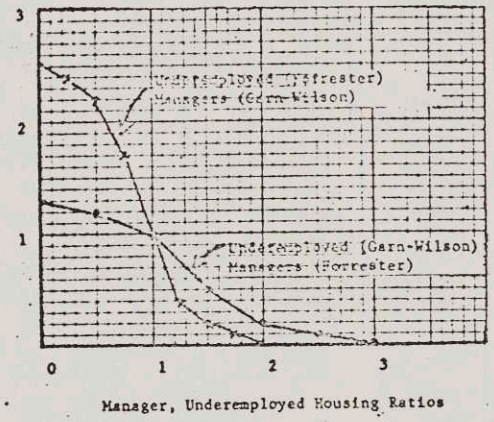


Fig. 4. Manager arrival housing multiplier (MAHM) and underemployed housing multiplier (UHM) [1, figs. A-3, A-28].

assumptions about the shape and nature of these multipliers can result in important differences in the outcomes. Even assuming, however, that the model is thought to be accurate, we are still left with the problem of evaluating model outputs. Furthermore, it is clear that evaluative criteria depend upon the perspective of the evaluator. A conclusion that overall system performance of an individual city is improved by a particular course might be considerably less optimal from, say, a national perspective, or from the perspective of identifiable groups within the city system. We suggest that these possibilities should be considered when evaluating *Urban Dynamics*. At the same time, we must recognize the need for broadening our empirical knowledge of urban systems, their critical variables, and the nature of their interrelationships so that the models we employ provide reasonable guidance in the selection of policies to improve life in both the cities and the country.

Appendix

Figs. 1–3 are a representation of the changes made in the sensitivity of new enterprise, premium housing, and worker housing to the fraction of the land area of the city occupied. Fig. 4 illustrates the switch made in the sensitivity of underemployed and manager arrivals to housing conditions. Finally, the changes made in the taxes needed are given with Forrester's values in parentheses as follows:

tax per management person (dollars/person/year) (TMP)
= 300 (150) [1, eq. 126.1,C]
tax per underemployed person (dollars/person/year) (TUP)
= 150 (300) [1, eq. 126.3,C].

References

[1] J. W. Forrester, *Urban Dynamics*. Cambridge, Mass.: M.I.T. Press, 1969.
[2] ——, *Industrial Dynamics*. Cambridge, Mass.: M.I.T. Press, 1961.

agricultural historians, achitectural historians, and art historians" (p. 238). It is singular that the historian of technology fails to be included in the litany. What Glassie might have succinctly stated is that the field needs the attention of American scholars working in the genre of the Welshman, J. Geraint Jenkins (see my review of his *Traditional Country Craftsmen* [New York, 1966] in *Technology and Culture* 8 [1967]: 104–105). Until that happens material folk culture as presented by Glassie will remain largely an enigma to the historian of technology.

The format of the book presents a problem. Footnotes are at the bottoms of pages, commendably. The bibliography is long and will suggest and aid further study. Illustrations and cuts are crisp throughout the text. In spite of this the book suffers, for there is no table of contents, no list of illustrations, no chapters and chapter headings as such, and finally, no index. Glassie sounds the clarion for an interdisciplinary attack upon a threatened field. These mechanical additions to his book would have helped the cause.

PETER C. WELSH*

Urban Dynamics. By Jay W. Forrester. Foreword by John F. Collins. Cambridge, Mass.: M.I.T. Press, 1969. Pp. 285; tables; charts; graphs. $12.50.

This study is the result of a sophisticated methodological approach to metropolitan problems, and it offers policy recommendations which deviate markedly from current programs. The author rejects continued low-income housing construction in favor of slum demolition and encouragement of industry in order to revive the economy of the city. This conclusion rests on a specific method of analysis and the validity of the theory derived from it; the following remarks will emphasize this aspect of the book.

Urban Dynamics is a study of urban growth from the perspective of systems analysis. It is based on the assumption that complex social systems are inadequately understood by normal human thought processes. This is because the intuitive way we learn to think is formed within the context of linear relationships, where an action (cause) leads directly to an observable result (effect). By contrast, the city is a nonlinear, "counter-intuitive" entity made up of the constant flow of many interacting variables, wherein a single cause may have unforeseen and contradictory consequences throughout the system. When applied to such a system, intuitive, linear thought will focus on symptoms rather than basic causes of problems. Policies arrived at by this ap-

* MR. WELSH is the assistant to the Director of Museums, Smithsonian Institution. He was formerly curator of the section on Growth of the United States. He is an authority on the development of tools and their relationship to other aspects of American culture.

proach are, the author asserts, short range in effect, and either neutral or actually detrimental in the long run. Thus there is a need to examine the city by means of a computer model which can handle the multiple interactions between various parts of the system, and in so doing provide a theory of urban growth and structure.

Such a model is simulated by a digital computer which, on the basis of mathematical equations entered into it, produces flow charts illustrating the interaction of various urban components. Two models are presented in this study, a growth model representing 250 years of urban development, and an equilibrium model used to test the results of certain policies over a projected fifty-year period. They do not represent a specific city but are meant to illustrate processes common to all urban areas. The urban area generated is conceived as "a closed dynamic system," a self-contained and self-regulating entity which evolves its own development and problems but is little affected by, and has small impact upon, the surrounding environment. The main contact between the outside environment and the urban system is the movement of people into and out of the area. This flow is determined by the attractiveness of the city, relative to the surrounding environment, in terms of (1) jobs provided by industry, (2) housing, and (3) population mix. These three variables, industry, housing, and people, are the main interacting components of the city.

In the growth model, the life cycle of an urban area is charted. The charts plot interactions and different ratio levels among the three basic components, now subdivided into nine variables: new enterprise, mature business, declining industry, premium housing, skilled-worker housing, underemployed housing, managerial-professionals, skilled workers, and underemployed. Beginning with empty land, the city develops to full land occupancy in the first 100 years, at which time new enterprise and premium housing have peaked, while the managerial-professional and skilled-worker population as well as worker housing are at a near maximum. The next 150 years see a realignment of internal variables which finally emerge into an equilibrium stage of stagnation marked by slums, underemployment, flight of industry to the suburbs, high tax rates, and increasing welfare rolls. What has happened after 250 years of growth and stagnation is that the city has increased its attractiveness for underemployed and become less attractive to new enterprise which could provide jobs, thus raising the underemployed to the skilled-worker class. The success of a city, the author contends, is not its ability to concentrate the economically less successful into areas of little economic hope but its effectiveness in providing upward economic mobility for the underemployed.

Which programs are most likely to achieve this goal? Using the equilibrium model, computer runs show the neutral results of a job training program and a tax subsidy, and the detrimental effect of low income housing construction. In regard to the latter, a program providing housing for 5 percent of the underemployed per year for fifty years has a detrimental effect on worker housing, the labor population, the underemployed/job ratio, the tax rate, new enterprise, and mature business. Another group of computer runs, based on a different set of policy variables, provides results which convince the author that the most efficient way to revive the urban economy is to pursue a program of slum clearance and replacement with new business enterprise.

