Econometric Theory, **31**, 2015, 85–92.
doi:10.1017/S0266466614000206

STRUCTURAL MODELS AND ECONOMETRICS

TRYGVE HAAVELMO
University of Oslo

Editors' Foreword

This article *Structural Models and Econometrics* by Trygve Haavelmo has remained unpublished until now. The paper was retrieved from Haavelmo's archival remains after his death in 1999. It had been presented at ESEM-16 in Uppsala 1954. After the first draft of the paper Haavelmo wrote a brief appendix which was appended to the paper distributed at the meeting.

It is a rare case of an econometric paper written by Haavelmo some years after he put econometric research aside upon his return to Oslo from the USA in 1947. We set out briefly below the context in which the paper was written. Notwithstanding the simplicity of the paper's framework, its relevance to the role of structural modeling and issues of specification in econometrics remains as alive as ever.

At the Innsbruck meeting in 1953 it was decided that the next meeting would be in Uppsala with Herman Wold in charge of the arrangement. Wold wrote shortly afterward to Haavelmo setting out that the meeting would benefit from concentrating discussion on a main theme which, in Wold's view, ought to be 'recursive versus simultaneous systems.' Haavelmo responded, without commenting on Wold's proposal, that he would like very much to take part in the meeting and being currently concerned with theories of long-run growth he would suggest that as a topic for his own contribution. As Frisch also had confirmed interest Wold wrote a joint letter to Frisch and Haavelmo, stating that he wanted René Roy, Hicks, Tinbergen, and Zeuthen on the program committee with one of the two Norwegians. Frisch's seniority and his record as the forceful center of several of the interwar meetings could not easily be bypassed but Wold still made it clear that he preferred to have Haavelmo on the committee in view of the chosen theme for the meeting. Wold furthermore expressed in the letter the opinion that Haavelmo did not share, at least not fully, the 'formalistic approach to the problems of the Chicago School.' Frisch and Haavelmo agreed that Haavelmo would serve on the committee.

Wold came to Oslo for a preparatory talk with Haavelmo half a year before the meeting. They agreed that they would introduce separate sessions, Haavelmo with a paper on *Structural Models*, and Wold on *The Possibilities and Limitations of Recursive Systems*.[1] It was a small meeting, with only 17 papers presented, among them papers by Allais, Malinvaud, and Shackle. Present at the meeting were also Richard Stone, then Vice-President of the Society, and John Chipman.

At the meeting Haavelmo commented upon Wold and vice versa. Richard Stone and Henri Theil commented, Stone siding with Haavelmo and Theil with Wold.

Wold's last comment on Haavelmo was an unfair and rather self-serving long tirade: '... Professor Haavelmo has not at all tried to clear up the issues I find obscure in the interdependent system, ... the rationale of mixing causal relations with identities or equilibrium relations remains as obscure as ever,... Neither is his argument very convincing as a sweeping criticism against recursive systems and least-squares regression. At any rate it runs contrary to many fruitful applications of regression analysis' (*Econometrica* 22, 206).

Three months after the meeting Wold wrote to Haavelmo to state that his paper would be published in René Roy's *Cahiers d'économetrie* and inquired where Haavelmo would publish his paper. Haavelmo responded briefly that he was not going to publish his Uppsala paper beyond what would be included in the report from the meeting, which indeed was very little. The first citations of Haavelmo's Uppsala paper are by Hoover and Juselius (*Econometric Theory*, Volume 31, Number 2, 2015) who discuss the contribution of the paper in relation to Haavelmo's earlier work and by Juselius (*Econometric Theory*, Volume 31, Number 2, 2015) who translates the model in the appendix of the paper into a modern CVAR formulation.

Haavelmo's paper is self-explanatory and requires no comment. It has interest for the historian for its careful reformulations and elaborations on some key points of his tenets in the *Probability Approach*. Perhaps the aspect of the paper that is most relevant to ongoing research is that some of the ideas relate to the method of indirect inference—using one generating mechanism (perhaps the wrong mechanism) to learn about parameters in another generating mechanism, using simulation methods that characterize the limits of the procedure in terms of the parameter of central interest—thereby opening up the prospect of valid inference from misspecified systems. As Haavelmo presciently puts it in closing his article:

> *"There is nothing mysterious about operating with two different stochastic models ... one which we are finally interested in and another applicable to certain data to be used for estimation purposes. But what* <u>*is*</u> *important is, of course, not to force certain data into an alien model."*

Olav Bjerkholt and Peter C. B. Phillips

1. MEANING OF STRUCTURAL RELATIONS

Economic theory is concerned, to a very large extent, with hypotheses and theories that have the form of functional relations between two or more economic variables. Such relations may be purely definitional, that is, relations that define new variables in terms of others, or relations that represent an axiomatic, conventional or "obviously reasonable" constraint upon the variability of the variables considered. Other relations may have a technological foundation. And, finally, there is the large class of economic relations that are supposed to reflect patterns of economic behavior.

