INTERNATIONAL
JOURNAL OF
AGRARIAN AFFAIRS
Vol. V, No. 6, March 1973

Contemporary Issues

Produced by the
University of Oxford Institute of Agrarian Affairs
in conjunction with the International Association
of Agricultural Economists

Price 75p net

OXFORD UNIVERSITY PRESS
LONDON

By MARIA DEL CARMEN NIETO-OSTOLAZA[1]

BALANCE BETWEEN STATISTICAL INFORMATION AND MODELS IN AGRICULTURAL ECONOMIC RESEARCH[2]

I. *The new trends in the theory of statistics*

THEORETICAL statistics has gone through three important stages in its historical development: the development of political arithmetic in the seventeenth and eighteenth centuries when demographic data began to be used; the formulation of the laws of frequency distribution and statistical interdependence in the nineteenth century, and the mathematical developments of the first half of the twentieth century.

Looking at that period of time we can see that the development of statistical theory to a certain extent went in the direction of problems or questions with smaller and smaller empirical content; the science had been born because of these empirical problems but was paying more attention to purely mathematical questions.[3]

On the other hand, though with a lag with respect to statistics, the econometric models and econometric planning have begun to be developed. These models and techniques have as their main aim the capacity to predict and the possibility of decision-making at a time when economic activity develops extraordinarily fast with the historical circumstances that today a good number of countries try to develop economically using more or less rigid planning models established beforehand.

These latter tendencies have made themselves felt in economic theory and in the politic-economic affairs of different countries. They have caused students of these problems to turn back to what not long

[1] M. C. Nieto-Ostolaza is coordinador de los Departamentos de Economía y Sociología Agrarias, Instituto Nacional de Investigaciones Agronómicas, Ministerio de Agricultura, Madrid, España.

[2] This article was one of a group on the subject of obtaining statistics suitable for modern types of quantitative analyses applied to agricultural economic problems which the Editor had invited specialists in various fields to contribute to the journal. Unfortunately the authors of the others did not find it possible to complete them. The Editor deeply regrets that this has resulted in a very long delay in publishing this paper.

[3] See M. G. Kendall, 'On the Future of Statistics, a Second Look', *Journal of the Royal Statistical Society*, Series A (General), vol. 131, part 2, 1968.

ago was considered of secondary importance: the problems relating to the need for data, that is to the most 'elementary stages' of statistics.

II. *Information and decision-making*

Economic data are most frequently used in the decisions made by the government and by business firms which generally bring about important consequences for both society and the individual. But decision-making is open to various types of uncertainty. The first of these is the confrontation of the goal, as proposed, with the decision of those who are opposed to it, which at times is Nature herself and at other times an individual or a group of individuals. Next, there is the uncertainty whether all the necessary information has been gathered and its quality correctly evaluated. Finally, doubts may arise as to whether or not the right conclusions have been reached once the information has been integrated into a satisfactory theory.

On the other hand, using the terminology of systems analysis for the problem in question, we might consider a whole agrarian information system as a vast operation forming part of data processes in which such input is transformed into information by means of numerous processors—acting as researchers, for example. This way, the output of a processor will become the input of another. That is to say, data will become information, as the results of a process, when they adopt an adequate form that meets the requirements of the person receiving them. Naturally, there will exist a requirement feed-back similar to what takes place in an automatic control system. In Fig. 1 we have represented the real agrarian system by a rectangle, and within this system a set of functions placed in circles which are joined to each other by the information that they receive and process. In this way, the output of information from one of them will be significant when it is received as information input by another set in order to perform the pertinent function within the system. For example, one of the functions of data collection made by an official statistics bureau consists of reducing the mass of data on the agrarian system to information which can be received and used for other sets of functions of the system, that is, research, advising, and decision-making. As can be seen in Fig. 1, this flow of information supplies the main channel through which observations of the real agrarian system reach those who make the fundamental decisions. In this way, the research has an input of data, mainly statistical, which are transformed into output, that is, models of agrarian transformation, which are in turn input for the

function of advising, by means of proposals, reports, etc. On the other hand, the function of research can also feed back to the function of data collection by supplying the conceptual structures for the development of data accumulations. The feed-back of advising to the researcher would adopt, for example, the form of the requirements in more adequate models, applied theory, concepts, etc. The whole of the function of research might be seen as divided into various intellectual and professional subunits of study with inputs and outputs of information from one to the other and, of course, with inputs and outputs of information of the research function as a whole. Within this real agrarian system the function of agrarian economic research would be a subunit of great importance in the research.