Some of the questions that can be asked of Forrester's method concern the uses of data and the extent to which the computer model corresponds to reality. What kind of information or evidence formed the basis for the mathematical equations which were fed into the computer? It will not satisfy some to be told that the sources of information have been "people with practical experience in urban affairs . . . from the insights of those who know the urban scene firsthand, from my own reading in the public and business press, and from the literature on the dynamics of social systems." It is even more unsatisfactory to find that only three of those practical people are named, and that out of a total of six references five are to the author's own works. The dismissal of historical evidence and the historical dimension is made explicit when the author asserts that, with few exceptions, the stagnation of a city does not depend on the city's history (p. 106), and that "today's problems extend from the present into the future." The failure to recognize that urban reality is three-dimensional, consisting of *past*, present, and future, and that every city is to some extent unique because of its past, is a fault which will cause some to question the validity of Forrester's model and theory. Finally, when readers of this journal learn that "the model does not, and need not, deal with changing technology" (presumably on the grounds that technology exists outside the system in the limitless environment), they will rightfully ask what kind of reality this systems analysis model is supposed to represent.

PARK DIXON GOIST*

Agricultural Development and Economic Growth. Edited by Herman M. Southworth and Bruce F. Johnston. Ithaca, N.Y.: Cornell University Press, 1967. Pp. xv + 608. $12.00.

This book is not about technology as such but does contain much valuable information on important and related areas in agricultural development. Some thirteen subjects are covered in detail, including development theory, social barriers to change, infrastructure, education,

* DR. GOIST teaches urban history and American studies at Case Western Reserve University.

more objectively. His excellent study of Dwight Moody, the popular and influential evangelist, adds immeasurably to our understanding of the revival movement which occurred during the late 19th century and which so deeply stirred a majority of the middle class as they attempted to adjust their thought to the urban and industrial forces which were changing so drastically the life of the country. Moody emerges as an appealing if naive individual. A successful businessman with little religious training and almost no formal education, many of the techniques and methods he initiated form the basis of the gospel crusades of today. Although his message no longer fully appeals, he was one of the great organizational innovators of his age.

ALAN SEABURG, *Unitarian Universalist Association, Boston*

MICHAEL FITZGIBBON HOLT, *Forging a Majority: The Formation of the Republican Party in Pittsburgh, 1848–1860.* 408 pp. Yale University Press. $10.00.

This historically significant book is full of surprises for both general reader and historian. For example, the Republican party in Pittsburgh was not founded on the demand for a protective tariff nor on the moral abhorrence of slavery. Instead, author Michael Holt, assistant professor of history at Yale, holds that "Pittsburgh Republicans cared more for the rights of white men than of Negroes . . . their appeals were aimed at the unfair power of the minority South and its aggressions against the rights of the Northern majority, rather than at slavery." Their objection to the expansion of slavery was aimed at preserving the territories for white men, not at helping the Negro slave. According to Holt it was the local religious, ethnic and social factors rather than regional ones that determined the voting patterns of Pittsburgh. Chief among these were the intense Protestant-Catholic and immigrant-native born antagonisms. If these conditions were paramount in the formation of the GOP in Pittsburgh who can say they played no significant part elsewhere in the North?

EUGENE W. JONES, *Angelo State University*

JAY W. FORRESTER, *Urban Dynamics.* 280 pp. MIT Press. $12.50.

The author of this systems analysis study of urban development assumes that by itself the human mind is incapable of grasping the multitude of interactions among the variables—industries, housing and people—which constitute the dynamics of a complex urban system. Thus, using a model of an urban area simulated by a digital computer to measure the results of various existing and suggested programs, the author concludes that the means most likely to provide upward economic mobility for the underemployed in American cities is to discontinue building low-income housing and concentrate instead on replacing slums with new industries and businesses. Such conclusions will cause some readers to question the market-regulating bias of the author and the lack of any real historical perspective in his systems analysis model.

PARK DIXON GOIST, *Case Western Reserve University*

W. B. FOWLER, *British-American Relations, 1917-1918: The Role of Sir William Wiseman*. 334 pp. Princeton University Press. $9.50.

In brisk, unobtrusive prose, Fowler describes the feat of Sir William Wiseman, a 32-year-old British intelligence officer, in converting his almost fortuitous acquaintance with Colonel House into a vital wartime liaison between Lloyd George's London and Wilson's Washington. House and Wiseman, an uncommonly compatible team, together smoothed impressive obstacles to cooperation among the proud principals in the anti-German coalition. In the absence of effective formal structures of wartime diplomacy, British impatience with the pace of American military and fiscal assistance—combined with Wilson's studied detachment from the long-range political goals of the Allies—made constant demands on the vast fund of empathy between the two nations. Exploiting House's trust, and Wilson's too (a lengthy Appendix records Wiseman's wartime conversations with the President), the young Englishman served his nation mainly by his shrewd estimates of American motives and his cautionary advice against headstrong British behavior. Throughout he insisted on the need for accommodation to the new fact of American power, as well as to the peculiarities of Wilsonian idealism, in order to preserve and realize essential British interests.

GEOFFREY BLODGETT, *Oberlin College*

ALFRED FRANKENSTEIN, *After the Hunt: William Harnett and Other American Still Life Painters 1870-1900*. Rev. ed. 201 pp. University of California Press. $16.50.

Nearly 25 years ago Frankenstein began his study of the *trompe l'oeil* paintings of William Harnett. The present book describes the extended investigations by which he established the definitive criteria for the genuine Harnetts, separated the true Harnetts from the many forgeries and tracked down the work of other illusionistic still-life artists of the late 19th

THE WORLD SYSTEM

Models, Norms, Applications

Edited by ERVIN LASZLO

GEORGE BRAZILLER *New York*

$7.95

THE WORLD SYSTEM
Models • Norms • Variations
Edited by ERVIN LASZLO

"In the remaining decades of this century, mankind's problems will be increasingly complex in detail and global in scope. They will also be increasingly critical for human survival and civilization. Any attempt to isolate issues and apply short-range remedies will continue to fail by reason of the growing interdependency of all vital processes on this planet.

"General systems theory is the branch of science specifically designed to cope with complexity. It makes sense to attack our present and future problems through concepts and principles developed in this theory. And it is not accidental that more scientists and humanists are now turning to systems theory for solutions than ever before; whenever problems are due to interrelated processes on multiple levels, the systems approach has a selective advantage over all others.

" 'World system' is the conceptualization fittest to handle mankind's current needs. Different contents can be assigned to this concept, but its basic scope and nature is clear: the world system is that sphere of multi-level interdependencies which unites the planet's human population with its organic and inorganic environment. The future of this system is at stake; and the resolution of the problems threatening it determines the fate of mankind, its culture and civilization.

"World system modeling is a new art. It makes use of scientific data,

continued on back flap

continued from front flap

system dynamics principles, and computer simulated projections. In its present stage of development (represented by the work of Forrester, Meadows, and collaborators), it raises a multitude of fundamental issues. Foremost among these are questions concerning the completeness of the variables, the accuracy of the represented system dynamics, the incorporation of normative or 'soft' data, the evaluation of the findings, the uses and misuses of existing models, the conceptual and behavioral reorientation presupposed or suggested by the models, and the understanding of the methods and principles by which they can be further refined.

