SYMPOSIUM ON MODEL SAMPLING EXPERIMENTS

A symposium on "Model Sampling Experiments" was held at the 24th Annual Conference of the Indian Society of Agritultural Statistics in Madras on December 31, 1970 under the Chairmanship of Prof. M.C. Chakrabarti, Head of the Department of Statistics, Bombay University. The Rapporteur for this symposium was Dr. P. Narain*. Four papers from varied disciplines, *viz.*, economics, genetics, sampling techniques and computor science, were contributed at this symposium. Extended summaries of the papers contributed by the speakers are given in the following paragraphs.

A. L. Nagar ** : *Simulation in Economics* :

2. An economist or a policy maker is often faced with the problem of measuring the consequences of changing the magnitudes of certain economic variables on some other economic variables, for example, how it would affect the total income, the total net investment etc., in an economy if the government expenditure is lowered or raised by a certain amount, or if the rate of interest is changed in a certain direction. The only data available to the economist for this purpose are those generated in the past. Using this past experience, economic hypotheses are set up, tested and altered, if required, over the passage of time.

3. The basic procedure is to explain the structure of the economy with the help of a set of inter-dependent functional relationships (economic hypotheses) between several relevant variables which may be linear or non-linear. The object of an econometric simultaneous equations model is to describe the structure of a particular economy, to study effects of changes in exogenous (independent) variables on endogenous (dependent) variables, and to forecast the magnitudes of endogenous variables for the period of prediction. The structure is completely determined as soon as the

* Institute of Agricultural Research Statistics, New Delhi.

** Delhi School of Economics, University of Delhi.

parameters of the model are arrived at by employing a suitable method of estimation. However, before proceeding to make forecasts it may be useful to exmine as to what extent the model can reproduce, for the sample period, the economic system which it is supposed to describe. This can be done by simulating the economic system. The simulation paths of the endogenous variables may be traced out by considering alternative specifications of the values taken by the exogenous variables.

4. Dr. Nagar then went on to discuss briefly the simulation procedure adopted by him in context with the Brooking's econometric model for the United States economy (Nagar[1], 1969). Simulated solutions of a reduction in federal income tax rates were made. Quarterly tax reductions were derived by shifting the tax rate parameters. This model involved about 350 equations in more than 1,000 variables. A condensed version of this model with only 112 equations was, however, studied with the help of quarterly time series data available for the period 1953-62. Since the structural equations involved were non-linear in the endogenous variables, iterative methods were employed. The solutions depend on the least square estimates of the structural coefficients involved in the equations as well as the properties of the structural disturbances which are usually assumed as independently and identically distributed as a multivariate normal law with zero means and a specified covariance matrix. If these disturbances are replaced by their mathematical expectations *i.e.* by zeros, then the method of solution of the problem is known as 'Non-stochastic simulation'. However, in the Brooking's model these disturbances were added to the corresponding intercept terms in each equation by generating random numbers possessing specified stochastic properties. This enabled a study of the complete sampling distribution of the solutions in each period. Such a procedure is called "Stochastic simulation".

5. Dr. Nagar concluded his talks by describing the procedure of generating random normal disturbances with zero means and specified covariance matrix used by him in obtaining the solutions for Brooking's model.

D.P. Narain: *A genetic simulation study of recessive lethals in finite populations*

6. The use of "Model Sampling" which is often referred to as "Artificial Sampling" or "Monte Carlo Method" dates back to the

time when the concepts of probability were first formulated. In tackling the problems raised by gamblers, mathematicians often simulated gambling as such when exact mathematical formulation and subsequent solution of the problem was not forthcoming. Later on, these simulations were related to the problems of stochastic processes, sequential analysis and the solutions of differential equations. However, it was not until the advent of modern electronic computers that sufficient interest in the technique of simulation was revived, This was due to the fact that quite a large number of variables involved in the problem could be simultaneously handled in a simulation programme on the computer. The role of population size in genetic selection theory is one example where the exact mathematical formulation of the problem is not always possible due to a large number of variables involved. Given the initial conditions, the problem can be simulated on the computer. It is then categorised as a problem in genetic simulation.