The diagram is simply a reference structure by which we can examine the process of problem-solving and its division of functions. It is obvious that all the important decisions which affect the real agrarian system are not taken as a result of the performance of a data process divided in this way. There will be divisions superimposed over the functional divisions and a great number of mixed information flows. For example, once the models and the theory have been established, a great number of statistical data will flow directly to those who make decisions or to their advisers, by some means different from the research function. But the figure gives us a picture of the logic which serves as basis for the structure by which information affecting the real agrarian system is accumulated in order to be able to plan and administrate.

The diagram divides those decisions which affect agrarian transformation into two types:

(*a*) planning decisions made before acting;
(*b*) operational decisions made in the course of daily administration.

The advising mechanism which is backed by the function of research and by the collection of data, as represented in the design, is principally aimed at planning decisions. A planning decision can set down the policy strategy for the total flow of tactical operative decisions as well as daily routine decisions. These last, appearing as transactions in the files of business firms and the government, form part of the basis of the next cycle of data collection, analysis, synthesis, and decisions.

Within the structure of agrarian political power, the key planning decisions can be made in the public sector, but many of them will be

made autonomously in the private sector, out of the reach of public planners. But in both of these sectors, public and private, when the task of planning becomes complicated and needs an analysis as a basis for a rational judgement, those who make decisions will seek information and advice from individuals and organizations with specialized services. At the same time, the tendency of many countries, with respect to agriculture, is to have these services supplied by private firms or by government bureaux, which should be ultimately at the

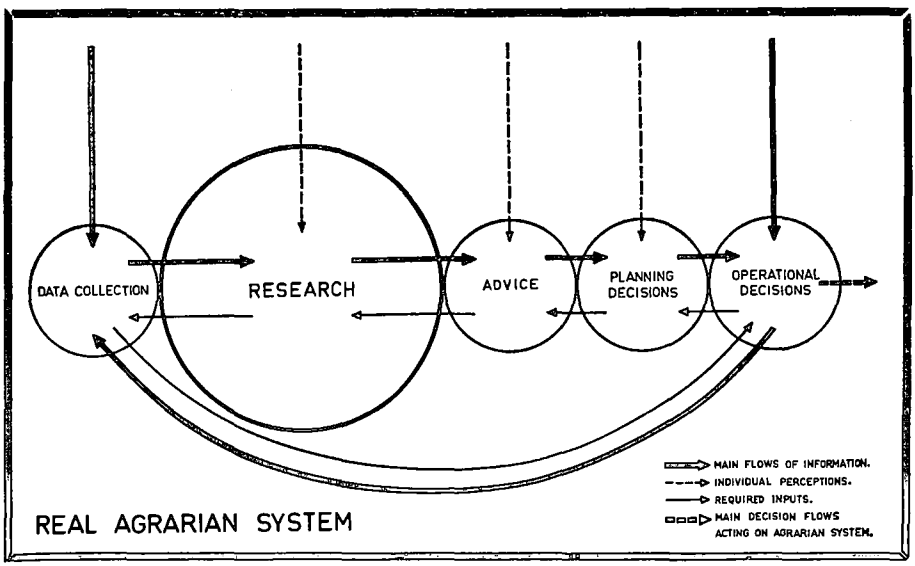

FIG. 1

service of the community. This way, the accumulation and processing of data which are of interest to society will probably be the monopoly of an official statistics bureau, and basic research work will be made principally in universities, while the function of advising tends to be increasingly under the direct or indirect control of a planning bureau belonging to the government.

The whole of the process involved in reaching a planning decision may be described as 'planning', whether the task is divided or not among the one making the decision, the adviser, the researcher, and the one who gathers the data. It might be interesting to note that planning ends at the very moment when a selection among alternatives is made or that it begins after the basic data have been gathered

and a theoretical structure, contrasted with the information, has been established.

The function of accumulating data might be considered as a data bank. The function of research would adopt the form of the elaboration of models, mainly mathematical ones, and their respective contrast. The function of advising would consist in simulating real situations with the models and the information from the data bank, while those who make planning decisions would determine the goals and the constraints which would influence the selection of alternatives. Now that we have shown the importance of the function of research for a rational decision in the agrarian sector, and especially of economic research, since it formulates models and contrasts theories, we will deal with the quality of the information received by the said function of research.

III. *Economic data and economic theory*

The economic data which are available are, in general, subproducts of concrete activities of the state and of business firms and therefore they do not describe those phenomena of interest to economic researchers. Consequently, there is an inadequacy between the information received by economists and the inputs they need. Richard Stone[1] has called this an inadequacy between the supply and demand of information in economic research. This fundamental inadequacy affects all available information and thus reflects even upon non-quantitative information and has become one of the main obstacles for the progress of scientific research.

