



Measurement Without Theory

Tjalling C. Koopmans

The Review of Economic Statistics, Vol. 29, No. 3 (Aug., 1947), 161-172.

Stable URL:

<http://links.jstor.org/sici?sici=0034-6535%28194708%2929%3A3%3C161%3AMWT%3E2.0.CO%3B2-W>

The Review of Economic Statistics is currently published by The MIT Press.

Your use of the JSTOR archive indicates your acceptance of JSTOR's Terms and Conditions of Use, available at <http://www.jstor.org/about/terms.html>. JSTOR's Terms and Conditions of Use provides, in part, that unless you have obtained prior permission, you may not download an entire issue of a journal or multiple copies of articles, and you may use content in the JSTOR archive only for your personal, non-commercial use.

Please contact the publisher regarding any further use of this work. Publisher contact information may be obtained at <http://www.jstor.org/journals/mitpress.html>.

Each copy of any part of a JSTOR transmission must contain the same copyright notice that appears on the screen or printed page of such transmission.

JSTOR is an independent not-for-profit organization dedicated to creating and preserving a digital archive of scholarly journals. For more information regarding JSTOR, please contact support@jstor.org.

MEASUREMENT WITHOUT THEORY¹

TJALLING C. KOOPMANS

THE EMPIRICAL APPROACH

WHEN Tycho Brahé and Johannes Kepler engaged in the systematic labor of measuring the positions of the planets, and charting their orbits, they started with conceptions and models of the planetary system which later proved incorrect in some aspects, irrelevant in others. Tycho always, and Kepler initially, believed in uniform circular motion as the natural basic principle underlying the course of celestial bodies. Tycho's main contribution was a systematic accumulation of careful measurements. Kepler's outstanding success was due to a willingness to strike out for new models and hypotheses if such were needed to account for the observations obtained. He was able to find simple empirical "laws" which were in accord with past observations and permitted the prediction of future observations. This achievement was a triumph for the approach in which large scale gathering, sifting, and scrutinizing of facts precedes, or proceeds independently of, the formulation of theories and their testing by further facts.

The book by Burns and Mitchell,² discussed here, approaches the problems of cyclical fluctuations in economic variables in the same empirical spirit. The book has two main purposes: first, a detailed exposition, with experimental applications, of the methods of measuring cyclical behavior, developed by the National Bureau of Economic Research; secondly, a search, with the help of these methods, for possible changes in cyclical behavior of economic variables over time, whether gradual, abrupt, in longer cycles, or otherwise.

The approach of the authors is here described as empirical in the following sense: The

various choices as to what to "look for," what economic phenomena to observe, and what measures to define and compute, are made with a minimum of assistance from theoretical conceptions or hypotheses regarding the nature of the economic processes by which the variables studied are generated.

In fact, Burns and Mitchell are more consistently empiricist than Kepler was. The latter made no secret of his predilection for the principle of circular motion until observations spoke decisively for the elliptical orbit. He held other speculative views as to the role of the five regular solids and of musical intervals in the proportions of the planetary system, which now appear as irrelevant. Burns and Mitchell do not reveal at all in this book what explanations of cyclical fluctuations, if any, they believe to constitute plausible models or hypotheses.

The undertaking commands respect, and the precedent holds great promise: For, in due course, the theorist Newton was inspired to formulate the fundamental laws of attraction of matter, which contain the empirical regularities of planetary motion discovered by Kepler as direct and natural consequences. The terms "empirical regularities" and "fundamental laws" are used suggestively to describe the "Kepler stage" and the "Newton stage" of the development of celestial mechanics. It is not easy to specify precisely what is the difference between the two stages. Newton's law of gravitation can also be looked upon as describing an empirical regularity in the behavior of matter. The conviction that this "law" is in some sense more fundamental, and thus constitutes progress over the Kepler stage, is due, I believe, to its being at once more elementary and more general. It is more elementary in that a simple property of mere matter is postulated. As a result, it is more general in that it applies to all matter, whether assembled in planets, comets, sun or stars, or in terrestrial objects — thus explaining a much wider range of phenomena.

It appears to be the intention of Burns and

¹ This article will be reprinted as part of Cowles Commission Papers, New Series, No. 25. I am indebted to several friends, including Dr. A. F. Burns, for comments on an earlier draft. These comments have helped me to bring out more clearly the issues raised in this review, for which, of course, I remain exclusively responsible. T.C.K.

² Arthur F. Burns and Wesley C. Mitchell, *Measuring Business Cycles* (National Bureau of Economic Research, Studies in Business Cycles, No. 2, New York, 1946).

Mitchell — in any case it is the opinion of the present reviewer — that their book represent an important contribution to the “Kepler stage” of inquiry in the field of economics. It is concerned exclusively with cyclical fluctuations. Its hypotheses are concerned with the character of such fluctuations, rather than with the underlying economic behavior of man.

