PREFACE

This study is intended as a contribution to econometrics. It represents an attempt to supply a theoretical foundation for the analysis of interrelations between economic variables. It is based upon modern theory of probability and statistical inference. A few words may be said to justify such a study.

The method of econometric research aims, essentially, at a conjunction of economic theory and actual measurements, using the theory and technique of statistical inference as a bridge pier. But the bridge itself was never completely built. So far, the common procedure has been, first to construct an economic theory involving *exact* functional relationships, then to compare this theory with some actual measurements, and, finally, "to judge" whether the correspondence is "good" or "bad." Tools of statistical inference have been introduced, in some degree, to support such judgments, e.g., the calculation of a few standard errors and multiple-correlation coefficients. The application of such simple "statistics" has been considered legitimate, while, at the same time, the adoption of definite probability models has been deemed a crime in economic research, a violation of the very nature of economic data. That is to say, it has been considered legitimate to use some of the *tools* developed in statistical theory *without* accepting the very *foundation* upon which statistical theory is built. For *no tool developed in the theory of statistics has any meaning*—except, perhaps, for descriptive purposes—*without being referred to some stochastic scheme.*

The reluctance among economists to accept probability models as a basis for economic research has, it seems, been founded upon a very narrow concept of probability and random variables. Probability schemes, it is held, apply only to such phenomena as lottery drawings, or, at best, to those series of observations where each observation may be considered as an independent drawing from one and the same "population." From this point of view it has been argued, e.g., that most economic time series do not conform well to any probability model, "because the successive observations are not independent." But it is *not* necessary that the observations should be independent and that they should all follow the same one-dimensional probability law. It is sufficient to assume that the *whole set* of, say n, observations may be considered as *one* observation of n variables (or a "sample point") following an n-dimensional *joint* probability law, the "existence" of which may be purely hypothetical. Then, one can test hypotheses regarding this joint probability law, and draw inference as to its possible form, by means of *one* sample point (in n dimensions). Modern statis-

tical theory has made considerable progress in solving such problems of statistical inference.

In fact, if we consider actual economic research—even that carried on by people who oppose the use of probability schemes—we find that it rests, ultimately, upon some, perhaps very vague, notion of probability and random variables. For whenever we apply a theory to facts we do not—and we do not expect to—obtain exact agreement. Certain discrepancies are classified as "admissible," others as "practically impossible" under the assumptions of the theory. And the *principle* of such classification is itself a theoretical scheme, namely one in which the vague expressions "practically impossible" or "almost certain" are replaced by "the probability is near to zero," or "the probability is near to one."

This is nothing but a convenient way of expressing opinions about real phenomena. But the probability concept has the advantage that it is "analytic," we can derive new statements from it by the rules of logic. Thus, starting from a purely formal probability model involving certain probabilities which themselves may not have any counterparts in real life, we may derive such statements as "The probability of A is almost equal to 1." Substituting some real phenomenon for A, and transforming the statement "a probability near to 1" into "we are almost sure that A will occur," we have a statement about a real phenomenon, the truth of which can be tested.

The class of scientific statements that can be expressed in probability terms is enormous. In fact, this class contains all the "laws" that have, so far, been formulated. For such "laws" say no more and no less than this: The probability is almost 1 that a certain event will occur.

Thus, there appears to be a two-fold justification for our attempt to give a more rigorous, probabilistic, formulation of the problems of economic research: First, if we want to apply statistical inference to testing the hypotheses of economic theory, it *implies* such a formulation of economic theories that they represent *statistical* hypotheses, i.e., statements—perhaps very broad ones—regarding certain probability distributions. The belief that we can make use of statistical inference without this link can only be based upon lack of precision in formulating the problems. Second, as we have indicated above, there is no loss of generality in choosing such an approach. We hope to demonstrate that it is also convenient and fruitful.

The general principles of statistical inference introduced in this study are based on the Neyman-Pearson theory of testing statistical hypotheses.

Chapter I contains a general discussion of the connection between abstract models and economic reality.

Chapter II deals with the question of establishing "constant relationships" in the field of economics, and with the degree of invariance of economic relations with respect to certain changes in structure.

In Chapter III we discuss the nature of stochastical models and their applicability to economic data.

In Chapter IV it is shown that a hypothetical system of economic relations may be transferred into a statement about the *joint probability law* of the economic variables involved, and that, therefore, such a system can be regarded as a statistical hypothesis in the Neyman-Pearson sense. A brief exposé of the Neyman-Pearson theory of testing statistical hypotheses and estimation is given at the beginning of this chapter.

Chapter V deals, essentially, with the following problem of estimation: Given a system of stochastical equations, involving a certain number of parameters, such that the system is actually satisfied by economic data when a certain set of values of the parameters is chosen, is then the system also satisfied for *other* values of the parameters? If that be the case, no unique estimate of the parameters can be obtained from the data. (This is, in the case of linear relations, the now well-known problem of multicollinearity.) Mathematical rules for investigating such situations are given.

Chapter VI contains a short outline of the problems of predictions. Some examples are presented to illustrate essential points.

* * *

The idea of undertaking this study developed during my work as an assistant to Professor Ragnar Frisch at the Oslo Institute of Economics. The reader will recognize many of Frisch's ideas in the following, and indirectly his influence can be traced in the formulation of problems and the methods of analysis adopted. I am forever grateful for his guiding influence and constant encouragement, for his patient teaching, and for his interest in my work.

The analysis, as presented here, was worked out in detail during a period of study in the United States, and was first issued in mimeographed form at Harvard in 1941. My most sincere thanks are due to Professor Abraham Wald of Columbia University for numerous suggestions and for help on many points in preparing the manuscript. Upon his unique knowledge of modern statistical theory and mathematics in general I have drawn very heavily. Many of the statistical sections in this study have been formulated, and others have been reformulated, after discussions with him. The reader will find it particularly useful in

connection with the present analysis to study a recent article by Professor Wald and Dr. H. B. Mann, "On the Statistical Treatment of Linear Stochastic Difference Equations," in ECONOMETRICA, Vol. 11, July-October, 1943, pp. 173–220. In that article will be found a more explicit statistical treatment of problems that in the present study have only been mentioned or dealt with in general terms.

I should also like to acknowledge my indebtedness to Professor Jacob Marschak, research director of the Cowles Commission, for many stimulating conversations on the subject. I wish further to express my gratitude to Professors Joseph A. Schumpeter and Edwin B. Wilson of Harvard University for reading parts of the original manuscript, and for criticisms which have been utilized in the present formulation. Likewise, I am indebted to Mr. Leonid Hurwicz of the Cowles Commission and to Miss Edith Elbogen of the National Bureau of Economic Research for reading the manuscript and for valuable comments.

Of course, the author alone should be blamed for any mistake or incompleteness.

TRYGVE HAAVELMO

New York, June, 1944

CHAPTER I

Abstract Models and Reality

1. Introduction

Theoretical models are necessary tools in our attempts to understand and "explain" events in real life. In fact, even a simple description and classification of real phenomena would probably not be possible or feasible without viewing reality through the framework of some scheme conceived a priori.

Within such theoretical models we draw conclusions of the type, "if A is true, then B is true." Also, we may decide whether a particular statement or a link in the theory is *right* or *wrong*, i.e., whether it does or does not violate the requirements as to inner consistency of our model. As long as we remain in the world of abstractions and simplifications there is no limit to what we might choose to prove or to disprove; or, as Pareto has said, "Il n'y a pas de proposition qu'on ne puisse certifier vraie sous certaines conditions, à déterminer."[1] Our guard against futile speculations is the requirement that the results of our theoretical considerations are, ultimately, to be compared with some real phenomena. This, of course, does not mean that every theoretical result, e.g., those of pure mathematics, must have an immediate practical application. A good deal of the work in pure theory consists in deriving rigorous statements which may not always have a direct bearing upon facts. They may, however, help to consolidate and expand the techniques and tools of analysis and, thus, increase our power of attacking problems of reality.

When statements derived from a theoretical model are transferred to facts, the question of "right" or "wrong" becomes more ambiguous. The facts will usually disagree, in some respects, with any *accurate* a priori statement we derive from a theoretical model. In other words, such exact models are simply false in relation to the facts considered. Can we have any use for models that imply false statements? It is common to answer this question by observing that, since abstract models never correspond exactly to facts, we have to be satisfied when the discrepancies are not "too large," when there is "a fairly good correspondence," etc. But on second thought we shall realize that such a point of view is not tenable. For we must then, evidently, have a rule for deciding in advance *when* we shall say that our a priori statements are right or wrong. That is, such rules will have to be part of our

[1] *Manuel d'économie politique*, 2nd ed., p. 9.

models. Our models, thus expanded, then lead to somewhat broader statements which, when applied to facts, will be either true or false.

Still, whatever be the theory, it cannot remain true in regard to a certain set of facts if it ever *implies* a false statement about those facts. We shall then find that it is practically impossible to maintain any theory that *implies* a nontrivial statement about certain facts, because sooner or later the facts will, usually, contradict any such statement. Therefore, we shall not only have to be satisfied with broader statements than the ones usually implied by an "exact" model, but we shall also have to adopt a particular kind of model, namely such models as permit statements that are not *implications*, but merely have a certain chance of being true. This will lead us to a probabilistic formulation of theories that are meant to be applied.

Expressions like "the theory is almost true" simply do not make sense unless specified in some such manner as we have indicated. Therefore, when we say that an "exact" theory is "almost true" it seems that we must mean that this theory, although wrong as its stands, in practice can replace *another* model which, first, would lead us to somewhat broader statements and, second, would permit even these broader statements to be wrong "on rare occasions."

Thus, the question of whether or not an exact theoretical model is "almost true" is really the same question as whether or not some other model that claims less is actually true in relation to the facts, or at least does not contradict the facts. It is with models of the latter type that we have to concern ourselves when we want to engage in testing theories against facts. As already mentioned, we shall see that this leads us to adopting a probabilistic formulation of theories to be applied.

These remarks apply, more or less, to all types of economic theory, whether quantitatively formulated or not. But we shall not follow up the study of theory versus facts in this broad sense. In all that follows we shall be concerned with a particular, but very important, class of economic theories, namely those where the theoretical model consists of a system of (ordinary or functional) equations between certain economic variables. A few remarks may be made as to the common sense of this type of economic theory.

Broadly speaking, we may classify such quantitative economic relations in the three groups:

I. Definitional identities,
II. Technical relations,
III. Relations describing economic action.

The first group is exemplified by such relations as: Total expenditure = price multiplied by quantity bought, total output = output per worker times the number of workers, and similar types of "bookkeeping iden-

tities." To the second group belong, e.g., technical production functions, and other natural or institutional restrictions which are usually taken as data in economic planning. In the third group we find the broad class of relations describing the behavior of individuals or collective units in their economic activity, their decisions to produce and consume.

In such relations two sorts of quantities occur, viz., the *variables* under investigation, and the *parameters* introduced in the process of analysis. (The terms "variables" and "parameters" are relative to the particular problem in question, they cannot be defined in any absolute sense.) In relations of type I the parameters, if any, are given by *definition,* while in relations of type II or III the parameters are *at our disposal* for the purpose of adapting such hypothetical relations to a set of economic variables. From the point of view of economic theory this distinction applies in particular to relations of type III; it applies perhaps less to those of type II, inasmuch as the choice of form and of parameters in technical relations may be regarded as the task of other sciences.

Let us consider in particular the relations of type III. Certainly we know that decisions to consume, to invest, etc., depend on a great number of factors, many of which cannot be expressed in quantitative terms. What is then the point of trying to associate such behavior with only a limited set of measurable phenomena, which cannot give more than an incomplete picture of the whole "environment" or "atmosphere" in which the economic planning and decisions take place? First of all, we should notice that "explanations" of this kind are only attempted for such phenomena as themselves are of a quantitative nature, such as prices, values, and physical volume. And when economic decisions are of the type "more" or "less," "greater" or "smaller," they must have *consequences* for some *other* measurable phenomena. Thus, if a man starts to spend more of his (fixed) income on a certain commodity, he must spend less on other things. If a manufacturer wants to increase his production, he must buy more means of production. If his profit increases, this must have measurable consequences for his spending-saving policy; and so forth. It would certainly be very artificial to assume that these quantities themselves do not influence the decisions taken, and that there should be no system in such influences. It is, then, only a natural step to attempt an approximate description of such influences by means of certain behavioristic parameters.

At least this is one type of "explanation." Other types may be chosen. But whatever be the "explanations" we prefer, it is not to be forgotten that they are all our own artificial inventions in a search for an understanding of real life; they are not hidden truths to be "discovered."

2. "Exact Quantitative Definitions of the Economic Variables"

This phrase has become something like a slogan among modern economists, but there sometimes appears to be some confusion as to what it actually means. The simple and rational interpretation would seem to be that, since the most important facts we want to study in real economic life present themselves in the form of numerical measurements, we shall have to choose our models from that field of logic which deals with numbers, i.e., from the field of mathematics. But the concepts of mathematics obtain their quantitative meaning implicitly through the system of logical operations we impose. In pure mathematics there really is no such problem as quantitative definition of a concept *per se*, without reference to certain operations.

Therefore, when economists talk about the problem of quantitative definitions of economic variables, they must have something in mind which has to do with real economic phenomena. More precisely, they want to "give exact rules how to measure certain phenomena of real life," they want to "know exactly what elements of real life correspond to those of theory." When considering a theoretical set-up, involving certain variables and certain mathematical relations, it is common to ask about the actual *meaning* of this and that variable. But this question has no sense within a theoretical model. And if the question applies to reality it has no precise answer. The answer we might give consists, at best, of a tentative description involving words which we have learned to associate, more or less vaguely, with certain real phenomena,

Therefore, it is one thing to build a theoretical model, it is another thing to give rules for choosing the facts to which the theoretical model is to be applied. It is one thing to choose the theoretical model from the field of mathematics, it is another thing to classify and measure objects of real life. For the latter we shall always need some willingness among our fellow research workers to agree "for practical purposes" on questions of definitions. It is never possible—strictly speaking—to avoid ambiguities in classifications and measurements of real phenomena. Not only is our technique of physical measurement unprecise, but in most cases we are not even able to give an unambiguous description of the *method* of measurement to be used, nor are we able to give precise rules for the choice of *things to be measured* in connection with a certain theory. Take, for instance, the apparently simple question of measuring the total consumption of a commodity in a country during a given period of time. Difficulties immediately arise from the fact that the notions of a "commodity," "consumption," etc., are not precise terms; there may be dispute concerning their content or quantitative measure. And this applies to all quantities that represent practical measurements of real objects.

3. "Observational," "True," and Theoretical Variables; an Important Distinction

Even though our actual knowledge of economic facts is based on rough classifications and approximate measurements, we feel that we often "could do better than this," that, in many cases, it *would* be possible to give descriptions and rules of measurement in such a way that two or more independent observers applying these rules to a described group of objects would obtain practically the same quantities. Often, when we operate with such notions as national income, output of certain commodities, imports, exports, etc., we feel that these things have a definite quantitative meaning and could possibly be measured rather accurately, but—for financial reasons or lack of time—we are not able to carry out the counting and measurement in the way we should really like to do it. And we also usually feel that these problems of measurements are somewhat different from those of searching for "explanations." When we speak of certain known facts to be "explained" we think, in many cases, of some more correct or controlled measurements of facts than those that happen to be given by current economic statistics. From experience in various fields we have acquired empirical knowledge as to sources of errors and the degree of precision connected with current types of statistical observation technique. At least as the situation is at present in the field of economic statistics, we almost always know that we could do better, if we could only find the necessary time and money. When we speak of the "true" values of certain observable phenomena, as compared with some approximate statistical information, the distinction we have in mind is probably something like the one we have described above in somewhat vague terms.

In pure theory we introduce variables (or time functions) which, by construction, satisfy certain conditions of inner consistency of a theoretical model. These theoretical variables are usually given names that indicate with what actual, "true," measurements we hope the theoretical variables might be identified. But the theoretical variables are not *defined* as identical with some "true" variables. For the process of correct measurement is, essentially, applied to each variable *separately*. To impose some functional relationship upon the variables means going much further. We may express the difference by saying that the "true" variables (or time functions) represent our ideal as to accurate measurements of reality "as it is in fact," while the variables defined in a theory are the true measurements that we should make if reality were actually in accordance with our theoretical model.

The distinction between these three types of variables, although somewhat vague, is one of great importance for the understanding of

the connection between pure theory and its applications. Let us try to explain the matter in a different way that is, perhaps, clearer.

One of the most characteristic features of modern economic theory is the extensive use of symbols, formulae, equations, and other mathematical notions. Modern articles and books on economics are "full of mathematics." Many economists consider "mathematical economics" as a separate branch of economics. The question suggests itself as to what the difference is between "mathematical economics" and "mathematics." Does a system of equations, say, become less mathematical and more economic in character just by calling x "consumption," y "price," etc.? There are certainly many examples of studies to be found that do not go very much further than this, as far as economic significance is concerned. But they hardly deserve the ranking of contributions to economics. What makes a piece of mathematical economics not only mathematics but also economics is, I believe, this: When we set up a system of theoretical relationships and use economic names for the otherwise purely theoretical variables involved, we have in mind some actual *experiment*, or some *design of an experiment*, which we could at least imagine arranging, in order to measure those quantities in real economic life that we think might obey the laws imposed on their theoretical namesakes. For example, in the theory of choice we introduce the notion of indifference surfaces, to show how an individual, at given prices, would distribute his fixed income over the various commodities. This sounds like "economics" but is actually only a formal mathematical scheme, until we add *a design of experiments* that would indicate, first, what real phenomena are to be identified with the theoretical prices, quantities, and income; second, what is to be meant by an "individual"; and, third, how we should arrange to observe the individual actually making his choice.

There are many indications that economists nearly always have some such design of ideal experiments in the back of their minds when they build their theoretical models. For instance, there is hardly an economist who feels really happy about identifying current series of "national income," "consumption," etc., with the variables by these names in his theories. Or, conversely, he would often find it too complicated or perhaps even uninteresting to try to build models such that the observations he would like to identify with the corresponding theoretical variables would correspond to those actually given by current economic statistics. In the verbal description of his model, "in economic terms," the economist usually suggests, explicitly or implicitly, some type of experiments or controlled measurements designed to obtain the real variables for which he thinks that his model would hold. That is, he

has in mind some "true" variables that he would like to measure. The data he actually obtains are, first of all, nearly always blurred by some plain errors of measurements, that is, by certain extra "facts" which he did not intend to "explain" by his theory. Secondly, and that is still more important, the economist is usually a rather passive observer with respect to important economic phenomena; he usually does not control the actual collection of economic statistics. He is not in a position to enforce the prescriptions of his own designs of ideal experiments.

One could perhaps also characterize the difference between the "true" and the "observational" variables in the following way. The "true" variables are variables such that, if their behavior should contradict a theory, the theory would be rejected as false; while "observational" variables, when contradicting the theory, leave the possibility that we might be trying out the theory on facts for which the theory was not meant to hold, the confusion being caused by the use of the same names for quantities that are actually different.

In order to test a theory against facts, or to use it for predictions, either the statistical observations available have to be "corrected," or the theory itself has to be adjusted, so as to make the facts we consider the "true" variables relevant to the theory, as described above. To use a mechanical illustration, suppose we should like to verify the law of falling bodies (in vacuum), and suppose our measurements for that purpose consisted of a series of observations of a stone (say) dropped through the air from various levels above the ground. To use such data we should at least have to calculate the extra effect of the air resistance and extract this element from the data. Or, what amounts to the same, we should have to expand the simple theory of bodies falling in vacuum, to allow for the air resistance (and probably many other factors). A physicist would dismiss these measurements as absurd for such a purpose because he can easily do much better. The economist, on the other hand, often has to be satisfied with rough and biased measurements. It is often his task to dig out the measurements he needs from data that were collected for some other purpose; or, he is presented with some results which, so to speak, Nature has produced in all their complexity, his task being to build models that explain what has been observed.

The practical conclusion of the discussion above is advice that economists hardly ever fail to give, but that few actually follow, viz., that one should study very carefully the actual series considered and the conditions under which they were produced, before identifying them with the variables of a particular theoretical model. (We shall discuss these problems further in Chapter II.)

4. *Theoretical Models, Hypotheses, and Facts*

Let $x_1', x_2', \cdots, x_n'$, be n real variables, and let $(x_1', x_2', \cdots, x_n')$, or, for short, (x'), denote any particular set of values of these variables. Any such set may be represented by a point in n-dimensional Cartesian space. Let S be the set of all such points, and let "A" be a system of rules or operations which defines a subset S_A of S. (S_A might, for example, be a certain n-dimensional surface.) The rules "A" ascribe to each point (x') a property, viz., the property of belonging to S_A or not belonging to S_A. If we allow the n variables x' to vary only under the condition that (x') must belong to S_A, this forms a theoretical model for what the variables x' can do.

Similarly, consider n time functions $x_1'(t), x_2'(t), \cdots, x_n'(t)$. Let F be the set of all possible systems of n time functions, and let "B" be a system of rules or operations that defines a subclass F_B of F. Any system of n time functions will then have the property of either belonging to F_B or not belonging to F_B. The system of rules "B" defines a model with respect to n time series.

Thus, a theoretical model may be said to be simply a restriction upon the joint variations of a system of variable quantities (or, more generally, "objects") which otherwise might have any value or property. More generally, the restrictions imposed might not absolutely exclude any value of the quantities considered; it might merely give different *weights* (or probabilities) to the various sets of possible values of the variable quantities. The model in question would then usually be characterized by the fact that it defines certain restricted subsets of the set of all possible values of the quantities, such that these subsets have nearly all of the total weight.

A theoretical model in this sense is, as it stands, void of any practical meaning or interest. And this situation is, as we have previously explained, not changed by merely introducing "economic names" for the variable quantities or objects involved. The model attains economic meaning only after a corresponding system of quantities or objects in real economic life has been chosen or described, in order to be *identified with those in* the model. That is, the model will have an economic meaning only when associated with a design of actual experiments that describes—and indicates how to measure—a system of "true" variables (or objects) $x_1, x_2, \cdots, x_n$ that are to be identified with the corresponding variables in the theory.

As a consequence of such identification all the permissible statements that can be made within the model with respect to the theoretical variables or objects involved are automatically made also with respect

to the actual, "true" variables. The model thereby becomes *an a priori hypothesis* about real phenomena, stating that every system of values that we might observe of the "true" variables will be one that belongs to the set of value-systems that is admissible within the model. The idea behind this is, one could say, that Nature has a way of selecting joint value-systems of the "true" variables such that these systems are as if the selection had been made by the rule defining our theoretical model. Hypotheses in the above sense are thus the joint implications—and the only testable implications, as far as *observations* are concerned —of a theory *and* a design of experiments. It is then natural to adopt the convention that a theory is called true or false according as the hypotheses implied are true or false, when tested against the data chosen as the "true" variables. Then we may speak, interchangeably, about testing hypotheses or testing theories.

If a certain set of value-systems of the variables is *excluded* in the model then any one system of observed values that falls into this excluded set would be sufficient to reject the hypothesis (and, therefore, the theory) as false with respect to the "true" variables considered. But as we have mentioned, the model may be (and we believe that to be practical it has to be) such that it does not exclude any system of values of the variables, but merely gives different weights or probabilities to the various value-systems. These weights then need a practical interpretation in order that the model shall express a meaningful hypothesis with respect to the corresponding "true" variables. According to experience it has very often been found fruitful to interpret such weights as a measure of actual "frequency of occurrence." If the total weight ascribed to all the possible value-systems is finite, we can then say that the practical meaning of a set of value-systems that has a weight almost equal to zero according to the model is a hypothesis saying that Nature has a way of selecting joint value-systems of the corresponding "true" variables that makes it "practically impossible" that a system of observed values should fall within such a set. For the purpose of *testing* the theory against some other alternative theories we might then agree to deem the hypothesis tested false whenever we observe a certain number of such "almost impossible" value-systems. That is, at the risk of making an error, we should then prefer to adopt another hypothesis under which the observations made are not of the "almost impossible" type.

If we have found a certain hypothesis, and, therefore, the model behind it, acceptable on the basis of a certain number of observations, we may decide to use the theory for the purpose of predictions. If, after a while, we find that we are not very successful with these predictions,

we should be inclined to doubt the validity of the hypothesis adopted (and, therefore, the usefulness of the theory behind it). We should then test it again on the basis of the extended set of observations.

It has been found fruitful to introduce a special calculus for deriving such types of hypotheses. This is the calculus of probability. Later on we shall study at length the common sense of applying this calculus for the derivation of hypotheses about economic phenomena.

Now suppose that we have a set of observations that all confirm the statements that are permissible within our model. Then these statements become facts interpreted in the light of our theoretical model, or, in other words, our model is acceptable so far as the known observations are concerned. But will the model hold also for future observations? We cannot give any a priori reason for such a supposition. We can only say that, according to a vast record of actual experiences, it seems to have been fruitful to believe in the possibility of such empirical inductions.

* * *

In the light of the above analysis we may now classify, roughly, the main problems that confront us in scientific quantitative research. They are:

1. The construction of tentative models. It is almost impossible, it seems, to describe exactly how a scientist goes about constructing a model. It is a creative process, an art, operating with rationalized notions of some real phenomena and of the mechanism by which they are produced. The whole idea of such models rests upon a belief, already backed by a vast amount of experience in many fields, in the existence of certain elements of invariance in a relation between real phenomena, provided we succeed in bringing together the right ones.

2. The testing of theories, which is the problem of deciding, on the basis of data, whether to maintain and use a certain theory or to dismiss it in exchange for another.

3. The problem of estimation, which, in the broadest sense, is the problem of splitting, on the basis of data, all a priori possible theories about certain variables into two groups, one containing the admissible theories, the other containing those that must be rejected.

4. The problem of predictions.

The problems 2, 3, and 4 are closely bound to a probabilistic formulation of hypotheses, and much confusion has been caused by attempts to deal with them otherwise. In a probabilistic formulation they can all be precisely defined, and much of the confusion in current economic research can then be cleared away. These problems will be the subjects of Chapters IV, V, and VI.

Many economists would, however, consider the problems 2–4 as details. Their principal concern is in a sense a more fundamental one, namely the question of whether we might have any hope at all of constructing rational models that will contribute anything to our understanding of real economic life. In the next chapter we shall try to clarify some of the main arguments and points in this discussion.

CHAPTER II

The Degree of Permanence of Economic Laws

If we compare the historic developments of various branches of quantitative sciences, we notice a striking similarity in the paths they have followed. Their origin is Man's craving for "explanations" of "curious happenings," the observations of such happenings being more or less accidental or, at any rate, of a very passive character. On the basis of such—perhaps very vague—recognition of facts, people build up some primitive explanations, usually of a metaphysical type. Then, some more "cold-blooded" empiricists come along. They want to "know the facts." They observe, measure, and classify, and, while doing so, they cannot fail to recognize the possibility of establishing a certain order, a certain system in the behavior of real phenomena. And so they try to construct systems of relationships to *copy* reality as they see it from the point of view of a careful, but still passive, observer. As they go on collecting better and better observations, they see that their "copy" of reality needs "repair." And, successively, their schemes grow into labyrinths of "extra assumptions" and "special cases," the whole apparatus becoming more and more difficult to manage. Some clearing work is needed, and the key to such clearing is found in *a priori reasoning*, leading to the introduction of some very general—and often very simple—principles and relationships, from which *whole classes* of apparently very different things may be deduced. In the natural sciences this last step has provided much more powerful tools of analysis than the purely empirical *listing of cases*.

We might be inclined to say that the possibility of such fruitful hypothetical constructions and deductions depends upon two separate factors, namely, on the one hand, the fact that there *are* laws of Nature, on the other hand, the efficiency of our analytical tools. However, by closer inspection we see that such a distinction is a dubious one. Indeed, we can hardly describe such a thing as a law of nature without referring to certain *principles of analysis*. And the phrase, "In the natural sciences we have stable laws," means not much more and not much less than this: The natural sciences have chosen very fruitful ways of looking upon physical reality. So also, a phrase such as "In economic life there are no constant laws," is not only too pessimistic, it also seems meaningless. At any rate, it cannot be tested. But we may discuss whether the relationships that follow from our *present* scheme of economic theory are such that they apply to facts of real economic life. We may discuss problems which arise in attempting to make comparisons between reality and our present set-up of economic theory. We

may try to find a rational explanation for the fact that relatively few attempts to establish economic "laws" have been successful. I think that considerable effort should first be spent on clarifying these restricted problems.

In the following we propose to deal with some of the fundamental problems that arise in judging the degree of persistence over time of relations between economic variables. For the sake of simplicity we shall often operate here with the notion of "exact" rather than "stochastical" relationships. We can do this because the main points to be discussed do not seem to be principally related to the particular *type* of relations that we might hope to establish. The problems to be discussed are more directly connected with the general question of whether or not we might hope to find elements of invariance in economic life, upon which to establish permanent "laws."

5. *What Do We Mean by a "Constant Relationship"?*

When we use the terms "constant relationships," or "unstable, changing relationships," we obviously refer to the behavior of some *real* economic phenomena, as *compared* with some behavior that we *expect* from theoretical considerations. The notion of constancy or permanence of a relationship is, therefore, *not* one of pure theory. It is a property of real phenomena *as we look upon them* from the point of view of a particular theory. More precisely, let $x_1', x_2', \cdots, x_n'$, be n theoretical variables, restricted by an equation

$$f(x_1', x_2', \cdots, x_n'; \alpha_1, \alpha_2, \cdots, \alpha_k) = s', \tag{5.1}$$

where the α's are constants, and where s' is a shift possessing certain specified properties. (5.1) does not become an economic theory just by using economic terminology to *name* the variables involved. (5.1) becomes an economic theory when associated with a rule of actual measurement of n economic variables, $x_1, x_2, \cdots, x_n$, to be *compared with* $x_1', x_2', \cdots, x_n'$, respectively. The essential feature of such a rule of measurement is that it does *not a priori* impose the restriction (5.1) upon the variables to be measured. If we did that, we should fall back into the world of abstract theory, because one of the variables would follow from the measurement of the $n-1$ others and the properties assigned to s'. The rule of measurement is essentially a technical device of measuring *each variable separately*. It is a *design of actual experiments*, to obtain the "true" variables as described in Section 3.

All value-sets of the n theoretical variables x' in (5.1) have a common property, namely the property of satisfying that equation. We are interested in whether or not the "true" variables $x_1, x_2, \cdots, x_n$, have the same property. Let $(x_1, x_2, \cdots, x_n)$ be any one of the results obtain-

able by our design of experiments, and let s be a variable defined implicitly by

$$f(x_1, x_2, \cdots, x_n; \alpha_1, \alpha_2, \cdots, \alpha_k) = s, \tag{5.2}$$

where f is the same as in (5.1). If then s has the same properties as s' in (5.1) whatever be the system of experimentally observed values of $x_1, x_2, \cdots, x_n$, we say that the observable "true" variables x_i follow a constant law.

Therefore, given a theoretical relation, a design of experiments, and a set of observations, the problem of constancy or invariance of an economic relation comes down to the following two questions:

(1) Have we actually observed what we meant to observe, i.e., can the given set of observations be considered as a result obtained by following our design of "ideal" experiments?

(2) Do the "true" variables actually have the properties of the theoretical variables?

A design of experiments (a prescription of what the physicists call a "crucial experiment") is an essential appendix to any quantitative theory. And we usually have some such experiments in mind when we construct the theories, although—unfortunately—most economists do not describe their designs of experiments explicitly. If they did, they would see that the experiments they have in mind may be grouped into two different classes, namely, (1) experiments that *we should like to make* to see if certain real economic phenomena—when *artificially isolated* from "other influences"—would verify certain hypotheses, and (2) the stream of experiments that Nature is steadily turning out from her own enormous laboratory, and which we merely watch as passive observers. In both cases the aim of theory is the same, namely, to become master of the happenings of real life. But our approach is a little different in the two cases.

In the first case we can make the agreement or disagreement between theory and facts depend upon *two* things: the facts we choose to consider, as well as our theory about them. As Bertrand Russell has said: "The actual procedure of science consists of an alternation of observation, hypothesis, experiment, and theory."[1]

In the second case we can only try to adjust our theories to reality as it appears before us. And what is the meaning of a design of experiments in this case? It is this: We try to choose a theory and a design of experiments to go with it, in such a way that the resulting data *would be* those which we get by passive observation of reality. And to the extent

[1] *The Analysis of Matter*, New York, 1927, p. 194.

that we succeed in doing so, we become master of reality—by passive agreement.

Now, if we examine current economic theories, we see that a great many of them, in particular the more profound ones, require experiments of the first type mentioned above. On the other hand, the kind of economic data that we actually have belong mostly to the second type. In economics we use a relatively small vocabulary to describe an enormous variety of phenomena (and sometimes economists use different names for the same phenomenon). The result is that many different things pass under the same name, and that, therefore, we are in danger of considering them as identical. And thus, theories are often being compared with data which cannot at all be considered as observations obtained by following the design of experiment we had in mind when constructing the theory. Of course, when a theory does not agree with the facts we can always say that we do not have the right kind of data. But this is an empty phrase, unless we can describe, at the same time, what *would be* the right kind of data, and how to obtain them, at least in point of principle. If every theory should be accompanied by a carefully described design of experiments, much confusion on the subject of constant versus changing economic "laws" would be cleared up.

This description of the problem of stability or permanence of economic relations is a very broad one. It may give a preliminary answer to very superficial critics of the possibility of developing economics as a science. But it does not answer the many profound problems of details which confront us when we really try to investigate why economics, so far, has not led to very accurate and universal laws like those obtaining in the natural sciences.

Let us first once more look upon the general argument: "There are no constant laws describing phenomena of economic life." Above we said that this argument was meaningless. We shall support this statement a little further. It is not possible to give any precise answer to the argument, because it does not itself represent a precise question. But let us try to understand what the argument means. Suppose, first, we should consider the "class of all designs of experiments," the results of which "we should be interested in as economists." Here, of course, we get into difficulty immediately, because it is probably not possible to define such a class. We do not know all the experiments we might be interested in. Consider, on the other hand, the class of all possible economic theories (of the type we are discussing here). By each design of experiments there is defined a sequence of actual measurements. Consider, for each such measurement, the subclass of theories with which the measurement agrees. For a sequence of measurements we get a

sequence of such subclasses of theories. Now, if these classes of theories did *not* have any nontrivial property in common, we might say that the measurements obtained by the design of experiments used do not follow any law. But does this statement really say anything? Obviously, very little. Because it is a statement about classes of things which are completely *un*defined. No matter how much we try and fail, we should never be able to establish such a conclusion as "In economic life there are no constant laws."

