The Logic of Econometric Business-Cycle Research

Author(s): Tjalling Koopmans

Source: *Journal of Political Economy*, Vol. 49, No. 2 (Apr., 1941), pp. 157-181
Published by: The University of Chicago Press
Stable URL: https://www.jstor.org/stable/1824427
Accessed: 16-09-2019 14:13 UTC

THE JOURNAL OF POLITICAL ECONOMY

Volume XLIX APRIL 1941 *Number 2*

THE LOGIC OF ECONOMETRIC BUSINESS-CYCLE RESEARCH

TJALLING KOOPMANS
Princeton University

I. INTRODUCTORY REMARKS

IN REVIEWING Tinbergen's first League of Nations publication on statistical testing of business-cycle theories, Mr. Keynes expressed the desire for a more systematic exposition of the logic of the methods applied in econometric business-cycle research.[1] The present paper is intended as a contribution in this direction.

As may be expected in a treatment of logical problems, I shall need a certain amount of space for writing down some self-evident statements. The argument will lead further to a number of observations of which a majority are implicit or explicit in Tinbergen's two publications.[2] However, I am offering the present discussion with the conviction that a consideration of the logical framework of the methods concerned may prove helpful in realizing the exact nature of the controversies that have arisen.

Business-cycle analysis differs from the analysis of economic

[1] J. M. Keynes, "Professor Tinbergen's Method," *Economic Journal*, XLIX (1939), 588.

[2] J. Tinbergen, *Statistical Testing of Business Cycle Theories*, Vol. I: *A Method and Its Application to Investment Activity;* Vol. II: *Business Cycles in the United States of America, 1919–1932* (League of Nations, 1939).

equilibria in that it is likely to draw a greater benefit from statistical induction. The fact that it deals with short-run movements increases the possibilities of extracting from statistical observations information regarding the relations underlying those movements. These possibilities and their limitations form the subject of this article. I propose to consider quite abstractly the question as to what extent results reached by econometric business-cycle research derive from statistical observation and to what extent they are dependent on other information or hypothesis. In order to concentrate on this question I shall restrict as much as possible the description of methods, for which I may refer to Tinbergen's two publications.

It is becoming more and more widely recognized that the generation of business cycles may differ in essential features from one country and period to another. The problem to be dealt with, therefore, may be narrowed down to that of finding a quantitative explanation of cyclical movements occurring in a given country during a given period in which no important, or only readily recognized, changes in economic structure took place. Further, in so far as testing of business-cycle theories appears possible, it means testing the relevance of such theories with respect to country and period considered.

I shall enumerate the elements of the logical situation facing the student of that problem.

II. THE DATA

The first element is the availability of a considerable number of statistical time series, to be referred to as the data,[3] each series representing some measurable economic phenomenon or variable which plays a role in cyclical fluctuations. The variables may be aggregate sums of money held at a certain moment or spent during a certain unit of time for some general type of economic purposes, or index numbers of prices quoted or of quantities produced, con-

[3] The word "data" has sometimes been used in the sense of what is called below "external variables," in contexts concerned with the theoretical economic problem of explaining the movements of economic variables from those of the "data" so defined. The present use of the word seems justified by a context concerned with a problem of statistical induction.

sumed, or held for a certain type of purpose, etc. Numerous problems are connected with the choice and definition of the variables. So far as these are economic problems, their nature will become clear in what follows. The more technical problems connected with the construction of index numbers may be disregarded, partly on account of their technical nature and partly because they have not been raised in the discussions referred to.

The following three characteristics of the data are important in our context. First, they include variables of two types, internal and external. The former are those phenomena whose fluctuations the investigator wishes to explain; the latter are those other phenomena which he may also require as elements in that explanation, while excluding the possibility that their fluctuations are influenced in turn by what he may regard as internal variables. Examples of external variables are natural phenomena like rainfall, temperature, and also such economic phenomena in other countries as are assumed to be independent of the economic events in the country considered.

In this stage only the distinction between internal and external variables is introduced, and the occurrence of both types among the data is stated. The question of whether or not a given variable must be regarded as the object of explanation comes up in a later stage.

Next, the data are unique. Their extension is limited by their availability and by the limits set to the time period under consideration by wars or other abrupt changes in economic structure. It is not possible to run the same country once more through the same period while making changes only in the initial phase of the cycle or only in the impact of certain noneconomic factors like the yield of crops.

Finally, the various series composing the data show a high degree of interrelation. By that expression—while inviting suggestions for a better substitute—I mean that in many ways a "good fit" or a "high correlation" can be obtained by picking out a few variables and approximating the fluctuations of one of them through simple arithmetical operations applied to the others with appropriate coefficients.

The combination of uniqueness and manifold interrelation of data is, it appears, not often encountered in the sciences in which statistical methods have first been developed. This is why the fundamental difficulties and limitations arising from this combination must be considered and formulated in explicit connection with the application of these methods to economic problems.

