

A new type of socio-economic system

Guy H Orcutt *

By permission of MIT Press reproduced below is the full text of Orcutt G (1957) 'A new type of socio-economic system', *Review of Economics and Statistics*, 39(2), 116-123. The original page numbering is indicated throughout the text in square brackets. In the original all footnotes appeared on the page in which they were alluded to in the text; here all footnotes appear on the final page.

[116]

Existing models of our socio-economic system have proved to be of rather limited predictive usefulness. This is particularly true with respect to predictions about the **effects of alternative governmental actions** and with respect to any predictions of a long-range character. It is even the case with respect to **very short-run forecasting**. In addition, it is recognized that current models of our socio-economic system have an unduly narrow reach in that they have little to say about such fundamental things as the size and location of the population of individuals, of households, or of firms.

It is also true, but not so widely noticed, that **current models of our socio-economic system only predict aggregates and fail to predict distributions of individuals, households, or firms in single or multi-variate classifications**.

The severe difficulties of testing hypotheses and of estimating relations by use of highly aggregative time series are by now fairly widely understood by economic statisticians and are beginning to be more adequately recognized and faced by the economic profession in general.¹ These difficulties and the resulting failure to achieve satisfactory testing or estimation at a highly aggregative level have been among the elements leading to the large interest now exhibited in formulating and testing hypotheses about the behavior of such elemental decision-making units as individuals, households, and firms. As a result, research efforts in the behavioral sciences have yielded and show promise of yielding very substantial amounts of knowledge about such elemental decision-making units. However, existing models of socio-economic systems are neither built in terms of such units nor are they well adapted to making use of knowledge about such units.

There is an inherent difficulty, if not practical impossibility, in aggregating anything but absurdly simple relationships about elemental decision-making units into comprehensible relationships between large aggregative units such as industries, the household sector, and the government sector. Strictly speaking, the difficulties involved in adequate aggregation of relationships about elemental decision-making units are not just technical ones that are capable of solution by better logicians. This type of difficulty is indeed present and formidable enough. A more basic difficulty is that such aggregation cannot be correctly made without a reasonable model of the same socio-economic system stated in terms of the behavior and interaction of the

elemental decision-making units. Then, and only then, could ways be found of aggregating relationships without a disastrous loss of accuracy of representation.

Aggregation of relationships about elemental decision-making units is fairly easy if the relationships to be aggregated are linear. Under these circumstances aggregation may be useful if a limited number of variables appears over and over again. However, if nonlinear relationships are present, then stable relationships at the

[117]

micro level are quite consistent with the absence of stable relationships at the aggregate level. The following simple numerical example may illustrate this point.

Let us suppose that we have 100 individuals, each of whom produces an output, Y , and has an input, X . Let the relation of Y to X be the same for each of the 100 individuals, so that $Y = 0$ when $X = 0$ and $Y = 1$ whenever $X = 1$ or $X = 2$. Now, given the values of X for each of the 100 individuals, it is clear that the sum of the Y 's will have a definite value. However, it is equally clear that the sum of the X 's is not enough to specify the sum of the Y 's. Thus, if each of the 100 X 's equals one, then the sum of the Y 's also will be 100. **If, however, 50 X 's each equal 0 while the other 50 X 's each equal 2, then the sum of the Y 's will be 50 despite the fact that the sum of the X 's is still 100.** The sad truth is that even in this very simple situation the aggregate value of Y **depends on the distribution of X values**. It also is true that the behavior of decision-making units is known to abound in nonlinearities and discontinuities of many sorts.

This paper represents a first step in meeting the need for a new type of model of a socio-economic system designed to capitalize on our growing knowledge about decision-making units. Many more steps will be required and the labors of many individuals will be needed. Nevertheless, it seems reasonable to claim that models of the type suggested in this paper could perform a useful function, by facilitating and improving predictions about aggregative aspects of our socio-economic system, by facilitating and improving testing of hypotheses about behavior of individuals, households, and firms, and by furnishing guidance in the selection of research efforts.

The most distinctive feature of this new type of model is the key role played by actual decision-making units of the real world such as

the individual, the household, and the firm. In each time period, some types of behavior of each individual unit are conceived of as being functionally dependent on prior events, and other types of behavior of each individual unit are conceived of as being determined by one or more random drawings from one or more discrete probability distributions.

