Cowles Foundation Paper 32

Reprinted from the Proceedings of the International Statistical Conferences, 1947, Vol. III.

STATISTICAL INFERENCE FROM NON-EXPERIMENTAL OBSERVATION: AN ECONOMIC EXAMPLE*

by Jacob Marschak

Cowles Commission for Research in Economics
University of Chicago (United States)

Suppose we want to predict the extent of shortage (or surplus) which will occur if the government fixes the price of a commodity at a certain level. At this price level, supply of the commodity may be equal to demand, but it may also exceed or fall short of demand. When I say that we want to predict the extent of shortage, or of surplus, of the commodity *i.e.*, the difference between demand and supply, resulting from fixing the price at a given level, I use the term prediction in a sense different from that in which Mr. Stone used the term a few minutes ago. When discussing M. Fréchet's report on the results of the questionnaire on the scope of probability theory in social science, Mr. Stone said that, because of the stochastic character of economics, economic prediction is not possible. That is true under a narrow definition of prediction, a definition which if applied rigorously, would exclude prediction from almost all empirical sciences, inasmuch as almost all empirical laws are, strictly speaking, stochastic relations similar to the economic relations that I shall treat in my example.

Let us define *prediction* as the estimation of the probability distribution of a certain random variable or of the joint probability distribution of a set of random variables (the "predictand" set), given the values of certain other variables (the "predictor" set). In that sense, the shortage which we shall denote by y, can be predicted if we can estimate its conditional distribution, given p^*, the price fixed by the government. The best known, though not the only possible technique of estimating any distribution is the "parametric" estimation: one or more parameters of the distribution are estimated on the basis of statistical data, after having used some previous knowledge independent of those data, to specify in advance other parameters, and the form itself, of the distribution function. Similarly we can speak of the *parametric estimation of a stochastic relation* such as the relation

$$f(y, p^*) = v, \tag{1}$$

where v is a non-observable random "disturbance" with zero-mean.

* This article will be included in Cowles Commission Papers, new series, no. 32.

The form of the distribution function of v, as well as the form of the function f relating v to the observable variables y, p^*, have to be specified *a priori* before proceeding to the estimation of the parameters of these functions on the basis of observations. A particular case of (1) is the *prediction equation* of y for a given p^*, such as

$$(2) \qquad y = R(p^*) + v,$$

where v is a "prediction error", distributed normally with zero mean and with unknown variance σ_{vv} while $R(\ )$, a "regression function," is specified with respect to its form (*e.g.*, linear) but not with respect to all its parameters. It follows that, for every fixed value of p^*, the predictand y will have a normal distribution characterized by the expectation (mean) of y given p^*,

$$(3) \qquad \mathcal{E}(y \mid p^*) = R(p^*),$$

and by the variance of y given p^*, *i.e.*, the square of "the standard error of the estimate,"

$$(4) \qquad \sigma^2 (y \mid p^*) = \sigma_{vv}.$$

Hence the estimation of the unknown parameters of the regression function $R(\ )$, and of the unknown parameter σ_{vv} of the distribution of v, will suffice to estimate the stochastic "prediction relation" (2), and also to predict the variable y.

I shall now try to show that useful prediction requires different procedures according to whether experiments are or are not possible. Economics will be shown to require techniques not developed in experimental sciences. Specifically: the economist may have to predict the results of introducing price control, not on the basis of experiments with price control but on the basis of observations which were made in the absence of price control.

Denote by x^d_t the demand, and by x^s_t the supply, at time t. Hence, shortage $y_t = x^d_t - x^s_t$. Let p_t be the price which is either the free market price or the government fixed price as the case may be. Suppose that for any t, the behavior of buyers and that of sellers is described, respectively, by

$$(5) \qquad x^d_t = u^d_t + L(p_t, z_t; \alpha^d): \quad \text{demand equation; and}$$

$$(6) \qquad x^s_t = u^s_t + L(p_t, z_t; \alpha^s): \quad \text{supply equation.}$$