Oskar Morgenstern[2] has shown this circumstance by means of a diagram (Fig. 2) which we consider of interest for a fuller understanding of the problem, although the author has admitted that it may not meet with general approval.

Circle A represents the whole of statistical information; B represents other data which are not measurable as, for example, historical information; and C represents economic theory. The intersection of circles A and C, of B and C and, if it exists, of A, B, and C is scientific information proper, which is made up of quantitative observation (the

[1] See Richard Stone, *The Role of Measurement in Economics*, Cambridge U. Press, London, 1951, p. 10.

[2] See Oskar Morgenstern, *On the Accuracy of Economic Observations*, second edition. Princeton U. Press, 1963, pp. 88 and following, and Enrique Fuentes Quintana, 'Ciencia Económica y Realidad', *Estadística Española*, abril–junio 1959, n. 3, Instituto Nacional de Estadística, Madrid.

intersection of A and C) and description (intersection of B and C). Data are therefore a much more ample term than that of scientific information. Only the latter information is related with the theory. Data become scientific information, thanks to this connection, otherwise they are nothing more than possible foundations for the construction of a theory. The majority of quantitative economic data are of the A type, excluding C, that is they are represented by the unmarked

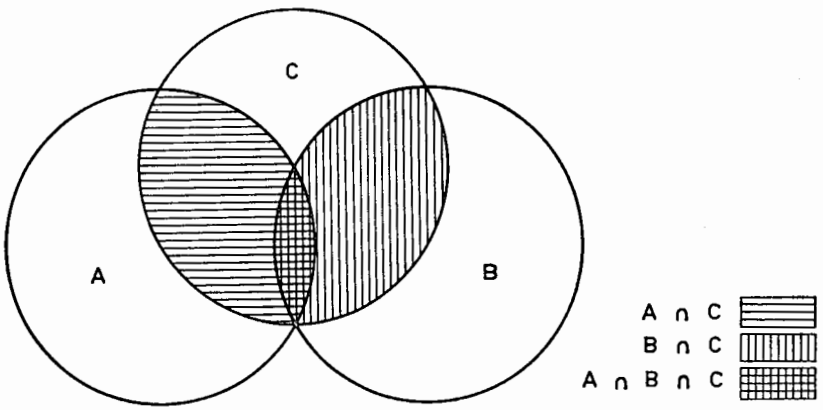

FIG. 2

area of circle A. Although they may clarify our theoretical hypothesis, they do not do it in an automatic way. These data do not describe to us any concrete 'story' or, on the contrary, they may describe many different and conflicting 'stories'. Both assumptions are really equivalent to the lack of an accepted theory.

Quantitative observation and description are planned processes in which the initial stimulus has come from some existing theory although it may only be a rudimentary one. This is true even in the case of planning to produce observations, quantitative or not, aimed at demolishing the initial theory in order to replace it by a better structure. This is the normal process of scientific progress.

In the field of Economics, continuous processes automatically produce an enormous amount of statistical data which are often useful for the previous objective. They are the statistics of production, prices, commerce, etc. At times, even a single phenomenon produces its own quantitative records as a subproduct. Therefore, there are a great many economic statistics which, as such, do not form part of the scientific information. They may reach that level when they are

connected to the theory. These statistics, as well as their counterparts in the field of non-quantitative descriptions, are potential scientific information.

In the Natural Sciences few of the data will accumulate in an automatic way. Almost all of them are the product of a planned action, guided by a theory. A great number of measures are taken without the help of a formal theory, especially in the case of newly developed fields. In general, it is difficult to decide beforehand which are the new measures that a future theory will demand. But the theory will always come last. Even when data are not strictly related to the theory, they are produced systematically by experimental scientists. Here lies an ulterior problem of the Social Sciences, including Economics, as opposed to the Experimental Sciences.

IV. *Possible causes of error in economic statistics*

Among the principal causes of errors very often included in economic statistics, the following have been pointed out, among others:[1]

1. In the first place, the absence of an economic design which, in any process of measurement, has to come before the elaboration of a statistical design in order to avoid the usage of concepts with dubious frontiers and to make international comparisons possible. On the other hand, if an economic design is made before the statistical one, it might mean a considerable saving to the cost of research by making a detailed account of the information and precision requirements in the computation that is needed for the use of the corresponding statistics. But economic statistics in general are not, as we have said before, the results of designed experiments but rather, subproducts or the result of the activities of the government and business firms. Therefore, they often measure something that is not part of the phenomenon which the economist is interested in.
2. The falsification of information, on the part of the elements observed and in some cases on the part of the observers themselves, who might try to present results that will be favourable to them in order to obtain certain objectives in non-statistical activities.