The probabilities associated with alternative behaviors or responses are treated as dependent on conditions or events prior to the behavior.² Thus, these probabilities vary over time as the system develops or as external conditions change, and the model presented is a recursive type which progresses by short, but discrete, steps. Solution of models of the type presented here will involve extensive calculations, and it is only the advent of very powerful computing facilities that makes this kind of model an exciting possibility.

Predictions about aggregates will still be needed but will be obtained by aggregating behavior of elemental units rather than by attempting to aggregate behavioral relationships of these elemental units. That is, aggregates will be obtained from the simulated models in a fashion analogous to the way a census or survey obtains aggregates relating to real socioeconomic systems. Given a satisfactory model of the socio-economic system developed in terms of elemental decision-making units, aggregation of relationships would become more nearly feasible. Such aggregation might well be interesting and useful, but it would no longer be a necessity.

DESCRIPTION OF NEW TYPE OF MODEL

This new type of model consists of various sorts of interacting *units* which receive *inputs* and generate *outputs*. The *outputs* of each unit are, in part, functionally related to prior events and, in part, are the result of a series of random drawings from discrete probability distributions. These probability distributions specify the probabilities associated with the possible outputs of the unit. The appropriate probability distributions are determined by in-

[118]

puts into the unit and the *operating characteristics* of the unit. They therefore change from period to period as new inputs occur.

The *units* of this new type of model may, if desired, be large aggregates such as markets or industries, but in general they are elemental decision-making entities such as individuals, families, firms, labor unions, and governmental units. There are thus a very large number of each of a relatively few different types of elemental units or entities. The exact number of types of units used will be a matter of choice and will be somewhat dependent on the operating characteristics selected to describe units and on the available data and knowledge. The number of

units of each particular type in the model will be set, insofar as possible, equal to, or at least proportional to, the number of the corresponding units in the real socio-economic system being described.

An *input* into a unit is anything which enters into, acts upon, or is taken account of, by the unit. Inputs thus include what are commonly called economic inputs, but the concept is broader since they may include such things as rainfall, information, social pressures, age, etc. Inputs may have been produced as previous outputs of other units or they may derive from the physical environment.

An *output* from a unit is anything which stems from, or is generated by, the unit. It thus includes economic outputs, but may also include such things as expression of opinions, actions of all sorts, birth of a child, marriage, divorce, location, and death. An output of a unit may be also an input into the same unit as in the case of the birth of a child.

There are a variety of outputs which are possible for each type of unit in the model. The *operating characteristics* of any unit are equations, graphs, or tables which either determine outputs or the probabilities of possible outputs by the unit as a function of the previous inputs into the unit. For example, if death in an interval of time is taken as a possible output of a particular individual, then one operating characteristic of this individual might be a relation specifying the probability of its death as a function of its age, sex, race, marital status, and occupation. This usage of the term "operating characteristic" is similar to that frequently intended when reference is made to the operating characteristics of a condenser, a resistor, a light bulb, a vacuum tube, etc.

Operating characteristics are, in general, regarded as stable aspects of units. Units having identical operating characteristics are considered to be of the same type. However, it must not be expected that units of the same type will have identical outputs. In part, this will be a consequence of differing inputs. But even units having identical operating characteristics and receiving identical inputs will not in general have identical outputs. This follows because many of the operating characteristics and the inputs only determine the probabilities associated with each possible output. Actual outputs are then determined by one or more random drawings from the specified probability distributions.

Many of the operating characteristics are conceived of as specifying probabilities of various outputs rather than precise outputs, because this is the form taken by much of our knowledge about small decision-making units. Thus, even after taking account of as many factors or inputs as is feasible, there almost always remains a considerable amount of individual variation. However, a substantial amount of useful regularity

is often discovered in the relative proportions in which large numbers of individuals produce alternative outputs under conditions which appear homogeneous. It is, of course, because of this that insurance companies do so well. It is this fact that makes a probabilistic approach seem highly desirable. The use of a probabilistic approach, based upon knowledge similar to that contained in mortality tables, reflects the state of knowledge about small decision-making units. It does not imply much, if anything, about the underlying nature of reality. Furthermore, use of a probabilistic approach where it seems most suitable does not exclude use of exact or functional types of relationships where they seem most suitable.