We shall consider a much more restricted problem, namely this: How far do the hypothetical "laws" of economic theory in its present stage apply to such data as we get by passive observations? By passive observations we mean observable results of what individuals, firms, etc., *actually do* in the course of events, not what they *might* do, or what they *think* they would do under certain other specified circumstances. It would be superficial to consider this problem merely as a question of whether our present economic theory is *good* or *bad;* or, rather, that is not a fruitful setting of the problem. We have to start out by analyzing what we are actually trying to achieve by economic theory. We have to compare its designs of idealized experiments with those which would be required to reproduce the phenomena of real economic life that we observe passively.

In such a discussion we soon discover that we have to deal with a manifold of different questions. Let us try to review the most important ones:

(a) Are most of the theories we construct in "rational economics" ones for which historical data and passive observations are *not* adequate experiments? This question is connected with the following:

(b) Do we try to construct theories describing what individuals, firms, etc., *actually do* in the course of events, or do we construct theories describing schedules of alternatives at a given moment? If the latter is the case, what bearing do such schedules of alternatives have upon a series of decisions and actions actually carried out?

(c) Why do we not confine ourselves only to such theories as are directly verifiable? Or, why are we interested in relations for which Nature does not furnish experiments?

(d) Very often our theories are such that we think certain directly observable series *would* give adequate experimental results for a verification, provided *other things* did not change. What bearing may such theories have upon reality, if we simply neglect the influences of these "other things"? This, again, is connected with the following problem:

(e) Are we interested in describing what *actually does* happen, or are we interested in what *would happen* if we could keep "other things" unchanged? In the first case we construct theories for which we hope

Nature itself will take care of the necessary *ceteris paribus* conditions, knowing, e.g., that this has been approximately so in the past. In the second case we try to take care of the *ceteris paribus* conditions ourselves, by statistical devices of clearing the data from influences not taken account of in the theory (e.g., by multiple-correlation analysis).

(f) From experience with correlation of time series we know that it is often possible to establish very close relationships between economic variables for some particular time period, while the relationships break down for the next time period. Does this fact mean that we cannot hope to establish constant laws of economic life?

These questions, being taken more or less directly out of current discussions on problems of economic research, are, as can be seen, hopelessly overlapping; nor does any one of them form a precise analytical problem. We, therefore, ask: Can these problems be covered, at least partly, by analysis of a set of simplified and more disjunct problems? In the following we shall try to do so by studying three different groups of problems, which we may call, for short,

I. The reversibility of economic relations,
II. The question of simplicity in the formulation of economic laws,
III. The autonomy of an economic relation.

6. *The Reversibility of Economic Relations*

In the field of economic research the application of relations of pure theory to time series or historic records has become something like taboo. Many economists, not sufficiently trained in statistical theory, have, it seems, been "scared away" by such critical work as, e.g., that of G. U. Yule.[2] They have come to think that there is something inherent in economic time series *as such*, which make these data unfit for application of pure economic theory. The general argument is something like this: In economic theory we operate with hypothetical schedules of decisions, which individuals, firms, etc., may take in response to certain *alternatively* fixed conditions (e.g., adaptation of quantity consumed to a given price change). But economic time series showing actual results of decisions taken are only historic descriptions of a *one-way* journey through a sequence of ever-shifting "environments," so that it is not possible to make actual predictions by means of the schedules of alternatives given by pure economic theory.

In trying to analyze this problem more precisely, we notice first that the general argument above does not deny the possibility that relations deduced from economic theory may prove very persistent and accurate

[2] E.g., "Why Do We Sometimes Get Nonsense-correlations between Time Series?" *Journal of the Royal Statistical Society*, Vol. 89, 1926, pp. 1–64.

when applied to facts. The argument implies only that the types of data represented by economic time series are not those which would result from the designs of experiments prescribed in economic theory. Here we should, first of all, think of the difficulties that arise from the fact that series of passive observations are influenced by a great many factors not accounted for in theory; in other words, the difficulties of fulfilling the condition "Other things being equal." But this is a problem common to all practical observations and measurements; it is, in point of principle, not a particular defect of economic time series. If we cannot clear the data of such "other influences," we have to try to introduce these influences in the theory, in order to bring about more agreement between theory and facts. Also, it might be that the data, as given by economic time series, are restricted by a *whole system* of relations, such that the series do not display enough variations to verify each relation separately. These problems we shall discuss at length in the next two sections. Again, there is the problem of errors of measurements proper. But this problem also is a general one, and not one peculiar to economic time series.

If these difficulties are put aside, is there still some property peculiar to economic time series that makes them unfit for the application of relations deduced from pure economic theory? Even by a careful inspection it is difficult to see what such a property could be, because, if we can construct any general laws at all, describing what individuals *actually do*, and if we have a series of observations of what the individuals actually *have done* in the past, then, necessarily, the theoretical law would fit these observation series. If, therefore, we see here a problem at all, I think it arises, mostly, from a confusion of two different kinds of relations occurring in economic theory, namely (1) those intended to describe what the individuals *actually do* at any time, and (2) those describing a schedule of alternatives at a given moment, *before* any particular decision has been taken. Relations of the first type are, usually, derived from a *system* of relations of the second type. To make the discussion on this point more concrete we shall consider a simple example of consumers' demand for a single commodity.

Suppose that an individual consumes n different commodities, and let $x_1, x_2, \cdots, x_n$ denote quantities of these n commodities. And let $p_1, p_2, \cdots, p_n$ be their corresponding prices. Assume that the individual has constant money income. According to the general theory of consumers' choice, we may write

$$x_i = f_i(p_1, p_2, \cdots, p_n) \qquad (i = 1, 2, \cdots, n), \tag{6.1}$$

where f_i are some demand functions. Assume now that all prices, except

one, say p_1, are constant, and consider the corresponding quantity, x_1, of commodity No. 1. We may then write

$$x_1 = f(p_1). \tag{6.2}$$

What does this function mean, under the assumptions made above? It may mean two different things.

One interpretation is that, whenever p_1 has a particular value, say p_1', the individual chooses to buy a quantity $x_1' = f(p_1')$ of commodity No. 1.

Another interpretation is this: Suppose that the individual is in a position where he pays the price p_1^0 and consumes a quantity x_1^0. He considers in *that position* the possible changes in his consumption of commodity No. 1 that he would choose in response to various changes in the price *from* p_1^0. If the price be changed from p_1^0 to p_1', say, he will buy $x_1' = f(p_1')$; if the price be changed from p_1^0 to p_1'' say, he will buy $x_1'' = f(p_1'')$; and so forth. That is to say, he has a schedule of alternatives with respect to the *next* price change as *judged* from his *present position* (x_1^0, p_1^0). To indicate that his schedule may depend upon his present position, we might write

$$x_1 = f^0(p_1), \tag{6.3}$$

where f^0 satisfies $x_1^0 = f^0(p_1^0)$.

It is clear that these two types of demand schedules are of different nature, and, furthermore, that the first one claims more than the second one. For the first one requires the assumption that there is a unique relation between consumption and prices according to which the individual acts *irrespective* of the position he happens to be in at the moment the decision has to be taken. The second only says that the individual has a schedule of alternatives with respect to the *next* price change, as judged from his present position (x_1^0, p_1^0). *After* he has taken a decision in response to a price change, so that he no longer is in the position (x_1^0, p_1^0), he might change his schedule of alternatives, because from the new position he might "see things differently."

If the individual has a fixed demand schedule that is independent of the point on it where he is at any given moment [i.e., a schedule of type (6.2)], then, of course, a historical record of prices and corresponding quantities consumed would represent points on this demand schedule, and we could use it for predicting the consumption for any given value of the price (under the assumption, as before, that other prices did not change). On the other hand, if the demand schedule depends upon the actual position of the individual, there might, *for each such actual position*, be a perfectly well-defined schedule of alternatives,

which, if we knew it, would allow us to predict the quantity that would be bought if the price were changed from p_1^0 to p_1', say. But as soon as the new position (p_1', x_1') is actually reached, we might need *another* schedule, f' say, to predict the quantity bought if the price were changed from p_1' to p_1'', say. The two situations are illustrated graphically in Figures 1 and 2.

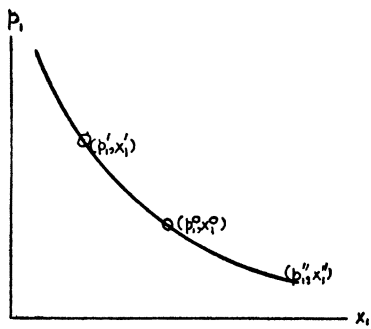

FIGURE 1.—Reversible Demand Schedule.

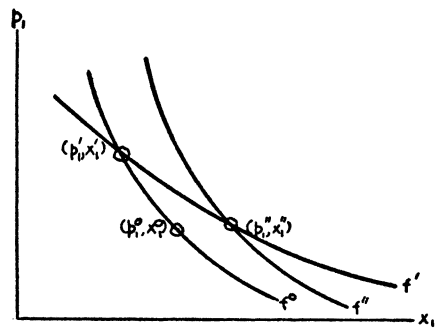

FIGURE 2.—"Milieu"-affected Demand Schedule. Irreversible Demand Process.

In Figure 2 a historical record of the actual positions (p_1^0, x_1^0), (p_1', x_1'), etc., would *not* form points on any fixed demand curve. And if we should fit some curve through these points of actual positions, such a curve could not be used for predicting the effect of the next price change. To find the demand schedule of the individual at a given moment we should have to *interview* him, asking him what he *would do* if the price were changed alternatively by certain amounts.

We might consider Figure 1 as a static scheme, while Figure 2 represents a dynamic one, because in Figure 1 the *sequence* of price changes is irrelevant, while in Figure 2 it is essential. However, we do not here emphasize so much the time succession of the price-quantity changes as the fact that the actual *carrying out* of a planned decision may bring the individual into a new "milieu," so to speak, where he feels differently from the way that he thought he would feel before he got there.

If, actually, a set-up like that in Figure 2 is nearer to reality than that in Figure 1, then, naturally, an attempt to use the scheme in Figure 1 would fail. On the other hand, if the theory operates with "milieu"-bound schedules like those in Figure 2, then historical records of actual price-quantity combinations are simply not the data that are relevant to the theory.

An irreversible scheme like that in Figure 2 may often be reduced to a reversible one by introducing *more variables*. We might, e.g., assume

that the demand schedules in Figure 2 change in a *regular manner* with the initial positions with which they are associated. Let the variables $\bar{x}_1$, $\bar{p}_1$ be the quantity and price that represent *actual positions* of the individual, and let (x_1, p_1) be any point on the demand schedule through $(\bar{x}_1, \bar{p}_1)$. It might be that the individual's behavior could be described by a relation of the type

$$x_1 = F(p_1, \bar{x}_1, \bar{p}_1), \tag{6.4}$$

where F is such that

$$\bar{x}_1 = F(\bar{p}_1, \bar{x}_1, \bar{p}_1). \tag{6.4'}$$

This function would then be compatible with the time series for actual prices and quantities consumed. More specifically, each pair of successive points representing actual positions would satisfy (6.4); i.e., if $(\bar{x}_1^0, \bar{p}_1^0)$ and $(\bar{x}_1', \bar{p}_1')$ be two such successive points, we should have

$$\bar{x}_1' = F(\bar{p}_1', \bar{x}_1^0, \bar{p}_1^0). \tag{6.5}$$

We could then determine the parameters of F from the actual time series, and then, by (6.4), we could calculate the demand schedule for any given initial point $(\bar{x}_1, \bar{p}_1)$.

This scheme would probably be too simple. In general we should probably have to introduce as variables, not only the instantaneous position of the individual, but also the whole sequence of past positions, as well as the lengths of the time intervals between the price changes. And the situation would, of course, be still more complicated when all the other prices also varied. This was excluded in our discussion above. Whether or not it be actually possible in this way to fit historical records into schemes of reversible relationships is a question which cannot be answered a priori. We have to try to find out.

Beside difficulties of the type discussed above, which seem—in point of principle—very simple and clear ones, I do not see that economic time series have any other "mystic" property that makes them incompatible with economic theory.

7. *The Question of Simplicity in the Formulation of Economic Laws*

Let y denote an economic variable, the observed values of which may be considered as results of planned economic decisions taken by individuals, firms, etc. (e.g., y might be the annual consumption of a certain commodity within a certain group of individuals, or the annual amount they save out of their income, etc.; or, it might be the rate of production in a monopolized industry, or monthly imports of a certain raw material, etc., etc.). And let us start from the assumption that the

variable y, is influenced by a number of *causal* factors. This viewpoint is something that is deeply rooted in our way of reasoning about the things we observe in reality. We do not need to take the notions of cause and effect in any metaphysical sense. What we mean is simply that the individuals, firms, etc., are bound in their planning and decisions by a set of conditions that are *data* in the process of adaptation. Within the limits of these given conditions the adaptation process consists in choosing what is deemed the "best" decision, in some sense or another. And we assume that the individuals have a system of preference schedules which determine "best decisions" corresponding to any given set of choice-limiting conditions. We, therefore, have the following scheme:

$$\text{(7.1)} \qquad \left\{\begin{array}{l}\text{Given conditions}\\ \text{(the independent}\\ \text{variables)}\end{array}\right\} \rightarrow \left\{\begin{array}{l}\text{System of}\\ \text{preference}\\ \text{schedules}\end{array}\right\} \rightarrow \left\{\begin{array}{l}\text{"Best decision"}\\ \text{(the dependent}\\ \text{variables)}\end{array}\right\}.$$

If the system of preference schedules establishes a correspondence between sets of given conditions and "best decisions," such that for each set of conditions there is one and only one best decision, we may "jump over" the middle link in (7.1), and say that the decisions of individuals, firms, or groups, are *determined* by the system of given choice-limiting conditions (the independent variables).

In point of principle there may, perhaps, appear to be some logical difficulties involved in operating with such *one-way,* or causal relationships. In fact, modern economists have stressed very much the necessity of operating with relations of the mutual-dependence type, rather than relations of the cause-effect type. However, both types of relations have, I think, their place in economic theory; and, moreover, they are not necessarily opposed to each other, because a system of relations of the mutual-dependence type for the economy *as a whole* may be built up from *open* systems of causal relations within the various *sectors* of the economy. The causal factors (or the "independent variables") for one section of the economy may, themselves, be dependent variables in another section, while here the dependent variables from the first section enter as independent variables. The essential thing is that, while for the economy as a whole everything depends upon everything else, so to speak, there are, for each individual, firm, or group, certain factors which *this* individual, firm, or group *considers as data.* The notion of causal factors is of a relative character, rather than an absolute one.

Let us, therefore, accept the point of view that decisions to produce, to consume, to save, etc., are influenced by a number of quantitatively defined relative causal factors $x_1, x_2, \cdots$. Our hope in economic theory and research is that it may be possible to establish constant and rela-

tively *simple* relations between dependent variables, y (of the type described above), and a relatively *small* number of independent variables, x. In other words, we hope that, for each variable, y, to be "explained," there is a relatively small number of explaining factors the variations of which are practically decisive in determining the variations of y. (The problem of simplicity of the *form* of a relationship is usually far less important than that of the number of variables involved, because, if we know there is a functional relationship at all, it is, usually, possible to approximate it, e.g., by expanding the function in series.)

Whether or not such simple relations can be established must be decided by actual trials. A priori it can neither be shown to be possible nor proved impossible. But we may do something else, which may give us some hint as to how optimistic or pessimistic we have reason to be: we can try to indicate what would have to be the actual situation in order that there should be *no hope* of establishing simple and stable causal relations.

First of all, it is necessary to define what we mean by the "influence" of an economic factor. This expression, as used in the economic literature, seems to have several different meanings. We shall distinguish between two different notions of "influence," which we shall call *potential* influence, and *factual* influence respectively. We shall first define these two concepts in a purely formal way.

Let y' be a theoretical variable defined as a function of n independent "causal" variables $x_1, x_2, \cdots, x_n$, e.g.,

$$y' = f(x_1, x_2, \cdots, x_n), \tag{7.2}$$

where f is defined within a certain domain of the variables x. The *potential* influence of the factor x_i upon y' we shall define as $\Delta_i y'$ given by

$$\Delta_i y' = f[x_1, x_2, \cdots, (x_i + \Delta x_i), \cdots, x_n] - f(x_1, x_2, \cdots, x_n), \tag{7.3}$$

where Δx_i is a positive magnitude such that $x_i + \Delta x_i$ is within the domain of definition of f. It is clear that this quantity $\Delta_i y'$ will, in general, depend upon the variables x as well as upon the value of Δx_i. And, furthermore, what we shall mean by a large or a small Δx_i depends, of course, upon the units of measurement of the variables x. To *compare* the size of the influence of each of the variables x we have, for any point $(x_1, x_2, \cdots, x_n)$, to choose a set of displacements $\Delta x_1, \Delta x_2, \cdots, \Delta x_n$, which are considered to be of *equal size* according to some *standard of judgment.* (E.g., one particular such standard would be to define the increments Δx_i at any point in the space of the variables x as constant and equal percentages of $x_1, x_2, \cdots, x_n$ respectively.) For a given sys-

tem of displacements $\Delta x_1, \Delta x_2, \cdots, \Delta x_n$, the potential influences are, clearly, formal properties of the function f.

Now, let us define the notion of *factual* influence of x_i upon y'. In contrast to the potential influence, the factual influence refers to a *set* of values of y' corresponding to a set of value systems of the variables $x_1, x_2, \cdots, x_n$, chosen according to some *outside principle*. Let

$$\begin{array}{l} y_1', \ x_{11}, \ x_{21}, \ \cdots, \ x_{n1}, \\ y_2', \ x_{12}, \ x_{22}, \ \cdots, \ x_{n2}, \\ \cdots\cdots\cdots\cdots \\ y_N', \ x_{1N}, \ x_{2N}, \ \cdots, \ x_{nN}, \end{array} \tag{7.4}$$

be a set of N such value systems. By the factual influence of x_i upon y' within this set of value systems we mean, broadly speaking, the *parts* of $y_1', y_2', \cdots, y_N'$ that may be *ascribed* to the variations in x_i. This could be defined quantitatively in various ways. One way would be the following: Let us replace the variable x_i in (7.2) by a constant, c_i say, so determined that

$$\begin{aligned} Q_i = \sum_1^N{}_j [f(x_{1j}, x_{2j}, \cdots, x_{ij}, \cdots, x_{nj}) \\ - f(x_{1j}, x_{2j}, \cdots, c_i, \cdots, x_{nj})]^2 \\ = \text{minimum with respect to } c_i, \end{aligned} \tag{7.5}$$

assuming that such a minimum exists. The *factual* influence upon y' of the variable x_i in the system (7.4) could then, for example, be defined as: Constant $\sqrt{Q_i^{(\min.)}}$.

From the definitions above it is clear that the potential influence of a factor may be large, while—at the same time—the factual influence of this factor in a particular *set of data* may be zero or very small. And, conversely, the factual influences may be very large even if the potential influence is small (but not identically zero).

This distinction is fundamental. For, if we are trying to explain a certain observable variable, y, by a system of causal factors, there is, in general, no limit to the number of such factors that might have a *potential* influence upon y. But Nature may limit the number of factors that have a nonnegligible *factual* influence to a relatively small number. Our hope for simple laws in economics rests upon the assumption that we may proceed as if such natural limitations of the number of relevant factors exist. We shall now discuss this a little more closely.

Suppose that, out of a—possibly infinite—number of factors $x_1, x_2, \cdots$, with a potential influence upon y, we pick out a relatively small number, say $x_1, x_2, \cdots, x_n$, and consider a certain function

$$y^* = U(x_1, x_2, \cdots, x_n) \tag{7.6}$$

of these variables. Suppose that, *if* all the other factors, $x_{n+1}, x_{n+2}, \cdots$, (assuming them to be denumerable) *did not vary*, we *should* have $y = y^*$ for every observed value-set $(y, x_1, x_2, \cdots, x_n)$. Would the knowledge of such a relationship help us to "explain" the actual, observed values of y? It would, provided the *factual* influence of all the unspecified factors together were very small as compared with the factual influence of the specified factors $x_1, x_2, \cdots, x_n$. This might be the case even if (1) the unspecified factors varied considerably, provided their potential influence was very small, or if (2) the potential influences of the unspecified factors were considerable, but at the same time these factors did not change much, or did so only very seldom as compared with the specified factors.

On the other hand, suppose that all the factors $x_1, x_2, \cdots, x_n, x_{n+1}, \cdots$, or at least a *very large number* of them, were of the following type: (1) Each factor x has a considerable potential influence upon y; (2) each x varies *usually* very little, but *occasionally* some great variations occur. Since there are a great many factors x, we might then still have great variations going on almost all the time, *in one factor or the other*. To pick out a small number of factors x, assuming the rest to be constant, would then be of very little help in "explaining" the actual variations observed for y, i.e., relations of the form (7.6) would show very little persistence over time if y were substituted for y^*, simply because the *ceteris paribus* conditions, $x_{n+1} = \text{constant}$, $x_{n+2} = \text{constant}$, etc., would be no approximation to reality. From the point of view of "explaining" reality, we might then say that it would be practically impossible to construct a theory such that its associated design of experiments would approximate that followed by Nature. From the point of view of *verifying* certain simplified relations of theory we might say that, under the situation just described, it would be impossible to find data for such a purpose by the method of passive observation.

What is the actual situation as we know it from experience in economic research? Do we actually need to consider an enormous number of factors to "explain" decisions to produce, to consume, etc.? I think our experience is rather to the contrary. Whenever we try, a priori, to specify what we should think to be "important factors," our imagination is usually exhausted rather quickly; and when we attempt to apply

our theory to actual data (e.g., by using certain regression methods), we often find that even a great many of the factors in our a priori list turn out to have practically no factual influence.

Frequently, our greatest difficulty in economic research does *not* lie in establishing simple relations between actual observation series, but rather in the fact that the observable relations, over certain time intervals, *appear to be still simpler* than we expect them to be from theory, so that we are thereby led to *throw away* elements of a theory that would be sufficient to explain apparent "breaks in structure" later. This is the problem of autonomy of economic relations, which we now shall discuss.

8. The Autonomy of an Economic Relation

Every research worker in the field of economics has, probably, had the following experience: When we try to apply relations established by economic theory to actually observed series for the variables involved, we frequently find that the theoretical relations are "unnecessarily complicated"; we can do well with fewer variables than assumed a priori. But we also know that, when we try to make predictions by such simplified relations for a new set of data, the relations often break down, i.e., there appears to be a *break in the structure* of the data. For the new set of data we might also find a simple relation, but a *different* one. Even if no such breaks appear, we are puzzled by this unexpected simplicity, because, from our theoretical considerations we have the feeling that economic life is capable of producing variations of a much more general type. Sometimes, of course, this situation may be explained directly by the fact that we have included in our theory factors which have no potential influence upon the variables to be explained. But more frequently, I think, the puzzle is a result of confusing two different kinds of *variations* of economic variables, namely hypothetical, *free* variations, and variations which are *restricted* by a system of simultaneous relations.

We see this difference best by considering the rational operations by which a theoretical system of relations is constructed. Such systems represent attempts to *reconstruct*, in a simplified way, the mechanisms which we think lie behind the phenomena we observe in the real world. In trying to rebuild these mechanisms we consider *one* relationship *at a time.*

Suppose, e.g., we are considering n theoretical variables x_1', x_2', $\cdots$, x_n', to be compared with n observational variables x_1, x_2, $\cdots$, x_n, respectively. We impose certain relations between the n theoretical variables, of such a type that we think the theo-

retical variables, so restricted, will show some correspondence with the observed variables.

Let us consider one such particular relation, say $x_1' = f(x_2', \cdots, x_n')$. In constructing such a relation, we reason in the following way: *If* x_2' be such and such, x_3' such and such, etc., *then* this implies a certain value of x_1'. In this process we do not question whether these "ifs" can actually occur or not. When we impose more relations upon the variables, a great many of these "ifs," which were possible for the relation $x_1' = f$ separately, may be impossible, because they violate the other relations. After having imposed a whole system of relations, there may not be very much left of all the hypothetical variation with which we started out. At the same time, if we have made a lucky choice of theoretical relations, it may be that the possible variations that are left over agree well with those of the observed variables.

But why do we start out with much more general variations than those we finally need? For example, suppose that the Walrasian system of general-equilibrium relations were a true picture of reality; what would be gained by operating with this general system, as compared with the simple statement that each of the quantities involved is equal to a constant? The gain is this: In setting up the different general relations we conceive of a *wider set of possibilities* that might correspond to reality, were it ruled by one of the relations only. The simultaneous system of relations gives us an *explanation* of the fact that, out of this enormous set of possibilities, only one very particular one actually emerges. But once this is established, could we not then forget about the whole process, and keep to the much simpler picture that is the actual one? Here is where the problem of *autonomy* of an economic relation comes in. The meaning of this notion, and its importance, can, I think, be rather well illustrated by the following mechanical analogy:

If we should make a series of speed tests with an automobile, driving on a flat, dry road, we might be able to establish a very accurate functional relationship between the pressure on the gas throttle (or the distance of the gas pedal from the bottom of the car) and the corresponding maximum speed of the car. And the knowledge of this relationship might be sufficient to operate the car at a prescribed speed. But if a man did not know anything about automobiles, and he wanted to understand how they work, we should not advise him to spend time and effort in measuring a relationship like that. Why? Because (1) such a relation leaves the whole inner mechanism of a car in complete mystery, and (2) such a relation might break down at any time, as soon as there is some disorder or change in any working part of the car. (Compare this, e.g., with the well-known lag-relations between the Harvard

A-B-C-curves.) We say that such a relation has very little *autonomy*,[3] because its existence depends upon the simultaneous fulfilment of a great many other relations, some of which are of a transitory nature. On the other hand, the general laws of thermodynamics, the dynamics of friction, etc., etc., are highly autonomous relations with respect to the automobile mechanism, because these relations describe the functioning of some parts of the mechanism *irrespective* of what happens in some *other* parts.

Let us turn from this analogy to the mechanisms of economic life. Economic theory builds on the assumption that individuals' decisions to produce and to consume can be described by certain fundamental behavioristic relations, and that, besides, there are certain technical and institutional restrictions upon the freedom of choice (such as technical production functions, legal restrictions, etc.).

A particular system of such relationships defines one particular theoretical *structure* of the economy; that is to say, it defines a theoretical *set* of possible simultaneous sets of value or sets of time series for the economic variables. It might be necessary—and that is the task of economic theory—to consider various *alternatives* to such systems of relationships, that is, various alternative *structures* that might, approximately, correspond to economic reality at any time. For the "real structure" might, and usually does, change in various respects.

To make this idea more precise, suppose that it be possible to define a *class*, Ω, of *structures*, such that *one member or another* of this class would, approximately, describe economic reality in *any practically conceivable situation*. And suppose that we define some nonnegative *measure* of the "size" (or of the "importance" or "credibility") of any subclass, ω in Ω, including Ω itself, such that, if a subclass contains completely another subclass, the measure of the former is greater than, or at least equal to, that of the latter, and such that the measure of Ω is positive. Now consider a particular subclass (of Ω), containing all those—and only those—structures that satisfy a particular relation "A." Let ω_A be this particular subclass. (E.g., ω_A might be the subclass of all those structures that satisfy a particular demand function "A.") We then say that the relation "A" is *autonomous* with respect to the subclass of structures ω_A. And we say that "A" has a

[3] This term, together with many ideas to the analysis in the present section, I have taken from a mimeographed paper by Ragnar Frisch: "Statistical versus Theoretical Relations in Economic Macro-Dynamics" (Mimeographed memorandum prepared for the Business Cycle Conference at Cambridge, England, July 18–20, 1938, to discuss J. Tinbergen's publication of 1938 for the League of Nations.)

degree of autonomy which is the greater the larger be the "size" of ω_A as compared with that of Ω.

The principal task of economic theory is to establish such relations as might be expected to possess as high a degree of autonomy as possible.

Any relation that is derived by combining two or more relations within a system, we call a *confluent* relation. Such a confluent relation has, of course, usually a lower degree of autonomy (and never a higher one) than each of the relations from which it was derived, and all the more so the greater the number of different relations upon which it depends. From a system of relations, with a certain degree of autonomy, we may derive an infinity of systems of confluent relations. How can we actually distinguish between the "original" system and a derived system of confluent relations? That is *not* a problem of mathematical independence or the like; more generally, it is not a problem of pure logic, but a problem of actually *knowing something* about real phenomena, and of making realistic assumptions about them. In trying to establish relations with high degree of autonomy we take into consideration various *changes* in the economic structure which might upset our relations, we try to dig down to such relationships as actually might be expected to have a great degree of invariance with respect to certain changes in structure that are "reasonable."

It is obvious that the autonomy of a relation is a highly relative concept, in the sense that any system of hypothetical relations between real phenomena might itself be deducible from another, still more basic system, i.e., a system with still higher degree of autonomy with respect to structural changes.

The construction of systems of autonomous relations is, therefore, a matter of intuition and factual knowledge; it is an art.

What is the connection between the degree of autonomy of a relation and its observable degree of constancy or persistence?

If we should take constancy or persistence to mean simply invariance with respect to certain hypothetical changes in structure, then the degree of constancy and the degree of autonomy would simply be two different names for the same property of an economic relation. But if we consider the constancy of a relation as a property of the behavior of *actual observations*, then there is clearly a difference between the two properties, because then the degree of autonomy refers to a class of *hypothetical* variations in structure, for which the relation *would be* invariant, while its actual persistence depends upon what variations *actually occur*. On the other hand, if we always try to form such relations as are autonomous with respect to those changes that are *in fact most*

likely to occur, and if we succeed in doing so, then, of course, there will be a very close connection between actual persistence and theoretical degree of autonomy. To bring out these ideas a little more clearly we shall consider a purely formal set-up.

Suppose we have an economic system, the mechanism of which might be characterized by the variations of n measurable quantities $x_1, x_2, \cdots, x_n$. Suppose that the structure of this mechanism could be described by a system of $m<n$ equations,

$$f_i(x_1, x_2, \cdots, x_n) = 0 \qquad (i = 1, 2, \cdots, m). \tag{8.1}$$

$(n-m)$ of the variables—let them be $x_{m+1}, x_{m+2}, \cdots, x_n$—are assumed to be given *from outside*. From the system (8.1) it might, e.g., be possible to express each of the first m variables uniquely in terms of the $n-m$ remaining ones. Let such a solution be

$$\begin{aligned} x_1 &= u_1(x_{m+1}, x_{m+2}, \cdots, x_n), \\ x_2 &= u_2(x_{m+1}, x_{m+2}, \cdots, x_n), \\ &\cdots\cdots\cdots\cdots\cdots \\ x_m &= u_m(x_{m+1}, x_{m+2}, \cdots, x_n). \end{aligned} \tag{8.2}$$

The system (8.2) would describe the covariations of the variables just as well as would the original system (8.1). But suppose now that there should be a change in structure of the following type: *One* of the functions f_i in (8.1), say f_1, is replaced by another function, say f_1', while all the other relations in (8.1) remain unchanged. In general, this would change the whole system (8.2), and if we did *not* change the system (8.2) [e.g., because we did not know the original system (8.1)], some or all of its relations would show lack of constancy with respect to the observations that would result from the *new* structure. On the other hand, the last $m-1$ equations in (8.1) would—by definition—still hold good, unaffected by the structural change. It might be that, as a matter of fact, one or two particular equations in (8.1) would break down *very often*, while the others remained valid. Then any system (8.2) corresponding to a *fixed* system (8.1) would show little persistence with respect to the actual observations.

In this scheme the variables $x_{m+1}, x_{m+2}, \cdots, x_n$, were, in point of principle, free: they might move in any arbitrary way. This includes also the possibility that, e.g., all these free variables might move as certain well-defined functions of time, e.g.,

$$
\begin{aligned}
x_{m+1} &= g_1(t), \\
x_{m+2} &= g_2(t), \\
&\cdots\cdots \\
x_n &= g_{n-m}(t).
\end{aligned} \tag{8.3}
$$

As long as this should hold, we might be able to express the variables $x_1, x_2, \cdots, x_m$, as functions of $x_{m+1}, x_{m+2}, \cdots, x_n$, in many different ways. For example, it might be possible to express x_1 as a function of x_n, say

$$x_1 = F(x_n). \tag{8.4}$$

But could this relation be used to judge the effect upon x_1 of various *arbitrary* changes in x_n? Obviously not, because the very existence of (8.4) rests upon the assumption that (8.3) holds. The relation (8.4) might be highly unstable for such arbitrary changes, and the eventual persistence observed for (8.4) in the past when (8.3) held good, would not mean anything in this new situation. In the next situation the original system (8.1), or even system (8.2), would still be good, if we knew it. But to *find* such a basic system of highly autonomous relations in an actual case is not an analytical process, it is a task of making fruitful hypotheses as to how reality actually is.

We shall illustrate these points by two examples.

First we shall consider a scheme which, I think, has some bearing upon the problem of deriving demand curves from time series.

Let x be the rate of per capita consumption of a commodity in a group of people who all have equal money income, R. Let p be the price of the commodity, and let P be an index of cost of living. Assume that the following demand function is actually true:

$$x = a\frac{p}{P} + b\frac{R}{P} + c + \epsilon, \tag{8.5}$$

where a, b, c, are certain constants, and ϵ is a random variable with "rather small" variance, and such that the expected values of x are

$$E\left(x \,\middle|\, \frac{p}{P}, \frac{R}{P}\right) = a\frac{p}{P} + b\frac{R}{P} + c. \tag{8.6}$$

Assume that (8.5) is autonomous in the following sense: For any arbitrary values of p/P and R/P, the corresponding value of x can be estimated by (8.6). Suppose we are interested only in variations that

are small relative to certain constant levels of the variables. Then we may approximate (8.5) by a linear relation in the following way: Let p_0, R_0, and P_0 be the average values of p, R, and P respectively. Then we have

$$
\begin{aligned}
x &= a\frac{p_0 + (p - p_0)}{P_0 + (P - P_0)} + b\frac{R_0 + (R - R_0)}{P_0 + (P - P_0)} + c + \epsilon \\
&= a\frac{p_0 + (p - p_0)}{P_0}\cdot\frac{1}{1 + \dfrac{P - P_0}{P_0}} \\
&\quad + b\frac{R_0 + (R - R_0)}{P_0}\cdot\frac{1}{1 + \dfrac{P - P_0}{P_0}} + c + \epsilon \qquad (8.5') \\
&\simeq a\frac{p_0 + (p - p_0)}{P_0}\left(1 - \frac{P - P_0}{P_0}\right) \\
&\quad + b\frac{R_0 + (R - R_0)}{P_0}\left(1 - \frac{P - P_0}{P_0}\right) + c + \epsilon \\
&= \frac{a}{P_0}p - \frac{ap_0}{P_0^2}P + \frac{ap_0}{P_0} - \frac{a(p - p_0)(P - P_0)}{P_0^2} \\
&\quad + \frac{b}{P_0}R - \frac{bR_0}{P_0^2}P + \frac{bR_0}{P_0} - \frac{b(R - R_0)(P - P_0)}{P_0^2} + c + \epsilon.
\end{aligned}
$$

If the deviations $(p-p_0)$, $(P-P_0)$, and $(R-R_0)$ are small compared with p_0, P_0, and R_0, we may neglect product terms of these deviations. Then we obtain

$$x = Ap + BR + CP + D + \epsilon', \qquad (8.7)$$

where

$$A = \frac{a}{P_0},\quad B = \frac{b}{P_0},\quad C = -\left(\frac{ap_0}{P_0^2} + \frac{bR_0}{P_0^2}\right),\quad D = \frac{ap_0}{P_0} + \frac{bR_0}{P_0} + c,$$

and where ϵ' is a new residual term now also containing the errors made by the above approximation. For small variations of the variables, ϵ' may not be practically distinguishable from ϵ.