III. THE GENERAL WORKING HYPOTHESIS

The second element in the logical situation is the adoption, by the investigator, of the working hypothesis that causal connections between the variables are dominant in determining the fluctuations of the internal variables, while, apart from external influences readily recognized but not easily expressible as emanating from certain measurable phenomena, mere chance fluctuations in the internal variables are secondary in quantitative importance. Earthquakes, political events, and strikes are examples of the readily recognized but unmeasurable external factors. The term "causal connection" is used in the sense of a certain quantitative relationship having a character of necessity as opposed to pure chance. It does not necessarily imply a separation in time of "cause" and "effect." Finally, I speak of a working hypothesis regarding the prevalence of causal connections as distinct from the specific form these connections may be believed to assume with respect to any particular variable.

It is important to note that the dominance of causal connections is assumed with respect to aggregates or suitably weighted averages of measurable phenomena. Important and unrecognized accidental influences in individual economic decisions are not excluded but are assumed to balance approximately where a great number of economic subjects is concerned.

The foregoing hypothesis I shall call the general working hypothesis of econometric business-cycle research. It requires qualification in the following respect: The foregoing formulation does not mention the possibility of unmeasurable internal factors acting as a cause on other variables. Actually, expectations regarding future values of certain variables often affect present decisions made by economic subjects. In business-cycle literature, as well as in the daily press, this situation is reflected by frequent reference to "the state of confidence" as a factor. The working hypothesis introduced is not to be understood as excluding such possibilities. It implies only that, in such cases, "expectations" or "the state of confidence" or whatever other unmeasurable internal factor may be thought of as exerting an influence are themselves in the end again determined mainly by measurable internal and/or recognizable external phenomena, to a minor extent only by pure chance.

Indeed, this seems to be the only way in which sense can be given to the concept of an unmeasurable internal factor. And the logical content of the general working hypothesis reduces to the following three constituents: (1) the adoption of causality as the quantitatively dominant principle in explaining the fluctuations of economic variables; (2) the possibility of recognition of non-measurable external factors; and (3) the supposed existence, in economic life, of a sufficient number of actual or possible observation posts for statistical services as to make an understanding of cyclical movements possible from the study of the connections between measurable phenomena [and the nonmeasurable phenomena referred to under (2)], including such indirect connections as run through unmeasurable phenomena.

The first of these assumptions seems to be in agreement with the views of a majority of business-cycle theorists. Less comment is to be found on the other two points, perhaps because the distinction between measurable and nonmeasurable phenomena did not receive much attention as long as the observational side of business-cycle research remained undeveloped.

There is, therefore, a question of great interest as to how far the foregoing basic assumptions are themselves supported by the data. For reasons of exposition I shall postpone the discussion of this point until the end of this article. For the moment it is only necessary to point out that the general working hypothesis is, to the extent of present experience, not contradicted by the data. The high degree of interrelation among the several series leaves ample freedom for the construction of supposed causal connections between them.

IV. THE NECESSITY OF ADDITIONAL INFORMATION

In fact, it leaves too much freedom. Generally the data are such—and if the general working hypothesis holds they have to be such—that the fluctuations in one given variable may be reconstructed in more than one way by combining some other variables in a suitable manner. This is why the uniqueness of the data, or the impossibility of experimentation, is such a fundamental limitation to the body of conclusions which the data support. What kind of inferences can be drawn, then, on the basis of the two elements now introduced? It will be clear that the only unconditional inferences one may draw are negative. They state that this or that supposed causal connection is not in agreement with the facts reflected in the data. A very elementary example is what might be called the static explanation of share prices, i.e., the theory which would explain share prices as dividends capitalized by some market interest rate. This theory does not fit, e.g., the experience of Germany[4] for a long period before 1914 or of the United States[5] around 1929. Another example is the "acceleration principle," which, in many cases, is insufficient to explain as sole cause, or even to be a major element in explaining, fluctuations in general investment activity.[6]

The drawing of such unconditional negative conclusions from the data represents the most direct type of statistical testing of business-cycle theories or of elements in such theories. It is, however, inconclusive. To make a further choice from the numerous explanations not so discarded requires the introduction, as the third logical element in the situation, of certain additional information. That is, the only positive conclusions one may draw are conditional upon the validity of certain suitably chosen premises that make up the additional information.

The undefined word "information" seems most apt to cover the diversity of the statements that may serve as premises in the

[4] A. Donner, "Bestimmungsgründe der Aktienkurse," *Vierteljahrshefte für Konjunkturforschung*, Sonderheft 36 (Berlin, 1934).

[5] Tinbergen, *op. cit.*, II, 106.

[6] Tinbergen, "Statistical Evidence on the Acceleration Principle," *Economica*, V (1938), 164.

construction of positive conclusions. They may include observations not expressible as statistical time series, experiences from other countries or periods showing a similar economic structure, deductions from economic theory, or even mere working hypotheses having a certain degree of plausibility. They have in common only that they are used in addition to the two elements introduced above and that they are not held to be incorrect. Very often, however, they may on good grounds be supposed to approach the truth closely.

V. PRINCIPLES GOVERNING THE CHOICE OF THE ADDITIONAL INFORMATION

It will be useful to classify, through a number of principles, the considerations governing the choice of the additional information. The first principle naturally requires that the additional information should not imply statements which can be unconditionally rejected because they are contradicted by the data. This will be called the principle of statistical censorship on the additional information.