Here both demand and supply depend on price and, in addition, on a whole set of "*predetermined*" *variables* such as weather, tax-rates, past crops, etc. Some of these variables may appear in both equations, others in only one of the two equations. We denote the set of all predetermined given variables by a single symbol z_t. Both the demand and the supply functions are linear, as indicated by the symbol $L(\ )$; but they have different and, in general, unknown coefficient-sets, α^d and α^s respectively. For brevity, we shall also use a single symbol $\alpha \equiv (\alpha^d, \alpha^s)$. Finally, the behavior of both sellers and buyers is stochastic: we suppose both to be subject to "random shifts," or disturbances, u_t^d, u_t^s. These are non-observable random variances and covariances forming a matrix

$$\sigma \equiv \begin{Vmatrix} \sigma^{dd} & \sigma^{ds} \\ \sigma^{ds} & \sigma^{ss} \end{Vmatrix}.$$

Here σ^{ds} need not be zero since buyers' and sellers' behavior shifts may be correlated. The average "violence" of these disturbances is measured by σ^{dd}, σ^{ss}, respectively.

The parameters α of the demand and supply functions are properties of behavior or technology; so are the distribution parameters σ. All these parameters are necessary to describe the "structure" of the market. They are, however, not sufficient. The stochastic system (5), (6) consists of only two equations in the three variables (x_t^d, x_t^s, p_t) that are not determined by outside forces. Therefore, the values of these variables are not yet determined. The system becomes complete if supplemented by a third equation. Under price-control, this equation is

$$(7') \qquad p_t = p_t^*,$$

where p_t^* is some constant chosen by the government for the time t. In a free market we may have, instead,

$$(7'') \qquad x_t^s = x_t^d$$

if we assume that sellers adjust production instantaneously to wipe out the shortage or surplus, shown by decreasing or increasing stocks on hand—the simple static theory of economic textbooks which at best is a first approximation. (A somewhat more realistic, dynamic theory might introduce a positive lag λ, as in $x_{t+\lambda}^s = x^d$; or in $p_{t+\lambda} - p_t = k(x_t^d - x_t^s)$, $k > 0$.)

We can combine equations (7′) and (7″) into a single one,

$$(7) \qquad (x_t^d - x_t^s)\delta + (p_t - p_t^*)\varepsilon = 0,$$

$$\text{where } (\delta, \varepsilon) \begin{cases} = (1, 0) \text{ in free market;} \\ = (0, 1) \text{ under price control.} \end{cases}$$

The "structure" of the market is then described by the numerical values of the parameters (α, σ, δ, ε). We see that structure depends on men's behavior, technology, legal rules.

Now let me proceed to our problem of predicting the extent of shortage, $y_t = x^d_t - x^s_t$, as the result of fixing the price at $p_t = p_t^*$. Usually, observations will have been collected under the condition that $(\delta, \varepsilon) = (1,0)$. But prediction is to be made for the situation $(\delta, \varepsilon) = (0, 1)$. There is a change of structure—a *known* change to be sure—between the time of observed and the time of predicted events.

This is characteristic of the non-experimental case. I shall call it Case B. When experiments are possible I shall speak of Case A. In our economic problem, an imaginable experiment would consist in deliberately fixing the price at various values and observing the results. I shall briefly study this imaginable Case A before proceeding to the actual, non-experimental Case B.

Let us, then, assume that we can fix prices, from experiment to experiment, at varying values $p_1^*, p_2^*, \ldots p_T^*$. That is $(\delta,\varepsilon) = (0,1)$ for $t = 1, \ldots, T$. Our object is to predict the shortage, say, $y_{T+1} = x^d_{T+1} - x^s_{T+1}$, when, at time $T+1$, the price is fixed at p_{T+1}^*. This prediction is obtained by at least two, logically distinct procedures. Both occur in experimental sciences. I shall call the two procedures, the *specific experiment* and the *general experiment*, and denote them, respectively, by A1 and A2. Examples of specific experiment are the testing of model bridges, or the testing of airplanes in aerodynamic tunnels. Here the experiment consists in reproducing the specific structural change whose effect one wants to predict. The original structure is not the subject of investigation. General experiments, on the other hand, are designed, not to predict the results of a single specific structural change, but to make possible the prediction of the results of any possible structural change. Experiments made in physical laboratories to establish fundamental laws of physics or chemistry are of this kind. The time required to boil an egg can be estimated by boiling eggs; but, in principle, it can also be estimated by studying first the general properties of protein molecules, in which case not only the answer of the egg-boiling question, but of many other questions as well, will be prepared.