What we shall now show is that, if the data for p, P, and R, to be used for deriving the demand function have, for some reason or another, happened to move as certain regular functions of time, there may in *these* data exist *another* relation which has exactly the same *form* as (8.7), but different coefficients, and which may fit the data still better than (8.7) would do in general. And if we mistake this other relation for (8.7), we get merely a confluent relationship, and not an approximation to the demand function (8.5).

To see this let us write (8.5) as

$$x(t) = a\frac{p(t)}{P(t)} + b\frac{R(t)}{P(t)} + c + \epsilon(t). \tag{8.5''}$$

Assume now that the time functions $p(t)$, $P(t)$, and $R(t)$—for some reason—happen to be such that they satisfy the functional relations

$$\frac{p(t)}{P(t)} = k_1 p(t) + k_2 P(t) + k_0, \tag{8.8}$$

$$\frac{R(t)}{P(t)} = m_1 R(t) + m_2 P(t) + m_0, \tag{8.9}$$

where the k's and the m's are certain constants. A wide class of elementary time functions satisfy such functional equations. And whenever this is the case for the actual observations of p, P, and R, an equation of the form (8.7) could be fitted to the data. But we could not use the equation thus obtained for predicting the effect of an arbitrary price change, or an arbitrary income change, because this equation is not in general an approximation to (8.5) but merely a confluent result of (8.5), (8.8), and (8.9). It, therefore, does not hold, e.g., for price changes which violate (8.8), (8.9), or both.

In general, we have to be very careful in using a particular set of data to *modify* the form of relationships which we have arrived at on strong theoretical grounds. For example, in the case above we might be led to conclude that (8.7) might be a more correct "form" of the demand function than (8.5), or at least as good, while actually, when (8.8) and (8.9) are fulfilled, we may obtain a relationship of the form (8.7), which is not a demand function at all, and which breaks down as soon as $p(t)$, $P(t)$, and $R(t)$ take on some other time shape.

As an illustration to the question of autonomy of an economic relation with respect to a *change in economic policy*, let us consider the economic model underlying the famous Wicksellian theory of interest rates and commodity prices. (For the sake of simplicity and shortness we shall, however, make somewhat more restrictive assumptions than

Wicksell himself did. Our model does not do full justice to Wicksell's profound ideas.)

Consider a society where there are only three different economic groups: (a) individuals, (b) private firms, and (c) banks. We assume that: (1) All individuals divide their income into two parts, one part consisting of spending+increase in cashholding, the other part being saved, and all savings go into banks as (time) deposits. There is no other saving in the society. (2) All production in the society takes place in firms. The firms are impersonal organizations, guided in their production policy by profit expectations only. They can make new investments by means of bank loans only. They distribute all their profit to individuals. (3) Prices of goods and services of all kinds vary proportionally through time, and may be represented by a common variable, called the price level. (4) The banks have the power of expanding or contracting credit. We assume that there is only one money rate of interest, which is the same for all banks and the same for loans as for deposits. (This gives a rough description of the model we are going to discuss. It is hardly possible to give an exhaustive description of a model in words. The precise description is given implicitly through the relations imposed in the model.)

We are principally interested in the price effect of certain changes in the credit policy of the banks.

Let us introduce the following notations:

(1) $S(t)$ = total saving per unit of time,
(2) $I(t)$ = total investment per unit of time,
(3) $\rho(t)$ = bank rate of interest at point of time t,
(4) $P(t)$ = price level at point of time t,
(5) $R(t)$ = total national income per unit of time.

Now we shall introduce a system of fundamental relations describing the mechanism of our model. We consider linear relations, for simplicity.

First, we assume that there exists a market supply function for savings of the following form.

$$S(t) = a_0 + a_1\rho(t) + a_2P(t) + a_3\dot{P}(t) + a_4R(t). \tag{8.10}$$

This equation says that the supply of savings (bank deposits)—apart from a constant—depends upon the rate of interest, the total income, the price level, and the expectations regarding the future real value of money saved, as represented by the rate of change in the price level $\dot{P}(t)$. It might be realistic to assume that a_1 and a_4 are positive, a_2 and a_3 negative.

Next, we assume the following demand function for bank loans:

$$I(t) = b_0 + b_1\rho(t) + b_2P(t) + b_3\dot{P}(t), \tag{8.11}$$

where b_1 is negative and b_3 positive, while the sign of b_2 may be uncertain, a priori. b_3 would be positive because, when the price level is increasing, the firms expect to buy factors of production in a less expensive market than that in which they later sell the finished products, and this profit element is an inducement to invest.

Now, if the banks should lend to firms an amount equal to deposits, neither more nor less, i.e., if

$$I(t) = S(t), \tag{8.12}$$

then it follows from (8.10), (8.11), and (8.12), that to each value of $R(t)$, $P(t)$, and $\dot{P}(t)$, there would correspond a certain market equilibrium rate of interest, $\bar{\rho}(t)$, called by Wicksell the *normal* rate. That is, we should have

$$\begin{aligned}\bar{\rho}(t) &= \frac{b_0 - a_0}{a_1 - b_1} + \frac{b_2 - a_2}{a_1 - b_1}P(t) + \frac{b_3 - a_3}{a_1 - b_1}\dot{P}(t) - \frac{a_4}{a_1 - b_1}R(t) \\ &= A_0 + A_1P(t) + A_2\dot{P}(t) + A_3R(t),\end{aligned} \tag{8.13}$$

where $\bar{\rho}(t)$ is a value of $\rho(t)$ satisfying (8.10), (8.11), and (8.12), and where the A's are abbreviated notations for the coefficients in the middle term.

If the banks want, actively, to expand or contract currency (that is, if they want to change that amount of money outside the banks), they have to fix a rate of interest $\rho(t)$, which differs from $\bar{\rho}(t)$ as defined by (8.13). [Note that $\bar{\rho}(t)$ is by no means a constant over time.] From (8.10) and (8.11) we get

$$\begin{aligned}I(t) - S(t) = {} & (b_0 - a_0) + (b_1 - a_1)\rho(t) + (b_2 - a_2)P(t) \\ & + (b_3 - a_3)\dot{P}(t) - a_4R(t),\end{aligned} \tag{8.14}$$

which, for $\rho(t) = \bar{\rho}(t)$, reduces to

$$\begin{aligned}0 = {} & (b_0 - a_0) + (b_1 - a_1)\bar{\rho}(t) + (b_2 - a_2)P(t) \\ & + (b_3 - a_3)\dot{P}(t) - a_4R(t).\end{aligned} \tag{8.15}$$

Subtracting (8.15) from (8.14) we obtain

$$I(t) - S(t) = (b_1 - a_1)[\rho(t) - \bar{\rho}(t)], \tag{8.16}$$

which tells us that the amount of "money inflation," $I(t) - S(t)$, is (negatively) proportional to the difference between the actual bank rate of interest and the normal rate as defined by (8.13).

Assuming the "inflation" stream $I(t)-S(t)$ (taken as a barometer for total spending) to be accompanied by a proportional rise in the price level, we have

$$\dot{P}(t) = k[I(t) - S(t)] \qquad (k \text{ a positive constant}). \tag{8.17}$$

Combining (8.16) and (8.17) we obtain

$$\dot{P}(t) = k(b_1 - a_1)[\rho(t) - \bar{\rho}(t)], \tag{8.18}$$

which is a simplified expression for Wicksell's fundamental theorem about the price effect of a bank rate of interest that differs from the normal rate.

Accepting this theory (we are not interested in analyzing its actual validity any further in this connection, as we use it merely for illustration), what would be the degree of *autonomy* of the three equations (8.16), (8.17), and (8.18)?

Let us first consider the equation (8.16). Its validity in our set-up rests upon the two fundamental relations (8.10) and (8.11). In setting up these two equations we did not impose any restrictions upon the time shape of the functions $\rho(t)$, $P(t)$, and $R(t)$. Therefore, by hypothesis, *whatever* be the time shape of these functions, the corresponding time shapes of $I(t)$ and $S(t)$—and, therefore, also the time shape of $I(t)-S(t)$—follow from (8.10) and (8.11). [(8.16) is merely another way of calculating the difference $I(t)-S(t)$.] From (8.13) it follows that to each pair of time functions $P(t)$ [provided its derivative $\dot{P}(t)$ exists] and $R(t)$ there corresponds a time function $\bar{\rho}(t)$, while to each given time function $\bar{\rho}(t)$ there corresponds, in general, an infinity of time functions $P(t)$ and $R(t)$. The equation (8.16) is, therefore—by assumption—autonomous in the following sense: For any arbitrarily chosen time functions for $\rho(t)$ and $\bar{\rho}(t)$ the credit inflation $I(t)-S(t)$ can be calculated from (8.16).

We should notice that this property of (8.16)—if true—is not a mathematical property of the equation: it cannot be found by looking at the equation. It rests upon a hypothesis as to how the difference $I(t)-S(t)$ *in fact* would behave for various arbitrary changes in the interest rate and the normal rate. In another model we might obtain an equation of exactly the same form, but without the same property of autonomy. For example, assume that—as a consequence of some model, whatever be the particular economic reasoning underlying it—all the time functions above were bound to follow certain linear trends. In particular, suppose that we had $I(t)-S(t)=mt$, $\rho(t)-\bar{\rho}(t)=nt$. We should then have

$$I(t) - S(t) = \frac{m}{n}[\rho(t) - \bar{\rho}(t)], \tag{8.19}$$

which is of the form (8.16). But from (8.19) we could not calculate the effect upon $I(t)-S(t)$ of, say, various types of interest policy, because any changes in $\rho(t)$ that would violate the condition $\rho(t)-\bar{\rho}(t)=nt$ would break up the very foundation upon which (8.19) rests. The equation (8.19) *might* still hold after such a break, but that would have to follow from *another* model.

The equation (8.17) represents, per se, also an autonomous relation with respect to certain changes in structure. It is an independent hypothesis about the price level, saying that, whatever be the credit inflation $I(t)-S(t)$, we may calculate the corresponding rate of change in the price level. Here too, we cannot *know* how far this property of autonomy would in *fact* be true. It is an assumption, and it is a task of economic theory and research to justify it.

Let it be established that (8.16) and (8.17) are, in fact, highly autonomous relations. What is the situation with respect to the equation (8.18)? Obviously (8.18) would have a smaller degree of autonomy than either (8.16) or (8.17) separately, because the class of time functions satisfying (8.18) is—by definition—only the class of functions that satisfy (8.16) and (8.17) jointly.

So far we have not assumed any definite relations describing the credit policy of the banks. We have merely described the behavior of individuals and firms in response to a given bank rate of interest. Starting from certain assumptions as to the willingness to save and to invest, and assuming that an inflow of extra credit into the market causes a proportional change in the price level, we have obtained two structural relations (8.16) and (8.17). The variable $\rho(t)$ was considered as a free parameter. It might be, however, that the banks, over a certain period of time at least, choose to follow a certain pattern in their interest policy, or that they have to do so in order to secure their own liquidity. Over *this period* of time it might then be that we could add a new relation to the ones above, namely a relation describing—temporarily—the banking policy. Assume for instance, that the banks, over a certain period of time, act as follows: Whenever they realize that $I(t)-S(t)$ has become positive they start raising the interest rate, in order to protect their liquidity, and, conversely, they lower the rate of interest when they realize a negative $I(t)-S(t)$. Such a policy might be described by the relation

$$\dot{\rho}(t) = c[I(t) - S(t)], \tag{8.20}$$

where c is a positive constant. Because of (8.16) we have

$$\dot{\rho}(t) = c(b_1 - a_1)[\rho(t) - \bar{\rho}(t)]. \tag{8.21}$$

And combining (8.18) and (8.21) we have

$$\dot{P}(t) = \frac{k}{c}\,\dot{\rho}(t), \tag{8.22}$$

which apparently says that the price level moves in the *same* direction as the interest rate. But could we use this relation to calculate the "would-be" effect upon the price level of some arbitrary interest policy? Obviously *not*, because (8.22) holds only when $R(t)$, $I(t)$, $S(t)$, $P(t)$, $\rho(t)$, and $\bar{\rho}(t)$ are such time functions as satisfy, simultaneously, (8.13), (8.16), (8.17), and (8.20). Therefore, (8.22) is of no use for judging the effect of a change in interest policy. To obtain an equation for this purpose we might combine (8.13) and (8.18), which give a relation of the form

$$\dot{P}(t) + BP(t) = H_1\rho(t) + H_2R(t) + H_0, \tag{8.23}$$

where B, H_1, H_2, and H_0 are constants depending upon those in (8.13) and (8.18). Here there are—by hypothesis—no restrictions upon the time shape of the functions $\rho(t)$ and $R(t)$. We may choose such functions arbitrarily and solve the equation (8.23) to obtain $P(t)$ as an explicit function of $\rho(t)$ and $R(t)$.

But how could we know that (8.23) is the equation to use, and not (8.22)? There is no formal method by which to establish such a conclusion. In fact, by starting from another model with different assumptions, we might reach the opposite conclusion. To reach a decision we have to *know* or to imagine—on the basis of general experience—which of the two relations (8.22) or (8.23) would *in fact* be the most stable one if either of them were used as an autonomous relation.

* * *

To summarize this discussion on the problem of autonomous relations: In scientific research—in the field of economics as well as in other fields—our search for "explanations" consists of digging down to more fundamental relations than those that appear before us when we merely "stand and look." Each of these fundamental relations we conceive of as invariant with respect to a much wider class of variations than those particular ones that are displayed before us in the natural course of events. Now, if the real phenomena we observe day by day are really ruled by the simultaneous action of a whole system of fundamental laws, we see only very little of the whole class of hypothetical variations for which each of the fundamental relations might be assumed to hold. (This fact also raises very serious problems of estimating fundamental relations from current observations. This whole problem we shall discuss in Chapter V.) For the variations we observe, it is

possible to establish an infinity of relationships, simply by combining two or more of the fundamental relations in various ways. In particular, it might be possible to write one economic variable as a function of a set of other variables in a great variety of ways. To state, therefore, that an economic variable is "some function" of a certain set of other variables, does not mean much, unless we specify in what "milieu" the relation is supposed to hold. This, of course, is just another aspect of the general rule we laid down at the beginning of this chapter: The rule that every theory should be accompanied by a design of experiments.

CHAPTER III

Stochastical Schemes as a Basis for Econometrics

From experience we know that attempts to establish *exact* functional relationships between observable economic variables would be futile. It would indeed be strange if it were otherwise, since economists would then find themselves in a more favorable position than any other research workers, including the astronomers. Actual observations, in whatever field we consider, will deviate more or less from any exact functional relationship we might try to establish. On the other hand, as we have seen, the testing of a theory involves the identification of its variables with some "true" observable variables. If in any given case we believe, even without trying, that such an identification would not work, that is only another way of saying that the theory would be false with respect to the "true" variables considered. In order that the testing of a theory shall have any meaning we must first agree to identify the theoretical with the observable variables, and then see whether or not the observations contradict the theory.

We can therefore, a priori, say something about a theory that we think might be true with respect to a system of observable variables, namely, that it must *not exclude as impossible* any value system of the "true" variables that we have already observed or that it is practically conceivable to obtain in the future. But theories describing merely the set of values of the "true" variables that we conceive of as practically possible, would hardly ever tell us anything of practical use. Such statements would be much too broad. What we want are theories that, without involving us in direct logical contradictions, state that the observations will *as a rule* cluster in a limited subset of the set of all conceivable observations, while it is still consistent with the theory that an observation falls outside this subset "now and then."

As far as is known, the scheme of probability and random variables is, at least for the time being, the only scheme suitable for formulating such theories. We may have objections to using this scheme, but among these objections there is at least one that can be safely dismissed, viz., the objection that the scheme of probability and random variables is not general enough for application to economic data. Since, however, this is apparently not commonly accepted by economists we find ourselves justified in starting our discussion in this chapter with a brief outline of the modern theory of stochastical variables, with particular emphasis on certain points that seem relevant to economics.

9. *Probability and Random Variables*

The more recent developments in statistical theory are based upon the so-called modernized classical theory of probability. Here "probability" is defined as an absolutely additive and nonnegative *set-function*,[1] satisfying certain formal properties.[2]

Let us first take an example to illustrate this probability concept.

[1] See e.g. Stanislaw Saks, *Theory of the Integral*, New York, 1937; and Nicolas Lusin, *Les ensembles analytiques*, Paris, 1930.

We shall make frequent use of the following common notations and definitions from the theory of sets:

If A be a *set* of elements or objects, a, the symbol $a \in A$ is used to indicate that a is an element of A, or that a *belongs* to A.

Let (A) be a *family* of sets A, and let A_1 and A_2 be two members of (A). If every element of A_1 is also an element of A_2, we say that A_2 contains, or covers, A_1.

The symbol A_1+A_2 (called the logical sum of A_1 and A_2) indicates the set of all elements a which belong to *at least one* of the two sets A_1 and A_2. $A_1 \cdot A_2$ (called the logical product of A_1 and A_2) indicates the set of all those elements a which belong to *both* A_1 *and* A_2 (i.e., their common part). These notions of sum and product may be extended to any sequence of sets, finite or infinite.

If a product $A_1 \cdot A_2$ is empty, A_1 and A_2 are called *disjunct* sets.

If A_1 contains A_2, $A_1 - A_2$ is called the *difference* between A_1 and A_2, and denotes the set of elements that belong to A_1 but *not* to A_2.

A family of sets that is such that (1) the summation of any, at most denumerable, set of disjunct members of the family as well as (2) the subtraction $A_i - A_j$ of any two members where A_j is contained in A_i, give sets which belong to the family is called a *Borel corpus*. We denote it by $\{A\}$.

Suppose that we associate, with each member, A, of $\{A\}$, a finite number $F(A)$. Then $F(A)$ is called a *set*-function. (For example, if A be an interval on a straight line, its *length* is a set-function.) The function $F(A)$ is called *additive* if, for any arbitrary *disjunct* pair of sets A_i and A_j in $\{A\}$, we have

$$F(A_i + A_j) = F(A_i) + F(A_j).$$

$F(A)$ is called *absolutely* additive if, for any at most denumerable set of *disjunct* subsets $A_1, A_2, \cdots, A_n, \cdots$, in $\{A\}$, we have

$$F(A_1+A_2+\cdots+A_n+\cdots)=F(A_1)+F(A_2)+\cdots+F(A_n)+\cdots.$$

By the *measure* of a set A, belonging to a corpus $\{A\}$, we understand an absolutely additive set-function, $m(A)$, such that $m(A) \geqq 0$, and $m(A) = 0$ when A is empty. (Length, area, volume are simple examples of measures.)

[2] See e.g., J. Neyman, *Lectures and Conferences on Mathematical Statistics*, Washington, 1937, pp. 2–18; "L'estimation statistique traitée comme un problème classique de probabilité," *Actualités scientifiques et industrielles, 739, Conférence internationales de sciences mathématiques*, Paris, 1938, pp. 25–57; Paul Levy, *Théorie de l'addition des variables aléatoires*, Paris, 1937; S. S. Wilks, *Statistical Inference, 1936–37*, Princeton, N. J., 1937; *Mathematical Statistics*, Princeton, 1943.

Consider an ordinary die with six sides. For the purpose of probability calculus a die can be described as a set of 6 points on a straight line, $x=1, 2, \cdots, 6$. Consider now *all* the points on a straight line from $-\infty$ to $+\infty$. Over this set of points (i.e., over the whole real axis) we define a nonnegative real *measure-function* (or, a system of "weights") of the following type:

(1) To the point $x=1$ we ascribe a measure P_1, to the point $x=2$ we ascribe a measure P_2, etc., to the point $x=6$, finally, we ascribe a measure P_6, such that $P_i \geqq 0$, $i=1, 2, \cdots, 6$, and such that $P_1+P_2+\cdots+P_6=1$.

(2) If w be *any* subset of points (e.g., an interval) on the x-axis, the measure, $P(w)$, of the set w is defined as the *sum* of the measures, P_i, of those points, if any, among the 6 particular points $x=1, 2, \cdots, 6$, which belong to the set w. (For example, the measure of a set w defined by $1 \leqq x < 4$ would be $P_1+P_2+P_3$.)

(3) If w does not contain any of the points $x=1, 2, \cdots, 6$, then, for any such w, $P(w)=0$. [For example, if w is the interval $0 \leqq x \leqq \frac{1}{2}$, then $P(w)=0$.] $P(w)$, so defined, is called the *probability* that a point x belongs to the point-set w, or, for short, the probability of w. It follows that, if w is the whole real axis, then $P(w)=1$. If w contains just the point $x=1$, or $x=2$, or $\cdots$, or $x=6$, then $P(w)=P_1$, or P_2, or $\cdots$, or P_6 respectively.

Now let us consider n dice, Nos. 1, 2, $\cdots$, n, (or n hypothetical throws with the same die), *all* having the *same* system of probabilities $P_1, P_2, \cdots, P_6$. Let x_i be the result of one throw with the ith die, $i=1, 2, \cdots, n$ (i.e., $x_i=1$, or 2, or $\cdots$, or 6, with the probabilities $P_1, P_2, \cdots, P_6$, respectively, all other values of x_i having the probability zero). Consider any possible *system* $(x_1, x_2, \cdots, x_n)$ of values of the n variables x, one for each die. Any such sequence $x_1, x_2, \cdots, x_n$, can be represented by a *point* in n-dimensional Euclidean space. If we define the probability of any such point as the *product* of the probabilities of each of the x's *separately*, we may calculate the probability of an arbitrary point $(x_1, x_2, \cdots, x_n)$, or more generally, the probability of any arbitrary set of points in the n-dimensional linear space. It is easy to see that the system of *all* such probabilities satisfies conditions exactly similar to (1)–(3) above. The only difference is that we now consider points in n-dimensional space, instead of points on a straight line. For example, we might calculate the probability that exactly k (no matter which) out of the n variables x have the value 6, i.e., the probability of a point $(x_1, x_2, \cdots, x_n)$ having exactly k of its co-ordinates equal to 6. This probability is the sum of $n!/k!(n-k)!$ products, each equal to $P_6{}^k(1-P_6)^{n-k}$, or

$$\text{(9.a)} \qquad \frac{n!}{k!(n-k)!} P_6^k (1 - P_6)^{n-k},$$

which is, of course, also the probability of a *proportion* of "sixes" equal to k/n. From the formula (9.a) we may calculate the total probability of a *set* of points in the n-dimensional x-space, corresponding to a whole *system* of values of k, simply by summing up the probabilities (9.a) for these values of k. Hence we might also calculate, e.g., the probability, P, say, of $P_6 - \epsilon \leqq k/n \leqq P_6 + \epsilon$, where ϵ is any positive number. It follows from formula (9.a), as is well known, that if P_6 be a finite number, and if a positive ϵ be chosen, no matter how small, then P can be made as near to 1 as we please by choosing n sufficiently large.

What is the usefulness, if any, of such a purely formal apparatus, or, in other words, does it have any counterpart in the real world?

First of all, let us agree to assign a practical meaning to the theoretical notion "A probability near to 1." By this statement—when applied to real phenomena—we mean "practical certainty," that is, when we say—in the theory—that the probability of an event is near to 1, this means, in practical application, that we are "almost sure" that the event will actually occur.

Now let us apply this to the dice-example above. If the probability of a "six" be P_6 (not necessarily 1/6), then the probability calculus says that, when n is sufficiently large, the probability of a proportion k/n of "sixes" in n independent castings being near to P_6 is almost 1. Translated into practical language, this means: If we cast a die n_1 times, where n_1 is a large number, say $n_1 = 10{,}000$, and obtain a proportion k_1/n_1 of "sixes," then we are *practically sure* that in a new large number, n_2, of castings with this die, say $n_2 = 10{,}000$, the proportion k_2/n_2 of "sixes" will be near to k_1/n_1. Thus, for example, if we obtained $k_1/n_1 = 1/5$ for the first 10,000 castings, and, say, $k_2/n_2 = 2/5$ for the second 10,000 castings, we should be inclined to start investigations of the die and the casting procedure, because we should be almost sure, on the basis of a great many similar experiments in the past, that "something was wrong."

Purely empirical investigations have taught us that certain things in the real world happen only very rarely, they are "miracles," while others are "usual events." The probability calculus has developed out out of a desire to have a formal logical apparatus for dealing with such phenomena of real life. The question is not whether probabilities *exist* or not, but whether—if we proceed *as if* they existed—we are able to make statements about real phenomena that are "correct for practical purposes."

The above example may serve to illustrate the meaning of probabil-

ity, and of probability calculus. We shall now give a more general definition of probability.

Let A be a *set* (finite or infinite) of specified objects of any kind (e.g., a set of points in a certain region of space). Let $A_X = A \cdot A_X$ be a subset of A, consisting of all those elements of A which possess a certain property X among a *system* of properties X, such that the family of all the corresponding sets $A \cdot A_X$ form a Borel corpus $\{A \cdot A_X\}$, and such that $A \in \{A \cdot A_X\}$. Assume, furthermore, that we have defined a measure, $m(A \cdot A_X) \geqq 0$, within $\{A \cdot A_X\}$, such that $m(A) > 0$, and $m(A \cdot A_X) = 0$ when $A \cdot A_X$ is empty. The set A is then said to be *probabilized* (Neyman). A is called a *fundamental probability set.* For any element $A \cdot A_X$ of $\{A \cdot A_X\}$ we define

$$P(X \mid A) = \frac{m(A \cdot A_X)}{m(A)} \tag{9.1}$$

as *the probability of an element of A possessing the property X.* From the definition of a Borel corpus, and the definition of the measure $m(A \cdot A_X)$ it follows that

$$0 \leqq P(X \mid A) \leqq 1, \quad \text{and} \quad P(\overline{X} \mid A) + P(X \mid A) = 1,$$

where $\overline{X}$ is the property "*not X.*"

Any real variable, x, defined as a single-valued measurable *function of the elements in a probabilized set A*, is called a *random variable.* As a particular case $x = x^0 =$ constant may have the probability 1, while all other values of x have the probability 0. Then x is a constant in the stochastical sense. The values of x may be considered as properties of the elements of A.

A function, x, of the elements in the set A is measurable if the subset of A given by $x < c$ is measurable, in the probability measure defined, for every finite value of c. Therefore, whatever be the real numbers $c_1 < c_2$, the definitions of A and x determine uniquely the probability

$$P(c_1 \leqq x < c_2 \mid A). \tag{9.2}$$

And it is always possible to find c_1 and c_2 such that

$$0 < P(c_1 \leqq x < c_2 \mid A) \leqq 1. \tag{9.3}$$

For any fixed c_1, $P(c_1 \leqq x < c_2 \mid A)$ is a monotonically nondecreasing function of c_2, called the *integral probability law of x.*

The above definition of probability and random variables is practically equivalent to the following more direct definition: Let x be a real variable; its values can be represented by points on a straight line from $-\infty$ to $+\infty$. Let $\{w\}$ be a Borel corpus of measurable sets, w, on

this line, such that, in particular, $\{w\}$ contains the system of *all intervals* $c_1 \leqq x \leqq c_2$, where $c_1 < c_2$ may be any pair of real numbers. Let $P(w)$ be a *set-function* defined over $\{w\}$, such that $P(w)$ is (1) nonnegative, (2) absolutely additive, and (3) equal to 1 if w contains *all* points x from $-\infty$ to $+\infty$. Then this defines x as a random variable such that the probability of $(x \,\varepsilon\, w)$ is given by $P(w)$.

If there exists a nonnegative, Lebesgue-measurable function, $p(x)$, such that, for every interval (c_1, c_2) for which $P(c_1 \leqq x < c_2 \mid A)$ is defined, we have

$$P(c_1 \leqq x < c_2 \mid A) = \int_{c_1}^{c_2} p(x)dx, \tag{9.4}$$

where the integral is that of Lebesgue, then $p(x)$ is called the *elementary probability law* (or the probability density function) of x.

In statistics we usually have to consider *systems* of several random variables. There are two principal types of such systems, and—although they are not really different from the point of view of statistical methodology—the distinction between them helps when we want to compare a hypothetical model with actual observations.

The first type refers to a system of several random variables $x_1, x_2, \cdots, x_r$, associated with each element of a fundamental probability set. (For example, the fundamental probability set may be all persons who lived in the United States during the whole year 1940; x_1 may be personal income, x_2 may be private fortune, etc.) For each element of the fundamental probability set, the system of values $x_1, x_2, \cdots, x_r$, may be represented by a point, E_r, say, in r-dimensional space R_r. If w be any measurable set of points in R_r, we denote by

$$P(E_r \,\varepsilon\, w \mid A), \text{ or, for short, } P(w) \tag{9.5}$$

the probability that an arbitrary point E_r belongs to w. [In the following we shall use the shorter notation $P(w)$ in all cases where there is no danger of confusion as to what variable-space is considered.] $P(w)$, considered as a *function of the set* w, is called the *simultaneous integral probability law* of $x_1, x_2, \cdots, x_r$, within the fundamental probability set A.

It will be noticed that we use the same symbol P to indicate two different things, namely (1) a *number*, and (2) a *function*. If the argument, w, is a *fixed* set of points, w_0 say, then $P(w_0)$ means a number, namely the probability of w_0. If w is considered as an arbitrary, variable argument, then $P(w)$ means the probability *function*. We shall use particular letters or subscripts, etc., to indicate fixed sets in the variable-spaces in question, so no confusion can arise.

If there exists a nonnegative, Lebesgue-measurable function $p(x_1, x_2, \cdots, x_r)$, such that, for every w for which $P(w)$ is defined, we have

$$P(w) = \int\int_{(w)} \cdots \int p(x_1, x_2, \cdots, x_r) dx_1 dx_2 \cdots dx_r, \tag{9.6}$$

then $p(x_1, x_2, \cdots, x_r)$ is called the *joint elementary probability law* of $x_1, x_2, \cdots, x_r$.

Let $p_1(x_1)$, $p_2(x_2)$, $\cdots$, $p_r(x_r)$ be the elementary probability laws of the r variables x taken separately (i.e., the marginal distributions of the x's), within A. If then

$$p(x_1, x_2, \cdots, x_r) = p_1(x_1) \cdot p_2(x_2) \cdots p_r(x_r), \tag{9.7}$$

the variables $x_1, x_2, \cdots, x_r$, are said to be *stochastically independent.*

The second type of systems of random variables refers to *random sampling.* Suppose that we have a fundamental probability set, A, each element of which is characterized by the values of r random variables, $x_1, x_2, \cdots, x_r$. And suppose that we fix a certain *rule* by which to pick out a system of s elements from A. Let $(x_{11}, x_{21}, \cdots, x_{r1})$ denote the system of values of the first element picked, $(x_{12}, x_{22}, \cdots, x_{r2})$ that for the second element, and so forth. Let B_i denote the subset of A corresponding to *all a priori possible* value-systems $(x_{1i}, x_{2i}, \cdots, x_{ri})$ for the element to be picked as No. i $(i = 1, 2, \cdots, s)$. B_i may be considered as the fundamental probability set of the random variables $x_{1i}, x_{2i}, \cdots, x_{ri}$. The system

$$\begin{aligned} &(x_{11}, x_{21}, \cdots, x_{r1}), \\ &(x_{12}, x_{22}, \cdots, x_{r2}), \\ &\cdots\cdots\cdots \\ &(x_{1s}, x_{2s}, \cdots, x_{rs}), \end{aligned} \tag{9.8}$$

is called *a sample of size* s from the r-variate fundamental probability set (or "population") A, or, what amounts to the same thing, s samples of *one* observation each, namely one system of values $(x_1, x_2, \cdots, x_r)$ for each fundamental probability set B_i. The joint distribution of $(x_{1i}, x_{2i}, \cdots, x_{ri})$ may clearly change with i. The system (9.8) may also be considered as one sample of just *one* observation, namely one element picked from an rs-variate population, say B. Each element in B would then be characterized by a set of values of the rs random variables (9.8), and the probability distribution associated with B would be of rs dimensions. Each system of values (9.8) may be repre-

sented by a point, E, in rs-dimensional Euclidean space. Such a point, E, is called a *sample point*, or a point in the rs-dimensional *sample space*.

By random sampling we usually understand an experimental arrangement such that the various sets $E_i = (x_{1i}, x_{2i}, \cdots, x_{ri})$ $(i = 1, 2, \cdots, s)$, in (9.8) are mutually independent, i.e., such that, if the elementary probability laws exist,

$$p(E) = p_1(E_1) \cdot p_2(E_2) \cdots p_s(E_s). \tag{9.9}$$

The dependence or independence *within* each system $E_i = (x_{1i}, x_{2i}, \cdots, x_{ri})$ is usually "given by Nature."

When the (integral or elementary) probability law of a system of random variables is known, there are standard mathematical rules for deriving the probability laws of *functions* of these variables. (See, e.g., J. V. Uspensky, *Mathematical Probability*, New York, 1937.)

10. The Practical Meaning of Probability Statements

At the beginning of the preceding section we gave a simple illustration of the practical meaning of probability statements. We can now give a more general interpretation of such statements.

Suppose we should know that n observable variables $x_1, x_2, \cdots, x_n$, have the joint elementary probability law $p(x_1, x_2, \cdots, x_n)$. What are the practical statements we could make about a set of values $(x_1, x_2, \cdots, x_n)$ not yet observed? It has been found fruitful in various fields of research to use the observable "frequency of occurrence" of an event as a practical counterpart to the purely theoretical notion of probability. That is, if the elementary probability law p implies that the probability of a certain region or set, w say, in the n-dimensional x-space is $P(w)$, we take this to mean that by repeated actual observations of points $(x_1, x_2, \cdots, x_n)$ in the x-space the relative frequency of points falling into w would, for a very large number of points of observation, be close to $P(w)$.