The auspicious precedent in the history of celestial mechanics suggests that this is a promising procedure, which may expect to be rewarded in due course by further development of theory. Nevertheless, this reviewer believes that in research in economic dynamics the Kepler stage and the Newton stage of inquiry need to be more intimately combined and to be pursued simultaneously. Fuller utilization of the concepts and hypotheses of economic theory (in a sense described below) *as a part of the processes of observation and measurement* promises to be a shorter road, perhaps even the only possible road, to the understanding of cyclical fluctuations. Such a course, in addition, promises as by-products greater insight into noncyclical and even nondynamic problems of economics.

While a systematic argument in support of this position would surpass the bounds of a review, I shall attempt to adduce some of the arguments in the course of this discussion of the book. It is then my duty to point out in what respects, in my opinion, the present state of business cycle analysis differs from the situation in which Tycho and Kepler approached the phenomenon of planetary motion. I hasten to add that the parallel with the classical problems of celestial mechanics is not mentioned by, and may not have been in the minds of, the authors. It is merely the best example, known to the reviewer, of a case where the empirical approach paved the way for the discovery of fundamental laws.

The example has been selected because it is favorable to the empiricist position. Needless to say, the history of science knows of many cases in which “fundamental” hypotheses, more or less integrated into a theory of the phenomena studied, have played a much larger role. However, the spectacular success, achieved in the case here chosen as an example, has set a pattern which has ever since, con-

sciously or unconsciously, been in the minds of scientific workers in widely diverse fields.

MEASURES OF CYCLICAL “BEHAVIOR”

The authors formulate their objective in the following terms:

Whatever their working concepts,, all investigators cherish the same ultimate aim — namely, to attain better understanding of the recurrent fluctuations in economic fortune that modern nations experience. This aim may be pursued in many ways. The way we have chosen is to observe the business cycles of history as closely and systematically as we can before making a fresh attempt to explain them (p. 4).

The point of departure is a definition of business cycles, derived from experience, and to be tested in the light of further experience:

Business cycles are a type of fluctuation found in the aggregate economic activity of nations that organize their work mainly in business enterprises: a cycle consists of expansions occurring at about the same time in many economic activities, followed by similarly general recessions, contractions, and revivals which merge into the expansion phase of the next cycle; this sequence of changes is recurrent but not periodic; in duration business cycles vary from more than one year to ten or twelve years; they are not divisible into shorter cycles of similar character with amplitudes approximating their own (p. 3).

As is often the case in statistical work, a vast amount of data — represented here by (mostly monthly) observations of many economic variables over long periods — is to be reduced and summarized by computing a smaller number of “derived” measures, incorporating what is relevant and informative, omitting what is accidental or devoid of interest. The first eight chapters essentially consist in making reasoned choices as to what measures are relevant and informative. In that undertaking, the definition just quoted — itself the result of an earlier volume by Mitchell in the same series — is the main guide.

The *first* group of measures selected concerns location in time and duration. For each variable, lower and upper turning points are determined, as well as time intervals between them (expansion, contraction, trough-to-trough duration of *specific* cycles). In addition, turning points and durations are determined for

reference cycles, i.e., points around which the corresponding specific cycle turning points of a great many variables are concentrated. Leads and lags are found as differences between corresponding specific cycle and reference cycle turning points. All turning points are determined after elimination of seasonal variation but without prior trend elimination, using as much as possible monthly or otherwise quarterly data.

The *second* group of measures relates to movements of one variable within one cycle, which may be either a cycle specific to that variable, or a reference cycle. For the computation of these measures, each variable is expressed in per cent of its mean over the cycle concerned — a procedure which eliminates intercycle trend but preserves intracycle trend. For each cycle, a pattern of nine successive “standings” is then computed, i.e., a sequence of nine averages, indicated by Roman numerals, of which I, V, and IX are generally three-month averages centered at successive trough, peak, and trough months, respectively, and those numbered II, III, IV, and VI, VII, VIII are averages arising from subdivision, into three approximately equal parts, of the intermediate periods of expansion and contraction, respectively. The result is a specific cycle pattern, or a reference cycle pattern, of the variable concerned, depending on what kind of turning points were employed. These patterns are plotted on a time scale reflecting whatever inequality there is in duration between expansion and contraction. For specific cycles the following measures of amplitude are considered: “rise” ($V - I$), “fall” ($V - IX$), and “rise and fall” ($2V - I - IX$), both in absolute terms, and on a per month basis to indicate steepness of rise and fall. Reference cycle amplitudes are computed in a similar manner, but the three stages involved are not necessarily the reference stages I, V, IX, but rather such reference cycle stages (with constant Roman numeral) as are most frequently or closely coincident in timing with specific cycle stages I, V, IX.

The foregoing measures have been described for a single cycle. Averages of these measures for a sequence of cycles are likewise computed, and are qualified by presenting the average

deviation as a measure of variation between cycles.

The *third* group of measures expresses conformity of specific cycles of a variable to business cycles. These comprise ratios of average reference cycle amplitudes to average specific cycle amplitudes of the same variable, for expansions and contractions combined. They further comprise indexes of conformity expressing the proportion of all reference cycles covered in which the signs of ($V - I$), of ($V - IX$), and of ($V - I$)-per-month plus ($V - IX$)-per-month, respectively, are positive. In order to do justice to cases where specific cycles show regular lags or leads in relation to reference cycles, these measures are supplemented by similar conformity measures in which the reference cycle standings I, V, and IX are replaced by the three reference cycle standings described above, selected to reflect the average lag or lead shown by each type of specific turning point.

