Chapter XVI

THE ACCURACY OF ECONOMIC OBSERVATIONS [1]

By Oskar Morgenstern

Applications of theories such as those of linear programming use data that are subject to various errors. The significance of the far-reaching computations necessitated by these theories will depend on the knowledge of the errors. It is, therefore, necessary to form as precise ideas as possible about the accuracy of economic observations. So far there have been no tangible results regarding the quantitative estimation of errors in economic statistics, although attempts to improve the statistics in a general sense are made continuously wherever they are collected. But the accuracy actually required depends on the purpose of the statistics. Rough estimates may be all that are needed for one purpose; accuracy down to one-tenth of one per cent may be far too coarse for another. Accuracy of a given statistic can, therefore, not be uniformly good or bad. It can be judged only from the point of view of the theory which interprets the statistics and directs further logical and mathematical operations. Linear programming requires enormously large numbers of operations, and, at least for that reason, a high degree of accuracy is required. As a rule, however, present economic theory is not of a very fine-grain structure—a condition that may gradually be overcome. Furthermore, it is doubtful that there are many fields in economics with a theory powerful enough to make use of more than three or four significant digits. Published statistics, however, often seem to indicate that many more digits would be available. In addition this being frequently questionable, there would hardly be any theory available now to cope with such fine measurements.

It is noteworthy that little is known, except in an over-all way, about the extent of the errors in economic statistics. In the natural sciences a long tradition exists, and the study of errors has occupied a very

[1] A memorandum of this title was presented at the Conference on Linear Programming. In view of its length a separate publication has been undertaken [Morgenstern, 1950]. The following abstract serves only to indicate some of the main points that are discussed *in extenso* in that monograph.

This research is part of a project carried out under contract (N6 ONR27009) with the Office of Naval Research (Project NR 042086).

prominent place. Otherwise their progress would have been unthinkable. The difficulties of estimating errors of the data for the social sciences, however, cannot possibly be less than for the natural sciences. In fact, the sources of error are more numerous and the statistical problem is far more serious in the social sciences. Consequently the treatment of errors of observation has to be at least as rigorous as in the natural sciences. A factor not present in the latter field is unfortunately of great importance in social affairs: the deliberate lie, and the hiding and suppression of information. Statistical theory will have to evolve methods accounting for such possibilities which at present are ignored.

Space precludes a full enumeration of even the main sources of errors. As long as no quantitative measure is available, qualitative description is important. In economic statistics one of the most troublesome errors arises from the inevitable use of questionnaires. Furthermore, conflicts with the interests of private business arise when information is demanded. Sales prices and the volume of transactions are often closely guarded secrets, so that statements about these are often worthless. This is particularly true if the industry is highly cartelized or a monopoly.

Illustrations from various fields, such as foreign trade, employment, prices, indicate that the errors often are very large, even though they can be ascertained only in a rough manner. Variations in national income and especially in its composition are known only with a high degree of uncertainty. Figures such as these, however, enter significantly into input-output tables. Thus the uses to which the tables can be put are limited on two accounts: data and extent of numerical operations. The economic models that can be set up on the basis of information of this type (either with large known errors or with errors only imperfectly described) are naturally limited in scope and value.

Linear programming, or any other similar utilization of great masses of economic data, cannot be expected to make decisive practical progress until there is satisfaction that the data warrant the implied extensive and costly numerical operations. Therefore current and future collections of data suitable for linear programming, whether for economic or logistic purposes, should give particular attention to the numerical determination of the accuracy of the data. This work, furthermore, must be guided by the fact that large input-output tables are aimed at, and that their use, whether aggregations take place or not, will require, at any rate, many millions of numerical operations. These would lose all meaning unless performed with data of a standard of reliability that corresponds to the intricacy of the computations. Matrix inversions, for example, are performed on matrices in which the entries in each field are subject to errors (as must be the case). These errors often differ

widely from one field to another and in many instances are even unknown. They pose serious problems in addition to those of the inversion of large matrices themselves. Linear programming requires numerical operations of this kind.

In addition to observations made in statistical form there are economic events and phenomena that do not (as yet) lend themselves to statistical, numerical representation. An example is offered by expectations, where whatever information becomes available is also affected by error components. Variations in these data often have a direct bearing on the accuracy of statistical information which also should be taken into account. Another type of difficulty lies with those, possibly highly "accurate," data that lack functional meaning, such as official exchange rates of a country with exchange control. These would falsify, for instance, its foreign trade statistics and make them useless for input-output tables.

Economic measurements are peculiar in that they are most frequently made of *unique* phenomena. Sometimes the same event is observed simultaneously by different observers who are, however, seldom scientific observers. The great sharpening of measurements in the natural sciences is due primarily to the fact that the same event, say the velocity of light, has been measured time and again. But the transactions between two industries in a given year are ascertained only once by a single agency on the basis of questionnaires, with few internal checks that, if they exist at all, rest on the same type of data. It is clear that statistical theory has great tasks to accomplish in order to guide economists to the establishment of information suitable for such vast and important undertakings as linear programming. Applications to military data suffer perhaps less from these sources of errors, but this is probably compensated by the large number of different activities that ought to be recognized [2] for logistic purposes.

In summary, it is clear that the development of theories of linear programming and the establishment of more adequate economic models cannot progress very far without a thorough exploration of the nature of the observations at our disposal.

[2] When the number of activities is too large they have to be condensed to manageable proportions. This *aggregation* is itself a source of error and as yet is little understood.