In our economic case, a *specific experiment*, (A1) would run as follows. The experiment—conducted, say, in a small community—replaces the random variables p_t in (5), (6) by constants p_t^*, for t running from 1 through $T+1$. Therefore, if we define

(8) $$v_t = u^d_t - u^s_t \text{ and}$$

(9) $$\beta = \alpha^d - \alpha^s, \quad \text{we have}$$

(10) $$y_t = v_t + L(p_t^*, z_t; \beta), \qquad t = 1, \ldots, T+1.$$

Since v_t is, by (8), normally distributed with zero-mean, equation (10) is a prediction equation, exactly like (2). As in (3), (4), the conditional mean and conditional variance of the predictand y_t, are functions of given values of the predictor set (p_t^*, z_t):

$$\text{(11)} \qquad \mu_{y_t} \equiv \mathcal{E}(y_t \mid p_t^*, z_t) = L(p_t^*, z_t; \beta)$$

$$\text{(12)} \quad \sigma_{yy} \equiv \mathcal{E}\left\{ (y_t - \mu_{y_t})^2 \,\middle|\, p_t^*, z_t \right\} = \sigma_{vv} = \sigma^{dd} + \sigma^{ss} - 2\sigma^{ds}, \text{ by (8).}$$

We see that the distribution parameters μ_{y_t}, σ_{yy} of the predictand are obtained *without estimating the structural parameters* α, σ. Instead we obtain unbiased and efficient estimates of β and σ_{vv} directly from the observations $(y_t, p_t^*, z_t; t = 1, \ldots, T)$ by minimizing $\sum_1^T v_t^2$ with respect to β and σ_{vv}: (the method of least squares). We can thus estimate, for given p_{T+1}^* and z_{T+1}, the conditional mean of y_{T+1} by (11); (and the conditional variance $\sigma_{yy} = \sigma_{vv}$. That is, we can predict y_{T+1}, without estimating the structural parameters α, σ.

We can also use our economic example to illustrate the experimental procedure A2, the use of *general experiment*. Structural parameters are estimated; from these estimates, the effect of any given structural change is predicted as the need arises. The experiment itself will again consist (in our example) in replacing the random variables p_t produced in the free market, by fixed constants p_t^* $(t = 1, \ldots, T)$. But this time we measure not the shortages y_t directly, but, separately, the demand x_t^d and the supply x_t^s. The relation between the single random variable x^d (or x^s) and the predetermined variables p^*, z is again like that in (2). We can therefore estimate the structural parameters α^d, α^s by two least square regressions of, respectively, x^d and x^s on the predetermined variables p^*, z; and we estimate the structural parameters σ^{dd}, σ^{ds} from the regression residuals. To predict the shortage y_{T+1} for given p_{T+1}^*, z_{T+1}, we obtain the parameters β and σ_{vv}, and hence the parameters μy_{T+1}, σ_{yy}, by (9), (11), (12): that is, we are able to predict the effect of the transformation of the original free market structure $(\alpha, \sigma, 1, 0,)$ into the controlled market structure $(\alpha, \sigma, 0, 1)$. But in addition, we can now also predict the result of any other given change undergone by the original free market structure: for example to predict what happens if a technological change raises some of the coefficients α^s of the supply equation by a given percentage; or if an expected change in tastes or in population raises in a given proportion some of the demand coefficients α^d, or, possibly the violence of random demand shifts σ^{dd}, etc.