To facilitate ease of handling this type of model, all probability distributions are treated as discrete, and these discrete probabilities always refer to discrete units of time.

The basic interval of time will be a relatively short period, such as a week or a month. All stock variables will be treated as measured at

[119]

the beginning of each period, and all flow variables will be treated as though constant over each period. All change in them will be accomplished by discrete steps which take place between periods.

In general, inputs will be treated as not modifying any outputs or probabilities associated with outputs until at least one period after they become inputs. If in certain cases it seems desirable to treat some inputs as modifying outputs or probabilities associated with outputs during the same period in which they occur, this will be done subject to the limiting condition that the system as a whole is to remain a recursive system. This means that this type of model can always move forward in the generation of new outputs without the solution of any simultaneous equations. This feature will facilitate the use and interpretation of such models and may assist in giving them a causal interpretation. All sorts of interactions are possible, except that responses to outputs are treated in general as though they required one or more time periods to materialize.

The exact specification of types of outputs and inputs will, of necessity, depend on available knowledge, obtainable data, and specific objectives. Nevertheless, it seems highly desirable to aim at the inclusion of sufficient sorts of outputs and inputs to facilitate the testing of hypotheses and the making of various kinds of predictions relating to population size, and its distribution by age, sex, location, marital status, occupation, employment status, income, assets, and consumption.

Operating characteristics may relate the probabilities of alternative outputs to inputs by means of equations in which probabilities or parameters of probability distributions appear in

the role of dependent variables, and inputs or functions of inputs as independent variables. Use of an ordinary regression equation with assumed normality of errors would be a special case of this method. The predicted "expected" value of the dependent variable would be taken as the mean of the appropriate probability distribution of this variable. The standard deviation of this probability distribution would be set equal to the standard error of estimate of the regression equation. Given the assumption of normality, the probability distribution is thus completely specified. Operating characteristics also may relate the probabilities of alternative outputs to inputs by means of tables, such as mortality tables or tables giving the probability of a birth in a given time interval to a woman of specified age, marital status, number of previous births, etc. In this case, the values of the various inputs serve to locate the position in the table that contains the probability considered to be appropriate.

WAY OF OPERATING THIS TYPE OF MODEL

When models of the type proposed in this paper are actually constructed, how can they be used? How can models with so many interacting units actually be employed?

A variety of methods, or combinations of methods, may turn out to be feasible. It is too early to say which of these will be the best. However, the following method is feasible, readily comprehensible, and may serve to illustrate still further the proposed model. Using this approach the model would be simulated on a large electronic machine, such as the IBM 704 or the UNIVAC II, or some improved successor to these powerful giants. The units in the model are given initial characteristics in accord with whatever initial distributions are considered to match those of the real socioeconomic system being dealt with. Use of the initial conditions in connection with the relations and tables specified by the various operating characteristics yields the probabilities associated with alternative outputs of each unit. Actual drawings take place, and the selected outputs are produced. These outputs become inputs in the appropriate units. The second round then proceeds in a manner similar to the first; but this time, since the inputs have modified the characteristics associated with units, different probabilities or probability distributions are determined by the various operating characteristics. Random drawings again take place and serve to determine the specific outputs, these in turn are translated into inputs, and everything is ready for the third period's round of activity. At every instant of time, the characteristics associated with each unit, such as age, marital status, income, asset structure, location, etc., are precisely specified. The

[120]

operating characteristics associated with each unit, and the previous inputs into the unit, de-

termine the probability of occurrence associated with each of the various actions or types of behavior that the unit may produce. By a random sampling operation, in which these probabilities are used, the precise actions or types of behavior are determined for each unit.

That all of the necessary operations could, in fact, be carried out effectively by a large electronic calculator seems reasonably clear, when it is considered that they could all be carried out straightforwardly by a good record-keeper armed with a desk calculating machine and a table of random numbers. Only the time and cost involved serve to make high-speed electronic calculation a necessity.