"A significantly broad range of issues of current relevance is discussed here, from differing perspectives and often with divergent consequences. But the discussions are united in their diversity through the common language of general systems theory and the common aim of developing a humanistic body of scientific knowledge."

ERVIN LASZLO, Professor of Philosophy at the State University of New York at Geneseo, is Editor of the International Library of Systems Theory and Philosophy, and chairman of the Northeast Division of the Society of General Systems Research. His books include *The Systems View of the World* and *Introduction to Systems Philosophy*.

GEORGE BRAZILLER
One Park Avenue, New York 10016

JACKET DESIGN BY HELEN IRANYI

ACKNOWLEDGMENT

The editor and publisher wish to thank the following for permission to reprint Figures 1–4 in the essay titled, "Some Political Implications of the Forrester Model" by Alastair M. Taylor:

Prentice-Hall for Figure 1, from *A Framework for Political Analysis*, by David Easton © 1965.

Alfred A. Knopf, Inc. for Figure 2, from *Politics Among Nations*, by Hans Morgenthau © 1964, Third Edition.

John Wiley and Sons, Inc. for Figures 3 and 4, from *Systematic Political Geography*, by H. J. de Blij © 1967.

For information address the publisher:
George Braziller, Inc.
One Park Avenue, New York, N.Y. 10016

Standard Book Number: 0-8076-0695-2, cloth
0-8076-0696-0, paper

Library of Congress Catalog Card Number: 73-79050

First Printing
Printed in the United States of America

Preface

IN THE remaining decades of this century, mankind's problems will be increasingly complex in detail and global in scope. They will also be increasingly critical for human survival and civilization. Never before have so many people faced so many problems of such great complexity. Any attempt to isolate issues and apply short-range remedies will continue to fail by reason of the growing interdependency of all vital processes on this planet.

General systems theory is the branch of science specifically designed to cope with complexity. It makes sense to attack our present and future problems through concepts and principles developed in this theory. And it is not accidental that more scientists and humanists are now turning to systems theory for solution than ever before: whenever problems are due to interrelated processes on multiple levels, the systems approach has a selective advantage over all others.

"World system" is the conceptualization fittest to handle mankind's current needs. Different contents can be assigned to this concept, but its basic scope and nature is clear: the world system is that sphere of multilevel interdependencies which unites the planet's human population with its organic and inorganic environment. The future of this system is at stake; and the resolution of the problems threatening it determines the fate of mankind, its culture and civilization.

World system modeling is a new art. It makes use of scientific data, system dynamics principles, and computer-simulated projections. In its present stage of development (represented by the work of Forrester, Meadows, and collaborators), it raises a multitude of fundamental issues. Foremost among these are questions concerning the completeness of the variables, the accuracy of the represented system dynamics, the incorporation of normative or "soft" data, the evaluation of the findings, the uses and misuses of existing models, the conceptual and behavioral reorientation presupposed or suggested by the models, and the understanding of the methods and principles by which they can be further refined.

These are among the issues considered by contributors to the present volume. More specifically, Margaret Mead and Ervin Laszlo debate the practical use and effect of existing models; Alastair Taylor and Richard Falk explicate their implications for political thought and action; Henryk Skolimowski and Albert Wilson investigate the presuppositions of world system theories and trace the shift in scientific modes of thinking: Håkan Törnebohm outlines the structure of inquiry in studying research itself, i.e., "studies of studies"; and Ralph Burhoe elucidates the roles and functions of human values in the world system and calls attention to their isomorphy with the concepts and precepts of traditional religions. Jay Forrester assesses the problems and potentials of world system studies and replies to the main lines of criticism.

The papers fall into two broad classes: one class moves on the interface between theory and practice and concerns itself with defining the nature of the desired models and the range of their applications. Papers in this class comprise the material of Part One. The other class penetrates to the interface between theory and metatheory, examining the norms, methods, and presuppositions which guide world systems model-building. Papers in that class are grouped in Part Two.

A significantly broad range of issues of current relevance is discussed here, from differing perspectives and often with divergent consequences. But the discussions are united in their diversity through the common language of general systems theory and the common aim of developing a humanistic body of scientific knowledge. These are basic characteristics of systems philosophy, as it informs the thinking of the writer and the spirit of the International Library of Systems Theory and Philosophy. This volume resulted from the first of an annual series of Systems Philosophy Symposia,* devoted to the multidisciplinary discussion, through general systems theory, of current topics of human and philosophical interest.

E. L.

The Center of International Studies,
Princeton University

* Held September 29 and 30, 1973, at the State University of New York at Geneseo.

Contents

PART TWO

THEORY AND METATHEORY: NORMS, METHODS, AND PRESUPPOSITIONS

9

Comment

JAY W. FORRESTER

THE papers in this volume present fascinating, diverse, and informative viewpoints on how we may better model and thereby understand the world of which we are a part. With most of the comments I agree. Where I differ, misunderstandings seem to be the issue rather than incompatible philosophy.

Through the papers runs a persistent uneasiness about the apparent absence of psychological, sociological, and political variables in models such as that in my *World Dynamics* and the one described by Meadows in *Limits to Growth.* I say there is only an apparent absence because, at the implicit level, such variables are strongly present. In the models, the humanists miss a connecting linkage in terms of human values and psychological forces just as the economists miss a price system. But both changing values and changing prices are intervening variables in real life that connect the nature of the world to human reactions. They are not missing from the model but are swept up in the high degree of aggregation until their detailed terminology is submerged.

We see in the United States in this decade a falling birth rate. Some have suggested that the falling birth rate lies outside the scope of the present models and is to be attributed to the social and psychological variables that have been omitted. But from whence come the influences leading to the social and value changes? Are they not from the sense of crowding, from the material standard of living that is now absurdly high for a substantial percentage of the population, from the highway congestion, from the pollution, and, in short, from the pressures that are being reflected back from the natural barriers as population and industrialization begin to impinge on the limits of our surroundings? I suggest that psychological attitudes and social norms are shaped by physical circumstances and are the intervening variables between the condition of the world and the human responses. In the *World Dynamics* model, birth rate falls as crowding increases; the numerical values are such that the effect clearly does not come from physical crushing; the effect is primarily psychological. The effect of food on population need not be physical starvation, it can be the threat of hunger.

The issue here is more methodological than philosophical. One must always compromise between simplicity and completeness in constructing models. There is no right answer except in the context of the purpose of the particular model. The world models have focused on the broad sweep of major forces. At a later stage it will be appropriate to insert more connecting tissue. This is not to suggest that details of the connecting tissue will have no effect. The model in *World Dynamics*, by omitting the level variables that represent human values and attitudes, is not omitting such variables but is saying that the delays in accomplishing value change can be neglected for the particular purposes. Were those delays explicitly inserted, additional dynamic interactions would emerge. The changes would probably lead to worsening an already forbidding glimpse of the future. Far from providing a solution to the problems of mankind, the insertion of the psychological and political delays between the world condition and the human response would lead to more overshoot of population beyond the carrying capacity of the globe.

Psychological and social variables have not been omitted. They are subsumed in the variables already present. To the extent that we care to state a hypothesis about such intervening variables, they can be readily included in systems dynamics models. Such has been done in models more complex than represented in *World Dynamics*.