This somewhat lengthy, though still incomplete, enumeration of the various measures employed may serve to show the main preoccupation of the authors: faithful observation and summarizing of the cyclical characteristics of a large number of economic series. The toolkit of the theoretical economist is deliberately spurned. Not a single demand or supply schedule or other equation expressing the behavior of men or the technical laws of production is employed explicitly in the book, and the cases of implicit use are few and far between.

THE SPIRIT OF INQUIRY GROPING FOR GUIDANCE

As indicated above, I am here concerned mainly with evaluating this empiricist position taken by the authors, and with showing its implications and limitations. My *first argument*, then, is that even for the purpose of systematic and large scale observation of such a many-sided phenomenon, theoretical preconceptions about its nature cannot be dispensed with, and the authors do so only to the detriment of the analysis. It has already been mentioned that the later and more interesting part of the book (Chapters 9–12) is devoted to a search for possible changes in cyclical “behavior” over

time, with a view to qualifying the meaning of average measures of cyclical "behavior" computed from a sequence of cycles. This analysis employs the following seven series, arranged here as classified on page 372:

<i>Relating to</i>	<i>Series</i>
Durable goods market	1. Pig iron production
	2. Railroad freight car orders
Money market	3. Yields of high-grade railroad bonds ^a
	4. Call money rates ^a
Stock market	5. Railroad stock prices ^a
	6. Number of shares traded ^a
Volume of payments	7. Deflated bank clearings

There is no systematic discussion of the reasons for selecting these particular variables as most worthy of study. As a justification for this choice the following few lines are given on page 384:

These series cover processes that rank high among the activities stressed in theoretical studies of business cycles. Partly for this reason, partly because of the comparatively long stretch of time covered by these records, we regard our small sample as fairly satisfactory for the present purpose.

The choices made may have been the best possible ones. But "good" choices means relevant choices. What is relevant can only be determined with the help of some notions as to the generation of economic fluctuations, and as to their impact on society. In the light of such notions, wide fluctuations in call money rates may be unimportant if total employment is relatively stable. Fluctuations in the production of durable producers' goods would be less serious if they were approximately offset by opposite fluctuations in the production of consumers' goods. The choices as to what variables to study cannot be settled by a brief reference to "theoretical studies of business cycles." These issues call for a systematic argument to show that the best use has been made of available data in relation to the most important aspects of the phenomena studied.

Earlier in the book (pp. 71-76), some discussion is indeed devoted to the "meaning" of individual variables, in particular with a view to determining whether a single variable or

aggregate might be used to locate turning points of reference cycles. The shortness of the periods for which broad aggregates — like national income, an index of total production, or employment — are available rules out such series for all purposes requiring a long period of observation. But the use of a small number of aggregates is also warned against as being insufficient in principle. This question, it seems to me, admits of different answers in different cases, depending on the scope, the objective, and the underlying assumptions of each particular piece of analysis.

The lack of guidance from theoretical considerations is perceivable also in the choice of the measures computed from the variables selected. These are intended to be measures of cyclical "behavior." The use of the term "behavior" does not mean, however, that the authors intend to study the behavior of groups of economic agents (consumers, workers, entrepreneurs, dealers, etc.) whose modes of action and response, in the social organization and technological environment of the society studied, are the ultimate determinants of the levels of economic variables as well as their fluctuations. Instead, they study the "behavior" (in a more mechanical sense) of certain measurable joint effects of several of those actions and responses. This shift of attention from underlying human responses to their combined effects is a decisive step. It eliminates all benefits, described more fully below, that might be received from economic theory — by which I mean in this context the theoretical analysis of the aggregate effects of assumed patterns of economic behavior of groups of individuals. It also divorces the study of fluctuations from the explanation of the levels or trends around which the variables fluctuate, since such theoretical analysis is needed to bring out the common features in both groups of problems.

The rejection of the help that economic theorizing might give leaves a void. For now there is a need for some organizing principle to determine on what aspects of the observed variables attention should be concentrated. Here the definition of business cycles quoted above comes into operation. But it does not quite fill the gap. It does not become altogether clear why the cyclical forms of move-

^a Quoted on New York Stock exchange.

ment should receive such exclusive attention. With the great variety in types of movement in the real world, it is not even always clear what a cycle is. The gap left by the barring of explicit formal theory is thus filled with methodological quasi-theory concerned with delineating the object of study. There are lengthy discussions of questions like these: What is a turning point? When is a certain movement of a variable to be recognized as a specific cycle? (pp. 61-62). When are certain concurrent movements of a number of variables to be recognized as a business cycle? (pp. 87-94). In first instance the criteria employed are mechanical applications of clauses in the definition of business cycles quoted above, like limitations on the length of time between two successive turning points, or the rule that no "cycle" be divisible into shorter cycles with amplitudes approximating its own (even if those shorter cycles would escape recognition because of their shortness). Difficulties then arise in periods of war or important changes in economic policies. Reference is made to judgment and indeed to explanatory factors where those are clearly visible (tie-ups through weather or strike, economic effects of war, changing policies in the early "new deal" period). Arbitrary formal criteria are here combined with good though incidental pieces of causal analysis to answer what are, in frequent borderline cases, essentially irrelevant questions. The authors' insistence on seeing, counting, and measuring cycles before anything else reminds one of Kepler's preference for circular motion.