7. The genetic simulation can be performed at two levels :

(*i*) Generation of the gene frequency distribution of the selection process on computer at a given population size with the help of transition matrices. The selection on a character is simulated by the allocation of a selective advantage attached to a genotype.

(*ii*) Simulation of the selection experiment on computer involving the formation of gametes from diploid arrays, the formation of new diploid organisms, the prediction of genotypic values, the addition of an environmental component and finally the selection process applied to the array of variables so produced.

8. The former type of genetic simulation involving the use of transition matrices is described in quite general terms in Narain and Robertson[2] (1969). For studying the survival of recessive lethals in finite population, it may be assumed that the population consist of normal homozygotes at the locus as well as, possibly, lethal hetero-zygotes. The status of the population is specified by the number of heterozygotes, A population with no heterozygotes is said to be an "absorbing state" in the sense that a population in that state can only escape by mutation. The other states are known as "transient states". A population originally in such a state may, in the next generation, be found in any other

state including the absorbing state. The selection process is then described by the matrix of transition probabilities between states in a single generation.

9. Dr. Narain then went on to discuss the use of transition matrices in deriving the distribution of lethal gene frequency, the life statistics of individual lethals, the properties of the totality of lethals in the population and the allelism within as well as between generations. The results of generating the frequency distribution of lethal genes on the computer for specific population size and given mutation rates as well as the heterozygote dis-advantage were then described numerically. It was stated that the lethal gene frequencies in equilibrium are comparatively insensitive to selection on the heterozygote. The average as well as the subsequent expectation of life of lethals present in the population at any time were equal at equilibrium. For a population of 50 individuals the average age of lethals which are neutral in the heterozygote was estimated as about 10 gene-rations. This introduced a high correlation between observations in successive generations leading to a great difficulty in obtaining precise estimates of para-meters in any population. As a consequence results from a single population were of little value.

10. Dr. Narain concluded his talk by describing, briefly, the simulation experiment conducted on the computer to study the effect of linkage on the survival of recessive lethals. It was found that for lethals neutral in the heterozygotes, linkage had no detectable effect, at equilibrium, on the proportion of chromosomes containing a lethal or on the allelism between them. Individual lethals appeared to have a slightly higher expectation of life in the absence of crossing over.

R.P. Chakrabarty : Ratio estimators—Monte Carlo study and exact theory*

11. In sample surveys, ratio estimators are often used for estimating the population mean for a character under study or the population ratio utilising an auxiliary character which is positively correlated with the character under study. It is well known that the ratio method increases the precision of estimators in large samples under certain conditions depending on the relation between the

* The Fertiliser Association of India, New Delhi,

two characters and the coefficients of variation of the two characters. However, not much is known about the efficiency of ratio estimator and the stability of its variance estimator under a given model. Rao and Beegle[3] (1967) conducted a Monte Carlo study of the small-sample properties of the variance of the ratio estimator in comparison with the corresponding variance of an unbiased estimator assuming a linear regression with the auxiliary variable as normally distributed.

12. Dr. Chakrabarty then went on to develop the exact theory of this technique with the restriction that the auxiliary variable follows a gamma distribution. He presented extensive numerical results in regard to the stability of the variance estimator obtained under this model in comparison with the variance of the usual unbiased estimator. A comparison of these results with the Monte Carlo results mentioned above was also made. It was shown that the ratio estimator is generally efficient in small samples. The variance estimator of the ratio estimator was shown to be more stable than the variance estimator of the sample mean.

S. Ramani : A brief survey of Digital Simulation using electronic computors*

13. The technique of simulation is characterised by the lack of clear-cut approaches towards designing the model. Simulation involves the construction of a working mathematical or physical model presenting similar properties and relationships with the system under study. But there is no practical restriction on what the system may be. Through the use of mathematical simulation we can test plans, designs or ideas and vary them before the real system is operated. Through the use of high-speed electronic computors, the result of, say, a year's operation under alternative plans, can be obtained in a matter of hours. Simulation is thus a very powerful tool which cannot and should not be under-estimated. The simulation models can be classified into three categories *viz.* (*i*) data generation and data reduction, (*ii*) analog and digital simulation and (*iii*) stochastic or deterministic nature of simulators.