However, as a rule we are not particularly interested in making statements about such a large number of observations. Usually, we are interested in statements that could be made about a relatively small number of observation points; or, perhaps even more frequently, we are interested in a practical a priori statement about just one single new observation. Then it is of relatively little practical value to know that $P(w)$ is, let us say, 0.4, 0.5, or 0.6. For then we cannot have much confidence, either in the statement that the next observation point will fall into w or in the statement that it will fall outside w. In order to be able to make a useful statement, the situation must be such that there exists an "interesting" subset w for which the probability $P(w)$ is near to 1; or, in practical interpretation, such that "nearly every"

observation will fall into w. Then we could say that it would be a "miracle" if, in particular, the next observation should fall outside w. That is, we should be almost sure that this would not happen. Experience has shown that the purely hypothetical notion of probability distributions is a useful tool for deriving such practical statements.

Above we considered "frequency of occurrence" as a practical counterpart to probability. But in many cases such an interpretation would seem rather artificial, e.g., for economic time series where a repetition of the "experiment," in the usual sense, is not possible or feasible. Here we might then, alternatively, interpret "probability" simply as a measure of our *a priori confidence* in the occurrence of a certain event. Also then the theoretical notion of a probability distribution serves us chiefly as a tool for deriving statements that have a very high probability of being true, the practical counterpart of which is that "we are almost sure that the event will actually occur."

Much futile discussion has taken place in regard to the questions of what probabilities actually are, the type of events for which probabilities "exist," and so forth. Various types of "foundations of probability" have been offered, some of them starting from observable frequencies of events, some appealing to the idea of a priori belief or to some other notion of reality. Still other "foundations" are of a purely formal nature without any reference to real phenomena. But they all have one thing in common, namely, that they end up with a certain concept of probability that is of a purely abstract nature. For in all the "foundations" offered the system of probabilities involved are, finally, required to satisfy some logical consistency requirements, and to have these fulfilled a price must be paid, which invariably consists in giving up the exact equivalence between the theoretical probabilities and whatever real phenomena we might consider. In this respect, probability schemes are not different from other theoretical schemes. The rigorous notions of probabilities and probability distributions "exist" only in our rational mind, serving us only as a tool for deriving practical statements of the type described above.

When we state that a certain number of observable variables have a certain joint probability law we may consider this as a construction of a rational *mechanism*, capable of producing (or reproducing) the observable values of the variables considered. When we have observed a set of values of n observable variables $(x_1, x_2, \cdots, x_n)$ we may, without any possibility of a contradiction, say that these n values represent a sample point drawn from a universe obeying *some* unknown n-dimensional (integral) probability law. Whatever be the a priori statement we want to make about the values of n observable variables, we can

derive this statement from one of several (perhaps infinitely many) suitably chosen n-dimensional probability laws. The class of all n-dimensional probability laws can, therefore, be considered as *a rational classification of all a priori conceivable mechanisms* that could rule the behavior of the n observable variables considered.

Since the assignment of a certain probability law to a system of observable variables is a trick of our own, invented for analytical purposes, and since the same observable results may be produced under a great variety of different probability schemes, the question arises as to which probability law should be chosen, in any given case, to represent the "true" mechanism under which the data considered are being produced. To make this a rational problem of statistical inference we have to start out by an axiom, postulating that every set of observable variables has associated with it one particular "true," but unknown, probability law. Since the knowledge of this true probability law would permit us to answer any question that could possibly be answered in advance with respect to the values of the observable variables involved, the whole problem of quantitative inference may then in each case be considered as a problem of gathering information about some unknown probability law.

11. *Random Variables and Probability Distributions in Relation to Economic Data*

Through experience we have learned much about the type of real phenomena to which the schemes of probability theory are most successfully applied. (Later, we shall show that the field of application for probability schemes is much more general than is indicated in this section.) These phenomena we group under the name "random experiments." We cannot give a precise answer as to what is a random experiment, because it is not an abstract concept, but only a name applied to certain real phenomena. But we may indicate some of the essential properties that we ascribe to such experiments. First, the notion of random experiments implies, usually, some hypothetical or actual possibility of "repeating the experiment" under approximately the "same conditions." Second, it is implied that such repetitions may give varying results. And third, the inferences we draw from random experiments are essentially of the type: How *often* does a certain result occur?

Does this description apply to economic data?

Here, I think, it is useful—though not always actually possible—to make a distinction between two different classes of experiments, namely, on the one hand, those we plan and perform ourselves, as research workers, to investigate certain facts *already present;* on the

other hand, the experiments which, so to speak, are products of *Nature*, and by which the facts *come into existence*. To bring out this distinction more clearly, let us consider an example.

Suppose we try to "explain" the size and the variations of consumption of a given commodity, A, in a society or group consisting of N individuals or families. What we usually mean by "explanation" in such a case is that we want to pick out certain other measurable factors, the variations of which—by hypothesis or by experience—might be expected to "*influence*" the behavior of each individual, or family, etc., *in the same way*. Suppose we have specified a certain number of such factors, in the present case, for instance, price of the commodity A, prices of other commodities, individual (or family) income, the age of the individuals, etc. Let there be, all together, n such specified factors, $x_1, x_2, \cdots, x_n$; and let the actual consumption of the commodity A for a given individual (or family) be denoted by y. We neglect for the moment the errors of observation due to lack of precision in the definitions of what y and the variables x represent, as well as imprecision due to errors of measurement proper. In other words, we deal here with "true" variables as described in Section 3.

Let us assume, tentatively, that, for each individual, we could "explain" his consumption of A by an equation, say

$$y^* = f(x_1, x_2, \cdots, x_n), \tag{11.1}$$

where y^*, for each individual, is obtained by inserting in the right-hand side of (11.1) those values of the influencing factors x that are relevant to him. However, if we do this for each individual, we shall find—no matter what be the fixed function f—that our "explanation" is incomplete. More specifically, we shall find that two individuals, or the same individual in two different time periods, may be confronted with exactly the same set of specified influencing factors x [and, hence, they have the same y^*, by (11.1)], and still the two individuals may have different quantities y, neither of which may be equal to y^*. We may try to remove such discrepancies by introducing more "explaining factors," x. But, usually, we shall soon exhaust the number of factors which could be considered as *common* to all individuals, and which, at the same time, were not merely of negligible influence upon y. The discrepancies $y - y^*$ for each individual may depend upon a great variety of factors, these factors may be different from one individual to another, and they may vary with time for each individual.

In a purely formal way we may replace y^* by y in (11.1) and, instead, add a general shift, s, to take care of the discrepancies between y and y^*, i.e.,

$$y = f(x_1, x_2, \cdots, x_n) + s. \tag{11.2}$$

Suppose, e.g., we should know or assume that, for each set of values of the variables x, s (and, therefore, y) is a random variable having a certain probability distribution with zero mean (say). What is the meaning of such a scheme?

Let us pick out a subgroup of individuals from the total group of N, such that, for each member of this subgroup, the factors x are identically the same. When, nevertheless, the quantities y for the members of this subgroup are different, it means that the decisions of the individuals, even after fixing the values of $x_1, x_2, \cdots, x_n$, are still to some extent uncertain. The individuals do not all act alike. When we assume that s has, for each fixed set of values of the variables x, a certain probability distribution, we accept the *parameters* (or some more general properties) of these distributions as certain additional characteristics of the theoretical model itself. These parameters (or properties) describe the *structure* of the model just as much as do the systematic influences of $x_1, x_2, \cdots, x_n$ upon y. Such random elements are not merely some superficial additions "for statistical purposes."

When we describe s as a random variable with a certain probability distribution for each fixed set of values of the variables x, we are thinking of a class of hypothetical, infinite populations, each of which is completely described by the scheme (11.1) and by the characteristics of the distributions of s. The total number of individuals, N, actually present may then be considered as a mixed sample consisting of subsamples drawn from members of the hypothetical class of populations. There is no logical difficulty involved in considering the "whole population as a sample," for the class of populations we are dealing with does *not* consist of an infinity of different individuals, it consists of an infinity of possible *decisions* which might be taken with respect to the value of y. And all the decisions taken by all the individuals who were present during one year, say, may be considered as one sample, all the decisions taken by, perhaps, the *same* individuals during another year may be considered as *another* sample, and so forth. From this point of view we may consider the total number of possible observations (the total number of decisions to consume A by all individuals) as result of a sampling procedure, which *Nature* is carrying out, and which we merely watch as passive observers.

It is on purpose that we have used as an illustration an example of individual economic behavior, rather than an average market relation. For it seems rational to introduce the assumptions about the stochastical elements of our economic theories already in the "laws" of behavior for the single individuals, firms, etc., as a characteristic of their be-

havior, and then derive the average market relations or relations for the whole society, from these individual "laws." It will then, for example, in many cases be possible to show that, even under very weak assumptions about the distributions of the stochastical elements in these individual relations, the derived average or total relations for the whole market or the whole society will be characterized by certain compound stochastical variables (e.g., sums of individual error terms) which, by the laws of large numbers, will be approximately *normally* distributed.

As active research workers we may produce another type of random experiments. For instance, in the example above we might pick out, by some random process, a subgroup of all individuals actually present, and measure their y's and x's. From this subgroup we might draw inference as to the behavior of the whole group. But the connection between such a subgroup and the total group that we *might* have observed is different from that between this total group of individuals (or decisions) present and the hypothetical class of infinite populations from which the total group present is supposed to be drawn; for the first connection is, essentially, dependent upon our own choice of the random sampling procedure to be used. By choosing another random process we get another connection. And we might here gradually remove all possible sampling errors by increasing the size of the sample, so that, finally, we should obtain a *true picture* of the *sample of all* individuals present. But the uncertainty in the correspondence between this sample of all individuals and the hypothetical class of infinite populations still remains. *One* problem is to construct hypothetical probability models from which it is possible, by random drawings, to reproduce samples of the type given by "Nature." *Another* problem is to make exact measurements of these samples. The first task is essentially one of economic theory. The second is one of statistical observation technique and "classical" sampling theory. Of course, *after* the stochastic schemes have been chosen, there is no essential difference between the problems of statistical inference they present.

12. The Method of Splitting the Observable Variables into "Systematic Parts" and "Disturbances"

Observable economic variables do not satisfy exact relationships (except, perhaps, some trivial identities). Therefore, if we start out with such a theoretical scheme, we have—for the purpose of application—to add some stochastical elements, to bridge the gap between the theory and the facts. One much-discussed way of doing this is to adopt the convention that the observable variables considered are each made up

of two parts, viz., a *systematic* part which, by assumption, satisfies the exact relation considered, and an error part, or "disturbance," of a stochastical nature.[3]

Let $x_1', x_2', \cdots, x_n'$ be n theoretical variables satisfying, by assumption, a certain exact functional relationship. And let $x_1, x_2, \cdots, x_n$, be the corresponding observable variables to be considered. We then write $x_i = x_i' + x_i''$, $i = 1, 2, \cdots, n$, where the variables x'' are certain stochastical variables. In order that our relation between the variables x' should also tell something about the observable variables x we have to make certain additional assumptions about the distribution of the variables x''. Then our exact relation between the variables x' becomes in fact a stochastical relation in the variables x and x'', by substituting $x_i - x_i''$ for x_i'.

It is important to notice, however, that such a splitting of the variables is necessarily of a relative nature, depending on the particular system of theoretical equations with which we are concerned.

This can be brought out rather well by means of a theoretical illustration.

Consider for this purpose three ordinary dice, one black, one red, and one white, and let us perform the following series of experiments: First, we cast all three dice. We obtain as result three numbers, say x_b for the black die, x_r for the red, and x_w for the white. Let the *sum* of these three numbers be $X = x_b + x_r + x_w$. Next, we let the black die remain in its position from the first casting (of all three dice), but we cast again both the red and the white one. Let the result of this experiment be y_b $(= x_b)$, y_r, and y_w, and let $Y = y_b + y_r + y_w$. Now, finally, we let both the black and the red dice remain untouched, but we cast the white one again. Let the result of this experiment be z_b $(= y_b = x_b)$, z_r $(= y_r)$, and z_w, and let $Z = z_b + z_r + z_w$. Assume that we repeat this whole experiment N times. We obtain three series

$$
\begin{array}{ll}
X_1,\ Y_1,\ Z_1 & \text{(1st experiment)},\\
X_2,\ Y_2,\ Z_2 & \text{(2nd experiment)},\\
\cdots\cdots & \\
X_N,\ Y_N,\ Z_N & (N\text{th experiment}).
\end{array}
\tag{12.1}
$$

From the set-up of these experiments it is evident that the three series X, Y, Z, are correlated, because they have some common compo-

[3] This scheme is, e.g., the basis for Frisch's method of "Confluence Analysis." See Ragnar Frisch, *Statistical Confluence Analysis by Means of Complete Regression Systems*, Oslo, 1934. See also T. Koopmans, *Linear Regression Analysis of Economic Time Series*, Haarlem, De Erven F. Bohn N. V., 1937.

nents. Indeed, for any triple, say X_i, Y_i, Z_i (the result of the ith experiment), we have

$$\begin{aligned} X_i &= x_{bi} + x_{ri} + x_{wi}, \\ Y_i &= x_{bi} + y_{ri} + y_{wi}, \qquad (i = 1, 2, \cdots, N). \\ Z_i &= x_{bi} + y_{ri} + z_{wi}, \end{aligned} \tag{12.2}$$

Suppose now that we want to study the interdependences between the three variables X, Y, Z, separating as "disturbances" those factors which are not "common causes." From (12.2) we derive

$$\begin{aligned} Y_i - (y_{ri} + y_{wi}) &= X_i - (x_{ri} + x_{wi}), \\ Z_i - (y_{ri} + z_{wi}) &= X_i - (x_{ri} + x_{wi}), \qquad (i = 1, 2, \cdots, N), \\ Z_i - (z_{wi}) &= Y_i - (y_{wi}), \end{aligned} \tag{12.3}$$

where the expressions in brackets indicate "disturbances." The composition of the disturbances clearly depends upon *which relation* we are investigating. And to neglect this would make inefficient theory.

This schematic set-up has, I think, some relevance to many important problems in economics. E.g., let X, Y, and Z represent results of decisions taken in some economic planning. Then the scheme above may be looked upon in the following way: First X is determined by some considerations, which we do not investigate in this connection. Once this decision is taken, the decision Y is no longer quite free, it is "influenced" by X. But there are also other factors determining Y that have nothing to do with X, namely y_r and y_w. These factors, however, which act as disturbances in Y with respect to the "cause" X, are themselves partly systematic "causes" with respect to the decision Z after Y is chosen.

Let us consider an example from economic dynamics: The interrelation between investment and profit. Let $v(t)$ denote observed investment activity (per year) at point of time t, and let $z(t)$ be observed profit. Assume there are no errors made in registering these quantities. We make the following hypotheses: Investment activity at t depends upon profit realized at some previous time, say at $(t-\theta)$, while profit at t depends upon current investment at t. Letting $\epsilon_1(t)$ and $\epsilon_2(t)$ denote certain general random shifts, we may express these hypotheses by

$$v(t) = f[z(t - \theta)] + \epsilon_1(t), \tag{12.4}$$

$$z(t) = g[v(t)] + \epsilon_2(t), \tag{12.5}$$

where θ is positive, and where f and g are certain functions. Now it may be that, in (12.4), we have to allow for a considerable disturbance, $\epsilon_1(t)$, in $v(t)$ as compared with *that* part of $v(t)$, [namely $v(t) - \epsilon_1(t)$],

which is "explained" by $z(t-\theta)$. But this does *not* mean that only *this* part of $v(t)$ influences $z(t)$ through (12.5) [i.e., that we could replace $v(t)$ by $v(t)-\epsilon_1(t)$ in (12.5)]. Most certainly the *actual* investment [i.e., $v(t)$] has a more direct bearing upon the profit $z(t)$ than our hypothetically constructed "systematic part" of it [namely $v(t)-\epsilon_1(t)$].

The occurrence of such situations has very important consequences for the problem of *linking together conclusions* drawn from different relationships, as we shall see in the next section.

13. *Stochastic Equations versus Exact Equations*

The statement: "A set of variables satisfies a certain equation," has a different meaning according as it is applied to an abstract mathematical scheme or to variables observed in real life.

In an abstract mathematical scheme the statement means the following: Let $x_1', x_2', \cdots, x_n'$, be n real variables. Each set of values of these n variables may be represented by a point in n-dimensional Cartesian space. Let us denote by S the set of *all* points in this space, and let "A" be a rule by which to pick out a certain subset of points, S_A, of S. Let us *exclude* all points of S which do *not* belong to S_A. Then, if a function f exists that is not identically zero but is such that

$$f(x_1', x_2', \cdots, x_n') = 0 \tag{13.1}$$

for *all points belonging to* S_A, we say that the variables $x_1', x_2', \cdots, x_n'$ (the variations of which are limited by the rule "A") have the property of satisfying the equation $f=0$. Here the whole set S_A is given by definition through a logical operation A, and we may check whether the statement in (13.1) is right or wrong.

Similar statements about variables observed in real life are of a much more hypothetical character. When we make statements of the type (13.1) about a set of observable variables, say $x_1, x_2, \cdots, x_n$, we assume, so to speak, that Nature has a rule for picking out such observation points $(x_1, x_2, \cdots, x_n)$ in the x-space in such a way that none of these points contradict the hypothesis (13.1) when the variables x' are replaced by the variables x. We then say that (13.1) is *a law of Nature*. We try to establish such laws by testing the truth of (13.1) with respect to *past* observations. But even if they all satisfy (13.1), we cannot *know* that the *next* observation will do so. We usually, however, *think* that it will, because we have an enormous record of empirical cases showing that such empirical inductions have actually been very fruitful. At the same time, we have also learned that, in empirical research, it is useful to replace the expression "a set of variables satisfies a certain equation" by the expression "satisfies approximately"

such an equation. This means that, if we insert observation points $(x_1, x_2, \cdots, x_n)$ in the left-hand side of (13.1), we obtain, on the right-hand side, a certain *variable*, s.

Then—as we have already discussed above—if such an expression as "satisfies approximately" shall have a nontrivial meaning, we must change the hypothesis (13.1) in such a way that it expresses *what kind of approximation* we assume. One way of doing this is to change the hypothesis (13.1) to

$$f(x_1, x_2, \cdots, x_n) = s, \tag{13.2}$$

and ascribe to s certain general properties which should not be contradicted by data. We are particularly interested in such schemes as ascribe to s certain general properties of a *random variable*, first, because we have a large record of empirical cases showing that such schemes have been successfully applied to observed phenomena, and, secondly, because the theory of such schemes has been more developed than any other approximation schemes. And we find justification for applying them to economic phenomena also in the fact that we usually deal only with—and are interested only in—total or average effects of many individual decisions, which are partly guided by common factors, partly by individual specific factors (see Section 11).

In case s is assumed to be a random variable, we say that the variables x satisfy a *stochastic* equation (13.2). This is, of course, only a very particular type of stochastic equations. Here we have not "blamed" any particular element in our scheme for the fact that the observed variables $x_1, x_2, \cdots, x_n$, do not satisfy (13.1) exactly. We may operate with other schemes specifying more in detail *where* the stochastic elements come in. In general, we may lay down the following definition: If $x_1, x_2, \cdots, x_n$, be a set of observational variables, and if $\epsilon_1, \epsilon_2, \cdots, \epsilon_m$ be m random variables, and if a function, F, not identically zero, exists, such that for all observations

$$F(x_1, x_2, \cdots, x_n; \epsilon_1, \epsilon_2, \cdots, \epsilon_m) = 0, \tag{13.3}$$

then $x_1, x_2, \cdots, x_n$, are said to satisfy a stochastic equation. Thus, a stochastic equation in n variables may be an exact equation in $n+m$ variables.

Suppose that our observation material consists of $N \gg n$ points in the n-dimensional space of the variables x, and suppose that we ascribe to the joint probability distribution of $\epsilon_1, \epsilon_2, \cdots, \epsilon_m$, certain properties a priori. Now we insert, successively, the N observation points for the variables x in (13.3), and for each observation point we choose a

set of values of the ϵ's such that (13.3) is fulfilled. Thus, we get a sample of N points in the m-dimensional Cartesian space of the ϵ's. On the other hand, by ascribing a priori certain properties to the probability distribution of the ϵ's, and by excluding the possibility of obtaining certain samples of the ϵ's which *then* are "improbable" (in some sense or other, a question to be discussed later), we have set a probability limit to the subset of *admissible* samples of the ϵ's. Let this set of admissible sample points for the ϵ's be Q. Then we may say that, if the N observation points in the x-space are such that—under the condition (13.3)—it is possible to choose a sample of N sets of ϵ's which belong to the set Q, then we cannot reject the hypothesis that the n variables $x_1, x_2, \cdots, x_n$, satisfy the stochastic equation (13.3).

From a stochastic scheme of the form (13.3) we may derive certain *exact* equations, not containing the random variables ϵ, by giving one or more of the variables x a *new interpretation*. There are two important different types of such derived exact equations. The first type could be called "if-there-were-no-errors equations," the second, "expected-value equations."

The first type is obtained by assigning to the random variables ϵ in (13.3) certain *constant* values. In most cases we should formulate the stochastic equation in such a way, if possible, that these constant values of the ϵ's would be zero. Then, of course, if we require that

$$F(x_1, x_2, \cdots, x_n; 0, 0, \cdots, 0) = 0 \tag{13.4}$$

we impose a condition upon the variables x which, in most cases, will be violated by actual observations. Therefore, if (13.4) is imposed, one or more of the variables x must stand for—not what they actually are—but what they *would be* "if there were no errors." This kind of simplified exact equations, therefore, represents a hypothetical *correction* of the *individual observation points* in the x-space.

The second type of "exact" equations, on the other hand, represents *average* relations in a *group* of observations. Here we *do not simplify* the original scheme, but we confine ourselves to studying certain stochastic limit properties of the scheme. We may illustrate the difference by a simple example.

Consider a group of families of equal size and composition. Let r be family income, and let x be family spending, during a certain period of time. Assume all prices constant and the same for all families during this period. Still, among those families who have the *same* income, the amount spent, x, will vary from one family to the other, because of a great many neglected factors. Let us assume that the spending habits

of an infinite population of such families could be described by the following stochastic equation

$$\log_e x = k \log_e r + k_0 + \epsilon \qquad (k \text{ and } k_0 = \text{constants}), \tag{13.5}$$

where ϵ is a random variable, normally distributed with zero mean and variance $= \sigma^2$. From this stochastic scheme we may derive the following two "exact" equations:

First, let us imagine that we could, somehow, remove the forces which cause the discrepancies ϵ. In this hypothetical population all families with the same r would act alike, and we should have

$$\log_e x = k \log_e r + k_0. \tag{13.6}$$

Secondly, let the "errors" ϵ remain in the scheme, but consider only the average or *expected* consumption for those families who have the same income r. This gives

$$E(x \mid r) = \bar{x}(r) = e^{k_0} \cdot r^k \int_{-\infty}^{+\infty} e^{\epsilon} \frac{1}{\sqrt{2\pi}\sigma} e^{-\epsilon^2/2\sigma^2} d\epsilon = e^{k_0 + \frac{1}{2}\sigma^2} \cdot r^k, \tag{13.7}$$

where $E(x \mid r)$ means: Expected value of x, *given* r.

Therefore, what the *average* family in the scheme (13.5) does is *not* necessarily the same as what the families would all do if they acted alike.

It is particularly important to be aware of the difference between these two types of relations when we want to perform *algebraic operations* within stochastic equation systems. For instance, from the theoretical scheme (13.6) we may derive

$$x = e^{k_0} r^k. \tag{13.8}$$

But from $E(\log_e x \mid r) = k \log_e r + k_0$ we do *not* get $E(x \mid r) = e^{k_0} r^k$.

Therefore, when we perform such operations, we must keep in mind that we are using the hypothetical "if-there-were-no-errors" scheme, and *not* the "expected-value" scheme. Confusion on this point arises usually from the habit of dropping the operation symbol E (or the bar over x, etc.) in such equations as (13.7). Confusion arises in particular when we have a *system* of stochastic equations and apply algebraic elimination processes to the corresponding "expected-value" equations. The usual mistake here is that we identify the expected values of a variable in *one* equation with the expected values of the same variable in *another* equation. This may lead to nonsensical results. The following is an illustration:

Let x_1, x_2, x_3 be three observable variables, defined by $x_1 = \epsilon_1 + \epsilon_2$, $x_2 = k_1\epsilon_1$, and $x_3 = k_2\epsilon_2$, where ϵ_1 and ϵ_2 are two *in*dependent random variables with zero means. Then we have

$$(13.9) \qquad x_1' = E(x_1 \mid x_2) = \frac{1}{k_1} x_2, \quad x_1'' = E(x_1 \mid x_3) = \frac{1}{k_2} x_3.$$

Now, if we identify (by mistake) the two variables x_1' and x_1'', denoting them both by $\bar{x}_1$ say, we get $x_2 = (k_1/k_2)x_3$, which has *no meaning*.

CHAPTER IV

The Testing of Hypotheses

Statisticians have, often with much right, argued that the economists do not present their theories in such a form that these theories represent well-specified statistical hypotheses, and that, therefore, the statisticians simply do not "understand the language" of the economists. The economists, however, are not the only ones to be blamed. Indeed, the whole statistical theory was, until rather recently, in a state of much confusion. But this situation is now disappearing rapidly, through a very fruitful change of direction brought about by the fundamental work of J. Neyman and E. S. Pearson.[1] By introducing a few very general—and, in themselves, very simple—principles of testing statistical hypotheses and estimation, they opened up the way for a whole stream of high-quality work, which gradually is lifting statistical theory to a real scientific level. The working out of technical details on the basis of the general principles introduced by Neyman and Pearson is still only in its beginning. And very difficult technical problems are likely to arise. But the fundamental importance of the Neyman-Pearson principles lies in the fact that these principles specify clearly the *class of problems* that fall within the field of statistical theory and statistical inference. Thus, it has now become possible for the economist to see exactly *how* he has to formulate his theories if he wants the assistance of a statistician. It is of the greatest importance that the economist himself should know these principles of formulation, for then, even if he is not himself a statistical expert, he can at least ask intelligent statistical questions.

In the following we shall give a brief outline of the basic principles in the Neyman-Pearson theory of testing statistical hypotheses and estimation, and, thereafter, we shall use these principles for a *statistical formulation of hypotheses constructed in economic theory*. This will, it is hoped, clear up a few controversial issues in connection with the problem of statistical "verification" of economic relations.

14. An Outline of the Neyman-Pearson Theory of Testing Statistical Hypotheses and Estimation

Let $x_1, x_2, \cdots, x_n$, denote n random variables defined within a fundamental probability set A. And let $P(E_n \, \varepsilon \, w_n | A)$, or, for short, $P(w)$, be their joint integral probability law.

[1] See, in particular, *Statistical Research Memoirs*, Vol. I, 1936, Vol. II, 1938, London. Other references are given in the following text.

Any tentative statement, H, concerning the integral probability law $P(w)$ of the variables $x_1, x_2, \cdots, x_n$ [or concerning their elementary probability law, $p(x_1, x_2, \cdots, x_n)$ if this is assumed to exist], is called a *statistical hypothesis*. More precisely, let Ω_n, or for short, Ω, denote the set, or class, of *all possible* n-dimensional integral probability laws, and let ω be any specified *subset* of Ω (ω may, e.g., be the set of all n-variate normal distributions, or the set of all n-variate continuous distributions, or any other subset of Ω). A statement of the form

$$P(w) \,\varepsilon\, \omega \tag{14.1}$$

(read: The integral probability law of $x_1, x_2, \cdots, x_n$ *belongs to* the class ω) is called a statistical hypothesis.

The statement (14.1) might be *wrong*, and then the alternative is that

$$P(w) \,\varepsilon\, (\Omega - \omega). \tag{14.2}$$

Above, the only thing assumed to be *known for certain* was that $P(w) \,\varepsilon\, \Omega$, which is trivial. Usually, however, we know—or at least we *assume* that we know—more than this. Let Ω^0 denote a subset of Ω. And let ω^0 be any subset of Ω^0. If, on the one hand, we *know* or *assume* that the statement

$$P(w) \,\varepsilon\, \Omega^0 \tag{14.3}$$

is *true*, while, on the other hand, we admit that for any subset $\omega^0 \neq \Omega^0$, the statement

$$P(w) \,\varepsilon\, \omega^0 \tag{14.4}$$

may be wrong, then Ω^0 is called *the set of a priori admissible hypotheses* with respect to the probability law $P(w)$. (For example, Ω^0 might be the set of all n-dimensional probability laws for which the *elementary* probability law exists, and ω^0 might, e.g., be the set of all probability laws the elementary probability laws of which are symmetric about the mean.) In problems of testing a statistical hypothesis the specification of the set of a priori admissible hypotheses is, as we shall see, of fundamental importance.

A statistical hypothesis is called *simple* if it *specifies completely* the probability law $P(w)$. E.g., the statement

$$P(w) = \int\int_{(w)} \cdots \int \frac{1}{(\sqrt{2\pi}\,\sigma)^n} e^{-(1/2\sigma^2)\sum_1^n (x_i - \bar{x}_i)^2} dx_1 dx_2 \cdots dx_n, \tag{14.5}$$

where $\bar{x}_i$ $(i = 1, 2, \cdots, n)$ and σ are numerically specified constant parameters, is a simple hypothesis. Any hypothesis that is not simple is called *composite*. For example, if the value of the parameter σ or

some of the means $\bar{x}_i$ or all together are *not* uniquely specified, then (14.5) is a composite hypothesis.

A set of admissible hypotheses, Ω^0, is called *parametric*, if all the probability laws $P(w)$ belonging to Ω^0 are given by analytic expressions which differ from each other only with respect to the numerical values of a finite number of parameters. For example, all probability laws (14.5) such that $\bar{x}_i > 0$, $i = 1, 2, \cdots, n$, form a parametric set. A set which is not parametric is called nonparametric. If Ω^0 is parametric then the set ω^0 must be parametric. But if Ω^0 is nonparametric, ω^0 [in (14.4)] may or may not be parametric.

A *test* of a statistical hypothesis is a *rule* of *rejection* or *nonrejection* of the hypothesis, on the basis of a given *sample point*. Let $x_1, x_2, \cdots, x_n$, be n random variables, and let Ω^0 be the set of *all a priori admissible* hypotheses about their simultaneous integral probability law $P(w)$. For any particular member of the set Ω^0, and for any particular subset, w, of points in the sample space R_n, we might calculate the probability that a sample point, E, falls into w. If w be *fixed*, the probability of E falling into w ($= w^0$ say) will generally vary according to which member of Ω^0 is used to calculate it. What is an "improbable" part of the sample space with respect to one probability law in Ω^0 may be a more probable one for another probability law in Ω^0. And this fact, of course, forms the basis for testing any particular hypothesis within Ω^0 against the other a priori admissible ones.

Much controversy is found on this point in earlier literature, in particular because it was thought that a reasoning back from a sample point to its true population would involve the notion of "inverse probability." One often finds expressions such as "the most probable distribution" from which a given sample may have been drawn. Such a statement, of course, implies a certain probability distribution of the *hypotheses* within Ω^0. In most cases, however, such a model does not have much sense, because, when we draw a sample, we take it from a *fixed but unknown* member of Ω^0. The probability of any member of Ω^0 being the true one, i.e., the one we sample from, is, therefore, either 0 or 1, independent of what be the sample point obtained.

On the other hand, if we establish a rule by which to reject or not reject a hypothesis, and if the decision is made to depend uniquely upon the location of a sample point, we may speak of the probability of our *decision* being *right* or *wrong*, because the decision—being a function of the sample point E—is then a random variable.

Let us now formulate more precisely what is a test of a statistical hypothesis. Let Ω^0 be the set of all a priori admissible hypotheses as to the probability law $P(w)$ of the n random variables $x_1, x_2, \cdots, x_n$, and let $P(w) \in \omega^0$, where ω^0 is a subset of Ω^0, be the hypothesis, H_0,

to be tested. This means: We know for certain that, whatever be the sample point observed, the true probability distribution of the n random variables is one and only one member (fixed, but so far unknown) of the set Ω^0, and our hypothesis is that $P(w)$ belongs to a more restricted set of distributions, ω^0, within Ω^0. The class ω^0 may contain only one single member (a simple hypothesis) or several members (a composite hypothesis). In the last case all members of ω^0 are treated as equivalent, we are not interested in distinguishing between them.

Now, let W_0 be a set of points in the n-dimensional sample space R_n, such that, whenever a sample point falls into W_0, i.e., $E \,\varepsilon\, W_0$, we *reject* the hypothesis H_0, otherwise not. W_0 is then called a *critical region* (or more generally a critical set of points) for testing the hypothesis H_0, i.e., $P(w) \,\varepsilon\, \omega^0$, against the alternatives $P(w) \,\varepsilon\, (\Omega^0 - \omega^0)$. A critical region and a test are evidently just two different names for the same thing.

In particular cases a test of a hypothesis H_0 might be *decisive*, namely in cases where there exists a subset $\overline{W}_0$ of the sample space which has probability $=1$ according to H_0, but probability $=0$ according to any *other* member of Ω^0. Then, by means of one single sample point, we could decide—with a probability $=1$ of being correct—whether H_0 were true or false, by rejecting H_0 if and only if $E \,\varepsilon\, (R_n - \overline{W}_0)$. Also, suppose that the set Ω^0 of hypotheses H could be divided into a system of k *disjunct* subsets $\Omega_1^0, \Omega_2^0, \cdots, \Omega_k^0$, corresponding, one-to-one, with k nonoverlapping subsets $W_1, W_2, \cdots, W_k$, of the sample space, such that $P(W_i \mid H \varepsilon \Omega_j^0)$ be $=0$ when $i \neq j$, but $=1$ when $i=j$, $(i, j = 1, 2, \cdots, k)$. Then one single sample point would, at once, restrict the set of a priori admissible hypotheses to one of these subsets Ω_i^0. Such cases, although important, are trivial from the point of view of statistical theory. We may, therefore, assume the set Ω^0 to be so reduced in advance, that any subset, W, of the sample space having probability $=1$ according to *one* member of Ω^0, has also a positive probability with respect to all other members of Ω^0. The application of a test as defined above will then always involve some risk of erroneous decisions.