A similar group of questions, sometimes permitting only arbitrary answers, arises in deciding how to match the cycles specific to one variable with recognized business cycles, for certain comparisons. One of these questions is whether the variable concerned is to be treated on a positive (i.e., trough-to-trough cycle) or on an inverted (peak-to-peak) plan. We learn on page 115 the highly interesting fact that raw material stocks held by manufacturers tend to be positively related to business cycles, whereas stocks of finished products tend to be related invertedly. The authors do not at this stage ask for the motives or determining fac-

tors of this behavior of dealers or manufacturers. Instead they discuss formal rules to establish positive or inverted "behavior" of a variable on the basis of frequencies of concurrent or opposite directions of movement.

On the whole, the same measures are computed for all variables studied, irrespective of their economic nature. The importance of the economic phenomenon expressed by any particular variable is duly stressed (pp. 140-41) with reference to the interpretation to be placed on the measures computed, but is in general not permitted to influence the choice of measures used. An exception is found in the discussion of criteria for positive or inverted treatment, which contains hints of postulated behavior relationships (p. 117). This analysis would need to be made more explicit to remove the impression that somewhat scholastic distinctions are used in the discussion of how to relate specific and reference cycles (p. 118).

The notion of a reference cycle itself implies the assumption of an essentially one-dimensional basic pattern of cyclical fluctuation, a background pattern around which the movements of individual variables are arranged in a manner dependent on their specific nature as well as on accidental circumstances. (There is a similarity here with Spearman's psychological hypothesis of a single mental factor common to all abilities.) This "one-dimensional" hypothesis may be a good first approximation, in the same sense in which the assumption of circular motion provides a good first approximation to the orbits of the planets. It must be regarded, however, as an assumption of the "Kepler stage," based on observation of many series without reference to the underlying economic behavior of individuals. It is in this sense, I believe, that the authors refer (p. 3) to their definition of business cycles as "a tool of research, similar to many definitions used by observational sciences, and like its analogues subject to revision or abandonment if not borne out by observation." I believe that the authors would not object to the addition: "or by the logical consequences of observations of a wider range of phenomena."

SCANT GUIDANCE FOR MAKERS OF POLICIES

The examples given illustrate the authors' scientific "strategy," in which measurement and observation precede, and are largely independent of, any attempts toward the explanation of economic fluctuations. The plan of inquiry envisaged by the National Bureau is therefore to follow up the present methodological work by a series of monographs in which the techniques of measurement developed are applied comparatively to various industries, countries, or broad markets. Ultimately, it is intended to "weave the results established by the monographs together with existing knowledge into a theoretical account of how business cycles run their course."⁴

The wording of this statement of intentions still admits of the interpretation that even the ultimate objective of the authors is only a generalizing description of the typical course of a business cycle. However, I believe, and will assume for the purpose of this discussion, that more is meant, namely, a genuine explanation of economic fluctuations, i.e., an explanation in which only extra-economic phenomena are accepted as "data" without further inquiry, all relevant economic phenomena being subject to explanation in terms of assumed behavior patterns of men in a given institutional and technological environment. I am not sure whether a still further objective is included, which extrapolates the idea of explanation: the prediction, within the narrowest attainable limits of error, of the effects of stated hypothetical measures of economic policy on the level and movements of economic variables. However, I feel that such prediction is actually the most important objective of the analysis of economic fluctuations. The criterion of social usefulness of scientific analysis gives us the right to discuss the merits of any particular approach to the problem of economic fluctuation on the basis of the guidance it gives to economic policy, even if such guidance were not claimed by the authors.

Let us, then, now consider the question

⁴P. 22. A less ambitious "preview" of this final volume is promised shortly under the title *What Happens During Business Cycles: A Progress Report*, by Wesley C. Mitchell.

whether the development that led from the empirical regularities observed by Kepler to the general theory of gravity discovered by Newton might find a counterpart in similar discovery of the laws of economic motion on the basis of carefully described regularities. I shall mention and discuss a few important differences between the two scientific situations.

Newton's achievement was based, not only on the regularities observed by Kepler, but also on the experiments conducted on the surface of the earth by Galileo. Economists are not in a position to perform experiments with an economic system as a whole for the sole purpose of establishing scientific truth (although deliberate changes in parts of the system have been undertaken at various occasions for other than scientific purposes, and have incidentally added to our information). It is therefore not possible in many economic problems to separate "causes" and "effects" by varying causes one at a time, studying the separate effect of each cause — a method so fruitful in the natural sciences.