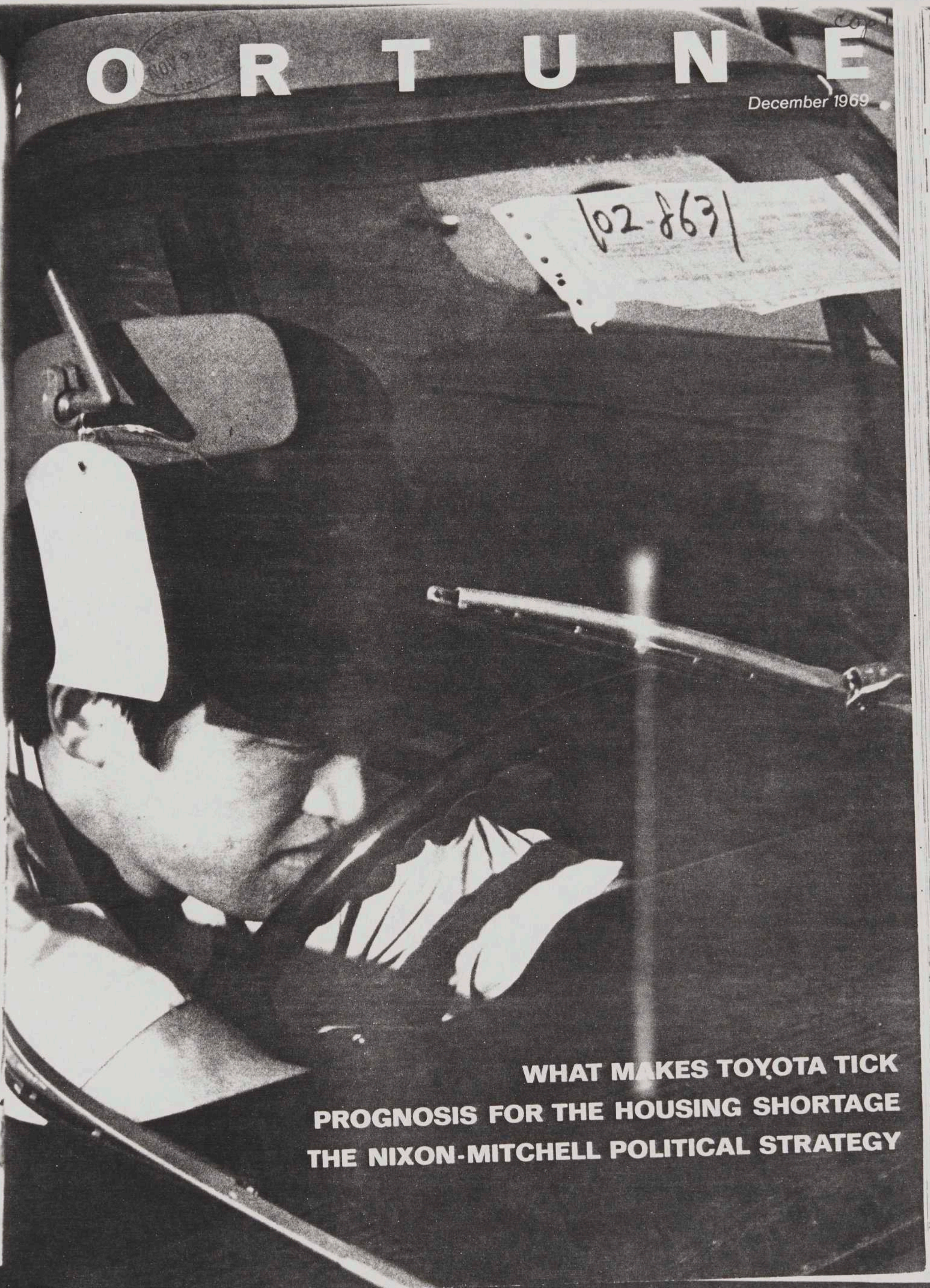
Another concern expressed about the world models relates to the aggregation of real variables. For example, the developed and underdeveloped nations are not separated. Here again the answer lies in purpose. If one were dealing with the relative struggle between the two groups, or if he were examining whether or not the income gap would be closed, the two would need to be separated and the differential forces between them would need to be represented. However, if the emphasis is on the total loading of the environment by the total population of the world and its total capital plant, then the issues arise not from the distinctions between but from the sum of the two. One must understand the models well enough to know the areas for which they are usable. Every practical model will have limits beyond which it is not useful. The proper attitude toward the world models is to look for what they can teach us and identify areas of inapplicability only so that they will not be used for the wrong purposes.

I was especially interested in the changing political perspectives outlined by Alastair Taylor as they have evolved in response to changing technology of power. The scope of power has gradually expanded from the family to the tribe, city, country, and now to the national alliances. But what does an understanding of the world system lead to? Will it be "one world" and an effective world government, or will the pendulum swing back to the independent nation?

At what level is the compromise to be made between population density and nature? We face the trade-off between quantity and quality. Every country is capable of supporting a sufficiently small population at a high standard of living and quality of life. But one of the freedoms is the freedom to choose between the size of the population and the conditions under which that population lives. Is the compromise to be made by a world authority that imposes the same balance on every culture? If not the choice must be decentralized. If the choice is made differently in various countries, then the standard of living cannot be the same. Some countries will choose a modest population coupled with a high material standard of living, national strength, and the disadvantage of maintaining the self-discipline to limit population. Other countries will, in effect, choose to avoid the trauma of self-discipline or perhaps belong to cultures that prefer a higher population density and will accept the corresponding reduction in material standard of living and in national strength.

Such reasoning suggests that each country must live within its own capability to a greater extent than today. The present accelerating pace of international trade is a device to allow growth to continue until the entire world simultaneously approaches shortages of all traded goods. Then we are apt to see hoarding for the future by resource-supplying nations and a consequent contraction of economic activity and standard of living in those countries that have expanded beyond their internal means.

The urgent task now is to face such issues squarely and to make an estimate of the most viable and realistic future. Much of today's actions are based on visions of impossible future utopias. Unrealistic expectations are a poor foundation on which to build the future. Hard choices must be made. The future is not to be free of pressures, but we have a range of choice in the combination of physical, social, psychological, and moral pressures under which we will live. Systems dynamics modeling can cope with such considerations as rapidly as we can think through the issues and identify the important relationships. From such an effort it will be possible to answer many of the questions raised in this book.


ORTUNE
December 1969
102-8631
WHAT MAKES TOYOTA TICK
PROGNOSIS FOR THE HOUSING SHORTAGE
THE NIXON-MITCHELL POLITICAL STRATEGY

FORTUNE

December, 1969

Fortune, December, 1969. Vol. LXXX, No. 7. Issued monthly, except two issues in May and August, by Time Inc., 540 N. Michigan Ave., Chicago, Illinois 60611. Second-class postage paid at Chicago, Illinois, and at additional mailing offices. Subscriptions: U.S., U.S. possessions, and Canada: one year $14; elsewhere, one year $20. Single copies $1.50. Address all subscriptions and correspondence concerning them to FORTUNE, 540 N. Michigan Ave., Chicago, Illinois 60611. Principal Offices: Time & Life Building, Rockefeller Center, New York, N.Y. 10020. James R. Shepley, President, Richard B. McKeough, Treasurer; John F. Harvey, Secretary. Authorized as second-class mail by the Post Office Department, Ottawa, Canada, and for payment of postage in cash. Member, Audit Bureau of Circulation.