Now, if the region of rejection W_0 should be the whole sample space R_n (or the whole space minus a part of it that has probability zero according to any member of Ω^0), then we should always (or almost always) reject H_0. This is evidently not what we want, because when we desire to test H_0, we imply that it might be correct, and in that case the test would constantly lead to wrong decisions. On the other hand, if $P(R_n - W_0 \mid H_0)$ and $P(W_0 \mid H_0)$[2] be both positive, we usually run a *two-way* risk of making an erroneous decision by the test.

[2] We recall that the general symbol $P(X \mid Y)$ means: The probability of X *given* Y, or, the probability of X calculated under the assumption that Y is true.

First, suppose that the hypothesis is actually *true* and, at the same time, the sample point *does* fall into W_0 (which is—by assumption—possible). Then we *reject* H_0, hence, we make an error. This is called an error of the *first kind.*

Second, suppose that the hypothesis is actually *wrong* (i.e., one of the alternative hypotheses is the true one), and, at the same time, the sample point does *not* fall into W_0. Then we *do not reject* H_0, hence, we make an error. This is called an error of the *second kind.*

For any given size of the sample we can make the probability of *one* or the *other* of these errors as small as we please, by an appropriate choice of W_0, but it is not possible to do so for both errors at the same time. We therefore have to make a compromise, depending upon the kind of *risk* we are willing to run, and this, again, depends upon the consequences which erroneous decisions may have in any particular case.

The whole problem of testing statistical hypotheses, and also that of estimation, consists of deducing "best critical regions" W_0, on the basis of certain *risk parameters*, which, themselves, are given by some outside considerations, and are taken as *data* in the statistical theory. We shall now indicate briefly the Neyman-Pearson approach to the solution of this problem. The fundamental principles of this approach rest upon the distinction between the two kinds of errors described above, a distinction suggesting itself by recognizing the simple fact that, when we desire to test a hypothesis, we imply that it *might be wrong*, and that, therefore, it is necessary to specify in *what sense* it might be wrong. The recognition and precise formulation of such elementary—apparently almost trivial—principles is often among the very greatest achievements of scientific thought.

Let us first consider the simple case when ω^0 consists of only one single probability distribution, say P_0, and let the set $\Omega^0 - \omega^0$ also contain just one single element, say $P_1 \neq P_0$. We want to test, on the basis of a sample point E, the hypothesis H_0, that the true probability distribution is P_0, the only alternative being that it is P_1. Let W_0 be a critical region such that the probability $P(W_0 | P_0)$ is exactly equal to α (say $\alpha = 0.05$). α is called the *level of significance*, or also, the *size* of the critical region W_0, and is an a priori chosen risk parameter. It tells us that, if we choose W_0 as a critical region for rejecting the hypothesis H_0, the probability that we shall reject the hypothesis when it is *true* (i.e., the probability of error of the first kind) is exactly equal to α. But there are in general many such different regions W_0 of the same size α. Now, if the hypothesis is *not* true, i.e., if the true distribution is P_1, we want, of course, to have as great a probability as possible of *rejecting* the hypothesis H_0, i.e., we want the probability $P(W_0 | P_1)$ of E falling

into W_0 when P_1 is true, to be as great as possible. This probability $P(W_0|P_1)$ is called the *power* of the test W_0 with respect to the alternative P_1. Let W_0^* be *that* region of size α for which this power is a maximum. We then obviously want to use this region W_0^* as our critical region, rather than any other region of size α. W_0^* is then called the *best* critical region for testing $P=P_0$ with respect to the alternative $P=P_1$.

Suppose now that we enlarge the set $\Omega^0-\omega^0$ to comprise a whole system of alternative probability distributions. Then, if W_0^* above is at the same time the best critical region for testing $P=P_0$ with respect to *every* element of the set of alternative hypotheses, W_0^* is called a *uniformly most powerful test.* In a few important cases it can be shown that such regions exist. But this holds only for certain types of hypotheses tested against certain restricted sets of alternatives. And if no such test exists, we have to choose some critical region of size α which is "as powerful as possible" with respect to the set of alternative hypotheses in question. And the choice of a "best" test will then be somewhat more subjective. It might be that we have in mind a certain system of *weights of importance* for the errors of the second kind, for the various elements in the set of alternative hypotheses. For example, if the hypothesis to be tested is that a certain parameter, θ, in a probability distribution (the *form* of which is known) is equal to a specified value, θ^0, the possible alternatives being all other values of θ from $-\infty$ to $+\infty$ say, it might be that, for some reason, we should consider it more important to detect the alternatives $\theta>\theta^0$ than the alternatives $\theta<\theta^0$. The problem of introducing such weight functions has been discussed by A. Wald.[3]

Above we have assumed that the hypothesis to be tested was a *simple* one, but the general idea is readily extended to composite hypotheses, although the technical difficulties of deriving critical regions of the type discussed here become more serious. Even the problem of determining regions W_0 that have the *same size* for *every member* of the set ω^0 to be tested may here present complicated mathematical problems, and sometimes *no* such region exists.[4]

Whatever be the principles by which we choose a "best" critical region of size α, the essential thing is that a test is always developed with respect to a *given fixed set* of possible *alternatives* Ω^0. If, on the basis of some general principle, a "best" test, or region, W_0' say, is developed for testing a given hypothesis $P \in \omega^0$ with respect to a set, Ω^0, of a

[3] A. Wald, "Contribution to the Theory of Statistical Estimation and Testing Hypotheses," *Annals of Mathematical Statistics*, Vol. 10, December, 1939, pp. 299–326.

[4] See, e.g., W. Feller, "Note on Regions Similar to the Sample Space," *Statistical Research Memoirs*, Vol. II, London, 1938, pp. 107–125.

priori admissible hypotheses, and if we shift the attention to *another a priori admissible set*, Ω', also containing ω^0, the same general principle will, usually, lead to *another* "best" critical region, say W_0''. In other words, if a test is developed on the basis of a given set of a priori admissible hypotheses, Ω^0, the test is, in general, valid only for this set, Ω^0. By extending the set of admissible hypotheses to include *new* alternatives *without changing* the critical region, one can always find alternatives such that, whatever be the fixed critical region chosen, its power with respect to some of the new alternative hypotheses is very poor. This is a more precise expression for such common phrases as: "What is the use of testing, say, the significance of regression coefficients, when, maybe, the whole assumption of a linear regression equation is wrong?" This is just the type of arguments we have discussed above. Usually, when we test the significance of regression coefficients, the alternative set of hypotheses, Ω^0, is only the system of regression equations of the *same form*, but with regression coefficients that are different from zero. Ω^0 does not include other *forms* of regressions (although this might very well be done).

In general, if a critical region W_0 for a given hypothesis H_0 be developed on the basis of a set, Ω^0, of a priori admissible hypotheses, and if the *true* hypothesis—instead of belonging to Ω^0 as assumed—actually belongs to $\Omega - \Omega^0$ (i.e., the set complementary to Ω^0), we have *lost the control* of errors, originally ascribed to the test. It might, of course, be that the power of the test, even with respect to these hypotheses "off the scheme," is still good, i.e., when one of these new alternatives is true instead of the hypothesis tested, the probability of the sample point falling into W_0 might be high. But this probability might also be very small, even smaller than α, which means that we should have an even *smaller* probability of rejecting the hypothesis tested when it is *wrong* than when it is *correct*.

The requirement of a specification of the set of a priori admissible hypotheses *before* constructing a test forces us to state explicitly what we assume known beyond doubt, and what we desire to test.

* * *

The problem of *estimation* is the problem of drawing inference, from a sample point, as to the probability law of the fundamental probability set from which the sample was drawn. The problem of estimation is closely connected with the problem of testing statistical hypotheses, in fact, estimation may be considered as a particular form of testing hypotheses.

Let $x_1, x_2, \cdots, x_n$, be n random variables with the (unknown) probability law $P(w)$. Let it be known that $P(w)$ belongs to a *parametric*

class of distributions, Ω^0, i.e., $P(w)$ is known *except* for the values of a certain finite number of parameters, $\theta_1, \theta_2, \cdots, \theta_k$, say. We may write this as $P(w) = P(w \mid \theta_1, \theta_2, \cdots, \theta_k)$, or, for short, $P(w \mid \theta)$, where the *function* P is known. A sample, E, is drawn from one of the members of Ω^0, but we do not know from which. The problem is to draw inference from E regarding the corresponding values of the parameters θ. Let these true unknown values be $\theta_1^0, \theta_2^0, \cdots, \theta_k^0$. Any system of values of the parameters θ may be represented by a point, θ, in the *parameter space*, i.e., a k-dimensional Euclidean space, where the axes represent the k parameters θ. The problem of estimation is to define a function which *associates every point, E,* in the sample space with a well-defined *set* of points θ in the *parameter* space. If this function is such that to each point E in the sample space there corresponds one and only one point θ in the parameter space, we speak of *point-estimation.* If, to each point E in the sample space, the estimation formula ascribes a *region* $I(E)$ [or more generally a set of points $I(E)$] in the parameter space, we speak of *interval-* (or set-) estimation. In the particular case of point-estimation $I(E)$ contains only one point θ for each E.

The interval (or set) $I(E)$ is, clearly, a random set, because it is a function of the sample point E. We may, therefore, speak of the *probability,* β say, of a set $I(E)$ *covering* the true parameter point θ^0, and we may choose the value of β according to the amount of risk we are willing to take, say $\beta = 0.95$. Since we do not know the true parameter point θ^0, β ought to be *independent* of θ^0, i.e., whatever be the true parameter point θ^0 of the distribution from which we draw the sample, the probability $P(\theta^0 \in I \mid \theta^0)$[5] should be the same. β is called the *confidence coefficient* for the estimate of θ^0, and the corresponding $I(E)$ is called a *confidence interval* (or, more generally, a confidence set) for the true parameter point.

Now consider the set of all a priori admissible parameter points corresponding to Ω^0. This set of parameter points may be considered as the set of *all simple hypotheses* contained in Ω^0, i.e., all hypotheses $\theta = \theta^0$, where θ^0 may be *any* point among the a priori admissible set of parameter points. (We now consider θ^0 as a variable point.) Assume that for *every simple* hypothesis $\theta = \theta^0$, in the a priori admissible set Ω^0, we construct, by some principle, *a "best" critical region* $W_0(\theta^0)$ of size α, as described above. $W_0(\theta^0)$ is the region (or set) of rejection of $\theta = \theta^0$. $R_n - W_0(\theta^0)$ is, therefore, the region of nonrejection or, for short, the region of acceptance of $\theta = \theta^0$, and its size is $1 - \alpha$. Let $1 - \alpha = \beta =$ the confidence coefficient for estimating the parameter point by means of a sample point. Let E_1 be any arbitrarily fixed sample point. Since we

[5] When using the notation $\theta^0 \in I$ we should remember that θ^0 is the constant element, while I is the random variable.

assume that the true hypothesis is contained in Ω^0, it is reasonable to require that, in our system of regions of acceptance, $R_n - W_0(\theta^0)$, there should be *at least one* region such that E_1 belongs to it. In general, E_1 will be an element of a whole system of regions of acceptance. Consider *all* the regions of acceptance of size β, of which E_1 is a member. To each region of acceptance, $R_n - W_0(\theta^0)$, corresponds a point in the *parameter* space, namely the point θ^0 representing the hypothesis θ^0 for which $W_0(\theta^0)$ is a region of rejection. To the system of *all* the regions of acceptance of which E_1 is a member, there corresponds, therefore, a *set* of parameter points, say $I(E_1)$. Since E_1 was arbitrary and, therefore, might be any point E in the sample space, this defines a function $I(E)$ for every E. This $I(E)$ clearly has the properties of a confidence set for estimating the parameters θ by means of a sample point E, because, whatever be the true parameter point θ^0, the probability that a sample point E falls into its corresponding region of acceptance is $1-\alpha=\beta=$ constant, and *whenever* E falls into the region of acceptance for $\theta=\theta^0$, then also $\theta^0 \in I(E)$. The probability that $I(E)$ covers the true parameter point, no matter what this is, is therefore equal to β.

The estimation problem may be formulated more generally. Let $x_1, x_2, \cdots, x_n$, be n random variables with the probability distribution $P(w)$, about which it is known only that it belongs to a certain a priori admissible set, Ω^0, of distribution functions. Ω^0 may be considered as the set of all a priori admissible *simple* hypotheses. For each of these simple hypotheses let there be constructed a certain region of acceptance, U, of size β, and let (U) be the family of all such regions corresponding to the set Ω^0. A sample point E_1 is given. Let $[U(E_1)]$ be the family of all those regions of acceptance of which E_1 is a member, and let $I(E_1)$ be the set of all simple hypotheses (contained in Ω^0) which correspond to the system of regions $[U(E_1)]$. Since E_1 might be any point E, there corresponds an $I(E)$ to every E. $I(E)$, thus defined, is a confidence set with the confidence coefficient β, i.e., the probability that $I(E)$ will contain the true member of Ω^0, no matter what this is, is equal to β.

15. General Formulation of the Problem of Testing Economic Relations

The Neyman-Pearson theory of testing statistical hypotheses is purely abstract, like any other theoretical scheme. The question which interests us here is therefore: Does this scheme represent a useful instrument by which to deal with the problem of verifying economic theories? Can it help us to understand better the nature of these problems, and to reach practical solutions of them? I think these questions may be

answered very much in the affirmative. The following discussion will, it is hoped, support this view.

We shall attempt to give a general, axiomatic, formulation of the problem of testing economic relations, using principles of the Neyman-Pearson theory.

A. Data relevant to econometric research

The objects of economic research are variations and covariations within groups of phenomena of economic life. Let $K_1, K_2, \cdots, K_n$, be such a group. K_1 may, e.g., mean a certain type of consumption goods, K_2 may denote the phenomenon called "price" of K_1, etc. Each K is just the *name* of a certain *category* of real phenomena conceived of as more or less equivalent, and distinct from those in other categories. Many kinds of variations and shifting conditions may unfold themselves within each such category. We are here interested in only such variations as are shown by a certain *measurable characteristic* of each K. Let these n measurable characteristics be denoted by $x_1, x_2, \cdots, x_n$, respectively, and let $(x_{1t_i}, x_{2t_i}, \cdots, x_{nt_i})$ be a set of values observed *jointly* for the n K's, t_i indicating "observation at point of time t_i," or simply observation No. i ($t_1, t_2, \cdots$ etc., need not be equidistant). Let

$$\begin{aligned} &(x_{1t_1}, x_{2t_1}, \cdots, x_{nt_1}), \\ &(x_{1t_2}, x_{2t_2}, \cdots, x_{nt_2}), \\ &\cdots\cdots\cdots\cdots \\ &(x_{1t_N}, x_{2t_N}, \cdots, x_{nt_N}), \end{aligned} \tag{15.1}$$

be a system of N such joint observations. Each column in (15.1) represents a series of measurements of "the same variable," e.g., a time series.

B. Fundamental assumption about the nature of economic data

The nN values $(x_{1t}, x_{2t}, \cdots, x_{nt})$, $t=t_1, t_2, \cdots, t_N$, in the system (15.1) of N value-sets, may be considered as a sample point E in the nN-dimensional sample space of nN random variables $(x_{1t}, x_{2t}, \cdots, x_{nt})$, $t=t_1, t_2, \cdots, t_N$, with a certain joint integral probability law $P(w)$. (w denotes an arbitrary point-set in the nN-dimensional sample space.) What this assumption means is the following: Consider the situation *before* the sample (15.1) was drawn, i.e., consider the system (15.1) as *nN empty cells*. And consider the whole set of alternative systems, each of nN elements, which, *a priori*, *might* fill the nN cells. The above assumption amounts to assuming—as a fact or by a hypothetical construction—that, *before* the sample was drawn, there was a set of such

systems satisfying the requirements of a fundamental probability set as defined in Section 9. This assumption is extremely general, as is seen from the definition of a random variable in Section 9.

It is indeed difficult to conceive of any case which would be contradictory to this assumption. For the purpose of testing hypotheses it is not even necessary to assume that the sample could actually be repeated. We make hypothetical statements before we draw the sample, and we are only concerned with whether the sample rejects or does not reject an a priori hypothesis. The above assumption covers also, as a particular case, the situation where, for certain cells in (15.1), there would actually be just *one* fixed system of numbers that could fill these cells, i.e., the case where—for some of the cells in (15.1)—certain fixed values of the corresponding x's have probability $=1$ (i.e., they are stochastically constant). This is of importance in many economic problems where some of the variables are considered as autonomously given.

C. The formulation of a theoretical stochastic scheme

There are two kinds of abstract schemes occurring in economic theory, namely, one type which we introduce merely as a matter of exercise in logical reasoning or as a model of an idealized economy (i.e., schemes for which a comparison with reality has no meaning), and another type which—although abstract—we think may have some bearing upon real economic phenomena. For our study here only the latter is relevant.

In constructing schemes of this latter type we nearly always have some real phenomena in mind, and we try to include in the scheme—in a simplified manner, of course—certain characteristic elements of reality. At the same time we realize that such schemes can never give a complete picture of reality. We must allow for certain discrepancies. In Chapter III we discussed how a stochastic scheme might be used for this purpose. Because of the very general definition of random variables, stochastic schemes represent an extremely general class of theoretical models. We shall, therefore, assume that the problem of testing economic relations consists in confronting certain specified stochastic models with a set of data (15.1).

Let

$$
\begin{array}{c}
x'_{1t_1},\ x'_{2t_1},\ \cdots,\ x'_{nt_1},\\
x'_{1t_2},\ x'_{2t_2},\ \cdots,\ x'_{nt_2},\\
\cdots\cdots\cdots\\
x'_{1t_N},\ x'_{2t_N},\ \cdots,\ x'_{nt_N},
\end{array}
\tag{15.2}
$$

denote a system of *theoretical* random variables to be compared with the corresponding observed variables in (15.1).

Further, let

$$\begin{matrix} \epsilon_{1t_1},\ \epsilon_{2t_1},\ \cdots,\ \epsilon_{mt_1}, \\ \epsilon_{1t_2},\ \epsilon_{2t_2},\ \cdots,\ \epsilon_{mt_2}, \\ \cdots\cdots\cdots\cdots \\ \epsilon_{1t_N},\ \epsilon_{2t_N},\ \cdots,\ \epsilon_{mt_N}, \end{matrix} \tag{15.3}$$

be another system of mN random variables introduced in the theoretical scheme as *auxiliary random parameters*, possessing certain specified joint distribution properties. (The ϵ's may also be introduced as counterparts to some real phenomena. See Section 11.)

Finally, let

$$\alpha_1,\ \alpha_2,\ \cdots,\ \alpha_k \tag{15.4}$$

be a set of constants.

Now we impose a system of restrictions,

$$\begin{aligned} f_{t_i}[&x'_{1t_i},\ x'_{1t_{i-1}},\ \cdots,\ x'_{1t_1};\ x'_{2t_i},\ x'_{2t_{i-1}},\ \cdots,\ x'_{2t_1};\ \cdots; \\ &x'_{nt_i},\ x'_{nt_{i-1}},\ \cdots,\ x'_{nt_1};\ (X_0);\ \alpha_1,\ \alpha_2,\ \cdots,\ \alpha_k; \\ &\epsilon_{1t_i},\ \epsilon_{2t_i},\ \cdots,\ \epsilon_{mt_i}] = 0, \qquad (i = 1, 2, \cdots, N), \end{aligned} \tag{15.5}$$

upon the quantities (15.2)–(15.4). Here f_{t_i} is a specified function for each value of i, $i=1, 2, \cdots, N$. (In particular all the f's might be the same, independent of t; then only the arguments of the function would change.) (X_0) is a short symbol for a *set* of *initial conditions*, i.e., the values of x_{jt} $(j=1, 2, \cdots, n)$, for $t=t_0, t_{-1}, t_{-2}, \cdots$. Such quantities may or may not enter into (15.5). If they do, we assume them to be *constants* having *known* values.

(15.5) is, for each point of time, $t=t_1, t_2, \cdots, t_N$, a stochastical relation, defining, implicitly, one of the variables, say x_{1t_i}', as a function of

(1) the *previous* values of that same variable,
(2) the simultaneous *and* the previous values of the *other* variables x',
(3) m random variables ϵ.

Let (15.5) be our economic theory to be tested, the random variables ϵ having certain prescribed distribution properties. The principal task of economic theory is to make a fruitful choice of the *forms* f.

In this general formulation, (15.5) with its associated assumptions about the ϵ's may represent a *static or* a *dynamic* theory. Assume, as above, that each equation (15.5) can be solved for x'_{1t_i}, $i=1, 2, \cdots, N$. The theory is then static if (1) only variables x' for the *same* point of time t_i enter into each of the equations (15.5), and, at the same time, (2) the $n-1$ random variables $x'_{2t_i}, x'_{3t_i}, \cdots, x'_{nt_i}$, and the m random variables $\epsilon_{1t_i}, \epsilon_{2t_i}, \cdots, \epsilon_{mt_i}$ $(i=1, 2, \cdots, N)$, are assumed to be sto-

chastically independent of the *previous* values of the variables x' and the *previous* values of the variables ϵ. Otherwise the theory is dynamic in the sense that "what happens at point of time t_i depends upon what happened previously."

(15.5) is, of course, an empty statement about the variables (15.2) *unless* we know something about the random variables ϵ in *addition* to (15.5), for—whatever be the variables x'—we could define such variables ϵ that (15.5) would be fulfilled. We must make some additional statement (however weak) about the properties of the joint *conditional* probability law of all the variables ϵ for *given values* of the $(n-1)N$ "independent" variables, which we assumed to be $x'_{2t}, x'_{3t}, \cdots, x'_{nt}$ $(t=t_1, t_2, \cdots, t_N)$. When that is done, it follows from (15.5) that the joint probability law of *all* the variables x' in (15.2) can not be just *any* distribution, it must belong to a (more or less) *restricted* class of probability laws.

As an example, suppose that (15.5) were of the form

$$x'_{1t_i} - \alpha_1 x'_{2t_i} - \epsilon_{1t_i} = 0 \qquad (i = 1, 2, \cdots, N), \tag{15.5'}$$

and suppose that the variables ϵ were assumed to be distributed *independently* of the variables x_{2t_i}'. And let $p_1(\epsilon_{1t_1}, \epsilon_{1t_2}, \cdots, \epsilon_{1t_N})$ be the joint elementary probability law of the N variables ϵ. Then it follows that, for *given* values of the variables x'_{2t_i}, the variables x'_{1t_i} have the joint elementary probability law $p_1[(x'_{1t_1}-\alpha_1 x'_{2t_1}), (x'_{1t_2}-\alpha_1 x'_{2t_2}), \cdots, (x'_{1t_N}-\alpha_1 x'_{2t_N})]$. And hence, whatever be the elementary probability law, p_2 say, of all the variables x'_{2t_i} themselves, the *joint* elementary probability law, p_3 say, of the $2N$ variables x' must have the form $p_3 = p_1 \cdot p_2$.

Thus, (15.5) together with any additional assumption made as to the distribution properties of the ϵ's, will imply that the nN-dimensional probability law of the nN random x' must belong to a certain restricted *subclass*, ω say, of the class of all possible nN-dimensional probability laws. At the same time, this is also, clearly, *all that our theory implies*, so far as possible observations of the *variables* x' are concerned. [The equations (15.5) say, of course, much more about the variables x' *and* the variables ϵ taken together, but—by assumption—there is no possibility of observing individual values of the ϵ's.] Now, if we add a new system of nN equations, namely, $x = x'$, i.e., if we identify each theoretical variable x' in the system (15.2) with the corresponding observed variable in (15.1), our theory leads to a statistical hypothesis, namely, the hypothesis that $P(w) \in \omega$. We shall formulate this a little more in detail.

D. The formulation of (15.5) *as a statistical hypothesis with respect to the probability law of the observable variables* (15.1)

Let $\mathcal{E}$ denote a point in the mN-dimensional sample space of the variables ϵ in (15.3). And let $D(\mathcal{E}\ \mathfrak{e}\ v)$, where v is the argument of the set-function D, denote the joint *conditional* integral probability law of the mN variables ϵ, *given* the values of the $(n-1)N$ variables $x'_{2t}, x'_{3t}, \cdots, x'_{nt}$ $(t=t_1, t_2, \cdots, t_N)$ (the "independent variables"). This distribution is at our disposal in formulating the theory. It, therefore, belongs—by hypothesis—to a certain set, S say, of mN-dimensional probability laws. In case we have specified the distribution D of the variables ϵ completely in our theory, S contains only one element.

We shall consider the general case where the values of the parameters α in (15.5) are not fixed by theory, but are at our disposal, i.e., we are prepared to accept any values of the α's. Then the definition of S, and the restrictions (15.5), define a certain class, ω say, of probability laws of the variables x'. This class ω we could imagine to be obtained by the following process:

Consider one single member D of the system S, and consider *all possible* joint distributions of the variables x', *subject to the restrictions* (15.5), for an arbitrarily fixed system of values of the α's. Repeat this process for (1) all possible value-systems of the parameters α *and* (2) for every member of the system S. *All* the joint probability laws of the variables x' obtained in this way together form the class ω.

We are interested in whether $P(w)$, i.e., the joint probability law of the nN *observable* variables x, belongs to ω. The hypothesis to be tested is, therefore,

(15.6) $\quad P(w)\ \mathfrak{e}\ \omega$; admissible alternatives: $P(w)\ \mathfrak{e}\ (\Omega - \omega)$;

where Ω is the set of *all* nN-dimensional probability laws.

This formulation of the problem of testing economic relations is very general. In order to develop nontrivial tests it is, however, necessary to impose further restrictions upon the sets Ω and ω (in particular, by restricting the set S of conditional probability laws of the random variables ϵ). We shall mention some important types of restrictions of the sets Ω and ω.

(1) Restriction of the random variables ϵ to variables following certain simple probability laws, or restriction of the system S to a certain parametric family of distributions, or even to one perfectly specified distribution.

(2) Restriction of the set of a priori admissible hypotheses to such a set, Ω^0, as the ω defined above, i.e., to the set of all probability dis-

tributions that are compatible with (15.5) for *at least one* system of values of the parameters α, and then restriction of the set of probability laws *to be tested* to a particular subset, ω^0 say, of this Ω^0, corresponding to one fixed system of values of the parameters α (e.g., test of significance). This means that we are sure—or that we accept without test—that the theory (15.5) is all right so far as the forms of the functions f are concerned.

(3) Restrictions imposed upon the variables x' by some *other* relationships in the economic theory *besides* (15.5). This is very often the case when we consider *systems* of economic relations, and it *must* be taken account of in formulating the set ω^0 above.

An interesting and important question in this connection is the following: Is a test of the hypothesis (15.6) also a test of the "correctness" of the form of the f's in (15.5)?

First of all, what *is* a "correct" system of functions f_t? A precise answer can be given to this question, namely: *Any* system of functions f, which is such that $[P(w)] \in \omega(f_{t_1}, f_{t_2}, \cdots, f_{t_N})$, where $\omega(f_{t_1}, f_{t_2}, \cdots, f_{t_N})$, or, for short, $\omega(f)$, denotes the set ω (or ω^0) corresponding to that system of f's, is a correct system of functions f. There will, therefore, in general be an infinity of "correct" theories (15.5). In particular, there might be various different systems of f's which—together with various assumptions about the distribution properties of the ϵ's—all lead to *identically the same* set of probability laws ω, i.e., *they are indistinguishable from the point of view of observations.* This, of course, does not mean that all "correct" forms of theories are equally *good*, or "interesting," e.g., for prediction purposes. The "goodness" of a stochastical relation, if it be a "correct" one, will in general be judged from the properties of the random variables ϵ which it contains. Usually we want these errors to be "small," in some sense or another.

Now, let $\omega(f^0, S)$ be a set of probability laws of the variables x', defined by a particular system, f^0, of functions in (15.5) and a set S of ϵ-distributions. Then, if a test W_0 of the hypothesis $P(w) \in \omega(f^0, S)$ should have *high power* with respect to *every* alternative *not* contained in $\omega(f^0, S)$, the test W_0 would, of course, also have a high power of detecting, in particular, a wrong choice of the forms f^0.

If we try, however, to test a hypothesis (15.6), the alternatives being, so to speak, "everything else" (i.e., the set of a priori admissible hypotheses is Ω), then, no matter what be the test chosen, there will always within this "everything else" be alternatives for which the power of the test is very poor. In case one of these alternatives were actually the true one, we should have only a very slight chance of rejecting the hypothesis tested. In all practical cases it is, therefore, necessary to be able to restrict, in advance, the set of admissible hy-

potheses Ω^0 as much as possible, having at the same time strong reasons to believe that the true hypothesis is not outside this Ω^0.

* * *

We have not here gone into any technical details as to the actual construction of tests, the theory of which was described briefly in Section 14. Our purpose has been to show how an economist should formulate testing problems for which he asks the help of a statistician. To give a more concrete illustration, however, we shall in the next section consider a simple, but rather important, example from economic statistics, namely the problem of testing a time series for *trend*, assuming that its additional variations are random variables of a simple type.

16. Example of Testing Hypotheses: A Simple Problem of Trend Fitting

Let y_t be an observable time series, where $t = 1, 2, \cdots, N$, denote N equidistant, discrete points of time. Suppose we *know*, or *believe without test*, that the following *model* (where E means "expected value of") is *true:*

$$y_t = kt + b + \epsilon_t \qquad (t = 1, 2, \cdots, N), \tag{16.1}$$

$$E(y_t \mid t) = kt + b \qquad (t = 1, 2, \cdots, N), \tag{16.2}$$

$$E(\epsilon_t) = 0, \quad E(\epsilon_t^2) = \sigma^2 \text{ (independent of } t), \tag{16.2'}$$

$$p(y_t \mid t) = \frac{1}{\sqrt{2\pi}\,\sigma} e^{-(1/2\sigma^2)(y_t - kt - b)^2}. \tag{16.3}$$

σ is assumed to be numerically known (for the sake of simplicity of our illustration in the following).

Consider N populations (or universes) corresponding to the N fixed values $1, 2, \cdots, N$, of t. For each t, y_t is normally distributed about the mean $(kt+b)$ with variance σ^2. For each value of t we assume that we draw exactly one value of y_t, such that these drawings are stochastically *independent*. The sampling distribution of these N drawings is, therefore,

$$p(y_1, y_2, \cdots, y_N) = \frac{1}{(\sqrt{2\pi}\cdot\sigma)^N} \exp\left[-\frac{1}{2\sigma^2}\sum (y_t - kt - b)^2\right] \tag{16.4}$$

($\sum$ means $\sum_{t=1}^{N}$ throughout this section).

All these things are assumed known, the only unknown elements in our set-up being the values of the constants k and b; i.e., we know that it is possible to choose k and b such that the observable series y_t satisfy our model.

By the method of least squares [or by the method of maximum likelihood applied to (16.4)] we obtain the following estimation formula for the parameter k:

$$\text{(16.5)} \qquad \text{Est. of } k = \hat{k} = \frac{\sum (t - \bar{t})(y_t - \bar{y})}{\sum (t - \bar{t})^2},$$

where $\bar{t}$ and $\bar{y}$ denote the observed arithmetic means of t and y respectively. $\hat{k}$ is, of course, a random variable in repeated samples (each of N drawings, the t's being the same all the time). Using (16.1) we have

$$\text{(16.6)} \quad \hat{k} = \frac{\sum (t - \bar{t})(kt + b + \epsilon_t - k\bar{t} - b - \bar{\epsilon})}{\sum (t - \bar{t})^2} = k + \frac{\sum (t - \bar{t})\epsilon_t}{\sum (t - \bar{t})^2}.$$

Thus, $E(\hat{k}) = k$, i.e., we have an unbiased estimate. *We want to test the hypothesis that* $k = 0$. What is the set of a priori admissible hypotheses, i.e., the set Ω^0? It is: The system of all probability distributions (16.4) obtained by letting k and b run (independently) through all values from $-\infty$ to $+\infty$, and *no other* alternatives. The hypothesis to be tested is that $k = 0$, b being anything from $-\infty$ to $+\infty$, i.e., the set ω^0 is the system of all probability distributions obtained from (16.4) by putting $k = 0$ and letting b take, successively, all values from $-\infty$ to $+\infty$. We, therefore, have a *composite* hypothesis to be tested.

To test $k = 0$ we have to choose a critical region of rejection W_0 in the N-dimensional sample space of the variables y such that the probability of a sample point falling into W_0, no matter what be the value of b, is equal to α (say 0.05) when the hypothesis $k = 0$ is *true;* and besides, the region W_0 should be such that the probability of a sample point falling into it when the hypothesis $k = 0$ is *false* is as great as possible, and independent of the value of b.

Let us for this purpose consider the sampling distribution of the estimate $\hat{k}$. From (16.6) it is seen that $\hat{k}$ is a linear function of the N independent normally distributed variables $\epsilon_1, \epsilon_2, \cdots, \epsilon_N$, the t's being a set of constants—by assumption. $\hat{k}$ itself is, therefore, also normally distributed with

$$\text{(16.7)} \qquad \text{mean} = k, \qquad \text{variance} = \frac{\sigma^2}{\sum (t - \bar{t})^2}.$$

The distribution of $\hat{k}$ is independent of b, and we have

$$\text{(16.8)} \quad p(\hat{k}) = \frac{\sqrt{\sum (t - \bar{t})^2}}{\sqrt{2\pi}\,\sigma} \exp\left[- \frac{\sum (t - \bar{t})^2}{2\sigma^2} (\hat{k} - k)^2 \right].$$

And corresponding to our hypothesis to be tested, $k = 0$, we have

$$\text{(16.8')} \qquad p_0(\hat{k}) = \frac{\sqrt{\sum (t - \bar{t})^2}}{\sqrt{2\pi}\,\sigma} \exp\left[- \frac{\sum (t - \bar{t})^2}{2\sigma^2} \hat{k}^2 \right].$$

Let us consider the following two equal "tails" of this distribution

$$(16.9)\quad \hat{k} < -K = \frac{-c\sigma}{\sqrt{\sum (t-\bar{t})^2}} \text{ and } \hat{k} > +K = \frac{+c\sigma}{\sqrt{\sum (t-\bar{t})^2}},$$

where c is a positive constant so determined that

$$(16.10)\qquad 1 - \int_{-K}^{+K} p_0(\hat{k})d\hat{k} = \alpha \qquad (= 0.05, \text{ say}).$$

The two intervals (16.9) together define a certain region of rejection W_0 in the sample space of the variables y, because $\hat{k}$ is, by (16.5), a single-valued function of the y's. The probability—when the hypothesis $k=0$ is actually true—that $\hat{k}$ should fall in either of the two intervals (16.9) is the same as the probability that the sample point falls into W_0, and this probability is exactly equal to α. On the other hand, what are the properties of this critical region if the hypothesis is wrong, i.e., if $k \neq 0$? It has been shown that the region of rejection W_0 corresponding to the two tails (16.9) has the following properties:[6]

Whenever the hypothesis $k=0$ is *wrong*, i.e., when $k \neq 0$, the probability that the sample point should fall into W_0 (i.e., the power of the test) is $> \alpha$, which means that the test is unbiased. And for *any other* unbiased critical region of size α the power is *smaller*.