The only operation that might be elaborated on usefully in this paper is that of random sampling. The main requirement for effectively carrying out the specified random sampling operations is a huge supply of random numbers. Since large electronic machines already have been used to produce millions of random digits, the procedure either would be to produce and store adequate quantities of such numbers on magnetic tapes or to introduce a sub-routine, for producing random numbers, into the complete program set-up for one of these electronic giants. Given the supply of random numbers, execution of random sampling operation might proceed as in the following case.

Assume that a unit may select one of K alternatives and that, given the characteristics of the unit and its previous inputs, the probabilities associated with each of these K alternatives are P_1, P_2, \dots, P_K , respectively. Then a range of whole numbers is associated with each alternative. The range of numbers associated with each alternative is chosen so as to be proportional to the probability of the specific alternative. Thus, if P_1 is .139, the range of numbers 1 through 139 might be used. And if P_2 is .105, the range of numbers 140 through 244 might be used. Then a specific one of the K alternatives is chosen by using a three-digit random number from a uniform probability distribution. If the number is in the range 1–139, the first alternative is specified. If it is in the range 140–244, the second is specified. If it is outside of these ranges then some other alternative is specified.

At the present time, the speed and capacity of electronic computers would still put economic limits on the number of units that could be handled in the above fashion. This means that it would be necessary to infer the properties of models with hundreds of millions of units from models having something like tens of thousands of units. It seems fairly certain that such extrapolation would definitely be feasible. In fact, it would seem to be very straightforward compared to the problem of making inferences from models with only a very small number of units. Furthermore, given the fantastic rate at which the power, capacity, and speed of calculating machinery is increasing, it does not

seem unreasonable to believe that within five to ten years it will be possible to operate such a model with substantially more units. Whether any significant gain would be achieved in going from models of, say, ten thousand units to models with millions of units is not so evident. The gain would depend on the extent to which such models are elaborated and the extent to which predictions are desired for very small sectors of the socio-economic system. The minimum number of units that reasonably might be used is the number needed to approximate adequately the initial joint distribution of units by characteristics of the real socio-economic system being represented. As long as the proportion of units in the various cells was maintained, any larger number of units presumably could be used without altering the expected value of aggregates. The variances associated with estimators of aggregates would be expected to vary inversely with the total number of units.

There is at least one alternative approach to solution of models of the type discussed in this paper which at first sight may seem preferable. This approach would be a head-on one, in which, having completely specified the model and having specified the aggregates of interest, one then proceeds to derive the probability distributions of these aggregates by purely deductive means from the model. In principle this is possible, and, in fact, a set of calculations that would achieve this could certainly be specified. It may be that this approach

[121]

could be, and will be, carried through by someone; and if so, we should all be grateful, since the important thing is effective implementation of such models and not the particular manner in which they are solved. The major reason that this approach is not suggested here is that I believe that, while this approach seems an obvious one and might yield somewhat more precise knowledge of the solution, it would in fact involve many times the computational effort than the one suggested in this paper. In view of the fact that even the attack suggested will involve a very substantial computational effort, the volume of computation required to reach a satisfactory solution cannot as yet be ignored.

The basic difficulty with the head-on deductive approach is that, in order to compute the probability distribution associated with each aggregate of interest, it would be necessary to calculate the probability of each possible way of reaching each possible value of each aggregate and then carry out the required summing of these probabilities. But in a system such as we envisage, the number of possible ways of reaching a given value or range of values for a given aggregate would certainly be fantastically large, since the number of paths that might be followed by each individual is already almost beyond comprehension, if many variables and many time

periods are involved. Since each possible variation of path of each and every unit will correspond to a different path by which the system generates aggregates, the problem of keeping track of all possible paths and their respective probabilities appears rather appalling. Nevertheless, it is probably true that, by appropriate mathematical techniques or by working with only the first few moments, ways can be found of drastically simplifying what at first appears to be an impossible computational problem. If this turns out to be the case, so much the better.

However, even if alternative approaches do turn out to be feasible, an approach based on simulation of the model does have important advantages which should not be discarded lightly. It is likely to be easier to modify as necessitated by changes in knowledge about operating characteristics of units. It can be made essentially unaffected by broad changes in the choice of aggregative outputs to be observed. It is not likely to require as many restrictive assumptions in order to facilitate solutions; and lastly, but perhaps not least significantly, it is intelligible to people of only modest mathematical sophistication.