Picture credits page 135

Overlooked Reasons for Our Social Troubles

by Jay W. Forrester

In this department last month, Professor John F. Kain of Harvard presented a critical commentary on an unusual and controversial book: Urban Dynamics *(M.I.T. Press), which reports strikingly unconventional conclusions derived from a computer study of a hypothetical urban area. This month the author of* Urban Dynamics, *a professor of management at M.I.T.'s Sloan School of Management, presents his own account of what his book has to tell us.*

From the city, the economy, and the environment come rising pressures on our social systems. Citizens, corporate executives, mayors, and national leaders strive to solve the problems, only to see matters worsen. Obviously, we do not understand how the structures and policies of our systems interact to *create* the troubles that surround us. In *Urban Dynamics*, using computer methods developed in my *Industrial Dynamics* (M.I.T. Press, 1961), I undertook to show how the structures and policies of an urban area turn growth into decline. Several popular proposals for remedying urban troubles (job training, financial subsidy to a city, and low-cost-housing programs) proved to lie somewhere between neutral and detrimental in their effects on a declining urban area. But policies directed to rebalancing population categories and jobs can start an internal revival, with increased upward mobility for low-income groups. This approach to policy design is applicable to any of our social systems.

Many people recoil at the thought of anyone's designing social systems. But we have no choice. We already live in social systems that have been designed—by national and state constitutions, laws, tax regulations, and traditions. If we lament the decline of our cities, the pace of inflation, or the increases in environmental pollution, we are asserting a preference for a different design. Corporate executives and legislative bodies design our systems by establishing policies and laws, but with only intuition and experience to guide their choices. Intuition and experience are demonstrably unreliable in efforts to cope with the complex systems that surround us.

New potential for enlightened choices

It is inhumane to go on trying to achieve humane objectives by means of policies that worsen the conditions they are meant to improve. In *Urban Dynamics*, I try to indicate ways of improving the functioning of social systems, which means improving the living conditions of human beings and making it possible for them to realize their potentialities more fully. The point that a cold-blooded computer model can have humane uses has escaped some readers of my book, but not all. Erich Jantsch, a scientist who specializes in long-range forecasting and planning, wrote in the British journal *Futures:* "In reality, Urban Dynamics—or Social Dynamics, as the method might be called even more generally—enhances the role of human creativity and inventiveness in an unprecedented way. By studying the consequences of alternative courses of action for entire social systems, man acquires a new potential for making enlightened choices . . ."

As everyone sees, our present social systems exhibit disturbing trends and stresses. Grave doubts surround the management of corporations, the environment, and the economy. For example, we need better to interrelate taxation, government expenditure, fiscal and monetary policy, economic output, unemployment, and inflation. Past failures in economic analysis and economic policy recommendations have been blamed on inadequate data, but a much more likely explanation lies in the inappropriate structures of the models used, the timid and fragmentary approaches to analysis, and the willingness merely to explain the past rather than try to understand the future. Likewise, the possibilities for sudden and irreversible changes in the ecological relationships of man to nature can be effectively explored neither by discussion nor by analysis of historical data. We must use the more powerful approaches that are now becoming available for dealing with our complex systems. We must, and can, anticipate changes that will evolve from presently known structures and processes, but that have no historical precedents.

> *"We must, and can, anticipate changes that will evolve from presently known structures and processes, but that have no historical precedents."*

Consequences burst forth

The pressures on our society will continue to rise until the fundamental questions can no longer be ignored. Through all of recorded history, our traditions, laws, and aspirations have been based on the dynamics of growth—growth in geographical frontiers, scientific knowledge, standard of living, population, and pollution. Our social systems contain the positive-feedback processes that generate exponential growth. Exponential growth has the characteristic that in its early stages it seems unimportant, appears to be getting nowhere, and is largely ignored. But then, in the last two or three doublings in growth, the process passes from insignificance to domination. The consequences of a long history of exponential growth suddenly appear to burst forth on an unprepared society.

Exponential growth cannot continue indefinitely, otherwise it would engulf the earth. *Urban Dynamics* shows the precipitous fall in standard of living and the changes in population mix that occur in the conventional urban area as it moves out of its growth stage into equilibrium. Similar prospects for major change and stress lie before our larger social systems. Growth will cease. Geographical frontiers have been exhausted. Natural resources are being used far faster than nature is recreating them. Ecological considerations probably exclude the possibility of even the present world population rising to the standard of living of the Western industrialized nations, so rising economic expectations will inevitably be frustrated, either in local stagnation or in a worldwide ecological disaster.

The present social malaise at all organizational levels is the first evidence of far greater pressures that will be generated by the worldwide suppression of growth processes. As with the urban area, there are many routes into the inevitable equilibrium. As we move toward that condition, we must, for our preservation, make wise choices about the kind of static earth we want, and adopt wise policies for attaining it.

The deceptiveness of systems

Urban Dynamics describes various characteristics of complex systems that lead us into self-defeating policies. These characteristics were first identified to explain, in management systems, the recurring choice of corporate policies that worsen the very troubles they are intended to correct. The same kinds of influences were rediscovered on the urban scene. They appear to be common to all our social systems.

▸ Complex systems are counterintuitive. They respond to policy changes in directions opposite to what most people expect. We develop experience and intuition almost entirely from contact with simple systems, where cause and effect are closely related in space and time. Complex systems behave very differently.

▸ Complex systems actively resist most policy changes. A new policy warps the entire system slightly, and so it presents a new ensemble of perceived information; the new information is processed through the new policy to produce nearly the old result.

▸ But influence points exist, often where least expected and often with a direction of influence opposite to that anticipated. These pressure points radiate new information streams that, when processed even through old attitudes and policies, produce new results.

▸ Complex systems tend to counteract programs that attempt to supplement and add to an action stream already in the system. For example, in *Urban Dynamics* a job-training program fails because the reactions within the system reduce the natural upward economic mobility, increase downward mobility, attract the unskilled, and in the end slightly enlarge the underemployed population.

▸ In a complex system the short-term response to a policy change is often opposite to the long-term effect. This treacherous behavior beguiles the executive and the politician into a series of steps, each appearing beneficial and each leading to deeper long-term difficulty.

▸ A system contains internal dynamic mechanisms that *produce* the undesirable behavior. If we ignore fundamental causes and simply try to overwhelm the symptoms, we pit great forces against one another, expending our energy to no avail.

▸ In a complex system, certain pressures go with each mode of behavior. To sustain a particular mode we must accept the corresponding pressures. The common tendency to alleviate one squeaky wheel after another constitutes incremental redesign that can move the system toward an undesirable and nearly irreversible mode of behavior.

In his review of *Urban Dynamics*, Professor Kain concentrates on another aspect of the book, its model, or theory, of system behavior in an urban area. Details of such a model change continuously as one addresses different questions or tests alternative assumptions. Although model details are of less long-term significance than method or the general character of systems, Kain worries details. If his doubts were justified, that might affect the particular conclusions of the book, but not the method.