The assumed existence of such functional relations may give rise to a great variety of different quantitative or qualitative statements. We may be interested in the fact that the numerical value of a certain variable is given when the values of certain other variables are known (e.g., that the demand for a certain commodity is determined by prices and income). But we may also be concerned only with the form of relations, or some general properties of their form, without making any specific distinction between "dependent" and "independent" variables (e.g., a set of indifference surfaces).

A study of functional relations does not, however, become economics merely by using "economic terminology." It remains pure mathematics until we introduce some idea of applicability to, or likeness with economic reality. The intricate connection between an abstract model and reality may perhaps never become fully clarified. In fact, it may not be possible to do it in a finite number of words or communicable symbols. But one of the essential ideas involved is probably the reference to a connection between theory and some actual or conceivable scheme of experimentation.

An experiment, no matter how simplified, is obviously concerned with a piece of reality. Therefore, to describe exhaustively the "conditions of the experiment" is an endless, perhaps an impossible, task. But we often manage to form pretty good ideas as to the essentials of what is going on. Briefly, we describe the conditions of an experiment by referring (a) to certain specific objects under observation and (b) to some reasonably complete and accurate description of the "environment" in which the experiment takes place. A typical example from analytical economics is a simple demand curve, involving only quantity and price of a commodity as variables. In order to specify the experimental conditions under which such a relation would hold we have, first, to assume that certain other essential factors be held constant, e.g., the consumer's income, prices of other commodities, age, size of family, and perhaps a few others. Then we have to describe the conditions of choice open to the consumer, e.g., that we are going to fix alternative prices (without thereby influencing the basic attitude of the consumer during the course of repeated observations). But there is obviously much more that needs to be said, even about such a relatively simple economic experiment in order to make the description reasonably complete. One needs only to think of all the things one would have to write in an instruction to even the most intelligent assistant observer in order to communicate to him a desired procedure of collecting appropriate data.

The validity of an observed interconnection between economic variables obviously depends on the persistence of a certain type of "experimental conditions." It has little meaning to talk about economic relations that exist without some notion of "environment" in which the relation may be expected to hold good. The totality of properties of the experimental conditions under which a particular economic relation is valid, is often called the structure of the economy considered, and the relation itself is called a structural relation. Thus, a structural economic relation is not actually a particular kind of economic relation, but rather any economic relation associated with and valid for a specified real economic structure that could conceivably be reproduced experimentally.

What is our interest in such relation? The study of structural relations may serve at least these three purposes:

1. To satisfy scientific curiosity.
2. To study the functioning of alternative structures that could have practical interest from the point of view of economic reform.
3. To explain current events in the actual economic structure under which an economy is at present operating.

The notion of some kind of experimental design is relevant in each case, even if the theorist is quite passive with regard to the economic reality around him. For it is an essential part of any form of "explanation" of facts that we want somehow to be able to construct or to reproduce the facts considered in order to understand them. A particularly important idea in this connection is the assumption that it may be possible to explain a more complicated economic structure, such as the functioning of a whole modern market economy, by piecing together relations derived from the consideration of several relatively simple partial experimental designs, or structures. This is for example obviously one of the purposes of studying separately demand and supply functions for specified commodities and for particular groups of buyers and sellers.

Among the various kinds of structural relations perhaps the most interesting ones are those for which the associated design of experiment consists in fixing a set of datum-parameters or "independent variables," the "outcome" of the experiment being the choice of a particular value of some dependent variable. In the following we shall confine our discussion to this kind of structural relations.

2. STOCHASTIC NATURE OF STRUCTURAL RELATIONS

Consider a structural equation derived from some theory of economic choice and which implies that, given the value of some specified variables (e.g., prices), there will be a unique choice of value of a "dependent" variable. If we think of a corresponding design of experiment, we should probably not in fact expect the dependent variable to remain absolutely rigid for repeated experimentation with the same set of values of the independent variables. A natural generalization of the theoretical relation is then to assume that the dependent variable is a stochastic variable for each set of values of the independent variable. Its distribution could depend on the values of the independent variables and on some additional parameters characteristic of the variability of economic choice under partially specified experimental conditions (cf. the design of experiment of the demand function mentioned above).