DETERMINING THE INFLUENCE OF PARAMETERS

One advantage of extremely simple models is the analytic possibilities which they afford in determining the way in which aggregate results are related to specification of parameter values.

The importance of determining the connection between parameter values and aggregate results is, of course, considerable, both for purposes of deriving policy implications and for purposes of determining which parameters are known with sufficient accuracy and which are not. Research can then be more effectively directed into areas in which it is critically needed.

In models of the type suggested in this paper, or even in relatively simple highly aggregative models, the possibility of analytically determining the influence of choice of parameter values may be remote. Nevertheless, experiments may be conducted in which parameter values are systematically altered and the resulting behavior of the model observed. By means of systematic experimentation and the use of multi-variate techniques, it will be possible to obtain linear or quadratic approximations to the true relationships between aggregative behavioral aspects of the model and values selected for the parameters. Such approximations could and probably would be centered on the specific parameter values considered the most realistic.

USES OF THIS TYPE OF MODEL

Models of the type suggested in this paper could perform a useful function by increasing the range of predictions that are feasible, by facilitating and

improving prediction, by facilitating and improving testing of hypotheses, and by furnishing guidance in selection of research efforts.

Models of the type suggested can increase the range of predictions which are feasible in

[122]

two sorts of ways. By making it possible to work with models incorporating a much wider range of behavior, they will directly assist prediction-making in areas that existing models of our socio-economic system do not deal with. Such models also can increase our predictive range by providing predictions of both single variate and multi-variate distributions, all quickly accessible in tabular or graphical form by spot interrogation.

Such models could facilitate and improve prediction about socio-economic aggregates by providing a method of bringing to bear knowledge about the elemental decision-making units that make up a socio-economic system. Such models could be used either for short-run or long-run forecasting by appropriate selection of initial conditions and by altering the number of periods the model is run. These models could be used either for unconditional forecasting or for predictions of what would happen given specified external conditions and governmental actions. The most that would be involved here would be substitution of certain things as given instead of having them generated by some process. All predictions could be obtained in the form of expected values plus some measure of uncertainty. Or, if desired, they could be in the form of confidence interval estimates. This is possible, since each time the model is started off with specified initial conditions and let run, it will generate one estimate of each aggregate of interest. Estimates on successive runs will be independent since all random sampling will be independent as between runs. Thus, by running a given model, with given initial and external inputs, more than once, it is readily possible to estimate the expected value and the variance associated with estimates of each aggregate. The choice of aggregates to be obtained has nothing to do with the operation of the model except for specification of what aspects of what units are to be added or averaged.

Models of the type suggested in this paper could facilitate and improve testing of hypotheses about elemental units by permitting testing of them at any level of aggregation. Such models also would improve the testing of such hypotheses by keeping the interrelated nature of the system in the consciousness of the investigator and by helping him satisfactorily to take it into account.

The role of such a model in guiding selection of research efforts would be similar in nature to that provided by other models of the socioeconomic system. They permit the researcher to see how small pieces can be fitted together and to see

where there are serious gaps or weaknesses. They enable him to produce a small piece that will contribute effectively to a useful whole. Since most research can be done effectively only in fairly small pieces, this is important. The main advantage of this sort of model in providing guidance in selection of research effort lies in the fact that the basic units are chosen to be elemental decision-making units of a sort not yet effectively incorporated into other available models of our socio-economic system.

ILLUSTRATIVE MODEL

The following model is included to lend concreteness to the previous discussion. A relatively simple model has been chosen as the most effective instrument for clarifying the ideas expressed in this paper. It has been chosen with the idea that it might be suggestive of ways in which useful and realistic models could be developed. Achievement of a realistic model of the socio-economic system obviously will require reinterpretation and reformulation of many existing research results, extensive research directed at filling in gaps, and considerable programming effort and computing time in connection with simulating the model on a large electronic machine. This is a large and long-range research program, and the most that can be hoped from this paper is that it will assist in stimulating its execution.

The model sketched here has three different kinds of units: individual males, individual females, and married couples.

The possible outputs of individual males and females are entrance into marriage and death of self. The inputs of each male and female after birth consist only of time.