Apart from this the choice is logically free, and further principles derive from the desire to give a maximum of scientific value to the conditional conclusions obtained. The principle of scientific economy in devising the additional information requires that no statement be included that could be derived from the data with the help of the other statements included—or, at least, on the basis of some less restrictive set of premises. Assertion of this principle is, of course, the more desirable the less one is a priori convinced of the validity of any statement in the premises that could possibly be dispensed with. Taken together with the principle of statistical censorship, the principle of scientific economy insures that the additional information is truly complementary to the data.

The next two principles also act in combination, but rather in the sense of opposed forces between which a balance is to be established. The principle of a solid basis requires that as nearly as possible only those statements are included which have a high degree of plausibility, whatever the grounds. The principle of providing a sufficient basis requires, on the contrary, that the additional information be sufficiently specific to allow quantitatively definite positive conclusions.

It is obvious that, whenever final conclusions, whether positive or negative, have to be formulated, a solid basis is of more import than a "sufficient" basis. In cases where both principles cannot be satisfied simultaneously, however, nothing prevents the tentative introduction of some plausible working hypotheses which make the basis sufficient, if the conditional nature of the conclusions so obtained is duly kept in mind. Actually, the econometrician not infrequently finds himself "playing with the data," trying out the consequences of some alternative assumptions. Generally it is a sound rule to increase the number of alternatives investigated as one feels less certain about the validity of each of them.

There is one further consideration which influences the choice of the additional information, though it does not deserve the name of a principle. It is the practice of aiming at mathematical simplicity and manageability, in the form of the functions representing the causal connections, and in the probability distribution of the random influences representing the minor chance element in the fluctuations of variables. Though of paramount importance in a certain stage of the development of econometric research, particularly if carried out with limited means, this practice as an end in itself has little justification in the long run. Mathematics (as involved here) being less beset with intrinsic difficulties than the other sciences concerned, the mathematician may justly be asked to give to economists the full benefit of the possibilities of reasoning provided by his craft.

VI. THE PREMISES IN WHICH THE ADDITIONAL INFORMATION IS GENERALLY EXPRESSED

The foregoing distinctions and considerations may help us in disentangling statistical and a priori elements in the foundation of econometric business-cycle research. I now proceed to demonstrate this by a discussion of the relevant features of Tinbergen's investigations.

The variables in the data will be denoted x_1, x_2 . . . . . It is desired to represent the connections between them by a number of equations equal to the number of variables supposed or found to be internal variables. A smaller number of these equations are trivial, as they represent only the definition of one of the quantities occurring in them by means of the others. Being self-evident,[7] they are not affected by the logical problems under consideration here. The other equations are each obtained by concentrating on one internal (or possibly internal) variable and trying to reconstruct its fluctuations from those of the other variables thought to be its immediate (or its most direct measurable) causes.

In Tinbergen's investigations the additional information employed to this end can frequently be given the general form of a list of three sets of premises as follows:

Set 1. All influences affecting the fluctuations of a certain variable, say x_1, other than those emanating from a specified set of variables, say $x_2 \dots x_n$, can be classified under one of these three headings: (*a*) influences that together give rise to a minor additive component of x_1 of a random nature, i.e., subject only to a certain law of probability specified in certain respects; (*b*) influences that together can be represented by an additive component of x_1, which is a "smooth" function of time (a "trend"); and (*c*) other influences, too large or too fluctuating to be included in (*a*) or (*b*), emanating from recognized unmeasurable external forces and affecting only a few specified observations.

Set 2. The influence exercised on the dependent variable x_1 by the specified (possible) determining variables $x_2 \dots x_n$ can be represented, in the region covered by the data, by mathematical functions of these variables, of a specified type. Frequently linear functions are specified with coefficients which are constant as the time varies. These functions may involve constant nonnegative time lags between the fluctuations of the determining variables and their effects on x_1.

Set 3. The time lags assumed have values within certain speci-

[7] Again disregarding the technical problems of index-number construction mentioned above.

fied limits; some of them have specified values (often zero!). Some coefficients have specified signs, or have values specified exactly or within certain limits, or stand in specified proportions to each other.

This scheme requires specialization and sometimes amendment for every particular variable considered. To make a first draft of this specialization is exclusively the economist's[8] task. This implies specification of the variables that may (but need not) have an influence on x_1. (Incidentally, it will now be clear that consideration of this task already governs the choice and definition of the variables constituting the data.) The economist will further reflect whether he feels no a priori suspicion against the assumption of a linear or other specified type of dependence of x_1 on its supposed determining variables. Finally, being aware of the principle of a sufficient basis, he will include in the third set of premises as many and as specific indications as he thinks compatible with the principle of a solid basis.

VII. APPLICATION OF THE PRINCIPLE OF STATISTICAL CENSORSHIP

The list of premises then goes to the mathematical statistician, who first applies the principle of statistical censorship. He will investigate whether at least one set of coefficients and lags exists which (1) has the properties specified in the third set of premises and (2) when combined with the series $x_2 \ldots . x_n$ into a "calculated series" x_1^* (according to the prescriptions given in the second set of premises) leaves only such "unexplained residuals" $x_1 - x_1^*$ in the variable x_1 as do not contradict the premises adopted in the first set.