Let us assume here, for the sake of illustration, that the generalized theory considered has the form

$$Y = f(X_1, X_2, \ldots, X_n; a; u), \tag{1}$$

where $X_1, X_2, \ldots, X_n$ are the specified "datum-variables" of the experiment and where u is a random variable with specified stochastic properties for each set of values of $X_1, X_2, \ldots, X_n$. The functional form f is assumed known while the theory does not specify the numerical value of a certain constant parameter a. According to (1) the dependent variable Y will be a stochastic variable for each fixed set of values of $X_1, X_2, \ldots, X_n$, the distribution of Y depending on the (unknown) value of the parameter a and on the stochastic properties of u. The theory about the variable Y as expressed by (1) could then be interpreted as follows:

Let there be prescribed a particular set of values of $X_1, X_2, \ldots, X_n$. Assume that the value of a is known. Assume further that, for the specified set of values of X_1, $X_2, \ldots, X_n$ and the specified value of a, the distribution of u is completely known. Then we can calculate the probability of Y falling within such bounds as we may be interested in. Suppose that bounds can be found which have these properties: (a) the probability of Y falling within these bounds is high (e.g., 99%); (b) the a priori statement that Y will fall within these bounds would be "interesting", if true. In that case we may interpret our theory as if it states bluntly that Y will necessarily fall within the bounds considered.

An actual design of experiment to which a theory of the kind described might be thought applicable could be as follows: we want to expose the individual or the group choosing the value of Y to a given set of data $X_1, X_2, \ldots, X_n$, relevant for the choice considered. We want to do this in such a manner that if we were to repeat the experiment many times keeping the X's fixed, the process of experimentation should have no influence upon the choice of Y. We want, as far as possible, to insure that the experiments take place under environmental conditions that do not violate the general basis of the theory of choice upon which (1) is built. Then, for such an experiment our theory claims that it would be possible to state in advance that Y will actually fall within certain "interesting" bounds.

One of the fundamental problems of econometrics can now be described as follows: suppose that we are going to arrange (or passively observe) some sort of economic experiment as described above. We know a priori that there exists a number a_0 such that if this number be substituted for a in (1) every statement that the theory (1) and its stochastic specification permits us to make about the variable Y will apply to a corresponding variable in the experiment. We do not, however, know the number a_0. The problem is somehow to obtain information regarding a_0. Among the various possibilities that may exist for getting this additional information we shall confine ourselves here to cases where the solution is sought through some procedure of statistical estimation.

3. THE PROBLEM OF ESTIMATING UNKNOWN PARAMETERS IN STRUCTURAL ECONOMIC RELATIONS

Consider the problem of obtaining an estimate of the parameter a in the stochastic structural equation discussed above. The most immediate idea in this connection would seem to be that we use data which could be assumed to have originated from a design of experiment identical with that to which the theory is meant to be directly applicable. More specifically, suppose that our theory (1) and its stochastic specification defines the probability density function of Y as

$$p(Y; X_1, X_2, \ldots, X_n; a). \tag{2}$$

Let X_{ij} denote a particular value of X_i, i = 1, 2, ..., n, j = 1, 2, ... , N, and let the random variable Y corresponding to a particular set of X-values $X_{1j}, X_{2j}, \ldots, X_{nj}$ be denoted by Y_j, j = 1, 2, .. . , N. On the assumption of "independent trials" the joint distribution of $Y_1, Y_2, \ldots, Y_N$ would be

$$\prod_{j=1}^{N} p\left(Y_j; X_{1j}, X_{2j}, \ldots, X_{nj}; a\right). \tag{3}$$

Let $y_1, y_2, \ldots, y_N$ denote a sample of the variables Y_j. That is, we assume that we have data from an experimental design identical with that for which the stochastic specification above is assumed to be valid. Then the problem of estimating a consists in deriving an estimate of the parameter a in (3) by means of a sample from <u>that same distribution</u>. This problem, though it may be technically difficult, is a straightforward problem of statistical estimation. <u>However, it is by no means certain that data having the probability distribution (3) are the only kind of data from which a could be estimated. In general, any kind of data following a probability distribution which depends in a known way upon a may serve as a means of estimating a, provided that the method of estimation is based on the appropriate stochastic specification of the data to be used for this purpose</u>. Failure to realize the possible difference between the stochastic structure for which a theory is valid and the stochastic structure of data available for the purpose of estimating unknown parameters of the theory has been a constant source of confusion in recent discussions on econometric methods. In order to obtain data corresponding to the design of experiment of a particular structural relation it is not enough to pick out some economic data bearing the same names as the variables of the theory. The relevant experimental design of such data may be quite different from that which forms the basis of the theory. On the other hand, it is by no means certain that such data could not be used to gain information relevant for the theory concerned, provided their true stochastic nature is properly specified.