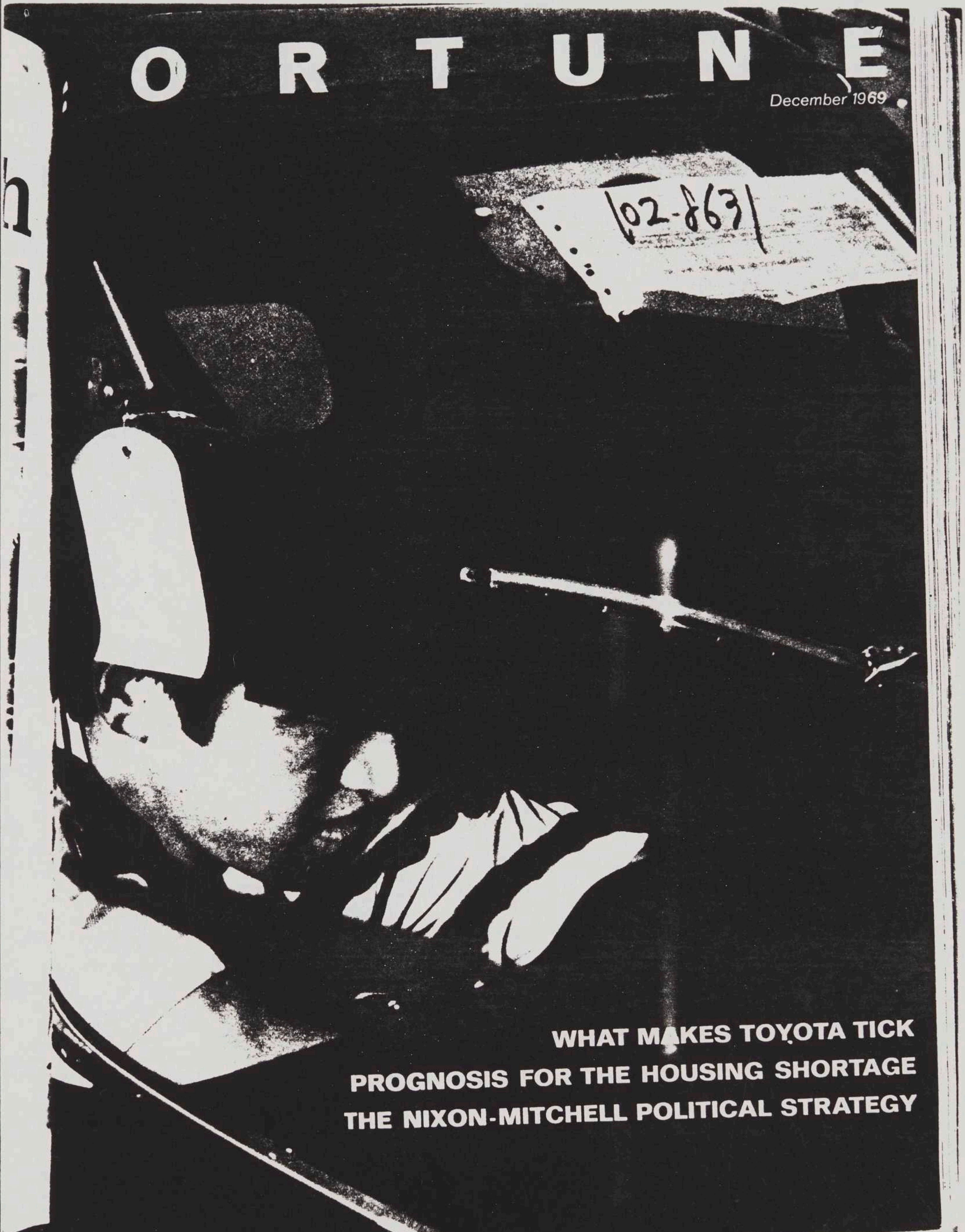
Almost the only concrete, testable statement Kain offers has to do with his doubtful premise that outside financial subsidy to a city would be used to reduce taxes rather than increase expenditure. He says: "If instead Forrester had used the outside support to reduce city taxes, the net effects would have been favorable to the hypothetical city." Here he is speaking explicitly of what the model will do. Only minutes are needed to make the suggested change in the model and test his assertion. This was done. There is no significant improvement. So even this unlikely use of a subsidy—to reduce taxes—is a waste of resources in the hypothetical city.

Regrettable perhaps, but inescapable

Even here where he has complete knowledge about the laboratory system and its governing policies, Kain should not be criticized for being unable, on the basis of intuition and judgment, to anticipate the effect of a policy change. But Kain and the social scientists he represents can be criticized for asserting with assurance the consequences of policy recommendations in our real-life systems when it has been repeatedly shown that intuition and judgment cannot yield such certainty even in the laboratory and with perfect information. Only after trying the policy change in a properly constructed, dynamic simulation model should one speak confidently about the consequences.

From his economist's viewpoint, Professor Kain primarily saw tax considerations in the book. As a test of his assertion that tax rates powerfully influence employment and population behavior in the model, the tax rate was changed to be constant and equal to the average outside tax rate. This change makes only a small improvement in the depressed condition of the city, an improvement not at all comparable to what results from the revival policies discussed in the book. Furthermore, the constant tax levy does not reduce the efficacy of the suggested revival policies, so no conclusions in the book would be altered.

Again we see the danger of continuing to base political decisions on intuitive judgments and "conventional wisdom." As I noted above, complex systems are counterintuitive. This perhaps regrettable but nonetheless inescapable fact is a main source of our present discontents. END


ORTUNE
December 1969
WHAT MAKES TOYOTA TICK
PROGNOSIS FOR THE HOUSING SHORTAGE
THE NIXON-MITCHELL POLITICAL STRATEGY

FORTUNE

December, 1969

Fortune, December, 1969. Vol. LXXX, No. 7. Issued monthly, except two issues in May and August, by Time Inc., 540 N. Michigan Ave., Chicago, Illinois 60611. Second-class postage paid at Chicago, Illinois, and at additional mailing offices. Subscriptions: U.S., U.S. possessions, and Canada: one year $14; elsewhere, one year $20. Single copies $1.50. Address all subscriptions and correspondence concerning them to FORTUNE, 540 N. Michigan Ave., Chicago, Illinois 60611. Principal Offices: Time & Life Building, Rockefeller Center, New York, N.Y. 10020. James R. Shepley, President, Richard B. McKeough, Treasurer; John F. Harvey, Secretary. Authorized as second-class mail by the Post Office Department, Ottawa, Canada, and for payment of postage in cash. Member, Audit Bureau of Circulation.