On the other hand, economists do possess more elaborate and better established theories of economic behavior than the theories of motion of material bodies known to Kepler. These economic theories are based on evidence of a different kind than the observations embodied in time series: knowledge of the motives and habits of consumers and of the profit-making objectives of business enterprise, based partly on introspection, partly on interview or on inferences from observed actions of individuals — briefly, a more or less systematized knowledge of man's behavior and its motives. While much in these theories is incomplete and in need of reformulation and elaboration (particularly in regard to behavior over time under conditions of uncertainty), such theory as we have is an indispensable element in understanding in a quantitative way the formation of economic variables. For according to that theory the relevant economic variables are determined by the simultaneous validity of an equal number of "structural" equations (of behavior, of law or rule, of technology). The very fact that so many relations are simultaneously valid makes the observation of any one of them difficult, and sometimes even impossible. For any ob-

served regularity between simultaneous and/or successive values of certain variables may have to be ascribed to the validity of several structural equations rather than any one of them. The mere observation of regularities in the interrelations of variables then does not permit us to recognize or to identify behavior equations among such regularities. In the absence of experimentation, such identification is possible, if at all, only if the form of each structural equation is specified, i.e., in particular, if we can indicate the set of variables involved in each equation, and perhaps also the manner in which they are to be combined. In each case, a preliminary study of the system of structural equations held applicable is required to decide whether the specifications regarding any particular equation are sufficiently detailed to permit its identification. Without such identification, measurement of the structural equation involved is not possible, and should therefore not be attempted.

One might object: why should measurement of the behavior equations of consumers, workers, entrepreneurs be necessary? If observed regularities are due to the simultaneous validity of several behavior equations, these regularities will persist as long as each of the underlying (unknown) behavior patterns persists. However, there are important arguments to counter this objection. Sheer scientific curiosity still urges us on to penetrate to the underlying structural equations. This curiosity is reinforced and justified (if you wish) by the awareness that knowledge of the behavior patterns will help in understanding or analyzing different situations, for instance, problems of secular trend, or cyclical problems in other countries or periods—in the same way (although one would not expect with the same exactness) in which the law of gravitation explains celestial and terrestrial phenomena alike. This point has particular relevance with regard to the different situations expected to arise in an impending future period of the same country that has been studied. Behavior patterns are subject to change: gradually through changing habits and tastes, urbanization and industrialization; gradually or unevenly through technological change; abruptly through economic policies or the economic

effects of political events. While one particular behavior pattern may be deemed fairly stable over a certain period, a much greater risk is involved in assuming that a whole system of structural equations is stable over time. An observed regularity not traced to underlying behavior patterns, institutional rules, and laws of production, is therefore an instrument of unknown reliability. The predictions it yields cannot be qualified with the help even of known trends in behavior or technology. It is of no help whatever in assessing the probable effects of stated economic policies or institutional changes.

There is no sign in the book of an awareness of the problems of determining the identifiability of, and measuring, structural equations as a prerequisite to the practically important types of prediction. Measurable effects of economic actions are scrutinized, to all appearance, in almost complete detachment from any knowledge we may have of the motives of such actions. The movements of economic variables are studied as if they were the eruptions of a mysterious volcano whose boiling caldron can never be penetrated. There is no explicit discussion at all of the problem of prediction, its possibilities and limitations, with or without structural change, although surely the history of the volcano is important primarily as a key to its future activities. There is no discussion whatever as to what bearing the methods used, and the provisional results reached, may have on questions of economic policy.

This, then, is my *second argument* against the empiricist position: Without resort to theory, in the sense indicated, conclusions relevant to the guidance of economic policies cannot be drawn.

CHANGES IN CYCLICAL "BEHAVIOR" OVER TIME

There is a highly interesting analysis in the last four chapters, already referred to, in which the following question is treated (phrasing by the reviewer): Is there evidence that such structural changes as have taken place during the period studied have led to changes in cyclical "behavior" of the variables studied? A search is made (Chapter 10) for secular

changes, in duration, amplitude (absolute and per month) and timing of the specific cycles, and in the pattern of reference cycles, of the seven American variables selected for intensive study. A hypothesis by Mills linking durations of business cycles in various countries to stages of industrialization, and the hypothesis of a break in average duration and amplitude of specific cycles of the seven American series due to the first world war are tested. A search is also made (Chapter 11) for long cycles in cyclical characteristics. Possible statistical connections with the long wave in building activity, and various long cycle hypotheses formulated by Wardwell, Kondratieff, Schumpeter, and Kitchin, respectively, are tested.

There appears to be a tendency in this chapter to select a hypothesis for testing because it has been stated in a scientific publication rather than on the basis of possible arguments in favor of it. Nevertheless, the hypotheses (granted that they concern the "behavior" of variables rather than of men) cover a wide range of possibilities. In particular the hypotheses of secular trend in cyclical characteristics, that of a break in structure due to war, and that of an influence of the long cycle in construction are of great theoretical and practical interest.