If we reject the hypothesis $k=0$ whenever $\hat{k}$ falls in either one of the intervals (16.9) we thus have a *best unbiased test* of the hypothesis $k=0$ corresponding to the level of significance α.

The probability that $\hat{k}$ should fall into either of the intervals (16.9) when $k \neq 0$, i.e., the power of the test, can be calculated as a function of k directly from (16.8). This *power-function*—let us call it $\beta(k)$—is simply

$$(16.11)\quad \beta(k) = 1 - \int_{-K}^{+K} \frac{\sqrt{\sum (t-\bar{t})^2}}{\sqrt{2\pi}\,\sigma} \exp\left[-\frac{\sum (t-\bar{t})^2}{2\sigma^2}(\hat{k}-k)^2\right] d\hat{k},$$

where K is given by (16.9).

Let us, as an example, take $N=9$, $\sigma=1$, $\alpha=0.05$, $c=1.96$ (from tables of the normal curve). We then have $\Sigma(t-\bar{t})^2=60$. If we introduce these numerical values, and change the variable of integration by the transformation $\kappa = (1/\sigma)\sqrt{\Sigma(t-\bar{t})^2}(\hat{k}-k)$, (16.11) becomes

$$(16.11')\qquad \beta(k) = 1 - \frac{1}{\sqrt{2\pi}} \int_{-1.96-\sqrt{60}\,k}^{+1.96-\sqrt{60}\,k} \exp\left[-\tfrac{1}{2}\kappa^2\right] d\kappa.$$

Values of $\beta(k)$ for different values of k then follow directly from tables of the normal distribution.

[6] See, e.g., Neyman, *Lectures and Conferences on Mathematical Statistics*, Washington, 1937, p. 29.

In Table 1 are given the results for a few values of k. The smooth curve in Figure 3 represents the continuous power function $\beta(k)$.

TABLE I

If the *true* value of k is	the probability that we shall *reject* $k=0$ by the test (16.9) (i.e., the power of the test) is
k	$\beta(k)$
0	0.05 $(=\alpha)$
± 0.1	0.12
± 0.2	0.34
± 0.3	0.64
± 0.4	0.87
± 0.5	0.97

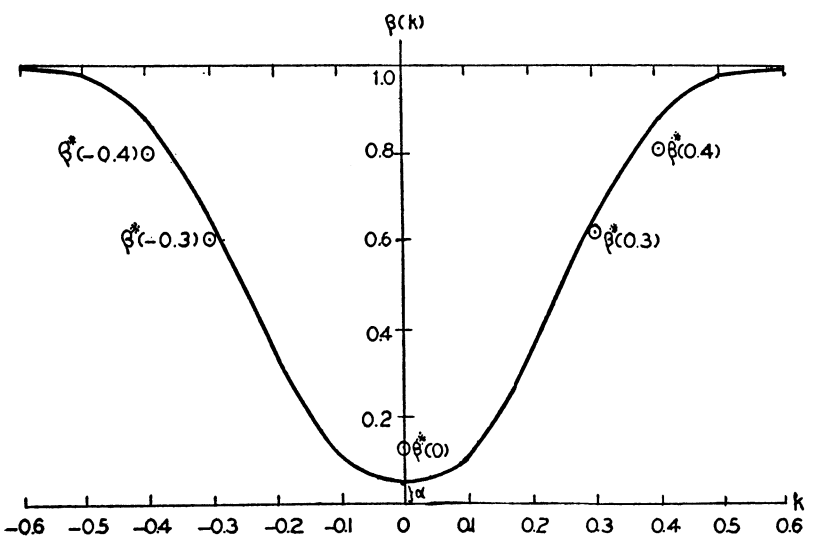

FIGURE 3.—The Power Function of the Test (16.9).
($N=9$, $\sigma=1$, $\alpha=0.05$, $c=1.96$)

Horizontal axis gives values of k ($k=0$ is the hypothesis tested; other values of k represent alternative hypotheses).

Vertical axis gives values of $\beta(k)$, representing the probability of $\hat{k}$ falling into the region of rejection (16.9) for the hypothesis $k=0$, when k is the *true* value of the parameter.

$\alpha=0.05$ represents the level of significance.

The encircled points ($\odot$) show the power of the same test (16.9) when the ϵ's are dependent as defined by (16.12).

This graph (the smooth curve) shows, for $k\neq 0$, the probability of rejecting—correctly—the hypothesis $k=0$ when it is false. The further away from $k=0$ we get, the greater is the probability that we shall reject $k=0$. $\beta(k)$ is the probability of *not* making an error of the second kind, considered as a function of the true k.

Now we shall consider an example showing what happens if the alternative which is actually *true* is *not* included in the set of a priori admissible hypotheses Ω^0 which was the basis for the above test.

One of the restrictions above was that the ϵ's in the N observations were stochastically *in*dependent. This was taken as a known fact, and not as a hypothesis which might be right or wrong. Suppose that we were *not* justified in doing so. As an example, let us assume that, *without our knowledge* and while proceeding *as if* our original scheme were correct, the actual series of ϵ's is of the following nature:

Let $\xi_0, \xi_1, \cdots, \xi_N$, be $N+1$ normally and independently distributed random variables, each with zero mean and variance $=\sigma^2=$ that of the ϵ's above. And let us consider a *new* series of ϵ's given by the formulae

$$\epsilon_t = \frac{1}{\sqrt{2}}(\xi_{t-1} + \xi_t) \qquad (t = 1, 2, \cdots, N). \tag{16.12}$$

Each of these new ϵ's taken separately then has mean $=0$, and the same variance σ^2 as the former ϵ-series. But ϵ_t and ϵ_{t+1} are now positively correlated (correlation coefficient $=\frac{1}{2}$).

Suppose now that we proceed *as if* we had to deal with the *original* ϵ-series instead of (16.12). By (16.6) $\hat{k}$ is still a linear function of normally and independently distributed variables, viz.,

$$\hat{k} = k + \frac{\sum (t - \bar{t})(\xi_{t-1} + \xi_t)}{\sqrt{2}\sum (t - \bar{t})^2}, \tag{16.13}$$

and, therefore, $\hat{k}$ is also now normally distributed with mean $=k$, and the variance of $\hat{k}$ is now that of the linear function

$$\frac{\sum (t - \bar{t})(\xi_{t-1} + \xi_t)}{\sqrt{2}\sum (t - \bar{t})^2}, \tag{16.14}$$

which gives

$$\sigma_{\hat{k}}^2 = \frac{1}{[\sum (t-\bar{t})^2]^2}\left[\sum (t-\bar{t})^2\sigma^2 + \sum_1^{N-1} (t-\bar{t})(t+1-\bar{t})\sigma^2\right]. \tag{16.15}$$

Taking, as in the previous example, $N=9$, $\sigma=1$, $c=1.96$, we obtain

$$\sigma_{\hat{k}}^2 = \frac{1}{60^2}(60 + 40) = \frac{1}{36}, \tag{16.16}$$

and, therefore, in analogy to (16.11′) we now get

$$\beta^*(k) = 1 - \frac{1}{\sqrt{2\pi}}\int_{-1.52-6k}^{+1.52-6k} \exp\left[-\tfrac{1}{2}\kappa^2\right]d\kappa. \tag{16.17}$$

$\beta^*(k)$ is the probability that $\hat{k}$—calculated by (16.5), the ϵ's being as defined by (16.12)—falls into either of the two intervals defined by

(16.9), this probability being considered as a function of the parameter k. As examples the values of $\beta^*(k)$ for $k=0$, $k=\pm 0.3$, and $k=\pm 0.4$ are plotted in Figure 3 (the encircled points). These values, as obtained from (16.17), are: $\beta^*(0)=0.13$, $\beta^*(\pm 0.3)=0.61$, and $\beta^*(\pm 0.4)=0.81$.

What do these results show? They show that the test (16.9) for an alternative hypothesis (namely dependent ϵ's) *not* included in Ω^0, may —incorrectly—reject the hypothesis tested (i.e., $k=0$), when it is true, *more frequently* than assumed (here 13 per cent instead of 5 per cent). That is to say, we had actually constructed the test such that we should reject the hypothesis $k=0$—when true—in only 5 per cent of the cases where the test is applied. But this no longer holds. The reason for this is easy to recognize: In order to make $\alpha=0.05$ in our first example (with independent errors) we had to fix a value of c such that the integral in (16.10) should be equal to 0.95. In the present case the integral over the *same range* [given by (16.9)] is, of course, smaller than 0.95, because the variance of the $\hat{k}$ we now have is *greater* (namely 1/36 instead of 1/60). 1 minus this integral is, therefore, greater than $\alpha=0.05$.

Also, we *thought* that we should be rejecting the false hypothesis $k=0$ in 87 per cent of those cases where $k=\pm 0.4$, while in fact we now do so only in 81 per cent of the cases, because, in constructing the test for the hypothesis $k=0$, we did *not* take account of the possibility that the ϵ's might be dependent.

The hypothesis $k=0$, as well as the alternative hypotheses about k in the last set-up, do not mean the same thing as in the first example with independent errors. In particular, the hypothesis tested (i.e., $k=0$) is *not* the one we set out to test, because it now includes the possibility of the errors being dependent. In other words: Even though the hypothesis $k=0$ might be true there is still *something wrong* with that case also—as compared with the hypothesis tested in the case of independent errors—namely the correlation between the ϵ's now present. It is interesting to note that the test above shows this to some extent, by rejecting the hypothesis $k=0$ in 13 per cent, and not 5 per cent, of the cases where $k=0$ is actually true. This result, however, is not a general one. The opposite may occur in other cases.

Of course, in the case above the mistake would not be so very bad, because *it so happens* that the power of the test is rather good also for the hypotheses *outside* Ω^0 which we just have considered. And in many important cases this might happen; that is to say, even if we develop a test only with respect to a certain very restricted class of a priori admissible hypotheses Ω^0, this test might—just by sheer luck, so to speak —be good also with respect to a much wider class of alternatives.

The example above illustrates, I think, a very useful method of proceeding in testing economic relations: We define first a certain set of a

priori admissible schemes, Ω^0, containing what we feel strongly to be the most important alternatives, and being at the same time such that it can be handled without prohibitive technical difficulties. Then, later, if—for some reason or another—we become suspicious as to the *completeness* of this Ω^0, we may study the power of the test for certain *outside* schemes not contained in Ω^0. For instance, it might be that a certain hypothesis outside Ω^0, i.e., a hypothesis rejected a priori, would, if it nevertheless were the true one, have important consequences for our decisions. *To see what risk we are taking* as to this hypothesis by using a test that simply *neglects* the possibility of this hypothesis being true, *we calculate the power of the test for this outside hypothesis.*

Of course, whatever be the test developed on the basis of a certain set, Ω^0, of a priori admissible hypotheses, it will always be possible to find hypotheses outside Ω^0, such that the power of the test with respect to these hypotheses is very poor; at least that is so if we want to have a test that is any good at all *within* Ω^0. To have some chance of reaching nontrivial conclusions we must assume a certain a priori knowledge, or be willing to take a certain amount of risk in order to restrict Ω^0. And the total risk involved in restricting Ω^0 is one which cannot be evaluated in probability terms. The choice of an a priori admissible set Ω^0 is, indeed, a matter of general knowledge and intuition.

The discussion above gives also, I think, a clearer interpretation of the general phrase, "Suppose the whole formal set-up of the theory is wrong, what is the use of testing significance of coefficients, etc.?" As a matter of fact, this question is, strictly speaking, always justified when we try to explain reality by a theoretical model. But if we follow this attitude to its bitter end, we shall never be able to accomplish anything in the way of explaining real phenomena.

17. The Meaning of the Phrase "To Formulate Theories by Looking at the Data"

All models of economic theory, however abstract they may be, probably arise from the consideration of some real economic phenomena. "Data" in the broad sense of empirical knowledge will, therefore, always to some extent influence our formulation of theories about them.

If we try to give only simplified and condensed *descriptions of empirical cases*, there is, of course, no risk in choosing a theory which "fits well." The risk comes in if we generalize, in the following sense: We specify an *empirical class* of phenomena (e.g., the class of all corresponding values of price and quantity sold of a certain commodity). We know empirically a certain number of members of this class. We form a *theoretical class* (e.g., a stochastic price-quantity relation) covering in particular the known members of the empirical class. We *hope*

that the theoretical class will cover *all* members of the empirical class. To construct such theoretical classes is, indeed, *the* problem of inductive science. And it involves risk of failures, which are beyond our control. A general discussion of "right" or "wrong" in connection with such empirical, inductive processes would take us into metaphysics.

But the phrase, "To formulate theories by looking at the data," has, among economic research workers, a narrower meaning, which it might be worth while to clarify. The common argument is as follows: Suppose we have a certain number of observations of simultaneous values of a system of economic variables. We have a broadly formulated economic theory about these variables, stating that there is *some* relation between the variables, without specifying the form of this relationship. We try out a great many different forms of relations, until we find one which "fits the data" (in some sense or another). Now, if we finally find a form of the relation which "fits well," is this in itself any verification of the "goodness" of that relation as a theory? Is not such a formula only a trivial restatement of facts?

Much discussion has taken place on this subject, e.g., in connection with the problem of testing business-cycle theories. For instance, a great many simplified dynamic models imply that each of the variables involved satisfies (apart from error terms) some linear difference equation of a certain order, with constant coefficients. It is clear that we may reach this same result by starting from different fundamental models, i.e., we might construct a great many models that are very different as to their basic assumptions or the type of economic mechanism they describe, and yet they may all imply that the variables, separately, satisfy certain linear difference equations as described. Now, if the *observed* series show some rather regular cycles, such difference equations may often be made to fit the series very well, by a proper choice of the coefficients. And if we accept this as a verification that the observed series actually satisfy such difference equations, we could say that the "correct" theory must belong to the *class* of models which lead to such difference equations. But we could not by this fitting alone pick out *the* "correct" theory from the class of admissible models. And if we choose one particular model, the fact that the corresponding difference equations in each variable may be made to fit the data gives no guarantee that just this model is *the* "correct" one. It is, therefore, generally argued that such good fits of "final" equations are not worth much from the point of view of verifying theories.

This argument, however, does not quite cover the real trouble point. In fact, if we could establish that the observed variables satisfied very closely a certain system of linear difference equations (say), we should have a strong and very useful restriction upon the *class* of a priori ad-

missible theoretical models. In general, whenever we can establish that certain data satisfy certain relationships, we add something to our knowledge, namely a restriction of the class of a priori admissible hypotheses. *The real difficulty* lies in deciding whether or not a given relation is actually compatible with the data; and the important thing to be analyzed is the reliability of the test by which the decision is made, since we have to deal with stochastic relations and random variables, not exact relations.

From this point of view there is, therefore, no justified objection against trying out various theories to find one which "fits the data." But objections may be made against certain *methods of testing the fit.* Let us examine this a little closer.

Consider a system of observable random variables, as in (15.1), and a relation to be tested, like (15.5). The theory defines a class, ω^0 say, of probability laws, and we want to test $P(w) \in \omega^0$. Now we have seen that, in order to develop a test of this hypothesis, we have to define a set, Ω^0, of a priori admissible hypotheses. Let (ω^0) be a system of different sets ω^0, corresponding to different relations to be tested, and such that each ω^0 is contained in Ω^0. For *any* one of these sets ω^0 we may test the hypothesis $P(w) \in \omega^0$, the set of a priori *admissible hypotheses* being constantly *the same*, namely Ω^0. It is clearly irrelevant *how* we happen to choose the hypothesis to be tested *within* Ω^0. In particular, the hypothesis might be one that suggests itself by *inspection of the data.* This is perfectly legitimate as long as *the set* Ω^0 *of admissible alternatives is a priori fixed and remains so.* For then we can calculate the power of the test used, and see what risk we run if we accept the hypothesis tested. What is *not* permissible is to let Ω^0 be a *function of the sample point.* Because then the test no longer controls the two types of possible errors in testing hypotheses. If Ω^0 be fixed on the basis of a sample point, and a test developed with respect to this set of admissible hypotheses, we have no idea whether the true hypothesis is actually contained in Ω^0 or not. We should have the untenable situation that the *method of testing* would itself be *varying at random* from one sample to the other.

The essential thing is, therefore, *not* the way in which we choose the hypothesis to be tested. Essential is what we *know or believe* to be the class of a priori admissible hypotheses, and what power our test has of rejecting the hypothesis tested, if a "really different" one among the alternatives be true.

CHAPTER V

Problems of Estimation

In Section 14 we described the general problem and the general principles of statistical estimation. More specific estimation problems arise in various fields of application. In the following we shall discuss a problem which is particularly relevant to economic research, namely that of estimating parameters in *systems of stochastic equations*.

A most dangerous—but often used—procedure in this field is to "fit each equation separately" without regard to the fact that the variables involved are, usually, assumed to satisfy, simultaneously, a number of *other* stochastic relations. If that is done, it is afterwards almost sheer luck if we have not created inner inconsistency in the system as a whole, such as, for instance, the assumption that some of the variables in *one* equation remain constant in repeated samples, while—because of another equation in the system—this is impossible. We shall illustrate this by an example later (see Section 21).

Even if no such inconsistency is created, the procedure of "fitting each equation separately" usually does *not* give the most efficient estimates of the parameters. For additional information about the parameters in one equation may be contained in the fact that, simultaneously, the variables satisfy another equation. And, what is even more important, we may fail to recognize that one or more of the parameters to be estimated might, in fact, be *arbitrary* with respect to the *system* of equations. *This* is the *statistical side* of the problem of *autonomous* relations, which we discussed in Section 8. It may be described in words as follows:

Suppose that a certain set of economic variables actually satisfies a *system* of (static or dynamic) equations, each of which we expect to have a certain degree of autonomy, so that we are interested in measuring the constant parameters involved (e.g., certain elasticities). From this equation system we can, by algebraic operations, derive an infinity of confluent systems. Suppose that, in particular, it is possible to derive an infinity of new systems which have exactly the same *form* as the original system, but with *different values of the coefficients* involved. (Usually this means that the number of parameters of the equation system may be reduced, as explained in Section 19.) Then, if we do not know anything about the values of the parameters in the original equation system, it is clearly not possible to obtain a unique estimate of them by any number of observations of the variables. And if we *did* obtain some "estimate" that appeared to be unique in such cases, it could only be due to the application of estimation formulae leading to

spurious or biased results. For example, the question of deriving both demand and supply curves from the same set of price-quantity data is a classical example of this type of problems.

This question (in the case of linear relations known as the problems of multicollinearity) is of great importance in economic research, because such research has to build, mostly, on *passive observations* of facts, instead of data obtained by rationally planned experiments (see Chapter II). And this means that we can obtain only such data as are the results of the economic system *as it in fact is*, and *not* as it *would be* under those unrestricted hypothetical variations with which we operate in economic theory, and in which we are interested for the purpose of economic policy. Considerable clarification on this point has been reached in recent years, following the pioneer work of Frisch.[1]

In the following we shall see that the investigation of this problem of indeterminate coefficients, as well as other questions of estimation in relation to economic equation systems, all come down to one and the same thing, namely, *to study the properties of the joint probability distribution of the random (observable) variables in a stochastic equation system.*

18. *General Formulation of the Problem of Estimating Parameters in Systems of Economic Relations*

We shall discuss one general class of static systems and one general class of dynamic systems.

A. Static systems

Let us denote by $\xi_{1j}, \xi_{2j}, \cdots, \xi_{mj}, \cdots, \xi_{nj}$ $(j=1, 2, \cdots, N)$, N "true" measurements of n economic variables. The subscript j indicates "observation No. j." The *actual* measurements of these variables might (and usually will) be subject to *errors of measurement proper*. Let the corresponding *actually observed* variables be x_{ij}, defined by

$$x_{ij} = G_{ij}(\xi_{ij}, \eta_{ij}) \qquad (i = 1, 2, \cdots, n; j = 1, 2, \cdots, N), \tag{18.1}$$

[1] R. Frisch, "Correlation and Scatter in Statistical Variables," *Nordic Statistical Journal*, Vol. 1, 1929, pp. 36–102; "Statistical Correlation and the Theory of Cluster Types" (joint authorship with B. D. Mudgett), *Journal of American Statistical Association*, Vol. 26, December, 1931, pp. 375–392; *Pitfalls in the Statistical Construction of Demand and Supply Curves* (Veröffentlichungen der Frankfurter Gesellschaft für Konjunkturforschung, Neue Folge, Heft 5), Leipzig, 1933, 39 pp.; *Statistical Confluence Analysis by Means of Complete Regression Systems*, Publication No. 5 from the Institute of Economics, Oslo, 1934; "Statistical versus Theoretical Relations in Economic Macro-Dynamics" (Mimeographed Memorandum prepared for the Business Cycle Conference at Cambridge, England, July 18–20, 1938, to discuss J. Tinbergen's Publications of 1938 for the League of Nations). See also J. Marschak, "Economic Interdependence and Statistical Analysis," in *Studies in Mathematical Economics and Econometrics, in Memory of Henry Schultz*, Chicago, 1942, pp. 135–150.

or, when solved for η_{ij},

$$\eta_{ij} = g_{ij}(x_{ij}, \xi_{ij}), \tag{18.1'}$$

where η_{ij} are random variables characterizing the errors of measurement, and where G_{ij} (and g_{ij}) are certain *known* functions. We introduce these functions G_{ij} for the following reason: If we wrote just

$$x_{ij} = \xi_{ij} + \text{error}, \tag{18.2}$$

the distribution of the errors would in general depend upon ξ_{ij}. If this be the case we assume it to be possible to write the error part as a *known* function of ξ_{ij} and a new random variable, namely η_{ij}, which is stochastically *in*dependent of ξ_{ij} (and also, of course, independent of ξ_{hk} when $h, k \neq i, j$). These transformations are expressed by the functions G_{ij} in (18.1). This leads us to

Assumption 1: The nN random variables η_{ij} $(i=1, 2, \cdots, n; j=1, 2, \cdots, N)$, have a joint *elementary*[2] probability law

$$p_1(\eta_{11}, \cdots, \eta_{nN}; \gamma_1, \gamma_2, \cdots, \gamma_q) \tag{18.3}$$

which is *known*, except—perhaps—for the values of q parameters $\gamma_1, \cdots, \gamma_q$, and which is independent of the variables ξ_{ij} and the variables ϵ defined below.

Assumption 2: The $(n-m)N$ quantities $\xi_{m+1,j}, \cdots, \xi_{nj}$ $(j=1, 2, \cdots, N; m<n)$, are considered as *constants in repeated samples.* The economic meaning of this is that these variables are autonomous parameters fixed by forces external to the economic sector under consideration.[3]

Assumption 3: The mN quantities $\xi_{1j}, \cdots, \xi_{mj}$ $(j=1, 2, \cdots, N)$ are random variables ("dependent variables") in repeated samples, and are *known to satisfy* m stochastical equations,

$$f_i[\xi_{1j}, \xi_{2j}, \cdots, \xi_{mj}; \xi_{m+1,j}, \cdots, \xi_{nj}; \alpha_1, \alpha_2, \cdots, \alpha_k; \epsilon_{1j}, \epsilon_{2j}, \cdots, \epsilon_{hj}] = 0 \tag{18.4}$$
$$(h \geqq m; i = 1, 2, \cdots, m; j = 1, 2, \cdots, N),$$

where $\alpha_1, \alpha_2, \cdots, \alpha_k$, are k *unknown* constants, and where $\epsilon_{1j}, \epsilon_{2j}, \cdots, \epsilon_{hj}$ $(j=1, 2, \cdots, N)$ are hN *random variables.* Here $\alpha_1, \alpha_2, \cdots, \alpha_k$ mean *all* the unknown constants in the whole system of equations (18.4).

[2] For the sake of simplicity we restrict ourselves, in this and the following sections, to cases where the *elementary* probability laws are assumed to exist. However, in point of principle, there would be no difficulty in reformulating our statements on the basis of *integral* probability laws.

[3] This assumption might, if it be desirable, be replaced by the assumption that the autonomous ξ's are themselves random variables. That would cause only small changes in the subsequent formulations.

There might actually be only a few of them present in each of the m equations. And similarly for the h ϵ's.

(18.4) represents an economic theory when certain restrictions are imposed upon the distribution of the ϵ's characterizing the stochastical model.

Assumption 4: The hN random variables $\epsilon_{1j}, \epsilon_{2j}, \cdots, \epsilon_{hj}$ $(j=1, 2, \cdots, N)$ have a joint elementary probability law

$$p_2(\epsilon_{11}, \cdots, \epsilon_{hN} \mid \xi_{m+1,1}, \cdots, \xi_{nN}; \beta_1, \beta_2, \cdots, \beta_r) \tag{18.5}$$

(i.e., the conditional distribution of the hN ϵ's, when the autonomous ξ's are *given*), which is *known*, except—perhaps—for the values of r parameters $\beta_1, \cdots, \beta_r$. By introducing a considerable number of β's, p_2 may be made to comprise a wide class of distributions.

The problem is: To estimate the values of $\alpha_1, \alpha_2, \cdots, \alpha_k$, on the basis of a sample point $(x_{11}, x_{21}, \cdots, x_{n1}, x_{12}, x_{22}, \cdots, x_{n2}, \cdots, x_{1N}, x_{2N}, \cdots, x_{nN})$ in the nN dimensional sample space of the observable variables x. And in order to do this it may, or may not, be necessary also to estimate the parameters γ and β in (18.3) and (18.5). We shall now see that this problem is a problem of statistical estimation as described under Section 14.

From the mN equations (18.4) we may (under certain conditions for solvability) express mN of the hN ϵ's as functions of the mN random variables $\xi_{11}, \cdots, \xi_{mN}$, and the $(h-m)N$ remaining ϵ's. These functions will, in general, involve the parameters α and the $(n-m)N$ autonomous ξ's. Introducing these expressions for mN ϵ's into (18.5), multiplying by the Jacobian of the transformation, and integrating over the $(h-m)N$ remaining ϵ's from $-\infty$ to $+\infty$, we obtain the joint elementary probability law of the mN ("dependent") variables $\xi_{11}, \cdots, \xi_{mN}$. Let this probability law be

$$p_3[\xi_{11}, \cdots, \xi_{mN} \mid \alpha_1, \alpha_2, \cdots, \alpha_k; \beta_1, \beta_2, \cdots, \beta_r; \xi_{m+1,1}, \cdots, \xi_{nN}]. \tag{18.6}$$

This is the *conditional* distribution of the mN random variables $\xi_{11}, \cdots, \xi_{mN}$, for *given* values of the autonomous ξ's. By assumption, this distribution is *in*dependent of the variables η defined by (18.1′). Therefore, the joint distribution of $(\xi_{11}, \cdots, \xi_{mN})$ and $(\eta_{11}, \cdots, \eta_{nN})$ is equal to

$$p_1 \cdot p_3. \tag{18.7}$$

Introducing the transformations (18.1′) in (18.7) and integrating the result with respect to the mN random variables $\xi_{11}, \cdots, \xi_{mN}$ from $-\infty$ to $+\infty$, we obtain the joint elementary probability law of the nN ran-

dom variables x_{ij} (the observed variables). Let this probability law be

$$\Phi[x_{11}, \cdots, x_{nN} \mid \xi_{m+1,1}, \cdots, \xi_{nN}; \alpha_1, \cdots, \alpha_k; \beta_1, \cdots, \beta_r; \gamma_1, \cdots, \gamma_q]. \tag{18.8}$$

We can now say: Our economic theory, so far as the observable variables x are concerned, is *indistinguishable from* (and it may even be equivalent to) *the statement that the observable variables x have the* joint probability law (18.8), where Φ is a known function. And the problem of estimating the unknown parameters is reduced to an ordinary problem of statistical estimation.

If, in particular, all the variables be observed *without* errors of measurement, our economic theory would be expressed by (18.6).

If, in particular, all the *autonomous* ξ's be measured without errors, we should—instead of (18.8)—have[4]

$$\Phi_1(x_{11}, \cdots, x_{mN} \mid \xi_{m+1,1}, \cdots, \xi_{nN}; \alpha_1, \cdots \alpha_k; \beta_1, \cdots, \beta_r; \gamma_1', \cdots, \gamma_q'), \tag{18.8'}$$

that is to say, a distribution with only mN instead of nN random variables x.

In (18.8) the $(n-m)N$ autonomous ξ's are unknown parameters which it might or might not be necessary to estimate in order to estimate the α's.

Clearly no more complete description of the interconnections between a certain number of random variables can be given than that which is contained in their joint probability law. *If*, therefore, *two different formulations* of an economic theory lead to identically *the same joint probability law* of the observable random variables involved, we can *not distinguish between them on the basis of observations.* (But the theories may not be equivalent in certain other respects.)

The joint probability law of all the variables covers also the particular case where the set of random variables can be split up into independently distributed subgroups of variables with *different* parameters to be estimated occurring in the distribution of each subgroup. And in all other cases the joint probability law of *all* the variables contains more information than that obtained from the probability laws of subgroups of variables. It is, therefore, clear that the *joint probability law* of *all* the observable random variables in an economic system is the only general basis for estimating the unknown parameters of the system.

[4] Here γ_i' denote the parameters in an mN-dimensional distribution instead of the nN-dimensional distribution (18.3).

B. Dynamic systems

We shall consider the following general type of dynamic economic systems (making similar assumptions to those above):

Let $\xi_1(t_i), \xi_2(t_i), \cdots, \xi_m(t_i), \cdots, \xi_n(t_i)$ be n time series defined at the points of time

$$t_N, t_{N-1}, \cdots, t_1, t_0, t_{-1}, t_{-2}, \cdots . \tag{18.9}$$

For the moment we shall neglect the problem of errors of measurement proper.

The $(n-m)$ series $\xi_{m+1}(t_i), \cdots, \xi_n(t_i)$, are assumed to be autonomous variables, they are assumed to remain fixed in repeated samples.

For each point of time (18.9) the quantities $\xi_1(t_i), \cdots, \xi_m(t_i)$ are random variables defined implicitly by a system of dynamic relations of the type

$$\begin{aligned}\xi_j(t_i) = F_{t_i}^{(j)}[&\xi_1(t_i), \xi_1(t_{i-1}), \cdots ; \xi_2(t_i), \xi_2(t_{i-1}), \cdots ;\\ &\xi_j(t_{i-1}), \xi_j(t_{i-2}), \cdots ; \cdots ; \xi_m(t_i), \xi_m(t_{i-1}), \cdots ;\\ &\xi_{m+1}(t_i), \xi_{m+1}(t_{i-1}), \cdots ; \cdots ; \xi_n(t_i), \xi_n(t_{i-1}), \cdots ;\\ &\alpha_1, \alpha_2, \cdots, \alpha_k; \epsilon_{1t_i}, \epsilon_{2t_i}, \cdots, \epsilon_{ht_i}]\end{aligned} \tag{18.10}$$

$$(i = 1, 2, \cdots, N; j = 1, 2, \cdots, m; h > m).$$

Or, expressed in words: Each of the "dependent" variables $\xi_1(t_i), \cdots, \xi_m(t_i)$, is a function of (1) the previous values of that same variable and (2) the simultaneous *and* the previous values of the other $n-1$ variables. The m functions $F_{t_i}^{(j)}$ may be different for each point of time t_i, but they have *known forms.*

The system (18.10) involves, altogether, k unknown constants $\alpha_1, \alpha_2, \cdots, \alpha_k$, some or all of which might be lacking in any particular one of the equations. For each point of time t_i the system involves, altogether, h random variables ϵ, which have certain known distribution properties. We refer all these h random variables ϵ to the same point of time as that for the variable to the left in (18.10), although the *actual events* from which they emerge might take place at different points of time. This is merely a simple transformation of variables in the joint probability law of all the ϵ's. If there happens to be functional relationship between the ϵ's at two different points of time (e.g., $\epsilon_{st_i} = \epsilon_{r,t_{i-1}}$), the dimensionality of the joint distribution of all the hN ϵ's can be correspondingly reduced.

(18.10) gives, altogether, mN equations. From these equations we can (under certain conditions for solvability) express the mN random variables $\xi_1(t_i), \cdots, \xi_m(t_i)$ $(i=1, 2, \cdots, N)$ as functions of: (1) initial

conditions, i.e., all (or some) of the values of $\xi_1(t_i), \cdots, \xi_n(t_i)$ for $i=0, -1, -2, \cdots$; (2) the values of the autonomous variables $\xi_{m+1}(t_i), \cdots, \xi_n(t_i)$, for $i=1, 2, \cdots, N$; (3) the hN random variables $\epsilon_{1t_i}, \cdots, \epsilon_{ht_i}$ $(i=1, 2, \cdots, N)$.

We shall assume the initial conditions to be *given* and constant in repeated samples. For short we denote the whole set of initial conditions by (ξ^0).

Let

$$(18.11)\quad p_2'[\epsilon_{1t_1}, \cdots, \epsilon_{ht_N} \mid \xi_{m+1}(t_1), \cdots, \xi_n(t_N); \quad (\xi^0); \beta_1, \beta_2, \cdots, \beta_r]$$

be the joint elementary probability law of all the hN random variables ϵ for given *initial conditions* of the ξ's and given values of the *autonomous* ξ's. (p_2' might or might not actually depend upon these quantities.) The β's are parameters which might or might not be known.

Since the mN random variables $\xi_1(t_i), \cdots, \xi_m(t_i)$ $(i=1, 2, \cdots, N)$, can be expressed as functions of the random variables ϵ, we can derive the joint distribution of the mN random variables $\xi_1(t_i), \cdots, \xi_m(t_i)$ $(i=1, 2, \cdots, N)$ in exactly the same manner as was discussed under A above. Let this probability distribution be

$$(18.12)\quad p_3'[\xi_1(t_1), \cdots, \xi_m(t_N) \mid \xi_{m+1}(t_1), \cdots, \xi_n(t_N); \quad (\xi^0); \alpha_1, \cdots, \alpha_k; \beta_1, \cdots, \beta_r].$$

If the measurements of the ξ's (but not those of the initial conditions) are subject to errors, we have an additional problem exactly similar to that discussed for static systems.