Generally, the statistician will conclude that the list of premises is contradicted by the data whenever large residuals remain in the observations not specified by the premises (*c*) of the first set as subject to incidental external influences, or whenever a

[8] The "economist" is introduced here as distinct from the "mathematical statistician." While the latter applies the type of reasoning and the procedures elaborated in statistical theory, the former is not supposed to be of the too academic type versed only in abstract deduction from the "economic motive." He is considered to have in addition an intimate knowledge of economic life and of the results of statistical investigations relating to similar countries and periods.

definite[9] cyclical movement remains in the residuals of these observations, with the use of any authorized set of coefficients and lags. Otherwise he will declare that the premises permit of an explanation of x_1 which yields a good fit to the data.

A situation sometimes leading to rejection of a list of premises on the foregoing model is that in which a good fit can be obtained only while one or more coefficients have "wrong signs." A rejection on this ground is likely to evoke a certain suspicion of the value and accuracy of sets of coefficients that all "happen" to come out of the calculations with right signs. Nevertheless, the procedure is quite sound, and the occurrence of wrong signs under such conditions[10] only demonstrates the paramount importance of a correct choice of the additional information. Knowing how easily a statistically undetectable omission of one relevant determining variable, or an incorrect specification of an a priori known lag, may under certain conditions distort the values and even the signs of the other coefficients, the investigator will devote a full share of his suspicion to the less technical part of the procedure: the choice of the premises.

VIII. CONDITIONS GOVERNING THE VALUE OF THE CONCLUSIONS

If, on the other hand, the statistician finds a good fit to be possible with authorized coefficients and lags, this is not a confirmation of the list of premises but a situation in which he can proceed to make a conditional conclusion. Such a conclusion takes the form of the calculation of certain "best estimates" of the coefficients and lags not specified in advance. These are such authorized values of coefficients and lags as give the "best fit" to the variable x_1, according to certain technical criteria.[11] In

[9] Since probability has been introduced in some part of the system, the criterion on which a given set of residuals is deemed to be sufficient evidence of incompatibility between premises and data is a matter of probability calculus. The statistician will accordingly have to fix the small risk of erroneous pronouncements of incompatibility he is willing to incur. The higher he puts his standard, the more frequently he will be unable to reach a conclusion, even in the presence of some indications of incompatibility.

[10] See, however, footnote 17.

[11] The observations specified as affected by incidental unmeasurable external factors are, of course, excluded in applying these criteria.

addition, certain margins of error may be indicated within which the estimate is likely to differ from the true coefficient or lag.[12]

The scientific value of such a conclusion depends on various circumstances, and a number of alternative cases will now be discussed.

It may turn out that the margins of error for some coefficients or lags not specified in advance are so wide that the conclusion lacks any value with respect to these quantities. If, as we now suppose, the best fit is a good fit, this can happen independently to one single coefficient only if the determining variable concerned shows very little fluctuation (other than a smooth trendlike movement). It may also happen to one single lag if the fluctuations of the variable concerned are so gradual that only small differences occur between immediately subsequent values. Apart from these two cases, it always happens to more than one parameter (coefficient or lag) at the same time and is an indication of "multiple collinearity" in the set of variables included. This is the situation in which, besides the equation under study, at least one other linear equation is, whether by chance or for some systematic reason, approximately satisfied by the same variables (occurring in it with the same lags) or by a smaller number of them. The simplest example is direct proportionality in the fluctuations of two determining variables. In all such cases large changes can be made to the coefficients and/or lags concerned without spoiling the good fit—if only these changes stand in certain definite proportions to each other.

This is what happened in Tinbergen's[13] and Polak's [14] attempts to explain the fluctuations in consumers' outlay in the United States. The trouble there was due to a high similarity in the movements of the income of labor and the nonspeculative income of entrepreneurs and owners of property, and also—differences in trend disregarded—in the movements of the cost-of-living

[12] This procedure is again subject to a small risk, to be fixed in advance, of the true difference exceeding the margins indicated.

[13] *Op. cit.*, II, 35.

[14] J. J. Polak, "Fluctuations in United States Consumption, 1919–1932," *Review of Economic Statistics*, XXI (1939), 1.

index. There is a way out of such a situation only in so far as more specific premises can be adduced to provide a more sufficient basis. Thus, in the example referred to, part of the uncertainty in the coefficients could be eliminated by adding to the premises that the marginal propensity to consume is always less than unity and that it is larger for the workers than for other classes of society. Even so, uncertainty remained about the relative influences of incomes and of the cost of living on consumption, and we must either have recourse to budget statistics or wait for a period in which retail prices move differently from incomes to arrive at a decision.

Finally, it may happen that narrow margins of error are found for all coefficients and lags not specified in advance. This, indeed, is the maximum remuneration of the efforts made. A successful conditional conclusion has then been reached.

This is the second, and more refined, type of statistical testing of business-cycle theories. It turns certain more or less plausible premises into perhaps unexpected conclusions. Quantitative precision is given to the influence of all the factors formerly known only qualitatively as possible causes. In particular, the influence of certain suspected causes may turn out to be absent or very small. This was the case with the "acceleration principle," even as a subsidiary cause, in a number of cases investigated by Tinbergen.