14. In the case of digital simulation, one of the most important problems is to generate a set of random numbers. An important criterion for generating the set is that the sequence of numbers yielded

*Indian Institute of Technology, Madras.

should be reproducible at any later stage. This is because, firstly, when testing out a computor programme the programme needs to be able to use the same series of numbers over and over again so that he can see unambiguously the effects of the corrections he has made. Secondly, when using a model to compare two systems, it is desirable to run the two system under as closely similar conditions as possible. It is, therefore, necessary to devise some deterministic method of generating a series of values that are sufficiently random in appearance which could be used for selecting the sequence with the desired properties. Such a series of pseudo-random numbers are generated by applying mathematical transformation over and over again to a suitably chosen starting value. This transformation is such that each term is produced from the previous term firstly by multiplying by a suitable constant and secondly by removing a certain number of digits from the leading end. The radom numbers are usually normalised so that they all lie in the range (0, 1).

15. Prof. Ramani then went on to describe the "rejection" method used for simulating random numbers on computor from a continuous distribution, and the problems involved in the designing of simulation experiments. In particular, he discussed the problem of determining precision of estimates obtained from a run made up of several successive periods of simulation. Successive periods possess some degree of mutual correlation due to the fact that the starting conditions of each period are the finishing conditions of the previous period. Such correlations can sometimes be eliminated by making use of natural periods in the system being simulated. In the absence of such natural periods, the effect of existing correlations can be allowed for by using a correction factor. The precision of a given sample can be improved by forcing a degree of dependence on values by careful choice of pseudo-random numbers used. This is done by the technique of "antithetic variable", which consists of performing two runs, identical apart from the choice of pseudo-random numbers and averaging the results. The two sets of pseudo-random numbers are chosen in such a way that they have a strong negative correlation coefficient. In some cases it is possible to compare the model with a simplified but similar model which can be solved mathematically. It is then possible to estimate the size of the error due to sampling by simulating the simplified model, using identical sampling numbers and then comparing the result with the known theoretical result. Such a technique is known as using a "control variable". For generating a separate stream of random numbers in each distribution,

it is sufficient to use the same random number generetor by using different starting values. However, care must be taken to ensure that these starting values (sometimes called seeds) come from widely separated parts of the random number sequence, otherwise unwanted correlations may be produced in the model resulting in erroneous results.

16. Prof. Ramani concluded his talk by discussing the role of simulation in system design, the use of computor programming for simulation and the various simulation programming languages.

17. Winding up the symposium, the Chairman Prof. Chakrabarti stressed the role of model sampling experiments using computers or otherwise, in the fields of physical, biological and social sciences. The papers presented in this Symposium demonstrated that intricate problems in economics, genetics and sampling techniques could be solved by simulating the models on the computor. The application of computer technology in this context is thus a very welcome development likely to benefit statisticians engaged in conducting investigations with model sampling.

REFERENCES

1. Nagar, A.L. (1969) : Stochastic Simulations of the Brooking's Econometeric Model. In the *Brooking's Model : Some Further Results* Ed. J. Duesenberry, G. Fromm, L.R. Klein and Ed. Kuh, North Holland Publishing Company, Amsterdam.
2. Narain, P. and A. Robertson (1969) : Limits and duration of response to selection in finite populations : the use of transition probability matrices. Indian J. Heredity, 1 : 79-97.
3. Rao, J.N.K. and L.D. Beegle (1967) : A Monte Carlo study of some ratio estimators. Sankhya Series B, *29* : 47-56.
4. Rao, J.N.K. and J.T. Webster (1966) : On two methods of bias reduction in the estimation of ratios. Biometrika, *53* : 571-77.