[1] O. Morgenstern, op. cit., and Angel Alcaide Inchausti, *Lecciones de Econometría y Métodos Estadísticos*. Ed. Copigraf, Madrid, 1966.

As to the first point, Professor Alcaide considers it the most serious handicap for the elaboration of economic statistics, specially when it is a question of finding out, directly or indirectly, the incomes or profits of families and firms, since it is believed that such information is meant for fiscal purposes. For this reason an attempt is being made to elaborate a theory of the sampling in a hostile environment which could be applied to innumerable relations of a social order.

With respect to the second point, Oskar Morgenstern states as an example that when the Marshall Plan was established, one of its principal exponents confided to him: 'We will produce statistics that will help us to obtain the greatest possible amount of money from the United States. We will simply make up those statistics which we lack but which are necessary for us to justify our demands.' These statistics, continues Morgenstern, went down in history as an authentic description of the economic conditions of that time and they may even be used in econometric models.

3. In general, statistical data are not compiled by the persons who are going to use them, and this might produce an elaboration of routine or intuitive statistics which is not what the economic researcher needs.
4. The difficulty of answering inquiries due to their length or to their faulty wording. Because of this, the questions should be stated in a way that will not give rise to ambiguous answers and they should be as discrete as possible.
5. The quality of the observer's preparation and the possibility of withholding part of the information, two causes that have already been mentioned, are of great importance when the number of activities that must be analysed in order to obtain quantitative data can be counted in thousands or millions. This way, errors that have originated due to the above-mentioned causes may accumulate.
6. The lack of a definition or a classification of the phenomenon which is being measured is another important cause of errors, although efforts are being made in many fields to remedy this.
7. Errors due to the use of calculating machines and statistical machines, such as perforators, classifiers and tabulators, and even typing errors made when publishing entire pages of statistical data.

8. The reference of statistical data to certain periods or intervals of time may be the cause of errors—due to a difficult periodicity of statistics. This happens, for example, in the case of agricultural production referred to natural years or to variations, during this period of time, in quality, customs, stock volume, etc.
9. Finally, there is the possibility of there being different centres of observation for identical economic fact. This happens, for example, in Spain, when it registers foreign commerce both through Customs and the Bureau of Foreign Currency.

V. *The quality of economic data and its improvement*

As we have said, the statistics necessary for the construction of economic models do not form a separate category. Almost all social and demographic statistics are or can be useful, in a direct or indirect way, for the establishment of models and the contrast of theories. But what is the quality of economic statistics?

In 1963 Oskar Morgenstern[1] presented a very complete empirical study in which he analysed the errors of various economic statistics—foreign commerce, prices, employment level, national income, rates of growth, and also agricultural statistics—as a sample of what should be done systematically in the future in order to have a more realistic idea of the quality of economic data in diverse fields. The study turned out to be rather pessimistic as to the average quality of the data available at that moment. Nevertheless it did offer various positive thoughts.

In the first place it pointed out a definite task, although it was foreseen as a slow process. It consisted of discouraging the government and important private agencies from presenting to the public statistics as if they were absolutely true. For example, the statements concerning monthly changes in the rate of growth of a nation were absurd, as were also those concerning variations over small intervals in the level of prices or in the cost of living. A more important step was to insist that economic statistics be published only with an estimate of their error, even if it is approximate.

It also stated the urgent need of forming special commissions to investigate particular fields of economic data and inform as to their quality. With regard to this, self-criticism, as far as data are concerned was a rather touchy affair. How could funds be obtained to collect statistics when the quality that was obtained was questioned? It was

[1] O. Morgenstern, op. cit.

difficult to maintain that statistics bearing a known error were not necessarily useless or made by incompetent persons. And when various agencies were involved in the elaboration of the same statistics as, for example, the Secretaría General Técnica del Ministerio de Agricultura and the Servicio Nacional de Cereales in Spain, the question of an error became a competitive affair between different agencies. Now, how could statistics be improved if they seemed correct at first sight? To pose the problem of quality in economic data in a realistic way was evidently to be half-way towards solving it. It became urgent to promote research work which would refine data and would fix a margin of quality, with an estimate of error.

Morgenstern's exposition in 1963 is still valid today. The process of improving data has no limits. Those who establish economic models are in ever-increasing need of an improved quality in the information flows they receive from statistical agencies. And just as it is normal in Physics to accompany data with information errors carefully determined, while efforts are constantly made to reduce errors, it is necessary in Economics to be aware of the true condition of the data so that the models may have a real significance.

Consequently, two useful measures to be taken for the improvement of statistical information are:

the publication of statistics with an estimate of their error even if it is approximate, and the creation of commissions to investigate the errors in the various economic data.