IX. TESTING THE LIST OF PREMISES AS A WHOLE AND TESTING ONE DUBITABLE PREMISE

To sum up, we have found that the statistician's "testing" of any draft list of premises presented leads him to make one of the following three statements (I invert the order of the second and third possibilities): (I) The list of premises is contradicted by the data. (II) The list of premises is not in contradiction with the data and provides a sufficient basis for the desired quantitative precision in the conclusions. (III) The list of premises is not in contradiction with the data but is an insufficient basis for the desired quantitative precision in the conclusions. Each of these are verdicts on the list of premises as a whole, and with none of them is the investigation closed. In each case it falls to the economist to initiate the next move.

In Case I, when the statistician has interposed his veto, he may of course be able to give valuable indications of changes in the list of premises which would make it acceptable. But he will always find more than one such possibility. Otherwise the data would support unconditional positive conclusions, and this we have already found to be untrue. Therefore, finding his draft rejected, the economist has to make up his mind what part of his premises to place beyond all doubt and in what part he is most prepared to make concessions.

A similar a priori relative valuation of premises forms the next step in Case II. When valuable conclusions have been derived from a given list of premises, it is time to remember the principle of scientific economy. It may now become important to detect whether the list of premises is not unduly specific; whether some part, and if so what part, of the premises might itself, with the help of the data, be made to follow from the others or at least from some other list of premises which is less restrictive, and therefore still more plausible a priori, than the original list.

In Case III, finally, the economist has to decide whether he can devise any additional premises likely to provide a more sufficient basis for the statistician to work on. If he finds he can, the enlarged list of premises comes up again for the statistician's judgment as described above. Below (see p. 180) I revert to the situation which arises when the economist's concern for a solid basis does not allow him in any way to go beyond the premises already found "insufficient."

This leaves us for the moment with Cases I and II. We now have to discuss the logical possibilities arising from a subdivision of the list of premises into two parts, one consisting of accepted premises, the other of dubitable premises. The former, of course, ought already to have been found compatible with the data. The latter are those which are rejected if the need arises (as happens in Case I) and made dependent if the possibility arises (as may happen in Case II).

In actual investigations such subdivisions usually exist from the beginning, often with dubitable premises in various degrees of presumed reliability. Nevertheless judgments regarding a list of premises as a whole have logical priority, and their prior discussion may have served to elucidate the division of responsibilities between economist and statistician.

In the situation now to be considered the statistician is authorized to take for certain the accepted premises. This allows him to make one of the following statements: (A) The dubitable premise[15] is wrong. (B) It is right because it is confirmed by the data, given the accepted premises. (C) It is not liable to test but if dropped reduces or nullifies the quantitative precision of the whole or part of the conclusions. (D) It is not liable to test but irrelevant to the quantitative precision of conclusions.

Comparison with the list of statements (I, II, III, p. 169) referring to a list of premises as a whole shows that on the new basis one new possibility has emerged. Before, there could be only rejection (I) or nonrejection (II and III) of the list of premises. Now in addition to rejection (A) and inconclusiveness of the test (C and D) there is the new possibility of a statistical confirmation (B) of the dubitable premise. If this occurs, the principle of scientific economy has been applied successfully. Conclusions formerly considered as depending on a premise open to some doubt have now been established on the basis of the accepted premises only.

The question answered in the case in which the test is inconclusive has also changed. While formerly the question was that of the sufficiency (II) or insufficiency (III) of the list of premises as a whole, now the question is that of the relevance (C or D) of the dubitable premise to the sufficiency or insufficiency of the whole list. Case D, irrelevance of the dubitable premise, though logically possible, is of infrequent practical occurrence. Whenever the irrelevance of any dubitable premise to the precision of the conclusions can be foreseen, this premise will not be tested statistically.

In order to illustrate the working of the tests under considera-

[15] In what follows I shall simply speak of the dubitable premise in the singular, as a few premises can always be considered as together constituting one premise.

tion I shall discuss a number of examples connected with various dubitable premises. Let the dubitable premise first be the specification of certain limits to coefficients and/or lags. It is found to be wrong if no coefficients and lags within the limits, combined with the several series as prescribed by the accepted premises, produce a good fit. It is found to be right, and hence superfluous as a premise, if only coefficients and lags within the limits produce a good fit. No decision is reached if values producing a good fit are found both inside and outside the limits. It is in such cases that the imposition of limits is likely to be required as a separate premise in order to give precision to the conclusions. Thus, Tinbergen's conclusion concerning the minor quantitative importance of the acceleration principle in determining the fluctuations in investment activity was in some cases dependent on the premise that the time lag involved in the influence of profits (via expectations) on investment is, at most, say one year.[16]

Suppose next that the accepted premises permit the possibility that a certain variable, say x_2, ranks among the determining variables of another variable x_1, whereas the dubitable premise denies an influence of x_2 on x_1. The latter premise is then rejected if the best fit is at once unsatisfactory without x_2 and good with x_2 included in the calculations. It is confirmed if the inclusion of the variable x_2 gives it a coefficient which is, apart from a narrow margin of error, equal to zero. It is left uncertain if, owing to multiple collinearity between x_2 and other possible determining variables, the coefficient of x_2 is subject to so large a margin of error that neither a difference from, nor virtual equality to, zero can be ascertained.[17]

[16] See *op. cit.*, I, 39, 48.