The problem of estimating the parameters in a dynamic system of the form (18.10) is, therefore, reduced to the problem of estimating the parameters of an mN- (or nN-) dimensional elementary probability law, by means of a sample point associated with this probability law.[5]

This way of condensing the statements implied in a system of stochastic relations may be extended to more general classes of economic schemes. And this procedure is not only convenient but, I think, necessary, if we want to make sure that the various assumptions made about the distribution properties of the random variables involved do not lead to inner contradictions, like those we mentioned in the introduction to this chapter.

* * *

We are now in a position to formulate precisely the two fundamental

[5] For explicit estimation formulae and confidence limits, etc., in the case of a system of linear stochastic difference equations see the article by H. B. Mann and A. Wald, "On the Statistical Treatment of Linear Stochastic Difference Equations," ECONOMETRICA, Vol. 11, July–October, 1943, pp. 173–220, in particular, Part II, pp. 192–216.

problems of estimation in economic research, namely (1) the problem of confluent relations, and (2) the problem of "best estimates":

I. The problem of confluent relations (or, the problem of arbitrary parameters)

If two stochastical equation systems lead to the same joint probability law of the observable random variables, they are *indistinguishable* (on the basis of observations). In particular, the systems might be such that they differ only with respect to the *values* of the (unknown) parameters involved. The problem of arbitrary coefficients is, therefore, included in the following general mathematical problem: Let

$$p(x_1, x_2, \cdots, x_s \mid \theta_1, \theta_2, \cdots, \theta_\kappa; z_1, z_2, \cdots, z_r)$$

be a function of s independent variables $x_1, x_2, \cdots, x_s$, involving κ unknown parameters θ, and r known parameters z. Let $\theta_1^0, \theta_2^0, \cdots, \theta_\kappa^0$, or, for short, θ^0, be a point in the κ-dimensional parameter space of the θ's. Does there, or does there not, exist at least one parameter point θ' $(\neq \theta^0)$, such that

$$(18.13)\quad p(x_1, \cdots, x_s \mid \theta_1^0, \cdots, \theta_\kappa^0; z_1, \cdots, z_r) \equiv p(x_1, \cdots, x_s \mid \theta_1', \cdots, \theta_\kappa'; z_1, \cdots, z_r)$$

for *all* values of the variables x? The answer to this question depends upon one or more of the following things: 1. The form of the function p. 2. The parameter point θ^0. 3. The values of the known parameters z.

If (18.13) be fulfilled, and if θ^0 be the "true" parameter point, then, no matter how many observations we have of the variables x, there is *no unique* estimate to be obtained for θ^0, because we cannot then distinguish between θ^0 and θ'. (The well-known problem of "multicollinearity" is, of course, included in this formulation as a very special case of the arbitrary parameter problem.)[6]

II. The problem of "best estimates"

Let

$$(18.14)\quad y = p(x_1, x_2, \cdots, x_s \mid \theta_1, \theta_2, \cdots, \theta_\kappa; z_1, z_2, \cdots, z_r)$$

be a parametric family of joint elementary probability laws of s random variables $x_1, x_2, \cdots, x_s$, involving κ unknown parameters $\theta_1, \theta_2, \cdots, \theta_\kappa$. If, for given values of the known parameters z, there be a one-to-one correspondence between the parameter points θ and the members of the κ-parametric family (18.14), and if θ^0 be the true

[6] Cf. the discussion by Mann and Wald on the problem of whether to deal with the "reduced" equations or the "original" equations, *op. cit.*, pp. 200–202.

parameter point, what is the best estimate of θ^0 to be obtained from a sample point $(x_1, x_2, \cdots, x_n)$?

* * *

The problem II is—at least in point of principle—a straightforward problem of statistical estimation, and there is no need, nor justification, for a separate discussion of that statistical problem here.

The same could, of course, also be said about problem I. It is a problem of pure mathematics. This problem, however, is of particular significance in the field of econometrics, and relevant to the very construction of economic models, and besides, this particular mathematical problem does not seem to have attracted the interest of mathematicians. In the following sections we shall, therefore, develop some mathematical tools of analysis for this particular purpose.

19. On the Reducibility of a Function with Respect to Its Number of Parameters

Let

$$y = f(x_1, x_2, \cdots, x_n; \theta_1, \theta_2, \cdots, \theta_\kappa) \tag{19.1}$$

be a real function of n real independent (i.e., *not functionally* related) variables $x_1, x_2, \cdots, x_n$, involving κ parameters $\theta_1, \theta_2, \cdots, \theta_\kappa$. [E.g., f might be the function (18.14) for fixed values of the z's.] Let θ^0 denote a point in the κ-dimensional parameter space of the θ's. And let $S(\theta^0)$ be the corresponding set of all points $(y, x_1, x_2, \cdots, x_n)$ in the $(n+1)$-dimensional variable space, that is to say, the set of all points $(y, x_1, x_2, \cdots, x_n)$ defined by (19.1) when $\theta = \theta^0$. Let θ' denote another parameter point $\neq \theta^0$, and let $S(\theta')$ be the corresponding set of points $(y, x_1, x_2, \cdots, x_n)$. If there exists at least one parameter point $\theta' \neq \theta^0$, such that

$$S(\theta^0) = S(\theta'), \tag{19.2}$$

or—what amounts to the same—such that

$$f(x_1, x_2, \cdots, x_n; \theta_1^0, \theta_2^0, \cdots, \theta_\kappa^0) \equiv f(x_1, x_2, \cdots, x_n; \theta_1', \theta_2', \cdots, \theta_\kappa') \tag{19.3}$$

identically, for *all* values of the variables x, we shall say that the parameter point θ has (a certain amount of) *arbitrariness with respect to the set* $S(\theta^0)$.

We may here distinguish between the following two cases:

(A): There exists a finite neighborhood of the point θ^0 such that *within* this neighborhood there is *no point* $\theta' \neq \theta^0$ satisfying (19.3), while outside, or on the border of, this neighborhood there may be one or more points θ' satisfying (19.3).

Example 1.

$$y = \theta_1^2 x_1.$$

Here, if $\theta_1 = \theta_1^0 > 0$ there is *no* point $\theta_1' \neq \theta_1^0$ in the range $\theta_1 > -\theta_1^0$ that satisfies (19.3), while, in the range $\theta_1 \leqq -\theta_1^0$, there is just one point θ_1' satisfying (19.3), namely $\theta_1' = -\theta_1^0$.

Example 2.

$$y = \theta_1 \sin(\theta_2 + \theta_3 x_1).$$

Here, if θ^0 be a parameter point, there are no parameter points in the immediate *vicinity* of θ^0 satisfying (19.3), but there is an infinity of *isolated* parameter points θ' satisfying (19.3), namely $\theta_1' = \theta_1^0$, $\theta_2' = (\theta_2^0 + 2\pi n)$, $\theta_3' = \theta_3^0$, $n = 1, 2, 3, \cdots$ ad inf.

Example 3.

$$y = v(\theta_1)x_1; \quad v(\theta_1) = (|\theta_1| + \theta_1 + 1) - 2[|\theta_1 - 1| + (\theta_1 - 1)].$$

Suppose that $\theta_1^0 = 2$, then $v(\theta_1^0) = 1$. Now, if $\theta_1 > 2$, then $v(\theta_1) < 1$, and no point $\theta_1' > 2$ will satisfy (19.3). Next, if $2 > \theta_1 > 0$, then $v(\theta_1) > 1$; therefore no points θ_1' such that $2 > \theta_1' > 0$ will satisfy (19.3). But if $\theta_1 \leqq 0$, then $v(\theta_1) \equiv 1$; hence, all points $\theta_1' \leqq 0$ will satisfy (19.3).

(B): If a finite neighborhood of θ^0 be chosen, no matter how small, there are points $\theta' \neq \theta^0$ in this neighborhood, satisfying (19.3).

Example:

$$y = (\theta_1 + \theta_2)x_1 + \theta_3 x_2.$$

We shall now derive certain general conditions under which (A) or (B) will occur.

For this purpose we consider the function f in (19.1) as a function of $s + \kappa$ *independent* variables, $x_1, x_2, \cdots, x_s, \theta_1, \theta_2, \cdots, \theta_\kappa$. We assume, throughout the rest of this section, that

(1) f is defined over a certain domain D_x of the s-dimensional x-space, and over a certain simply connected region D_θ of the κ-dimensional parameter space, and is a single-valued function for every point $x \in D_x$ and for every point $\theta \in D_\theta$.

(2) For every point $x \in D_x$, and for every interior point θ of D_θ, f has continuous first-order partial derivatives $\partial f / \partial \theta_i$ $(i = 1, 2, \cdots, \kappa)$ (i.e., continuous in the θ's).

Definition: The function $f(x_1, x_2, \cdots, x_s; \theta_1, \theta_2, \cdots, \theta_\kappa)$ is said to be *ν-fold reducible* $(\kappa \geqq \nu > 0)$ at the parameter point θ^0, where θ^0 is an interior point of D_θ, if there exist $\kappa - \nu$ functions $u_1(\theta_1, \theta_2, \cdots, \theta_\kappa)$, $u_2(\theta_1, \theta_2, \cdots, \theta_\kappa), \cdots, u_{\kappa-\nu}(\theta_1, \theta_2, \cdots, \theta_\kappa)$, not depending upon the point x, and a function $\bar{f}(x_1, x_2, \cdots, x_s; u_1, u_2, \cdots, u_{\kappa-\nu})$, having the following properties:

(a) $f(x_1, x_2, \cdots, x_s; \theta_1, \theta_2, \cdots, \theta_\kappa) \equiv \bar{f}(x_1, x_2, \cdots, x_s; u_1, u_2, \cdots, u_{\kappa-\nu})$ for every point $x \in D_x$, and for every point θ in an arbitrarily small but finite neighborhood of θ^0.

(b) $\partial u_i/\partial\theta_j$ $(i=1, 2, \cdots, \kappa-\nu; j=1, 2, \cdots, \kappa)$ *exist and are continuous* for every point θ within an arbitrarily small but finite neighborhood of θ^0.

(c) The Jacobian matrix $\partial(u_1, u_2, \cdots, u_{\kappa-\nu})/\partial(\theta_1, \theta_2, \cdots, \theta_\kappa)$ is of rank $\kappa-\nu$ at $\theta=\theta^0$.

If a function f has these properties at a parameter point θ^0, then, clearly, there exist infinitely many points θ' in the neighborhood of θ^0, such that (19.3) is satisfied, for if $u_1, u_2, \cdots, u_{\kappa-\nu}$ be fixed, ν parameters θ may be chosen arbitrarily in a certain neighborhood of θ^0 without changing the value of f, whatever be $x \in D_x$.

THEOREM 1. *If a function $f(x_1, x_2, \cdots, x_s; \theta_1, \theta_2, \cdots, \theta_\kappa)$ is ν-fold reducible at the parameter point θ^0, there exists a system of functions $\lambda_{ij}(\theta_1, \theta_2, \cdots, \theta_\kappa)$ $(i=1, 2, \cdots, \kappa; j=1, 2, \cdots, \nu)$ that are independent of the point x, and continuous in the neighborhood of θ^0, such that*

$$\begin{pmatrix} \lambda_{11}, & \lambda_{21}, & \cdots, & \lambda_{\kappa 1} \\ \lambda_{12}, & \lambda_{22}, & \cdots, & \lambda_{\kappa 2} \\ \cdot & \cdot & \cdots & \cdot \\ \lambda_{1\nu}, & \lambda_{2\nu}, & \cdots, & \lambda_{\kappa\nu} \end{pmatrix} \tag{19.4}$$

is of rank ν at $\theta=\theta^0$, and

$$\lambda_{1j}\frac{\partial f}{\partial\theta_1} + \lambda_{2j}\frac{\partial f}{\partial\theta_2} + \cdots + \lambda_{\kappa j}\frac{\partial f}{\partial\theta_\kappa} \equiv 0 \quad (j = 1, 2, \cdots, \nu), \tag{19.5}$$

for all points $x \in D_x$, and for all points θ in an arbitrarily small but finite neighborhood of θ^0.

Proof: Since the Jacobian matrix $\partial(u_1, u_2, \cdots, u_{\kappa-\nu})/\partial(\theta_1, \theta_2, \cdots, \theta_\kappa)$ is of rank $\kappa-\nu$ at $\theta=\theta^0$, it contains at least one $(\kappa-\nu)$-rowed determinant that is not zero at $\theta=\theta^0$. Since the numbering of the θ's is arbitrary, we may, without loss of generality, assume that $\partial(u_1, u_2, \cdots, u_{\kappa-\nu})/\partial(\theta_1, \theta_2, \cdots, \theta_{\kappa-\nu})$ is of rank $\kappa-\nu$ for $\theta=\theta^0$. Then the system

$$\begin{aligned} u_1(\theta_1, \theta_2, \cdots, \theta_\kappa) &= u_1, \\ u_2(\theta_1, \theta_2, \cdots, \theta_\kappa) &= u_2, \\ \cdots\cdots\cdots\cdots & \\ u_{\kappa-\nu}(\theta_1, \theta_2, \cdots, \theta_\kappa) &= u_{\kappa-\nu}, \end{aligned} \tag{19.6}$$

has a solution

$$\theta_i = \phi_i(u_1, u_2, \cdots, u_{\kappa-\nu}, \theta_{\kappa-\nu+1}, \cdots, \theta_\kappa) \qquad (i = 1, 2, \cdots, \kappa - \nu) \tag{19.7}$$

which is unique in a certain finite neighborhood of $\theta = \theta^0$, and such that $\partial\phi_i/\partial u_j$ $(i=1, 2, \cdots, \kappa-\nu;\ j=1, 2, \cdots, \kappa-\nu)$ and $\partial\phi_i/\partial\theta_k$ $(i=1, 2, \cdots, \kappa-\nu;\ k=\kappa-\nu+1, \cdots, \kappa)$ exist and are continuous functions of $u_1, u_2, \cdots, u_{\kappa-\nu}, \theta_{\kappa-\nu+1}, \cdots, \theta_\kappa$, in this neighborhood. (This follows from the classical theory of functional determinants.) Hence we have

$$\begin{aligned} &f(x_1, x_2, \cdots, x_s; \theta_1, \theta_2, \cdots, \theta_\kappa) \\ &\quad \equiv f(x_1, x_2, \cdots, x_s; \phi_1, \phi_2, \cdots, \phi_{\kappa-\nu}, \theta_{\kappa-\nu+1}, \cdots, \theta_\kappa) \\ &\quad \equiv \bar{f}(x_1, x_2, \cdots, x_s; u_1, u_2, \cdots, u_{\kappa-\nu}). \end{aligned} \tag{19.8}$$

By definition f has continuous partial derivatives $\partial f/\partial\theta_i$ $(i=1, 2, \cdots, \kappa)$; therefore, $\partial f/\partial\phi_i$ $(i=1, 2, \cdots, \kappa-\nu)$ exist and are continuous in the neighborhood of θ^0. Since also $\partial\phi_i/\partial u_j$ $(i=1, 2, \cdots, \kappa-\nu; j=1, 2, \cdots, \kappa-\nu)$ exist and are continuous as shown above, f has continuous partial derivatives $\partial f/\partial u_j$ $(j=1, 2, \cdots, \kappa-\nu)$ in a certain finite neighborhood of θ^0. But when $f \equiv \bar{f}$ also $\partial\bar{f}/\partial u_j$ $(j=1, 2, \cdots, \kappa-\nu)$ must exist and be continuous in the same neighborhood. We therefore have

$$\begin{aligned} \frac{\partial f}{\partial\theta_1} &\equiv \frac{\partial\bar{f}}{\partial u_1}\frac{\partial u_1}{\partial\theta_1} + \cdots + \frac{\partial\bar{f}}{\partial u_{\kappa-\nu}}\frac{\partial u_{\kappa-\nu}}{\partial\theta_1}, \\ \frac{\partial f}{\partial\theta_2} &\equiv \frac{\partial\bar{f}}{\partial u_1}\frac{\partial u_1}{\partial\theta_2} + \cdots + \frac{\partial\bar{f}}{\partial u_{\kappa-\nu}}\frac{\partial u_{\kappa-\nu}}{\partial\theta_2}, \\ &\cdots\cdots\cdots\cdots\cdots\cdots \\ \frac{\partial f}{\partial\theta_\kappa} &\equiv \frac{\partial\bar{f}}{\partial u_1}\frac{\partial u_1}{\partial\theta_\kappa} + \cdots + \frac{\partial\bar{f}}{\partial u_{\kappa-\nu}}\frac{\partial u_{\kappa-\nu}}{\partial\theta_\kappa}. \end{aligned} \tag{19.9}$$

(19.9) can be considered as a *singular* linear transformation of the $\kappa-\nu$ variables $\partial\bar{f}/\partial u_i$ $(j=1, 2, \cdots, \kappa-\nu)$, into the κ variables $\partial f/\partial\theta_i$ $(i=1, 2, \cdots, \kappa)$, the matrix of the transformation being—by definition—of rank $\kappa-\nu$ for $\theta=\theta^0$, and having continuous elements in the vicinity of θ^0. But then Theorem 1 follows immediately from the theory of linear dependence.

The converse of Theorem 1 may be shown, under certain weak additional restrictions upon the λ's.

We shall now prove a theorem which gives a sufficient condition for the *non*existence of a relation of type (19.3) in the vicinity of a parameter point θ^0, namely:

THEOREM 2. *If the functions* $\partial f/\partial\theta_1, \partial f/\partial\theta_2, \cdots, \partial f/\partial\theta_k$, *are linearly independent at the point* θ^0, *i.e., if, at the parameter point* θ^0,

$$\sum_{i=1}^{\kappa} \lambda_i{}^0 \frac{\partial f}{\partial \theta_i} \not\equiv 0 \tag{19.10}$$

whatever be the system of constants $\lambda_1{}^0, \lambda_2{}^0, \cdots, \lambda_\kappa{}^0$, *not all equal to zero, then there exists a finite neighborhood of the parameter point* θ^0 *such that, in this neighborhood, there are no parameter points* $\theta' \neq \theta^0$ *for which* (19.3) *is satisfied.*

Proof: First, it is easy to see that the linear *in*dependence of the functions $\partial f/\partial\theta_i$ at $\theta = \theta^0$ implies that the set $S(\theta^0)$ defined above contains at least κ different points $(y^{(j)}, x_1{}^{(j)}, x_2{}^{(j)}, \cdots, x_s{}^{(j)})$ $(j = 1, 2, \cdots, \kappa)$ such that if

$$y^{(j)} = f^{(j)} \qquad (j = 1, 2, \cdots, \kappa), \tag{19.11}$$

be the system of equations obtained by inserting successively these κ point in (19.1), the Jacobian

$$\begin{vmatrix} \dfrac{\partial f^{(1)}}{\partial\theta_1} & \dfrac{\partial f^{(1)}}{\partial\theta_2} & \cdots & \dfrac{\partial f^{(1)}}{\partial\theta_\kappa} \\ \dfrac{\partial f^{(2)}}{\partial\theta_1} & \dfrac{\partial f^{(2)}}{\partial\theta_2} & \cdots & \dfrac{\partial f^{(2)}}{\partial\theta_\kappa} \\ \cdots & \cdots & \cdots & \cdots \\ \dfrac{\partial f^{(\kappa)}}{\partial\theta_1} & \dfrac{\partial f^{(\kappa)}}{\partial\theta_2} & \cdots & \dfrac{\partial f^{(\kappa)}}{\partial\theta_\kappa} \end{vmatrix} \not\equiv 0 \qquad \text{for } \theta = \theta^0. \tag{19.12}$$

Since also, by definition, $\partial f/\partial\theta_i$ are continuous functions of the parameters θ, and the x's in (19.11) are constants, it follows from the theory of functional determinants that the system (19.11) can be solved for $\theta_1, \theta_2, \cdots, \theta_\kappa$, and the solution is *unique*, and therefore equal to $\theta_1{}^0, \theta_2{}^0, \cdots, \theta_\kappa{}^0$. Within a certain finite neighborhood of the parameter point θ^0 there are no other parameter points $\neq \theta^0$ satisfying (19.11). This proves Theorem 2.

From this follows immediately

THEOREM 3. *If* $\partial f/\partial\theta_1, \partial f/\partial\theta_2, \cdots, \partial f/\partial\theta_\kappa$ *are linearly independent for every point* θ^0 *in the interior of* D_θ, *then* (19.3) *is at most satisfied for parameter points between which there is a finite distance greater than a certain positive* ϵ.

Thus, in most practical cases the question of arbitrary coefficients can be answered by investigating whether or not the partial derivatives

of f with respect to the θ's are linearly dependent. But for this purpose we need—at least in the more complicated cases—a rule that decides, in a finite number of steps, whether or not such linear dependence exists. In the next section we shall give such a rule.

20. *The Gramian Criterion for Linear Dependence of Functions Extended to Functions of Several Variables*

Let

$$\begin{aligned} U_1 &= U_1(v_1, v_2, \cdots, v_n), \\ U_2 &= U_2(v_1, v_2, \cdots, v_n), \\ &\cdots\cdots\cdots\cdots \\ U_m &= U_m(v_1, v_2, \cdots, v_n), \end{aligned} \tag{20.1}$$

be m real functions of $n > m$ independent real variables, $v_1, v_2, \cdots, v_n$.

Assumption: $U_1, U_2, \cdots, U_m$, are *continuous* functions of the n variables v over a certain closed domain W in the v-space, defined by

$$\underline{a}_i \leqq v_i \leqq \bar{a}_i \qquad (i = 1, 2, \cdots, n) \tag{20.2}$$

where $\underline{a}_i$, $\bar{a}_i$ $(i=1, 2, \cdots, n)$ are $2n$ real numbers.

Consider the expression

$$s = c_1U_1 + c_2U_2 + \cdots + c_mU_m, \tag{20.3}$$

where the c's do *not* depend upon the v's, nor the a's in (20.2). If a system of c's, not all zero, can be found, such that

$$s \equiv 0 \tag{20.4}$$

for *all* values of $v_1, v_2, \cdots, v_n$ in the domain defined by (20.2), the m functions (20.1) are said to be linearly dependent in W.

Let us consider the integral

$$S = \int\int_{(W)} \cdots \int s^2 dv_1 dv_2 \cdots dv_n. \tag{20.5}$$

We have

$$S \geqq 0. \tag{20.6}$$

S is zero when and only when (20.4) is true. Therefore, if a set of c's, not all zero, exists for which (20.4) is satisfied, it must—at the same time—be such a set of c's as makes S a *minimum and equal to zero*. And, conversely, if there is a set of c's, not all zero, such that $S=0$, then (20.4) is fulfilled. The problem of linear dependence is, therefore, reduced to a study of the minimum of S with respect to the c's.

Since not all c's should be zero, we may assume that

$$\sum_{i=1}^{m} c_i^2 = 1. \tag{20.7}$$

We have to find the minimum of S under the side condition (20.7), or, what amounts to the same, to find the unrestricted minimum of

$$S' = S - \lambda \sum_{i=1}^{m} c_i^2, \tag{20.8}$$

where λ is a Lagrange multiplier. Let us introduce the following notations

$$M_{ij} = \int\int_{(W)} \cdots \int U_i U_j (dv_1 dv_2 \cdots dv_n) \tag{20.9}$$
$$(i = 1, 2, \cdots, m; \quad j = 1, 2, \cdots, m).$$

A set of c's minimizing (20.8) must satisfy the following system of linear equations

$$\begin{array}{lllll} (M_{11} - \lambda)c_1 + & M_{12}c_2 & + \cdots + & M_{1m}c_m & = 0, \\ M_{21}c_1 & + (M_{22} - \lambda)c_2 & + \cdots + & M_{2m}c_m & = 0, \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ M_{m1}c_1 & + M_{m2}c_2 & + \cdots + & (M_{mm} - \lambda)c_m & = 0. \end{array} \tag{20.10}$$

(20.10) has a solution of c's not all zero when and only when the determinant formed by the coefficients of the c's is *equal to zero,* i.e.,

$$\begin{vmatrix} (M_{11} - \lambda) & M_{12} & \cdots & M_{1m} \\ M_{21} & (M_{22} - \lambda) & \cdots & M_{2m} \\ \cdots & \cdots & \cdots & \cdots \\ M_{m1} & M_{m2} & \cdots & (M_{mm} - \lambda) \end{vmatrix} = 0. \tag{20.11}$$

Now S is a positive (semi) definite symmetric quadratic form. Therefore, all λ-roots of (20.11) are nonnegative. S has, therefore, a minimum $=0$, for other values of the c's than all zeros, when and only when the *minimal* λ-root of (20.11) is equal to zero. *A necessary and sufficient condition for the linear dependence of the m functions* (20.1) is, therefore, that

$$|M_{ij}| = 0, \tag{20.12}$$

where $|M_{ij}|$ is the determinant (20.11) for $\lambda = 0$.

21. An Illustration of the Problems of Estimation

We shall consider a simple linear supply-demand scheme, including certain random elements and an autonomously imposed sales tax.

Let $\xi_{1t}^{(D)}$ be the quantity demanded at point of time t, $\xi_{1t}^{(S)}$ the quantity supplied, ξ_{2t} the price per unit sold, and ξ_{3t} a sales tax per unit sold, fixed for each point of time independent of the quantity sold. Consider these variables at N equidistant points of time $t=1, 2, \cdots, N$. We shall assume it *known* that these variables satisfy the following system of random equations:

$$\xi_{1t}^{(D)} = \alpha_1\xi_{2t} + \epsilon_{1t} \qquad (t = 1, 2, \cdots, N), \tag{21.1}$$

i.e., a linear demand curve with random shifts ϵ_{1t};

$$\xi_{1t}^{(S)} = \alpha_2(\xi_{2t} - \xi_{3t}) + \epsilon_{2t} \quad (t = 1, 2, \cdots, N) \tag{21.2}$$

i.e., the supply is a linear function of (price minus tax) and a random shift ϵ_{2t}. Further, we impose the market relation

$$\xi_{1t}^{(D)} = \xi_{1t}^{(S)} = \xi_{1t} = \text{quantity sold at } t. \tag{21.3}$$

We assume *known* the following properties of the $2N$ random variables $\epsilon_{11}, \epsilon_{12}, \cdots, \epsilon_{1N}, \epsilon_{21}, \epsilon_{22}, \cdots, \epsilon_{2N}$: (a) They are *independently and normally* distributed and (b) their distribution does not depend upon ξ_{3t}. (c) All the N random variables ϵ_{1t}, $t=1, 2, \cdots, N$, have the same mean $\bar{\epsilon}_1$ and the same variance σ_1^2; likewise, all the N random variables ϵ_{2t}, $t=1, 2, \cdots, N$, have the same mean $\bar{\epsilon}_2$ and the same variance σ_2^2.

Further, we assume that there are errors of measurement in the observations of the quantity sold, ξ_{1t}, such that, instead of ξ_{1t}, we observe

$$x_{1t} = \xi_{1t} + \eta_{1t} \qquad (t = 1, 2, \cdots, N), \tag{21.4}$$

while the price ξ_{2t} and the tax ξ_{3t} are observed without errors, i.e.,

$$x_{2t} = \xi_{2t}, \qquad x_{3t} = \xi_{3t} \qquad (t = 1, 2, \cdots, N). \tag{21.5}$$

We assume that the N random variables $\eta_{11}, \eta_{12}, \cdots, \eta_{1N}$, are independently normally distributed with zero means and the same variance σ^2, and that their distribution does not depend upon the ξ's nor the ϵ's.

The N numbers $\xi_{31}, \xi_{32}, \cdots, \xi_{3N}$, are assumed to remain fixed in repeated samples.

Because of (21.3), *both* ξ_{1t} and ξ_{2t} will be random variables. Indeed, from (21.3), (21.1), and (21.2) we obtain (provided $\alpha_1 \neq \alpha_2$)

$$\begin{aligned} \xi_{1t} &= \frac{\alpha_1\alpha_2}{\alpha_2 - \alpha_1}\xi_{3t} + \frac{\alpha_2\epsilon_{1t} - \alpha_1\epsilon_{2t}}{\alpha_2 - \alpha_1}, \\ \xi_{2t} &= \frac{\alpha_2}{\alpha_2 - \alpha_1}\xi_{3t} + \frac{\epsilon_{1t} - \epsilon_{2t}}{\alpha_2 - \alpha_1}, \end{aligned} \tag{21.6}$$

which shows that both ξ_{1t} and ξ_{2t} are functions of the two independent random variables ϵ_{1t} and ϵ_{2t}.

If we could make experiments to study, separately, the demand function (21.1) and the supply function (21.2), we could reason in this way: (1) For a given value of ξ_{2t}, $\xi_{1t}^{(D)}$ is a random variable with expected value equal to $\alpha_1\xi_{2t}+\bar{\epsilon}_1$. (2) For given values of ξ_{2t} and ξ_{3t}, $\xi_{1t}^{(S)}$ is a random variable with expected value equal to $\alpha_2(\xi_{2t}-\xi_{3t})+\bar{\epsilon}_2$. And we could "fit each of the equations (21.1) and (21.2) separately" to the respective data obtained by the two series of experiments. *But in our case,* because of the market relation (21.3), we cannot assume ξ_{2t} to remain fixed in repeated samples. That would simply be inconsistent with the original assumption that the errors ϵ_1 and ϵ_2 are independent. To realize clearly all the implications of our scheme we have to consider the *joint probability distribution* of the observed variables x_{1t} and x_{2t}, *given* x_{3t}, $t=1, 2, \cdots, N$.

Introducing (21.4) and (21.5) in (21.1) and (21.2), we obtain

$$x_{1t} = \alpha_1 x_{2t} + \epsilon_{1t} + \eta_{1t}, \tag{21.1'}$$

$$x_{1t} = \alpha_2(x_{2t} - x_{3t}) + \epsilon_{2t} + \eta_{1t}, \tag{21.2'}$$

or

$$\begin{aligned} x_{1t} &= \frac{\alpha_1\alpha_2}{\alpha_2 - \alpha_1} x_{3t} + \frac{\alpha_2\epsilon_{1t} - \alpha_1\epsilon_{2t}}{\alpha_2 - \alpha_1} + \eta_{1t}, \\ x_{2t} &= \frac{\alpha_2}{\alpha_2 - \alpha_1} x_{3t} + \frac{\epsilon_{1t} - \epsilon_{2t}}{\alpha_2 - \alpha_1}. \end{aligned} \tag{21.6'}$$

x_{1t} and x_{2t} are jointly normally distributed, because they are linear functions of the normally distributed variables ϵ_{1t}, ϵ_{2t}, η_{1t}. We therefore have, for any fixed point of time t,

$$\begin{aligned} p_t(x_{1t}, x_{2t} \mid x_{3t}) = \frac{1}{2\pi\mu_1\mu_2\sqrt{1-\rho^2}} \exp\left\{-\frac{1}{2(1-\rho^2)}\left[\frac{(x_{1t}-\bar{x}_{1t})^2}{\mu_1^2}\right.\right. \\ \left.\left. - \frac{2\rho(x_{1t}-\bar{x}_{1t})(x_{2t}-\bar{x}_{2t})}{\mu_1\mu_2} + \frac{(x_{2t}-\bar{x}_{2t})^2}{\mu_2^2}\right]\right\}, \end{aligned} \tag{21.7}$$

where μ_1^2, μ_2^2 are the variances of x_{1t} and x_{2t} respectively, $\bar{x}_{1t}$, $\bar{x}_{2t}$ their mean values, and ρ their correlation coefficient, x_{3t} being *given*. From (21.6') we obtain

$$\bar{x}_{1t} = \frac{\alpha_1\alpha_2}{\alpha_2 - \alpha_1} x_{3t} + \frac{\alpha_2\bar{\epsilon}_1 - \alpha_1\bar{\epsilon}_2}{\alpha_2 - \alpha_1}, \tag{21.8}$$

$$\bar{x}_{2t} = \frac{\alpha_2}{\alpha_2 - \alpha_1} x_{3t} + \frac{\bar{\epsilon}_1 - \bar{\epsilon}_2}{\alpha_2 - \alpha_1}, \tag{21.9}$$

$$\mu_1^2 = E(x_{1t} - \bar{x}_{1t})^2 = \frac{1}{(\alpha_2 - \alpha_1)^2}(\alpha_2^2\sigma_1^2 + \alpha_1^2\sigma_2^2) + \sigma^2, \tag{21.10}$$

$$\mu_2^2 = E(x_{2t} - \bar{x}_{2t})^2 = \frac{1}{(\alpha_2 - \alpha_1)^2}(\sigma_1^2 + \sigma_2^2), \tag{21.11}$$

$$\rho = \frac{1}{\mu_1\mu_2} E(x_{1t} - \bar{x}_{1t})(x_{2t} - \bar{x}_{2t}) = \frac{\alpha_2\sigma_1^2 + \alpha_1\sigma_2^2}{(\alpha_2 - \alpha_1)^2\mu_1\mu_2}. \tag{21.12}$$

This shows that only the averages $\bar{x}_{1t}$ and $\bar{x}_{2t}$ depend upon t, while the other parameters are independent of t.

Since the random variables x_{1t}, x_{2t} for one value of t are distributed independently of those at another value of t, the joint distribution of the $2N$ variables $x_{11}, x_{12}, \cdots, x_{1N}, x_{21}, x_{22}, \cdots, x_{2N}$, is

$$p(x_{11}, \cdots, x_{1N}, x_{21}, \cdots, x_{2N}) = \Pi p_t \quad (t = 1, 2, \cdots, N). \tag{21.13}$$

Let us introduce new parameters by the transformations

$$a_1 = \frac{\alpha_1\alpha_2}{\alpha_2 - \alpha_1}, \tag{21.14}$$

$$a_2 = \frac{\alpha_2\bar{\epsilon}_1 - \alpha_1\bar{\epsilon}_2}{\alpha_2 - \alpha_1}, \tag{21.15}$$

$$b_1 = \frac{\alpha_2}{\alpha_2 - \alpha_1}, \tag{21.16}$$

$$b_2 = \frac{\bar{\epsilon}_1 - \bar{\epsilon}_2}{\alpha_2 - \alpha_1}. \tag{21.17}$$

Then (21.8) and (21.9) become

$$\bar{x}_{1t} = a_1x_{3t} + a_2, \tag{21.8'}$$

$$\bar{x}_{2t} = b_1x_{3t} + b_2. \tag{21.9'}$$

Introducing (21.7) in (21.13), and using (21.8') and (21.9'), we obtain

$$p = Ce^Q, \tag{21.18}$$

where

$$Q = -\frac{1}{2(1-\rho^2)} \sum \left[\frac{(x_{1t} - a_1x_{3t} - a_2)^2}{\mu_1^2} - \frac{2\rho(x_{1t} - a_1x_{3t} - a_2)(x_{2t} - b_1x_{3t} - b_2)}{\mu_1\mu_2} + \frac{(x_{2t} - b_1x_{3t} - b_2)^2}{\mu_2^2} \right], \tag{21.19}$$

and

$$C = \frac{1}{(2\pi\mu_1\mu_2)^N(1 - \rho^2)^{N/2}} \tag{21.20}$$

($\sum$ means, throughout this section, $\sum_{t=1}^{N}$).