One of the simplest type or differences between a given structural model and the model relevant to using certain data for estimation purposes is the occurrence of errors of measurement not accounted for in the original design of experiment.

To use such data for estimation purpose requires a reformulation of the theoretical model into a new stochastic scheme that is relevant to the data available. But there may be much more profound differences between the two kinds of structures considered, due to the fact that the design of experiment relevant to the data considered is essentially different from that which belongs with the structural relation in which we are finally interested.

As an illustration consider again a simple theory of demand corresponding to a design of experiment which consists in fixing alternative prices of the commodity involved, keeping income, other prices, etc. constant, and observing quantity demanded. We may be interested in the possibility of predicting the outcome of such experiments, e.g., for the reason that we expect to introduce a new economic policy of price fixation under which our theory would be directly applicable. Suppose that this theory, in a proper stochastic formulation, involves an unknown parameter. How do we estimate this parameter without actually arranging direct experiments to obtain data? It could be that an actually operating market mechanism which we have occasion to observe does in fact expose the consumers to something very close to the experimental design about which the theory is concerned. But even if the actual market operates quite differently, it is possible that the stochastic model which is relevant to the market data still in some way contains the unknown parameter and that the market data therefore could be used to estimate it. There is nothing mysterious about operating with two different stochastic models of demand, one which we are finally interested in and another applicable to certain data to be used for estimation purposes. But what is important is, of course, not to force certain data into an alien model.

4. HAAVELMO: APPENDIX TO PAPER ON STRUCTURAL MODELS

Let us consider a very simple illustration, which probably is not very good theory, but which is sufficient to bring out the principles about which we have been talking. Let x denote the quantity demanded of a certain commodity and p is price. Suppose that for a certain specified design of experiment we know that the following stochastic model is relevant:

(A) $y = a\,p + u$

(B) p can be deliberately fixed for experimental purposes

(C) For every fixed value of p u is an unobservable random variable with a known distribution which does not depend on the value of p. The u's are independent in repeated trials.

(D) $E(y) = a\,p + \text{constant}$

(E) a is an unknown parameter.

If we had a set of data $(x_1, p_1), (x_2, p_2), \ldots, (x_N, p_N)$ from an experimental set-up satisfying the requirements listed above, there are simple and well-known methods of estimating the value of a.

Suppose now that the data at hand do not correspond to the design of experiment related to the specification above. Suppose that the data instead are market time series related to the following stochastic scheme:

(a) $y(t) = a\ p(t) + w_1(t) + h\ w(t)$

(b) $p(t) = \beta\ p(t\text{-}1) + w_2(t) + k\ w(t)$

(c) w, w_1, w_2 are mutually and serially independent (unobservable) random variables with known distributions.

(d) a is the same unknown parameter as in (A). β, h, and k are unknown constants.

If one wants to introduce the notion of a "causal chain" in this scheme, the interpretation could be as follows: some outside event (newspaper reports on threat of war, etc.) fixes a sample value of the random variable w(t). The sellers fix the price p(t) partly as a function of p(t-1), partly as a function of the outside random event w(t), and finally through some other considerations represented by the random variable $w_2(t)$. When the price is fixed, the buyers are influenced partly by the same outside event as the sellers, e.g., to the extent hw(t), partly by the price p(t) , and, finally, by some random factor $w_1(t)$ characteristic of the normal stochastic reaction of buyers.

One important feature of such a model, as compared with the preceding one is, that now also p(t) will be a random variable and that, unless $k = 0$, $h = 0$, or both, we shall now in general find that, for a given value of p(t-1), $E\ y(t) \neq a\ p(t) +$ constant.

The problem of estimating a is, of course, still easily soluble (provided $\beta \neq 0$), but the estimation procedure has to be adjusted to the model relevant to the data at hand.

The model above is artificially simple but it contains a realistic element which I do not think it is easy to explain away by means of lags or assumptions of stochastical independence, viz. the fact that both the process of price fixation and the reaction of buyers in an actual market may be influenced in part by the same external random events.

NOTE

1. Haavelmo dropped the idea of another contribution to the meeting. His monograph *A Study in the Theory of Economic Evolution* (Amsterdam: North-Holland, 1954) was under publication and had aroused considerable interest already before it was published. Haavelmo was clearly motivated to follow up the book with a paper that, in the end, was never written.