[17] An interesting complication may arise if the accepted premises prescribe the sign of the coefficient measuring the influence of x_2 if that influence is present (which is denied by the dubitable premise). It may then happen that the best fit obtainable on inclusion of x_2 in the calculations gives a wrong sign for its coefficient, while—in contradistinction to the case of wrong signs considered on p. 167—the fit obtained after excluding x_2 is still satisfactory. Such a result has sometimes been considered a sufficient reason for accepting the dubitable premise. Thus Donner excluded the short-term interest rate from his explanation of share prices in Germany before 1914 because its coefficient did not get the anticipated sign. Tinbergen took a

X. THE ALTERNATIVE TO THE DUBITABLE PREMISE

Before going on to more complicated examples, some attention must be given to the logical relation between the accepted premises and the dubitable premise. Clearly the latter must be logically independent of the former, such that it might be true as well as untrue if the accepted premises are true. Now situations often arise in which one feels entitled, on a priori grounds, to specify an alternative or some alternatives to the dubitable premise indicating the only possibility (possibilities) which would remain if the dubitable premise were not true. It is not worth while drawing attention to this alternative if it simply consists of the negation of the dubitable premise, as in the foregoing two examples. It becomes a matter of primary importance, however, if the alternative can be given the form of a subsidiary premise which is more specific than the mere negation of the dubitable premise. This subsidiary premise, if it actually represents an alternative, stands in a relation of mutual exclusiveness to the dubitable premise. And the presumption that it forms the only alternative is expressed by including with the accepted premises a statement saying that either the dubitable premise or the subsidiary premise is true.

The importance of such a specification of an alternative in the statistical testing of hypotheses has been stressed very much by J. Neyman and E. S. Pearson in their impressive contributions

similar course when obtaining wrong signs for determining variables of minor importance in his study of the causation of investment activity (see *op. cit.*, I, 50). The very fact, however, that the exclusion of x_2 does not decisively deteriorate the good fit means that there is, owing to multiple collinearity between x_2 and one or more other possible determining variables (say x_3, x_4), a large purely statistical margin of uncertainty in the coefficient of x_2, a margin actually extending beyond zero into the region of admitted signs. A corresponding statistical uncertainty must then also arise in the coefficients or lags of x_3 and x_4. In this case, it is true, the exclusion of x_2 is tantamount to adopting that authorized value (viz., zero) for its coefficient which of all authorized values yields the relatively best fit. If not accompanied by a study of the effects of statistical uncertainty in the coefficient of x_2, however, it also means overlooking such uncertainty of the coefficients of x_3 and x_4 as may arise from the a priori possibility that x_2 actually has an influence with the prescribed sign. The dubitable premise would, in fact, be irrelevant only if the investigator were interested exclusively in such coefficients of other determining variables x_5, x_6 (if present) as happen to be unaffected by the uncertainty in the coefficients of x_2.

to that subject. It enables the statistician to shape his tests in such a manner as to increase the frequency of conclusive answers (Cases A and B), thus reducing the cases (C and D) of inconclusiveness.

XI. THE PREMISES OF LINEAR RELATIONS AND CONSTANT COEFFICIENTS

Realization of this point will doubtless render more fruitful the discussion of the possibilities and limitations of statistical induction in dynamic economics. I shall illustrate it in a discussion of the possibilities of testing the premises regarding linearity of relationships and constancy of coefficients and lags—premises which have been subjected to much criticism[18] on a priori grounds. Here a subsidiary premise specifying the alternatives is needed greatly, and its choice becomes a vital question. For, by dropping these premises altogether, one "would make it possible to fit any explanation to any facts," as Mr. Keynes pointedly remarks.[19] Without a subsidiary premise, therefore, only rejection of the dubitable premise would remain possible, and not confirmation.

The matter is still comparatively simple if the linear character of the influence of only one specified determining variable is open to doubt, while the alternative is a curvilinear influence of that variable with the shape of the curve not changing in the course of time. Tinbergen has in such cases resorted to the inspection of "partial scatter diagrams."[20] This technique, though usually leading to the detection of a curvilinear relationship when present, still does not give perfectly conclusive evidence about its absence, particularly if the number of observations is small. (It was not in the cases in which Tinbergen applied this technique.) An admittedly slight chance remains that the curvilinear influence of one determining variable gives rise to fluctuations in the dependent variable very similar to those of one other determining variable, or of a linear combination of some determining variables, and is in consequence ascribed to those variables. The number of such possibilities increases according as the alternative possi-

[18] Keynes, *op. cit.* [19] *Ibid.*, p. 563. [20] *Op. cit.*, I, 77.

bilities admitted by the subsidiary premise include unspecified types of curvilinear dependence with respect to more determining variables. A perfectly conclusive test then requires the inclusion, as additional variables, of the squares or other simple curvilinear functions of the variables concerned. Linearity is confirmed for all these variables if the coefficients of their squares or other functions are all about zero with small margins of error.