The most remarkable outcome of this whole group of tests is the extent to which mild traces of systematic change, of one type or another, in cyclical "behavior" are almost drowned by wide and apparently random variability between cycles. It is true that interesting particular changes are found. Money markets are found more susceptible to secular changes in cyclical behavior than industrial or security markets. The lead in the cyclical revival of pig iron production and freight car orders in early cycles is found to have disappeared in later cycles. The latter effect may be wholly or partly an automatic result of a diminishing rate of growth, given the fact that turning points are defined without prior elimination of secular trend.⁵ It would indeed be interesting to determine whether the gradual decrease in cyclical lead would remain if turning points were determined after trend elimination. If so, there is a parallel phenomenon in the gradual decrease in the responsiveness of

demand for railway rolling stock to changes in traffic and profitability, in the United Kingdom during a period preceding the first world war, apparent from one of Tinbergen's investigations.⁶

One of the results interpreted as a possible sign of longer cycles in cyclical behavior might be merely the effect of considerable random variation between cycles, combined with correlation between the various characteristics of a cycle. I am referring to the differences found between average characteristics of the first and last cycles of groups of successive cycles separated by severe depressions. For such averages are obtained by a process of selection of cycles that start and end, respectively, in especially deep depressions. The authors stress this selection effect when they deal with Schumpeter's hypothesis that each Juglar cycle contains three Kitchin cycles, but do not seem to give it sufficient emphasis in relation to their own grouping of cycles just described.⁷

However this may be, any systematic effects present are found to be greatly obscured and dominated by random variation of the characteristics of individual cycles. The authors themselves express surprise (p. 413) at the slight manifestations of structural change (other than mere growth, largely eliminated by the use of relatives to cycle means) in data covering a period known to have witnessed thoroughgoing changes in economic organization. They state their intention to press the search for secular changes in cyclical behavior in subsequent studies concerned with particular industries or markets.

ISOLATING THE SOURCES OF RANDOM VARIATION

The presence of random variability in economic data gives rise to methodological requirements which do not arise in the study of planetary motion. In the latter case, the phenomenon studied can for all practical purposes be treated as a deterministic process, with some randomness entering into the data only through

⁶ J. Tinbergen, *Statistical Testing of Business Cycle Theories: I. A Method and Its Application to Investment Activity*, Graph V.2 on page 120.

⁷ I do not understand the reasoning at the top of page 460, where evidence independent of selection is claimed.

⁵ See the discussion in Chapter 7, Section III.

errors of measurement. In dynamic economics, the phenomenon itself is either essentially a stochastic process or needs to be treated as such because of the great number of factors at work.⁸ Hence the analysis and interpretation of economic data call for the application of the methods of statistical inference.

The main problem of inference is the choice of "statistics," i.e., those functions of the observations — fewer in number than the observations themselves — which are to be used for estimation of parameters or for the testing of hypotheses. The question should therefore now be raised whether the authors' finding of strong domination of random variation over possible traces of systematic change in cyclical "behavior" is not at least partly due to the choice of the particular "statistics" studied. At the risk of becoming monotonous, I wish to state that explicit dynamic theory of the formation of economic variables is needed to throw light on this question. Most theories of this kind recently constructed have in common the attempt to describe the fluctuating economy by a complete system of structural equations which, as to their form, are stochastic difference equations. They are difference equations (embodying dynamic theory), in that they describe responses subject to time lags: past values of economic variables affect current actions of individuals. They are stochastic equations in that the behavior of any group of individuals, and the outcome of any production process, is determined in part by many minor factors, further scrutiny of which is either impossible or unrewarding. Such further scrutiny is not necessary provided that the analysis of each structural equation be pushed to the point where the joint effect of unanalyzed factors can indeed be regarded as random (if not necessarily independent) drawings from a reasonably stable probability distribution. To attain this end, it is often neces-

sary to introduce explicitly so-called "exogenous variables," representing the effects of wars, political events, population growth, economic policies, or technological developments which are not routine responses to economic conditions, etc.

Systems of this kind may possess a tendency for the variables to evolve in cyclical movements. Even if the random disturbances (or shocks) in individual equations possess a fairly stable distribution, however, there is no need for the ensuing cycles to be very regular or similar in duration or amplitude. Current values of economic variables are the cumulative effect both of a sequence of random shocks over the recent past, and of the impulses exerted by exogenous variables in the recent past.⁹ Because of this tendency to cumulation of effects, relatively small shocks may have considerable effects over time on such "cyclical characteristics" as duration and amplitudes of cycles. Also, different impulses exerted successively by the same exogenous variables may produce different cycles of quite diverse appearance.

Now any rigorous testing of hypotheses according to modern methods of statistical inference requires a specification of the form of the joint probability distribution of the variables. In principle, such specification does not need to take on a "parametric" form, as when linear, parabolic or exponential functions, or normal distributions, are specified — although parametric assumptions usually admit more accurate estimation or more powerful tests whenever they are justified. In any case, however, it is necessary to hypothesize in what manner randomness enters into the formation of economic variables. It is for this reason that the form of each structural equation should be specified and/or determined to the point where at least a conceptual isolation of the random influences at work is attained.