Picture credits page 135

Overlooked Reasons for Our Social Troubles

by Jay W. Forrester

In this department last month, Professor John F. Kain of Harvard presented a critical commentary on an unusual and controversial book: Urban Dynamics *(M.I.T. Press), which reports strikingly unconventional conclusions derived from a computer study of a hypothetical urban area. This month the author of* Urban Dynamics, *a professor of management at M.I.T.'s Sloan School of Management, presents his own account of what his book has to tell us.*

From the city, the economy, and the environment come rising pressures on our social systems. Citizens, corporate executives, mayors, and national leaders strive to solve the problems, only to see matters worsen. Obviously, we do not understand how the structures and policies of our systems interact to *create* the troubles that surround us. In *Urban Dynamics*, using computer methods developed in my *Industrial Dynamics* (M.I.T. Press, 1961), I undertook to show how the structures and policies of an urban area turn growth into decline. Several popular proposals for remedying urban troubles (job training, financial subsidy to a city, and low-cost-housing programs) proved to lie somewhere between neutral and detrimental in their effects on a declining urban area. But policies directed to rebalancing population categories and jobs can start an internal revival, with increased upward mobility for low-income groups. This approach to policy design is applicable to any of our social systems.

Many people recoil at the thought of anyone's designing social systems. But we have no choice. We already live in social systems that have been designed—by national and state constitutions, laws, tax regulations, and traditions. If we lament the decline of our cities, the pace of inflation, or the increases in environmental pollution, we are asserting a preference for a different design. Corporate executives and legislative bodies design our systems by establishing policies and laws, but with only intuition and experience to guide their choices. Intuition and experience are demonstrably unreliable in efforts to cope with the complex systems that surround us.

New potential for enlightened choices

It is inhumane to go on trying to achieve humane objectives by means of policies that worsen the conditions they are meant to improve. In *Urban Dynamics*, I try to indicate ways of improving the functioning of social systems, which means improving the living conditions of human beings and making it possible for them to realize their potentialities more fully. The point that a cold-blooded computer model can have humane uses has escaped some readers of my book, but not all. Erich Jantsch, a scientist who specializes in long-range forecasting and planning, wrote in the British journal *Futures:* "In reality, Urban Dynamics—or Social Dynamics, as the method might be called even more generally—enhances the role of human creativity and inventiveness in an unprecedented way. By studying the consequences of alternative courses of action for entire social systems, man acquires a new potential for making enlightened choices . . ."

As everyone sees, our present social systems exhibit disturbing trends and stresses. Grave doubts surround the management of corporations, the environment, and the economy. For example, we need better to interrelate taxation, government expenditure, fiscal and monetary policy, economic output, unemployment, and inflation. Past failures in economic analysis and economic policy recommendations have been blamed on inadequate data, but a much more likely explanation lies in the inappropriate structures of the models used, the timid and fragmentary approaches to analysis, and the willingness merely to explain the past rather than try to understand the future. Likewise, the possibilities for sudden and irreversible changes in the ecological relationships of man to nature can be effectively explored neither by discussion nor by analysis of historical data. We must use the more powerful approaches that are now becoming available for dealing with our complex systems. We must, and can, anticipate changes that will evolve from presently known structures and processes, but that have no historical precedents.

> *"We must, and can, anticipate changes that will evolve from presently known structures and processes, but that have no historical precedents."*

Consequences burst forth

The pressures on our society will continue to rise until the fundamental questions can no longer be ignored. Through all of recorded history, our traditions, laws, and aspirations have been based on the dynamics of growth—growth in geographical frontiers, scientific knowledge, standard of living, population, and pollution. Our social systems contain the positive-feedback processes that generate exponential growth. Exponential growth has the characteristic that in its early stages it seems unimportant, appears to be getting nowhere, and is largely ignored. But then, in the last two or three doublings in growth, the process passes from insignificance to domination. The consequences of a long history of exponential growth suddenly appear to burst forth on an unprepared society.

Exponential growth cannot continue indefinitely, otherwise it would engulf the earth. *Urban Dynamics* shows the precipitous fall in standard of living and the changes in population mix that occur in the conventional urban area as it moves out of its growth stage into equilibrium. Similar prospects for major change and stress lie before our larger social systems. Growth will cease. Geographical frontiers have been exhausted. Natural resources are being used far faster than nature is recreating them. Ecological considerations probably exclude the possibility of even the present world population rising to the standard of living of the Western industrialized nations, so rising economic expectations will inevitably be frustrated, either in local stagnation or in a worldwide ecological disaster.

The present social malaise at all organizational levels is the first evidence of far greater pressures that will be generated by the worldwide suppression of growth processes. As with the urban area, there are many routes into the inevitable equilibrium. As we move toward that condition, we must, for our preservation, make wise choices about the kind of static earth we want, and adopt wise policies for attaining it.

The deceptiveness of systems

Urban Dynamics describes various characteristics of complex systems that lead us into self-defeating policies. These characteristics were first identified to explain, in management systems, the recurring choice of corporate policies that worsen the very troubles they are intended to correct. The same kinds of influences were rediscovered on the urban scene. They appear to be common to all our social systems.

▸ Complex systems are counterintuitive. They respond to policy changes in directions opposite to what most people expect. We develop experience and intuition almost entirely from contact with simple systems, where cause and effect are closely related in space and time. Complex systems behave very differently.

▸ Complex systems actively resist most policy changes. A new policy warps the entire system slightly, and so it presents a new ensemble of perceived information; the new information is processed through the new policy to produce nearly the old result.

▸ But influence points exist, often where least expected and often with a direction of influence opposite to that anticipated. These pressure points radiate new information streams that, when processed even through old attitudes and policies, produce new results.

▶ Complex systems tend to counteract programs that attempt to supplement and add to an action stream already in the system. For example, in *Urban Dynamics* a job-training program fails because the reactions within the system reduce the natural upward economic mobility, increase downward mobility, attract the unskilled, and in the end slightly enlarge the underemployed population.

▶ In a complex system the short-term response to a policy change is often opposite to the long-term effect. This treacherous behavior beguiles the executive and the politician into a series of steps, each appearing beneficial and each leading to deeper long-term difficulty.

▶ A system contains internal dynamic mechanisms that *produce* the undesirable behavior. If we ignore fundamental causes and simply try to overwhelm the symptoms, we pit great forces against one another, expending our energy to no avail.

▶ In a complex system, certain pressures go with each mode of behavior. To sustain a particular mode we must accept the corresponding pressures. The common tendency to alleviate one squeaky wheel after another constitutes incremental redesign that can move the system toward an undesirable and nearly irreversible mode of behavior.

In his review of *Urban Dynamics*, Professor Kain concentrates on another aspect of the book, its model, or theory, of system behavior in an urban area. Details of such a model change continuously as one addresses different questions or tests alternative assumptions. Although model details are of less long-term significance than method or the general character of systems, Kain worries details. If his doubts were justified, that might affect the particular conclusions of the book, but not the method.

Almost the only concrete, testable statement Kain offers has to do with his doubtful premise that outside financial subsidy to a city would be used to reduce taxes rather than increase expenditure. He says: "If instead Forrester had used the outside support to reduce city taxes, the net effects would have been favorable to the hypothetical city." Here he is speaking explicitly of what the model will do. Only minutes are needed to make the suggested change in the model and test his assertion. This was done. There is no significant improvement. So even this unlikely use of a subsidy—to reduce taxes—is a waste of resources in the hypothetical city.

Regrettable perhaps, but inescapable

Even here where he has complete knowledge about the laboratory system and its governing policies, Kain should not be criticized for being unable, on the basis of intuition and judgment, to anticipate the effect of a policy change. But Kain and the social scientists he represents can be criticized for asserting with assurance the consequences of policy recommendations in our real-life systems when it has been repeatedly shown that intuition and judgment cannot yield such certainty even in the laboratory and with perfect information. Only after trying the policy change in a properly constructed, dynamic simulation model should one speak confidently about the consequences.