Frequently such tests require a larger number of observations than the estimation of "linear" coefficients from the premise of linearity. They have not been carried out by Tinbergen, and in the case of his investigation concerning the United States from 1919 to 1932 the number of observations would generally not have been sufficient. It should be stated, however, that a "majority" of the alternative possibilities would, if present, have been detectable by the technique of partial scatter diagrams. Some very interesting cases of curvilinear influences were, indeed, so detected.[21]

The matter becomes very complicated if the alternatives also held to be possible permit changes to occur in the course of time to the slope or shape of the curve or line representing the influence of a given variable. This seems to me the idea underlying part of Keynes's criticism. It is felt that the influence exerted by one variable on another may well be different in different circumstances. Here I appeal to economists to specify the criticism in order to make its relevance liable to statistical test. To facilitate the choice I list a few circumscribed alternatives to strict stability of the relations.

One may suspect that at some specified moment an abrupt change has taken place in some coefficients, because of a recognized change in economic organization. Thus, Tinbergen could obtain a good fit in explaining capital goods' prices in the United States only by allowing for a sudden decrease in the response of these prices to changes in production, coincident with the remonopolization of the iron and steel industry in 1923. In other cases one might prefer to assume a gradual and smooth change in one or

[21] See, e.g., *op. cit.*, II, 106.

more coefficients as time proceeds. There was, e.g., during the forty-year period preceding 1914, a gradual but distinct decrease in the responsiveness to various factors of the investment in railway rolling stock in the United Kingdom.[22] If a sufficient number of observations is available, such gradual changes might be detected, or their absence confirmed, by breaking up the period studied in two or more parts. Particularly when observations are not numerous, however, the more refined method of allowing for a gradual but not a priori defined change of coefficients in the explaining formula gives a greater possibility of extracting all relevant information from the data and is therefore more likely to give evidence in uncertain cases.

One may in other cases admit the possibility that the influence of one variable on another one depends on the value assumed by a third variable. This means that the influences of the several determining variables are no longer to be treated as additive. A special graphical technique dealing with this type of "joint causation" has been elaborated by Ezekiel and Bean.[23] Arithmetical procedures are, of course, also possible. An example is found in oil-tanker freight rates as influenced by shifts in the demand curve for transportation of oils. The supply of these services is very elastic so long as unemployed tankships are available but suddenly becomes inelastic as soon as the world tanker fleet is fully employed. Accordingly, the influence on rates of a shift of the demand curve from one given position to another one depends very much on the world's total tanker tonnage.[24]

Relationships of this type may be expected particularly in connection with bottlenecks in some part of the economic system. Other instances are, however, also conceivable, e.g., in connection with unmeasurable factors. The cumulated effect of a given dose of public works on entrepreneurs' investment may well depend on whether the works are considered as a temporary expedient or as a normal recovery in government demand.

[22] See the graph on p. 120 of Tinbergen, *op. cit.*, Vol. I.

[23] M. Ezekiel, *Methods of Correlation Analysis* (New York, 1930), chap. xx.

[24] Cf. T. Koopmans, *Tanker Freight Rates and Tankship Building* (Haarlem, 1939).

This statement qualifies Tinbergen's multiplier calculations based on "normal" entrepreneurial reactions as displayed over a complete cycle.[25]

The statistical technique dealing with lags is insufficiently developed. Hence we are not sure whether, and on what premises, constancy of lags may be tested with precision from the usual type of data, or whether some presumed type of systematic variation in lags may be detected easily and with reliability. Purely technical study is urgently required on this important point.

There are usually good chances for testing the clause in the premises specifying, for exclusion from the calculations, the observations largely affected by unmeasurable external factors—provided the premises are otherwise both solid and sufficient. If no observations thus affected have been included erroneously, the net effect of such external factors can always be estimated from the residuals which remain if the explaining formula is applied to the excluded observations. This answers the question whether perhaps too many observations have been excluded. Exclusion of too few, or of the wrong, observations usually manifests itself by large residuals in the observations erroneously included. Yet a small chance remains that no important residuals are found even though one or two actually "disturbed" observations have not been excluded. This is in the rare case in which certain co-ordinated changes in the coefficients of authorized determining variables are possible which alter the values of the explanatory series x_1^* only in such a manner as to absorb just the disturbances in those few observations. The presence of such a situation can always be recognized.

Not every premise liable to doubt is liable also to statistical test, even if all other premises are accepted without doubt. A feature of the premises listed on page 165 which is generally not liable to test is that no allowance has been made for possible errors of observation in the series representing the determining variables. If we suspect that such errors are not negligible, a corresponding change must be made to the list of premises.

[25] *Op. cit.*, II, 162.

I have worked out this case in a special investigation[26] to which I may refer here. It leads to larger margins of error in each coefficient.

XII. CONCLUSIONS

The foregoing review of possibilities may suffice to illustrate the nature of statistical testing of supposed relationships from a unique set of data. It shows that no single clear-cut answer can be given to our initial question: to what extent the results of econometric business-cycle research depend on the data and to what extent on additional information or hypotheses. The relative importance of data and additional information varies from one case to another. Their combination is a complicated process, the result of a continuous dialogue, of a game of give and take, between economist and statistician. The statistician's work is technical and bears no responsibility for the conclusions other than that it must avoid errors of reasoning. Its quality can be improved only in the directions of greater efficiency of, and greater diversity in, statistical tools, in order that conclusive answers can be found to a given type of question more frequently and that answers can be given to more types of questions. The economist must also avoid errors of judgment.