The authors do not discuss randomness in terms of definite distributional hypothesis, although the idea of random factors as one of the determinants of economic variables is

⁹ How long this "recent past" is to be taken depends on the degree of damping of the system, which in turn depends on the parameters or curves representing the several structural equations.

⁸ It has been stated by H. Hotelling ("Differential Equations Subject to Error and Population Estimates," *Journal of the American Statistical Association*, Vol. 22, 1927, pp. 283-314, quotation on p. 287), that celestial mechanics would for the same reason have developed as a statistical science, had the "solar" system to which the earth belongs contained several bodies of mass comparable to that of the sun. The full quotation is given and commented on by H. T. Davis, *The Analysis of Economic Time Series* (Bloomington, Indiana, 1941), see pp. 2-4.

clearly in their minds.¹⁰ They accordingly recognize (p. 392) that the analysis of variance tests applied by them to durations, amplitudes, time lags, are not rigorous, since such measures need not be independent in successive cycles. More important yet is the fact that those tests are not particularly powerful in discerning structural change under the welter of random variation. For on the one hand, these tests fail to take into account the influence of measurable exogenous variables, and to take advantage of the known time series of such variables — a possible advantage particularly important in periods of war or of new departures in economic policy. On the other hand, the basic cyclical measures they analyze are cumulative effects of random shocks, of which observations are limited to the number of cycles covered by the study. The additional information about the individual structural equations and the disturbances therein, contained in the more numerous original data, is thus lost.

In their defense of the application of analysis of variance, the authors mention that the original items of economic time series are even less independent serially than cycle durations or amplitudes. Probably they do not mean to imply a statement (which has often been fallaciously advanced) that the high serial correlation of economic time series precludes the use of such data (as distinct from "cyclical" measures derived therefrom) in any statistical tests or estimation procedures. Statistical theory is sufficiently flexible to face such situations. In the first place, it may be found that serial correlation in economic variables measured annually, say, is due only to their being determined by difference equations, with no serial correlation present in the disturbances (shocks) operating in individual equations — a situation which may be confirmed by tests based on the "residuals" obtained from fitting such equations. But even a situation of serially correlated disturbances — which is likely to prevail in any case in quarterly or monthly figures — is in principle equally amenable to statistical treatment. The mathematical and

¹⁰ This can be seen from the discussion of the causal interpretation of averages, particularly on page 506, where there is a groping for distinctions which only mathematical formulation can clarify.

computational difficulties inherent in such a situation pose technical problems which need to be overcome, to enable us to extract all information about the structure of our economy from statistical records.

The amplitudes, durations, and measures of conformity used by Burns and Mitchell are poor measures from this point of view. They waste an unknown but probably considerable amount of information contained in the original data. Their averages are unstable¹¹ because of the occurrence of borderline cases under the rules for recognizing cycles, because turning points are located without allowance for secular trend, and because of great variability between cycles.¹²

However, the extraction of more information from the data requires that, in addition to the hypotheses subject to test, certain basic economic hypotheses are formulated as distributional assumptions, which often are not themselves subject to statistical testing from the same data. Of course, the validity of information so obtained is logically conditional upon the validity of the statistically unverifiable aspects of these basic hypotheses. The greater wealth, definiteness, rigor, and relevance to specific questions of such conditional information, as compared with any information extractable without hypotheses of the kind indicated, provides the *third argument* against the purely empirical approach.

Let me wind up the argument with a statement combining exhortation and prophecy. In the monographs dealing with specific markets, in preparation or planned by the National Bureau of Economic Research, situations will frequently be encountered where the applicability of the behavior schedules of economic theory is more directly obvious, less beset with doubts on the score of unhomogeneity of commodities or individuals, and the connected difficulties of aggregation. Also, certain rela-

¹¹ See the discussion of conformity indexes on pages 184-85.

¹² In two cases, on page 425 and page 433, the exclusion of the "exceptional" reference cycle 1927-33 from an average makes a sufficient difference to the test comparison being made to be mentioned (recommended?) in the test. Addition to the averages of the 1933-38 cycle (which is not included in the tests discussed above) might well have a similarly large effect.

tionships between aggregates seem more strongly established *a priori* than others. The aggregate consumption function, a subject which the National Bureau is now investigating, so far stands on firmer ground than the investment schedule: consumption decisions are more of one kind than investment decisions. Among the latter decisions, inventory policies seem to be subject to a smaller number of considerations, more readily rationalized, than investment in productive equipment. Thus, the use of behavior schedules will inevitably force itself on the mind of an investigator dealing with some of the more specific partial subjects of dynamic economics. Such a development is both predictable and highly desirable. The combination of theoretical and statistical analysis into an explanation of cyclical fluctuations and an exploration of the means to influence them must necessarily proceed from detailed studies of individual relationships. Conversely, the statistical methods used in those detailed studies should recognize and take into account the fact that the specific relationship studied is part of a complete network of interrelations connecting the variables involved in many ways.