The distribution (21.18) is, therefore, characterized by 7 unknown parameters, namely a_1, a_2, b_1, b_2, μ_1, μ_2, ρ. If there exists a unique estimate of these 7 parameters, the question of uniqueness of the original parameters depends only upon the transformations (21.10)–(21.12) and (21.14)–(21.17). Now it is easy to see that these transformations establish a one-to-one correspondence between the old and the new parameters over the whole parameter space, except for a trivial set of measure zero (namely $\alpha_1=0$, or $\alpha_2=0$, or $\alpha_1=\alpha_2$). We therefore have to investigate the uniqueness of the parameters in (21.18). This can be done by means of the theorems in Section 19.

The partial derivatives of p [in (21.18)] with respect to the parameters are

$$
\begin{aligned}
&\frac{\partial p}{\partial a_1} = Ce^Q \frac{\partial Q}{\partial a_1}, \qquad \frac{\partial p}{\partial a_2} = Ce^Q \frac{\partial Q}{\partial a_2},\\
&\frac{\partial p}{\partial b_1} = Ce^Q \frac{\partial Q}{\partial b_1}, \qquad \frac{\partial p}{\partial b_2} = Ce^Q \frac{\partial Q}{\partial b_2},\\
&\frac{\partial p}{\partial \mu_1} = e^Q\left(\frac{\partial C}{\partial \mu_1} + C\frac{\partial Q}{\partial \mu_1}\right), \quad \frac{\partial p}{\partial \mu_2} = e^Q\left(\frac{\partial C}{\partial \mu_2} + C\frac{\partial Q}{\partial \mu_2}\right),\\
&\qquad\qquad \frac{\partial p}{\partial \rho} = e^Q\left(\frac{\partial C}{\partial \rho} + C\frac{\partial Q}{\partial \rho}\right).
\end{aligned}
\tag{21.21}
$$

According to Section 19 we are interested in whether these 7 partial derivatives are linearly dependent. If that should be the case, there would have to exist 7 λ's, $\lambda_1, \lambda_2, \cdots, \lambda_7$, which are independent of the variables x, not all zero at the same time, and such that

$$
\begin{aligned}
&\lambda_1 C\frac{\partial Q}{\partial a_1} + \lambda_2 C\frac{\partial Q}{\partial a_2} + \lambda_3 C\frac{\partial Q}{\partial b_1} + \lambda_4 C\frac{\partial Q}{\partial b_2} + \lambda_5\left(\frac{\partial C}{\partial \mu_1} + C\frac{\partial Q}{\partial \mu_1}\right)\\
&\qquad + \lambda_6\left(\frac{\partial C}{\partial \mu_2} + C\frac{\partial Q}{\partial \mu_2}\right) + \lambda_7\left(\frac{\partial C}{\partial \rho} + C\frac{\partial Q}{\partial \rho}\right) \equiv 0
\end{aligned}
\tag{21.22}
$$

for *all* values of the variables x. Since we do not know the true parameter values we are interested in whether there is any parameter point at all for which (21.22) is fulfilled.

From (21.19) and (21.20) we see that the left-hand side of (21.22) will be a second-degree polynomial in the variables x. (21.22) can be fulfilled only if the coefficients of equal terms in this polynomial vanish separately. Using this we verify easily that all the 7 λ's must be *equal to zero*, whatever be the true parameter point (except $\rho=\pm 1$, which is trivial), provided among the set of N constants $x_{31}, x_{32}, \cdots, x_{3N}$, there are at least two that are different.

All the 7 parameters of (21.18) can, therefore, in general be estimated.

We shall consider, in particular, the *maximum-likelihood estimates*[7] of the parameters in (21.18), i.e., the parameter values obtained by setting each of the 7 derivatives in (21.21) equal to zero and solving this system of 7 equations. We obtain the following equations defining the maximum-likelihood estimates (which we denote by $\hat{a}_1$, $\hat{a}_2$, etc.):

$$\sum (x_{1t} - \hat{a}_1 x_{3t} - \hat{a}_2) = 0, \tag{21.23}$$

$$\sum (x_{2t} - \hat{b}_1 x_{3t} - \hat{b}_2) = 0, \tag{21.24}$$

$$\sum (x_{1t} - \hat{a}_1 x_{3t} - \hat{a}_2) x_{3t} = 0, \tag{21.25}$$

$$\sum (x_{2t} - \hat{b}_1 x_{3t} - \hat{b}_2) x_{3t} = 0, \tag{21.26}$$

$$N\hat{\mu}_1^2 - \sum (x_{1t} - \hat{a}_1 x_{3t} - \hat{a}_2)^2 = 0, \tag{21.27}$$

$$N\hat{\mu}_2^2 - \sum (x_{2t} - \hat{b}_1 x_{3t} - \hat{b}_2)^2 = 0, \tag{21.28}$$

$$N\hat{\rho} = \frac{\sum (x_{1t} - \hat{a}_1 x_{3t} - \hat{a}_2)(x_{2t} - \hat{b}_1 x_{3t} - \hat{b}_2)}{\hat{\mu}_1 \hat{\mu}_2} = 0. \tag{21.29}$$

It is easy to verify that this system has, in general, a unique solution with respect to the 7 parameters $\hat{a}_1$, $\hat{a}_2$, $\cdots$ etc. For example, from the first 4 of these equations we obtain

$$\hat{a}_1 = \frac{m_{13} - m_1 m_3}{m_{33} - m_3^2}, \tag{21.30}$$

$$\hat{b}_1 = \frac{m_{23} - m_2 m_3}{m_{33} - m_3^2}, \tag{21.31}$$

where

$$m_{ij} = \frac{1}{N} \sum x_{it} x_{jt}, \qquad m_i = \frac{1}{N} \sum x_{it}. \tag{21.32}$$

These are the same results as we should obtain by writing the "confluent" relations (21.6′) in the form

$$x_{1t} = a_1 x_{3t} + a_2 + \text{error}, \qquad x_{2t} = b_1 x_{3t} + b_2 + \text{error}, \tag{21.6″}$$

[7] The method of maximum likelihood, commonly used by statisticians, was originally founded more or less upon intuition, but recently it has been shown by A. Wald that the method, under certain conditions, can be justified on the basis of modern theory of confidence intervals. See his articles, "A New Foundation of the Method of Maximum Likelihood in Statistical Theory," Cowles Commission for Research in Economics, *Report of Sixth Annual Research Conference on Economics and Statistics . . . 1940*, pp. 33–35, and "Asymptotically Most Powerful Tests of Statistical Hypotheses," *Annals of Mathematical Statistics*, Vol. 12, March, 1941, pp. 1–19.

and fitting each of these equations to the data by the method of least squares, treating x_{1t} and x_{2t}, respectively, as the dependent variable. This would, therefore, also be a correct procedure. But the results (21.30), (21.31) are *not* the same as those we should obtain by fitting the two *original* equations (21.1′) and (21.2′) separately, treating x_{1t} as dependent variable in both equations. For example, from (21.14), (21.16), (21.30), and (21.31) we obtain

$$\hat{\alpha}_1 = \frac{\hat{a}_1}{\hat{b}_1} = \frac{m_{13} - m_1 m_3}{m_{23} - m_2 m_3}, \tag{21.33}$$

while, if we should *fit* (21.1′) *directly* by the method of least squares, we should obtain

$$\alpha_1^* = \frac{m_{12} - m_1 m_2}{m_{22} - m_2^2}, \tag{21.34}$$

which is obviously different from (21.33) since (21.34) does not depend directly upon x_{3t}, while (21.33) does.

α_1^* in (21.34) is simply *not* an estimate of α_1, but something else. The point is this: Consider the equation (21.1′). From this equation we have $E(x_{1t} \mid x_{2t}) = \alpha_1 x_{2t} + E(\epsilon_{1t} \mid x_{2t})$. (21.34) would have been an estimate of α_1 if $E(\epsilon_{1t} \mid x_{2t})$ had been independent of x_{2t} and x_{3t}. But that is *not* the case here. And, therefore, the assumption upon which the least-squares "estimate" (21.34) is based, namely that $E(x_{1t} \mid x_{2t}) = \alpha_1 x_{2t}$ + constant, is here *simply wrong*. In fact, from the joint distribution (21.18) of x_{1t} and x_{2t}, and the transformations (21.10)–(21.12) and (21.14)–(21.17) we obtain easily that $E(x_{1t} \mid x_{2t})$ is a linear function of x_{2t} *and* x_{3t}, namely

$$E(x_{1t} \mid x_{2t}) = \frac{\alpha_2 \sigma_1^2 + \alpha_1 \sigma_2^2}{\sigma_1^2 + \sigma_2^2} x_{2t} - \frac{\alpha_2 \sigma_1^2}{\sigma_1^2 + \sigma_2^2} x_{3t} + \text{const.} \tag{21.35}$$

If we want to *predict* x_{1t}, *given* x_{2t} and x_{3t}, this formula (21.35) is the one to be used. For that purpose we may, if we like, write (21.35) as $E(x_{1t} \mid x_{2t}) = Ax_{2t} + Bx_{3t} + C$, and fit this equation directly to the data by the method of least squares. That gives the same result as if we first estimate all the coefficients in (21.35) by the method of maximum likelihood as described above, and then insert these estimates in (21.35).

Thus, we see that the method of least squares applied to the *original* equations (21.1′) and (21.2′) separately, *neither* gives correct estimation formulae for the coefficients, *nor* does it give the correct formulae for prediction. This shows the importance of studying the *joint* distribution of *all* the observable random variables in a system of stochastic relations.

CHAPTER VI

Problems of Prediction

A statistical prediction means simply a (probability) statement about the location of a sample point not yet observed. If we consider n random variables, say $x_1, x_2, \cdots, x_n$, and if we know their joint probability law we may, at least in point of principle, calculate the probability of a sample point $(x_1, x_2, \cdots, x_n)$ falling into any given region or point-set of the sample space, or we may prescribe a certain fixed probability level and derive a system of regions (or point-sets) which have this probability. In practice we should then usually be interested in that region which, at a given probability level, is the "smallest" (in some sense or another). Thus, if we actually knew the joint probability law of the variables to be predicted, the problem of deriving a prediction formula having certain desired properties would merely be one of probability calculus. And the question of choosing a "best" prediction formula would, largely, be a subjective matter, that is, a question of the type of "gambling" we should be willing to accept.

Usually, however, we do not know the probability law of the variables to be predicted. Then the problem of prediction becomes one closely connected with the problems of testing hypotheses and estimation. For we then have to draw inference concerning the probability law of the variables to be predicted from samples already observed. We shall attempt to give a more general and rigorous formulation of these problems.

22. General Formulation of the Problem of Prediction

Consider n sequences or time series of random variables x_{it} $(i=1, 2, \cdots, n)$ observable from $t=1$ on. Values, if any, of the variables prior to $t=1$ we shall here consider as given constants. Suppose that we can observe values of each series up to a certain point of time. Let $t=s_i$ be this point of time for the ith series. And suppose that the problem is to predict the results of later observations not yet made. We then have the following schedule of random variables to be considered.

$$x_{i,t} = x_{i,1}, x_{i,2}, \cdots, x_{i,s_i}, x_{i,s_i+1}, x_{i,s_i+2}, \cdots \qquad (i = 1, 2, \cdots, n). \tag{22.1}$$

$x_{1,t}, x_{2,t}, \cdots, x_{n,t}$, may for example be n related economic time series, $t=s_i$ denoting the latest point of time for which an observation of $x_{i,t}$ is, so far, available. We might want to predict the next value

in one or more of the series, or the second next, or both, or any other joint system of future values of the variables not yet observed. Consider any system of M variables chosen among the variables $x_{i,s_i+\tau}$ $(i=1, 2, \cdots, n; \tau=1, 2, 3, \cdots)$. Together with the $s_1+s_2+\cdots+s_n=N$ observed variables the variables to be predicted form a system of $N+M$ random variables. Let us, for simplicity, change notations of these variables, denoting the N observable variables by $x_1, x_2, \cdots, x_N$, and the M variables to be predicted by $x_{N+1}, x_{N+2}, \cdots, x_{N+M}$, so that there is a one-to-one correspondence between these variables and the $N+M$ variables $x_{i,t}$ considered.

The problem of prediction is then the problem of establishing certain functions of the observable variables $x_1, x_2, \cdots, x_N$, that may be used as guess values for the outcome of the future observations of $x_{N+1}, x_{N+2}, \cdots, x_{N+M}$.[1]

We shall assume that, whatever be $s_1, s_2, \cdots, s_n$, and whatever be the set of M future variables considered, the joint elementary probability law of the $N+M$ variables $x_1, x_2, \cdots, x_N, x_{N+1}, \cdots, x_{N+M}$ exists. (But it might not be—and usually is not—known.) Let this joint probability be denoted by $p(x_1, x_2, \cdots, x_N, x_{N+1}, \cdots, x_{N+M})$, or, for short, p. This probability law would usually be described implicitly by a system of stochastical relations between the variables considered, as explained in Chapters IV and V.

Let us for a moment suppose that p is known. From p we might then calculate the conditional elementary probability law of the M variables $x_{N+1}, \cdots, x_{N+M}$, *given* the N variables $x_1, x_2, \cdots, x_N$. Let this conditional probability law be denoted by $p_2(x_{N+1}, \cdots, x_{N+M} \mid x_1, x_2, \cdots, x_N)$, or for short, p_2. Let $p_1(x_1, x_2, \cdots, x_N)$, or for short, p_1, denote the joint probability law of the N observable variables. We may then write

$$p = p_1 \cdot p_2. \tag{22.2}$$

Let, further, E_1 denote any particular system of values—one for each—of the observable variables $x_1, x_2, \cdots, x_N$; and, similarly, let E_2 denote any system of values of the future variables $x_{N+1}, \cdots, x_{N+M}$. Any E_1 may be represented by a point in the N-dimensional sample space R_1 of the variables $x_1, x_2, \cdots, x_N$; and, similarly, any E_2 may be represented by a point in the M-dimensional sample space R_2 of the variables $x_{N+1}, \cdots, x_{N+M}$ to be predicted. Finally, let E denote a point in the sample space R of all $N+M$ variables.

Now, given any particular E_1, we may from p_2 calculate the conditional probability of E_2 falling into a prescribed point-set of the sample space R_2. This probability would usually be a function of E_1. Also,

[1] See, e.g., Harold Hotelling, "Problems of Predictions," *The American Journal of Sociology*, Vol. 48, July, 1942, pp. 61–76.

for any given E_1 and for any given level of probability, β say, we may derive a system of point-sets or regions in R_2, such that the probability of E_2 falling into any particular one of these sets is β. That is, we may predict, with probability $=\beta$ of being correct, that E_2 will fall into any particular one among these point-sets. Any such point-set in R_2 we shall call a *region of prediction*, and we shall denote such a region by W_2.

In general, however, not all the regions W_2 of probability β are equally "interesting." Usually (though not always) we are interested in that region, with probability β, which is the "narrowest," in some sense or another. Or, we might also be interested in predicting that the sample point E_2 will *not* fall within a certain region. In any case the choice of the probability level β and of the location of that region W_2, with probability β, which we want to use as a prediction formula will depend on the practical use we want to make of it. This choice is not a statistical problem. We shall simply assume that, whatever be the conditional probability law p_2, the purpose of our attempts to predict will lead us to one and only one region W_2 of predicting E_2, for every set of values of the "predictors" $x_1, x_2, \cdots, x_N$.

If, therefore, we knew p_2 the problem of prediction would merely be a problem of probability calculus, and not one of statistical inference from a sample. But in most practical cases p_2 is not known, and we then have to try to get information about p_2 from samples E_1 of the previous observations. The possibility of doing so rests upon a basic assumption, which can be formulated as follows: *The probability law p of the $N+M$ variables $x_1, x_2, \cdots, x_N, x_{N+1}, \cdots, x_{N+M}$ is of such a type that the specification of p_1 implies the complete specification of p and, therefore, of p_2.*

For instance, if p is characterized by a certain number of unknown parameters, then all these parameters must also be the characteristics of p_1 so that p_2 will contain no new parameters in addition to those occurring in p_1. This is only another, more precise, way of stating that, in order to be able to predict there must be a certain persistence in the type of mechanism that produces the series to be predicted.

Suppose now that the only thing known about p_1 is that it belongs to a certain specified class Ω_1 of elementary probability laws, and that, therefore, p_2 belongs to a certain corresponding class Ω_2. Let p_1^* denote any arbitrary member of Ω_1. And let $W_1(p_1^*)$ be a critical region, of size $(1-\alpha)$, in R_1, chosen according to some rule, such that the hypothesis $p_1=p_1^*$ is rejected when and only when E_1 falls into $W_1(p_1^*)$. Let there be established a system of such critical regions in R_1, one for every member p_1^* of Ω_1. If E_1 falls outside $W_1(p_1^*)$ then $p_1=p_1^*$ is not rejected. If the system of regions $W_1(p_1^*)$ is not to be trivial, any sample point E_1 will fall outside *some* of the regions $W_1(p_1^*)$. E_1 being an arbitrary sample point of the N observable variables, let $\omega(E_1)$ be the sub-

set of Ω_1 in whose critical regions [of size $(1-\alpha)$] E_1 does *not* fall. As explained in Section 14 it then seems reasonable to *estimate* the unknown probability law p_1 on the basis of E_1 by stating that $p_1 \in \omega(E_1)$. Now, we have assumed above that, for every member p_1^* of Ω_1 (or—what is the same—for every p_2^* of Ω_2) and for every set of values of $x_1, x_2, \cdots, x_N$, our choice of prediction formula leads to one and only one region of prediction W_2^*, of size β. To the subclass $\omega(E_1)$ there therefore corresponds a certain subclass of such regions of prediction. Let $K(E_1)$ be the (logical) sum of all the elements W_2^* of this subclass. It might then seem reasonable to predict E_2, on the basis of the sample point E_1, by stating that

(22.3) $$E_2 \text{ will fall into } K(E_1).$$

What is the probability of this statement being true? Let $g[K \mid p_1 \in \omega(E_1)]$, or, for short, $g(K)$ be the probability of E_2 falling into K when $p_1 \in \omega(E_1)$. And let $\bar{g}\{K \mid p_1 \in [\Omega_1 - \omega(E_1)]\}$, or, for short, $\bar{g}(K)$, be the probability of E_2 falling into K when p_1 is outside $\omega(E_1)$. The probability, $P(E_2 \in K)$, of (22.3) being true is then evidently

(22.4) $$P(E_2 \in K) = \alpha g(K) + (1 - \alpha)\bar{g}(K),$$

i.e., the probability of (22.3) being true is the probability of $\omega(E_1)$ covering p_1 times the probability that E_2 then falls into K plus the probability that $\omega(E_1)$ does not cover p_1 times the probability that E_2 then falls into K. Now, the probabilities $g(K)$ and $\bar{g}(K)$ will, in general, be functions of the true distribution p_1. But we may give inequalities for $P(E_2 \in K)$. Evidently $1 \geqq g(K) \geqq \beta$, while $0 \leqq \bar{g}(K) \leqq 1$. Therefore,

(22.5) $$1 \geqq P(E_2 \in K) \geqq \alpha\beta.$$

(For particular Ω_1's there might exist narrower limits.)

The procedure just described might also be looked upon in the following way: We have assumed that to every member p_1^* of Ω_1 there is a certain region of prediction W_2^* which we should use if p_1^* were the true distribution. If p_1^* is the true distribution the probability that $K(E_1)$ shall cover the corresponding region of prediction is evidently equal to α. Therefore, $K(E_1)$ may be considered as a confidence region, with confidence coefficient α, for estimating the location of the "ideal" region of prediction W_2 corresponding to the true hypothesis.

The usual problem in practice is, however, to derive regions of prediction for E_2 which, with a given probability level, are as "small" as possible. Then the regions K derived as described above might not necessarily be the "best" regions to choose. More precisely, if for a given β the regions W_2^* were the "smallest" regions (according to some measure), and if the confidence sets $\omega(E_1)$ were the "smallest" confidence

sets, the question of whether or not the corresponding $K(E_1)$, measured in the same measure as the regions W_2^* would be the "smallest" region of prediction would depend on the way in which the term "smallest" is defined with respect to $\omega(E_1)$. Or, expressed in simpler terms, the choice of a particular system of confidence sets for *estimating* p_1 depends on some system of weights of the type of errors that might be committed by stating that $\omega(E_1)$ will cover p_1. If, on the other hand, the purpose is to derive a region of *prediction* $K(E_1)$, a *different* weighting of the errors of estimate might be necessary in order to arrive at the desired weighting of the possible errors of prediction.

We see therefore that the seemingly logical "two-step" procedure of first estimating the unknown distribution of the variables to be predicted and then using this estimate to derive a prediction formula for the variables may not be very efficient. We shall discuss a simpler and more direct method of deriving prediction formulae that avoids the difficulties discussed above.

Let E_2 denote any point in the sample space R_2 of $x_{N+1}, \cdots, x_{N+M}$, and let $\overline{E}_2$ denote a point in R_2 to be used as a prediction of E_2. We consider the problem of defining $\overline{E}_2$ as a function of $x_1, x_2, \cdots, x_N$, in such a way that the probability will be high that $\overline{E}_2$ will be close to E_2 (in some sense or another). We shall call $\overline{E}_2$ a *prediction function.* If we state that E_2 will coincide with $\overline{E}_2$ and this does not occur, we commit an error the consequences of which will depend on the purpose of the prediction. Using an idea of A. Wald[2] we might assign a system of weights to the various possible errors. Let this system be defined by a weight function $Q(E_2, \overline{E}_2)$, such that $Q=0$ if $E_2=\overline{E}_2$ and $Q\geqq 0$ (and not identically zero) for all points $E_2\neq\overline{E}_2$. Q might be considered as the "loss" incurred if $E_2\neq\overline{E}_2$. The *expected value* r of this loss, in repeated samples, is given by

$$r = \int_R Q(E_2, \overline{E}_2) p dE, \tag{22.6}$$

the integral being taken over the whole sample space R of the $N+M$ variables $x_1, x_2, \cdots, x_N, x_{N+1}, \cdots, x_{N+M}$. We have to choose $\overline{E}_2$ as a function of $x_1, x_2, \cdots, x_N$, and we should, naturally, try to choose $\overline{E}_2(x_1, x_2, \cdots, x_N)$ in such a way that r (the "risk") becomes as small as possible.

Suppose there should exist a prediction function $\overline{E}_2(x_1, x_2, \cdots, x_N)$, depending on $x_1, x_2, \cdots, x_N$ only, such that for this particular function r would be at a minimum, independently of what be the true distribu-

[2] See A. Wald, "Contributions to the Theory of Statistical Estimation and Testing Hypotheses," *Annals of Mathematical Statistics*, Vol. 10, December, 1939, pp. 299–326.

tion p_1 (within Ω_1). Then we should naturally choose this function at the best prediction relative to the given weight-function Q. We might call such a prediction function "uniformly best (within Ω_1) relative to the given weight function."

In a few simple cases such prediction functions might exist. In general, however, we may expect that no uniformly best prediction function exists. Then we have to introduce some additional principles in order to choose a prediction function. We may then, first, obviously disregard all those prediction functions that are such that there exists *another* prediction function that makes r smaller for every member of Ω_1. If this is not the case we call the prediction function considered an *admissible* prediction function. To choose between several admissible prediction functions we might adopt the following principle, introduced by Wald: For every admissible prediction function $\overline{E}_2$ the "risk" r is a function of the true distribution p. Consider that prediction function $\overline{E}_2$, among the admissible ones, for which the *largest* value of r is at a *minimum* (i.e., smaller than or at most equal to the largest value of r for any other admissible $\overline{E}_2$). Such a prediction function, if it exists, may be said to be the *least risky* among the admissible prediction functions. The problem of deriving such prediction functions is closely related to the similar problem of deriving best estimates.[3]

23. Some Practical Suggestions for the Derivation of Prediction Formulae

From the discussion just concluded it is seen that the choice of a prediction formula cannot, in general, be made entirely on objective grounds. The choice of the weight function Q, for instance, is not an objective statistical problem. Also, the choice of a prediction formula when no uniformly best prediction formula exists is a more or less subjective matter. The advantage of the formal procedure we have outlined is, however, that it describes precisely where and how the subjective elements come into the picture, and what their logical consequences are. The apparatus described gives us more efficient tools for forming the prediction functions according to our wish. Thus, for instance, the notion of a weight function Q is useful in the sense that, if we should choose a prediction functon more or less arbitrarily (by a freehand method, let us say), the corresponding weight function that would make this arbitrary choice the "best" might be such that we would not accept it. That is, we should realize that the arbitrarily chosen prediction function was not very good after all.

[3] For a discussion of the problems of prediction within a model of linear stochastic difference equations see Mann and Wald *op. cit.*, pp. 192–202.

A practical rule, perhaps not generally recognized, in dealing with several time sequences simultaneously is the following: If we want to predict future values for one or more of the sequences it is usually necessary to derive the prediction formulae on the basis of the *joint* distribution of the observable elements in *all* the series. That is, we have to take into account, not only the serial, stochastical, dependence between successive observations in one and the same sequence, but also the interdependence, if any, between the various sequences considered. The situation is here similar to the situation in regard to estimation of unknown parameters, as discussed in Chapter V.[4]

The apparatus set up in the preceding section, although simple in principle, will in general involve considerable mathematical problems and heavy algebra. There are, however, important cases where more simple procedures will be sufficient. We should like to suggest one such procedure that might be applied with success in certain ordinary cases occurring frequently in econometrics and other types of statistical research.

Suppose we have a case where the following assumptions are fulfilled (using here the notations of Section 22):

1. The distribution p_1 of $x_1, x_2, \cdots, x_N$ is known to belong to a parametric family of distributions, involving the unknown parameters $\alpha_1, \alpha_2, \cdots, \alpha_k$, i.e., we may write $p_1 = p_1(x_1, x_2, \cdots, x_N; \alpha_1, \alpha_2, \cdots, \alpha_k)$, or, for short, $p_1[E_1; (\alpha)]$.

2. The distribution p of all the $N+M$ variables considered is obtained simply by substituting $N+M$ for N in p_1, N and M being arbitrary positive integers (except, perhaps, that N may have to be larger than a certain positive integer, say N_0). p_2 is, therefore, also known, except for the values of the parameters α.

3. It is established that the maximum-likelihood estimates of the α's derived from $p_1[E_1; (\alpha)]$ for an observed sample E_1 are unbiased and converge stochastically to the true parameter values with increasing N, and that these estimates are "good" estimates also for moderate size of N.

Consider the "conditional risk" $\bar{r}$ defined by

$$\bar{r} = \int_{R_2} Q(E_2, \overline{E}_2) p_2[E_2; (\alpha) \mid E_1]\, dE_2. \tag{23.1}$$

For fixed E_1 we may consider $\bar{r}$ as a function of $\overline{E}_2$. We might then proceed as follows, to derive the prediction function $\overline{E}_2 = \overline{E}_2(x_1, x_2, \cdots, x_N)$:

[4] For further discussion of this particular problem see the author's article, "Statistical Implications of a System of Simultaneous Equations," ECONOMETRICA, Vol. 11, January, 1943, pp. 1–12. See also the discussion by H. B. Mann and A. Wald, *op. cit.*, pp. 215–216.

I. Find that point $\overline{E}_2$ which, for a given set of α's and a given sample of $x_1, x_2, \cdots, x_N$, makes $\bar{r}$ a minimum (assuming that such a minimum exists). The point $\overline{E}_2$ corresponding to this minimum of $\bar{r}$ will, in general, be a function of the α's and the observable variables $x_1, x_2, \cdots, x_N$. Denoting this function by $\overline{\overline{E}}_2$ we may therefore write

$$\overline{\overline{E}}_2 = \overline{\overline{E}}_2(x_1, x_2, \cdots, x_N; \alpha_1, \alpha_2, \cdots, \alpha_k).$$

II. In the function $\overline{\overline{E}}_2$ insert for the α's their maximum-likelihood estimates $\hat{\alpha}_1, \hat{\alpha}_2, \cdots, \hat{\alpha}_k$, as derived from the observations $x_1, x_2, \cdots, x_N$ and the distribution p_1. The resulting prediction formula $\overline{\overline{E}}_2 = \overline{\overline{E}}_2(x_1, x_2, \cdots, x_N; \hat{\alpha}_1, \hat{\alpha}_2, \cdots, \hat{\alpha}_k)$ then contains only known elements and is, therefore, determined.

This procedure can be shown to lead to the same prediction formulae, in certain ordinary cases, as those which are already established as "best" on the basis of the general theory of statistical estimation. We shall give an example.

Consider a sequence of random variables defined by the recurrence formula

$$x_t = kx_{t-1} + \epsilon_t \qquad (t = 1, 2, \cdots), \tag{23.2}$$

where x_0 is a given constant, while k is an unknown constant, and where the ϵ's are independently, normally distributed with means equal to zero, and the same variances, equal to σ^2. Suppose we have observed x_t up to and including x_N and we want to predict x_{N+1} and x_{N+2}. Assume further that we have chosen a weight function of the following type:

$$\begin{aligned} Q = {} & a(x_{N+2} - \bar{x}_{N+2})^2 + 2b(x_{N+2} - \bar{x}_{N+2})(x_{N+1} - \bar{x}_{N+1}) \\ & + c(x_{N+1} - \bar{x}_{N+1})^2, \end{aligned} \tag{23.3}$$

where $\bar{x}_{N+1}$ and $\bar{x}_{N+2}$ denote the predicted values of x_{N+1} and x_{N+2}, and where $a>0$, b, and c are certain known constants, such that $ac>b^2$. (That is, the weight of an error in prediction is constant along an ellipse, with center at $\bar{x}_{N+1}, \bar{x}_{N+2}$.)

The joint distribution of x_{N+1} and x_{N+2}, *given* the preceding x's, is

$$p_2 = \frac{1}{2\pi\sigma^2} e^{-(1/2\sigma^2)Y}, \tag{23.4}$$

where

$$Y = (x_{N+1} - kx_N)^2 + (x_{N+2} - kx_{N+1})^2. \tag{23.5}$$

The conditional expectation of Q is then [see (23.1)],

$$\bar{r} = \int_{-\infty}^{+\infty} \int_{-\infty}^{\infty} \frac{1}{2\pi\sigma^2} Q e^{-(1/2\sigma^2)Y} dx_{N+1} dx_{N+2}. \tag{23.6}$$

Minimizing $\bar{r}$ with respect to $\bar{x}_{N+1}$ and $\bar{x}_{N+2}$ we obtain the following two equations for $\bar{x}_{N+1}$ and $\bar{x}_{N+2}$,

$$\begin{aligned} a\bar{x}_{N+2} + b\bar{x}_{N+1} &= ak^2x_N + bkx_N, \\ b\bar{x}_{N+2} + c\bar{x}_{N+1} &= bk^2x_N + ckx_N, \end{aligned} \tag{23.7}$$

which give

$$\begin{aligned} \bar{x}_{N+1} &= kx_N, \\ \bar{x}_{N+2} &= k^2x_N, \end{aligned} \tag{23.8}$$

independently of the values of a, b, and c. That is, the "best" prediction values relative to the weight function (23.3) are the expected values of x_{N+1} and x_{N+2}. But we do not know k. Its maximum-likelihood estimate $\hat{k}$ is, however,

$$\hat{k} = \frac{\sum_{t=1}^{N} x_t x_{t-1}}{\sum_{t=1}^{N} x_{t-1}^2}. \tag{23.9}$$

Our prediction formulae are therefore, according to the principle adopted,

$$\begin{aligned} \bar{x}_{N+1} &= \hat{k}x_N, \\ \bar{x}_{N+2} &= \hat{k}^2x_N. \end{aligned} \tag{23.10}$$

To judge the reliability of the prediction we may, e.g., consider the probability of $(x_{N+1}-\bar{x}_{N+1})$ and $(x_{N+2}-\bar{x}_{N+2})$ being within certain bounds, the variables $\bar{x}_{N+1}$ and $\bar{x}_{N+2}$ being defined by (23.9) and (23.10); or, we could simply study the values of the risk, as calculated from (22.6).

Conclusion

The patient reader, now at the end of our analysis, might well be left with the feeling that the approach we have outlined, although simple in point of principle, in most cases would involve a tremendous amount of work. He might remark, sarcastically, that "it would take him a lifetime to obtain one single demand elasticity." And he might be inclined to wonder: Is it worth while? Can we not get along, for practical purposes, by the usual short-cut methods, by graphical curve-fitting, or by making fair guesses combining our general experiences with the inference that appears "reasonable" from the particular data at hand?

It would be arrogant and, indeed, unjustified to condemn all the short-cut methods and the practical guesswork which thousands of economists rely upon in their daily work as administrators or as advisers to those who run our economy. In fact, what we have attempted to show is that this kind of inference actually is based, implicitly and perhaps subconsciously, upon the same principles as those we have tried to describe with more precision in our analysis. We do, however, believe that economists might get more useful and reliable information (and also fewer spurious results) out of their data by adopting more clearly formulated probability models; and that such formulation might help in suggesting what data to look for and how to collect them. We should like to go further. We believe that, if economics is to establish itself as a reputable quantitative science, many economists will have to revise their ideas as to the level of statistical theory and technique and the amount of tedious work that will be required, even for modest projects of research. On the other side we must count the time and work that might be saved by eliminating a good deal of planless and futile juggling with figures. Also, it is hoped that expert statisticians, once they can be persuaded to take more interest in the particular statistical problems related to econometrics, will be able to work out, explicitly, many standard formulae and tables. One of the aims of the preceding analysis has been to indicate the kind of language that we believe the economist should adopt in order to make his problems clear to statisticians. No doubt the statisticians will then be able to do their job.

In other quantitative sciences the discovery of "laws," even in highly specialized fields, has moved from the private study into huge scientific laboratories where scores of experts are engaged, not only in carrying out actual measurements, but also in working out, with painstaking

precision, the formulae to be tested and the plans for the crucial experiments to be made. Should we expect less in economic research, if its results are to be the basis for economic policy upon which might depend billions of dollars of national income and the general economic welfare of millions of people?