Their collaboration along the lines indicated is not merely a method, as Tinbergen modestly calls it in his title—perhaps because he employs a somewhat limited choice of premises. This is my main point: it is the only method by which the relevant information contained in statistical time series can be extracted and made available for giving such quantitative precision to the supposed relationships of business-cycle theory as it truly supports.

What, then, may be the outcome of this collaboration? Though the foregoing discussion has been concerned only with the explanation of one single internal variable, it will now be clear that the drawing-up of the list of internal variables is mainly the economist's work. Some influence of statistical evidence is not excluded even here.

[26] T. Koopmans, *Linear Regression Analysis of Economic Time Series* (Haarlem, 1937).

Let us suppose first that a basis of premises both solid and sufficient has been reached with respect to each variable to be explained. No new logical problem is involved in the combination of the explanations of the movements of individual variables into an explanation of the fluctuations of the economic system as a whole. This, indeed, is the object of mathematical business-cycle theory. For our present purpose it is sufficient to state that in the case supposed a quantitative explanation of economic fluctuations in the country and period concerned has been erected on premises indorsed by experience and judgment. A knowledge has been obtained permitting the reconstruction of the fluctuations of each internal variable during the period from those of the noneconomic causes and from the initial condition of the country —within the limits recognized as set by random influences of unspecified nature.

Only extrapolations of this knowledge would have a more than historical interest. The following two extrapolations are interesting. The first of these has been greatly developed by Tinbergen. This is the study of the influence which certain measures of business-cycle policy would have had on the stability of the economy in the conditions of country and period. It presupposes, in addition to the logical basis of the whole analysis already set out, that the effect of such measures on each of the causal processes, when present, can be assessed a priori in quantitative terms. Given these conditions, the rest is merely a matter of mathematics. Using the knowledge so obtained as a guide to actual policy again presupposes persistence of the main dynamic features of the economy in the future. Theoretical study of the dependence of the knowledge obtained on the particulars of the period concerned may be helpful in avoiding unsafe extrapolations.

Another extrapolation would be the idea of prediction of future developments. This would, however, be a much more hazardous undertaking. In the first place, it would presuppose a quantitatively much stricter persistence of past relationships in future conditions, or at least quantitatively precise recognizability of such changes as may occur to them. Second, such prediction could be nothing but conditional upon the actual realization of certain anticipated values of the external forces. Since the internal forces are little less important than the external factors in a country like the United States, this would still permit prevision within certain limits. There are, however, other reasons for thinking that such prevision is neither very desirable nor even quite conceivable. With the present organization of society, according as prevision would initially have a limited success, speculation would rush in to anticipate the predicted movements, thus defeating the premises on which the prediction was based and possibly aggravating the fluctuations. It would, therefore, seem better to develop the first extrapolation and reap its fruits before the second might become a dangerous possibility.

Fulfilment at least of the scientific conditions thereto seems not altogether excluded. Favorable to such a development is the circumstance that conclusions about the effect of certain policies on the stability of the economic system are less sensitive to an exact fulfilment of all of the premises. This brings us to the second case, more realistic at the present, in which a basis both solid and sufficient for quantitative precision in the conclusions could not be established in the explanation of one or more variables. About such is the situation in Tinbergen's study of the United States economy (even if we suppose complete acceptance of its premises). A major uncertainty is that concerning the relative influences of retail prices and incomes on consumption expenditure. Uncertainties may be present also in some other equations, but their extent, and particularly their relevance to the various conclusions, is not easily recognized since no margins of error have been calculated.

In such cases there is not one single causal reconstruction of the fluctuations of the period but a set of possible explanations covering a wider range of alternatives according as the gaps between the solid and the sufficient bases are wider. The important mathematical problem, then, is to detect and analyze the common characteristics of all possible explanations and to discover whether certain types of policy may have favorable effects on the stability of the economic system, whichever of the explanations might correspond best to reality. There are indications justifying the hope that valuable answers may still be found to such questions.

It remains to be seen how far the general working hypothesis introduced at the beginning of this article is confirmed by the data. This certainly is the case with the assumed dominance of causal connections in explaining the movements of the variables. Correlations as high as are obtained throughout would be extremely improbable if pure chance were the major determinant of the movements of each variable individually.[27]

Only the experience to be gained from wide application of the methods can give evidence on the other two general assumptions made. As has been said already, unidentified external forces will, if important, manifest themselves in most cases by poor fits incurable by authorized means. Failure to grasp the essential dynamic features of the economy from the study of aggregates or averages of measurable phenomena, as affected by any external factors and by measurable internal factors, might also appear in the end, according as fresh data would cross the regularities the investigator believed to have read from past experience. The setbacks of this nature experienced so far have throughout found explanations within the framework of the concepts and methods here described.

[27] "Pure chance," of course, may govern one or more external factors and thus ointly affect the connected movements of several variables.