This already lengthy review could well end here. However, I cannot forego the opportunity to append a few brief comments on various specific points of method raised in the volume.

INDEX NUMBERS, TIME UNITS, SMOOTHING, SINE CURVES VERSUS TRIANGULAR PATTERNS

The authors' preference for the study of many individual series rather than index numbers doubtless derives from their basic decision to place the large-scale study of facts before theoretical concepts and hypotheses regarding the formation of economic variables. But their arguments provide a challenge to those who believe that the most relevant phenomena of economic fluctuation can fruitfully, or even better, be analyzed through aggregates or index numbers. To withstand critical examination, this belief needs to be argued more cogently than is usually done. It will be necessary to specify the purposes index numbers are required to serve, and to show theoretically

and statistically to what extent these purposes are actually served efficiently and without undue loss of relevant information.

The authors' views that quarterly or monthly data contain much information which is lost by reduction to annual averages deserves strong sympathy. It is true that several of the particular measures on which the National Bureau concentrates are especially vulnerable to such reduction, as the authors amply demonstrate. But also if the purpose is one of estimating the parameters of structural equations, the presence and dynamic importance of relatively small time lags in many equations, as well as the shortness of available time series, makes the use of at least quarterly figures an important objective of the analysis of economic fluctuations.

The authors' rejection of the use of smoothing formulae is similarly appropriate. One could add to their arguments that, if explicit mathematical formulation of the distribution of the observations is introduced to guide the choice of estimation or test procedures, smoothing is found both to be wasteful of information and to complicate mathematical treatment, because it mixes up the effects of successive disturbances as well as blurs the time-shape of exogenous variables. In fact, one of the reasons why business cycle analysis is a difficult undertaking is that the economic system itself is such an effective smoothing agent of the random shocks to which it is exposed. The analytical problem is one of de-smoothing rather than smoothing.

Exception must be taken to a statement appearing on page 369:

When averages are struck for all cycles covered by a series, the erratic factors in the measures for single cycles have an additional opportunity to cancel out.

This is true generally, but is not applicable to the average standings at troughs (I, IX) and peaks (V). A selection effect is operative through the location of troughs and peaks at local minima and maxima of the curve, giving downward and upward biases to average trough and peak standings respectively, which will be especially pronounced if erratic disturbances persist for at least three months. This point is important because of the authors' statement

(p. 157) that a "triangular" cyclical pattern often gives a better approximation to reality than the sine-curve pattern (whose dominance in the literature the authors attribute particularly to the prevalent habit of smoothing time series before analysis (p. 343). An important theoretical question is involved: the rounded curve seems connected with the idea of a natural equilibrium level or trend line around which fluctuations take place; in particular, pure sine curves suggest linearity of the equations describing the economy, whereas less symmetric but still rounded curves are compatible with non-linear systems where no effective limits are placed on the range of the variables involved. However, the broken straight line pattern suggests one-sided movement as the natural condition of the economic system, reversed by capacity limits or other physical or incidental factors. Now the trough and peak standings are the crucial observations in making a choice between these two hypotheses. The selection effect mentioned produces a bias toward the triangular hypothesis, disqualifying average cyclical patterns as a means of testing the issue mentioned.

The authors are aware of the possibility of such bias,¹³ but seem to feel that it will be unimportant except in series with pronounced erratic movements.¹⁴ However, their graphs suggest a widespread occurrence of this bias. While all of the ten specific-cycle patterns in Chart 16 on page 56 show sharply defined kinks at the turning points, definite cusps are

developed most clearly in the series most subject to erratic fluctuations (shares traded, total exports, sugar meltings). This does not mean that the issue between rounded curves and triangular patterns is to be decided in favor of the former. Other evidence, less marred by methodological doubts, is adduced to show that at least a substantial proportion of cycles have kinked peaks and troughs: slightly over two-thirds of the turning points in five American series are not shifted in time if determined after trend elimination instead of before (p. 277).

CONCLUSION

To sum up: the book is unbendingly empiricist in outlook. Granted this basic attitude, it shows great perseverance and circumspection on the part of the authors in handling a vast amount of statistical data. In the latter part of the book, hypotheses of theoretical and practical relevance, referring to the characteristics of cyclical movements of the economy as a whole, are tested. But the decision not to use theories of man's economic behavior, even hypothetically, limits the value to economic science and to the maker of policies, of the results obtained or obtainable by the methods developed. This decision greatly restricts the benefit that might be secured from the use of modern methods of statistical inference. The pedestrian character of the statistical devices employed is directly traceable to the authors' reluctance to formulate explicit assumptions, however general, concerning the probability distribution of the variables, i.e., assumptions expressing and specifying how random disturbances operate on the economy through the economic relationships between the variables.

¹³ See page 334, footnote 30; page 346; and the third graph in the first column of Chart 47 on page 345.

¹⁴ The authors refer on page 347 to an opposite bias due to the mild smoothing involved in the use of three-month averages for trough and peak standings. However, this bias is likely to be smaller than the bias due to selection, owing to the smallness of the three-month period compared with the average duration of cycles.