From his economist's viewpoint, Professor Kain primarily saw tax considerations in the book. As a test of his assertion that tax rates powerfully influence employment and population behavior in the model, the tax rate was changed to be constant and equal to the average outside tax rate. This change makes only a small improvement in the depressed condition of the city, an improvement not at all comparable to what results from the revival policies discussed in the book. Furthermore, the constant tax levy does not reduce the efficacy of the suggested revival policies, so no conclusions in the book would be altered.

Again we see the danger of continuing to base political decisions on intuitive judgments and "conventional wisdom." As I noted above, complex systems are counterintuitive. This perhaps regrettable but nonetheless inescapable fact is a main source of our present discontents. END

JWF 3/4/71

The Commonwealth of Massachusetts

University of Massachusetts

Amherst 01002

School of Business Administration
Department of Management

March 2, 1971

Professor Jay W. Forrester
Sloan School of Management
Massachusetts Institute of Technology
Cambridge, Massachusetts 02139

Dear Professor Forrester:

We are preparing a book on Operations and Systems Analysis: A Simulation Approach and are selecting some of the outstanding works representing various types of models and approaches as its reading materials. Your paper, "Modeling the Dynamic Processes of Corporate Growth," in the <u>Proceedings of the IBM Scientific Computing Symposium on Simulation Models and Gaining</u>, provides an excellent example of modeling the corporate growth using the Industrial Dynamic concepts and principles. We would appreciate your permission to reproduce this article in our book which is scheduled for publication by Allyn and Bacon, Inc. of Boston in 1972. IBM has granted us the permission to use the material provided you agree to do the same. It is understood, of course, that full credit will be given to you and your publisher.

I am enclosing a permission form for your convenience. Your granting of permission at your earliest convenience will be greatly appreciated.

Sincerely,

Gordon Chen

Gordon Chen
Associate Professor of Management

Eugene E. Kaczka G.C.

Eugene E. Kaczka
Assoc. Professor of Management Science

GC/EEK
lcm

Enclosures

Professors Gordon Chen and Eugene Kaczka
School of Business Administration
University of Massachusetts
Amherst, Massachusetts 01002

Gentlemen:

I (We) hereby grant the permission requested for the use of the material listed below in your book.

By JWF

Date Mar. 5 '71 Firm

Author	Title
Jay W. Forrester	"Modeling the Dynamic Processes of Corporate Growth" *Proceedings of the IBM Scientific Computing on Simulation Models and Gaining*, IBM Corporation, New York, 1966.

10/7/75

JWF/FYI

Massachusetts Institute of Technology
Alfred P. Sloan School of Management
50 Memorial Drive
Cambridge, Massachusetts, 02139

September 22, 1975

Customer Service
McGraw-Hill Book Company
1221 Avenue of the Americas
New York, NY 10020

Gentlemen:

We are interested in purchasing a copy of Schumpeter's Business Cycles: A Theoretical, Historical, and Statistical Analysis of the Capitalist Process, 6th edition. The book was published in 1939.

At your earliest opportunity please advise us of the availability and cost of this book. Thank you very much.

Sincerely,

Emaline Cornett

(Miss) Emaline Cornett
Secretary to Professor Forrester

eac

The McGraw-Hill Bookstores

McGraw-Hill Book Company

1221 Avenue of the Americas
New York, New York 10020
Telephone 212/997-4100

Dear Customer:

If you wish to order any of the books listed on your attached letter, please send a check or money order which includes the price of the book, tax if applicable (8% in New York State, 5% in New Jersey) and postage/handling (~~45¢~~ .55 per book) for shipment in the United States.

If ordering from outside the United States, please send check or money order in U. S. dollars only.

PLEASE RETURN THIS CORRESPONDENCE WITH YOUR CHECK.

Sincerely yours,

McGraw-Hill Bookstore
Mail Order Division (C-2)
1221 Avenue of the Americas
New York, New York 10020

Attachment

Additional Information (if any): A Theoretical, Historical, and Statistical Analysis of the Capitalist Process, 6th ed

no Listing.

McGraw-Hill Book Company
1221 Avenue of the Americas
New York, New York 10020

NEW YORK OCT-2'75 N.Y.
U.S. POSTAGE 10
PB 521967

E40-253

Massachusetts Institute of Technology
Alfred P. Sloan School of Management
50 Memorial Drive
Cambridge, Massachusetts, 02139

D

Att. Emaline Cornett

JB

9490-044 (REV. 5-72)

From nathan

Mandel, Ernest. Late Capitalism chapter 4
London NLB Atlantic Highlands Humanities Press

Garvey, "Kondratieff's Theory of Long Waves"
RES 1943 Volume XXV

Rostow W.W. "The developing World in the Fifth Kondratieff Upswing" Annals AAPSS 420
July 1975

Abramovitz, M. "The Nature and Significance of Kuznets Cycles" Econ. Dev + Cult. Change April 1961
9, 225-48

Bowen W.G, Berry R.A., "Unemployment Conditions + Movements of money and Wage Level" RES May 1963 45
163-72

Kondratieff, N.D. "The Long Waves in Economic Life."
RES Nov. 1935

S...k.n, G. "Business Cycles - Aren't What they Used to Be and Never Were" Lloyds Bank Review
April 1972 N. 104 p 20.

✓B. Hickman, American Econ. Review May 1963
Postwar Retardation: Another Long Swing in the Rate of Growth p 490

✓Kelley, "Demographic Cycles and Economic Growth: The Long Swing Reconsidered", The Journal of Econ. History Dec. 1969 XXIX No. 4 p. 635

✓Temin, P. "The Last Great Depression and the Present One" Lessons for the Present from the Great Depression", AER May '76

✓Gordon R.A., Business Fluctuations Harper & Row

✓Hansen, Alvin Business Cycles & National Income

✓Fusfeld, D. Economics

✓Rostow, W.W. "Kondratieff, Schumpeter, & Kuznets: Trend Periods Revisited" JEH Dec '75

✓Schumpeter, "Analysis of Economic Change", RES 1935

✓Long, "Seventy Years of Building Cycles in Manhattan" RES, 1936

✓Kuznets, Equilibrium "Economics and Business Cycle Theory" QJE May 1930.

✓Souto, Commentary on Kuznets

✓Hawley J B. "Cycles of Institutional Development in Higher Education" Draf. So. Ill. U. Carbondale.

✓Cleveland H.B. Brittain WHB. "A World Depression" Foreign Affairs Jan '75

✓ The Bank Credit Analyst. May '73 Stock Market and Business Forecast

✓Rostow WW Letter to JWF

✓ Rostow W.W. WSJ Mar 8 1977 Caught by Kondratieff

✓Levy-Pascal E. "An Analysis of the Cyclical Dynamics of Industrialized Countries" CIA

✓ ~~Kali~~ Saeed, Khalid

✓ Cecilia Wong May 1977

✓Crocker, Uriel, ✓ The Cause of Hard Times
✓ Depression in Trade and the Wages of Labor
✓ Overproduction and Commercial Distress