It must be pointed out that deficiency in the processes involved in gathering data is less intense when sample methods are used to obtain economic statistics. Even though a sample might be bad and give rise to objections, its construction is subject to scientific scrutiny, and the problems that have to be solved when establishing a good sample are well known. Solutions are a function of the state of the sample theory and of its application to a given case. For this reason, sample statistics in Economics are becoming increasingly important although the possibilities have not been used to their full extent. Therefore, another useful measure is:

a greater use of sample procedures to obtain economic information.

Furthermore, sample statistics may be used as means of comparison. Indeed, checks on the quality of statistical data have generally been divided into two great categories.[1]

[1] See S. S. Zarkovich, *La calidad de los datos estadísticos*, FAO, Rome, 1968.

Within the first, various types of *a posteriori* techniques have been included, and in the other the use of sample methods.

The *a posteriori* techniques are used after the inquiry has been completed and the data have been tabulated and published. These techniques can therefore be easily applied to data belonging to very old inquiries. The *a posteriori* techniques do not call for any sort of field work related to the units included in the inquiry. On the other hand, checks on quality, made with the aid of sample methods, are essentially field inquiries done at a time that is not too distant from the reference period of the main inquiry. This way, checking by sampling will supply a sample of specific units with recent information concerning the characteristics that are being verified. This information is used to measure the quality of the main inquiry.

Among the *a posteriori* techniques used in assessing the quality of data are:

1. Comparisons with data that have come from independent sources.
2. Studies on consistency, seeing whether the data conform to a general accepted knowledge concerning the characteristics involved.
3. Studies on internal consistency, investigating to what degree the estimates of the various characteristics describe the same phenomena in the same way or degree in which they may be taken to be logically related to each other.
4. Finally, the method of following a group, which consists of the following: if a group of persons, as enumerated in a census, is followed after the census has been taken in order to establish all the changes that occur later, such as deaths or migrations, it will be possible to know the number of persons in the group at the date of the next census and to compare it with the size of the group as it appears in the new census.

In general, the conclusions reached by sample corroborations are based on the difference between the original data and those obtained in the check inquiry. This way, it becomes possible to reach numerical estimates, which are of greater interest than the descriptive conclusions that are the result of *a posteriori* techniques.

In both cases, when the check is completed, a report should be made of the results obtained, specifying the quality of the economic data.

The problem of the quality of statistical data in general has recently attracted the attention of the FAO,[1] which is trying to promote interest in the corroboration of quality as a means for finding the right course to be taken in the usage of data and in the way to improve the methods used to accumulate them.

VI. *Final considerations*

Agrarian economic research has been enriched in the last few years with new models: linear programming, quadratic programming, the theory of games, etc.[2] But although the models may be made simply by thinking, without having new and better data, their validity and application are subject to their confrontation with facts, quantitative or not, which gives us an idea of the degree of realism and utility in the models. Consequently, those who formulate and apply models need new and better data, a requirement which is easy to express but very difficult to obtain. They need at least to know the quality of the data they are using or to remedy the absence of diverse data, as in the case, for example, of various temporal series, by the usage of cross-section data.[3]

In the attempt to find a solution to the problem of statistical information Bayesian statistics has also been applied. Researchers have of late paid close attention to this as a way of avoiding the difficulties with which they are faced in inference, in econometric models.[4]

Those who create models should clearly state their different needs as far as data are concerned, although it is necessary to be aware of the problems involved in asking for a collection of data to compare alternative theories or elaborate models which are still considered as not operative.

In the future, the need of a close co-operation between those who make and use models and those who collect and prepare data seems likely. It is a difficult co-operation as far as planning is concerned, and it will have to be reached gradually by a better understanding of the mutual interests of both groups.

[1] See S. S. Zarkovich, op. cit.

[2] M. C. Nieto-Ostolaza, *A Bibliographical Study on Mathematical Programming and Theory of Games in Agriculture*, Anales del Instituto Nacional de Investigaciones Agronómicas, vol. xvi, Madrid, 1967.

[3] See Carl F. Christ, *Econometric Models and Methods*. John Wiley and Sons, Inc., 1966, pp. 102–9.

[4] A. Zellner, *Bayesian inference and Simultaneous Equation Econometric Models*, presented at the First World Congress of Econometry, held in Rome, Sept. 1965.

Still, the adoption of planning decisions in agriculture, at entrepreneurship level and in the public sector, in a scientific way within the actual state of our knowledge, cannot be made without the necessary balance between statistical information gathered from the real agrarian system and the models of research, since their use is conditioned by said information.