The possible outputs of married couples are male and female children and dissolution. The inputs of each married couple consist of the ages of the husband and wife and the presence

[123]

and ages of male and female children produced by the marriage.

The only operating characteristics ascribed to individual males and females are those having to do with death and marriage. Death of any individual male or female comes about as a result of a chance event in which the probability of death

during each month is given as a function of the age of the specific male or female in question. A different function is used for each sex. The age of each individual is obtained as the difference between the present date and the date of birth of that individual.

Marriage of a specific male to a specific female during a specified month occurs as a result of a chance event in which the male in question either remains single or else marries a specific female out of the group of unmarried females. There is a probability associated with each of his alternatives and its value is considered to be a function of the season, age of the male, age of the female, and relative number of marriageable males and females. In practice this and other matching problems would probably be handled by a two-step probability process.

Birth of no children, one boy, or one girl during a specific month occurs as a result of a chance event in which the probability associated with each alternative is considered to be a function of marital status of mother, age mother, number of previous births and interval since last birth, and the season. The possibility of multiple births is not introduced.

Dissolution of a couple automatically takes place if one or both die. Dissolution of a couple by divorce is specified to be a chance event. The probability of divorce in a specific month is given as a function of duration of the marriage.

Previous marriages are not assumed to influence probabilities associated with subsequent marriages or divorces.

In several respects, even such a simple model as this would already be more complete than existing formalized models dealing with population change. It would, of course, be desirable to introduce explicitly several other variables such as income and location. To do this would require extension of this oversimplified model to include business firms and government.

It also will be necessary to implement such models by introduction of explicitly and quantitatively stated initial conditions, operating characteristics, etc., and by actually simulating them on large-scale computers. Work by several individuals on these various aspects of implementation is in progress, but final success of the ideas sketched in this paper will require a long-term effort by many individuals of widely assorted abilities.

* The author is heavily indebted to many people for important suggestions and criticisms. These include Mrs. Alice Rivlin, Mr. Martin Greenberger, and Professors Dorfman, James Duesenberry, T.C. Koopmans, John Lintner, John Meyer, James Morgan, Robert Solow, and Daniel Suits. The author is also deeply indebted to the Carnegie Foundation and to the Ford Foundation for fellowships which made possible the study resulting in this article. However, the conclusions, opinions, and other statements in this article are those of the author and are not necessarily those of these two foundations or of individuals who have been helpful.

¹ These difficulties include fewness of observations, lack of independence between successive observations, multicollinearity, simultaneous and feed-back relationships between the variables, auto-correlated errors, errors of observation, missing data, index number and aggregation problems, and difficulties inherent in recognition or even specification of policy actions in terms of highly aggregative time series. The list of individuals primarily responsible for originally brining these problems to the attention of economists would include, among others, the following names: D. Cochrane, R. Frisch, T.

Haavelmo, M. G. Kendall, T. Koopmans, G. Orcutt, E. Slutsky, R. Stone, H. Theil, G. Tintner, H. Wold, and G. U. Yule. The number of individuals who have made significant contributions to the problems mentioned above is, of course, much larger and would include many statisticians who have been only remotely interested in economic time series. Nearly all the economists who have worked on various problems connected with using highly aggregative time series started out with the notion of finding ways of overcoming some particular difficulty. But despite the fact, or perhaps because of the fact, that statistical advances have been achieved, it has become increasingly apparent that insufficient evidence remains in highly aggregative economic time series for effective testing of economic hypotheses.

² For some interesting examples of the use of probability processes in economic models, see the following literature: David Rosenblatt, "On Some Stochastic Process Formulations of Individual Preference and Consumer Behavior," abstract in *Econometrica*, XXIV (July 1956); I. Blumen, M. Kogan, and J. McCarthy, *The Industrial Mobility of Labor as a Probability Process* (Cornell University, 1955); Robert Solow, *On the Dynamics of the Income Distribution* (Unpublished Ph.D. dissertation, Harvard University, 1951); Robert Summers, *An Econometric Investigation of the Size Distribution of Lifetime Average Annual Income* (Technical Report No. 31, prepared under contract N6ONR-sI33 [NR~47-004] for Office of Naval Research, Dept. of Economics, Stanford University, 1956).