Consider now the Case B: the non-experimental situation, usual in economics. In this case, no observations are available except those made in the free market, with prices p_t determined by supply and demand, and therefore depending on their random shifts u_t^d, u_t^s. Thus for $t = 1$, ..., T, the structure is $(\alpha, \sigma, 1, 0)$. But for $t = T+1$, the structure is $(\alpha, \sigma, 0, 1)$. Equation (10) is valid for $t = T+1$, but not for earlier time points. Estimates of β, σ_{vv} necessary, as before, to predict y_{T+1} from (10) cannot be obtained from past observations on (y_t, p_t^*, z_t) since, for $t = 1, ..., T$, we have $y_t = 0$, and p_t^* is not defined. During this past period, price p_t, was not fixed. Instead, it was a random variable having a joint distribution with the variable x_t (by which we shall now denote both demand and supply since both were equal); this joint distribution, conditional upon the values of the predetermined variables is, under the conditions of the problem, again normal. Denote by γ the parameters of this joint distribution of the observables x_t, p_t (that is, the set γ consists of the conditional means and variances and covariance of x_t and p_t.

Since p_t, x_t depend on u_t^d, u_t^s, the parameters γ are uniquely determined by the structural parameters σ, α. If the converse were always true, the structural parameters α, σ could always be determined from the knowledge of γ; while the parameters γ of the joint distribution of p_t, x_t can, of course, always be estimated from observations on these variables. Whether, in a particular case, the parameter sets α, σ (or some elements of them, *e.g.*, the coefficients α^d of the demand function) are in fact uniquely determined by γ, depends on the assumed economic theory, or "model". For example, if in the model (5), (6), the set of predetermined variables z is represented by a single variable (say, income tax-rate z') in the demand equation, and by another single variable (say, rainfall z'') in the supply equation, then it can be shown that only one structure can be compatible with a given distribution of x, p. In such a case we say that the model makes the whole structure identifiable. In some other case the model might make only a part of the structure—such as the demand coefficients α^d—identifiable; or no single parameters might be identifiable. This does not depend on the number of observations in hand but entirely on the model which the economist has assumed on the basis of some previous knowledge. The conditions of identifiability have been given by Koopmans and Rubin for an important class of cases. The concept itself was defined by Koopmans and has been further studied by L. Hurwicz (all of the Cowles Commission). In a less rigorous way economists have been aware of the problem for a long time but more especially since H a a v e l m o pointed it out in his book *Probability Approach to Econometrics*.

Suppose then that the chosen model is such as to make all the parameters of our structure identifiable. Then α and σ can be derived from γ; but γ can always be estimated from the observations on p, x, being the set of parameters of their distribution, hence α and σ can be estimated. Therefore, β and σ_{vv} can be estimated, because of (9), (12); and the shortage y can be predicted, since its conditional mean and its variance are determined by (11), (12).

Knowledge of the structure can help not only in predicting the result of the structural change just considered, *viz.*, the introduction of price control; it can be also used to predict the result of any other structural change, provided the change can be spelled out as a well-defined transformation of the original structure (α, σ, 1, 0) which prevailed when the observations were made. Thus, in our non-experimental Case B as well as in the case of general experiments A2, prediction is preceded by structural estimation. Then the knowledge of structure makes it possible to predict the effect, not only of one given structural change (as is the case with specific experiments, A1), but of any well-defined structural change. The estimation of structure, the establishing of a theory, whether with or without the help of experiments is therefore not only a matter of scientific curiosity, it has also the practical value of providing for a number of future contingencies.

The case of general experiments (A2) and the case of non-experimental observations (B) have in common the same aim of estimating the original structure, rather than producing the changed structure by specific experiments (A1). But this aim is reached by different methods. Experiments are common to A1 and A2. In the present example, they call for the application of the method of least squares, which arises as a special case of maximum likelihood estimation. In Case A1 the probability density of y given p^*, z, taken at the observed point $(y_1, \ldots, y_T)$, was maximized with respect to the estimated parameters, and this is equivalent to maximizing $-\sum_1^T v_t^2$ (the exponent of the normal probability density function considered), *i.e.*, to minimizing the sum of squares of the residuals v_t. In Case A2 the probability densities of the independent variables x^d, x^s were similarly maximized, each one separately. In the Case B, maximum likelihood estimates are obtained by maximizing the probability density not of a single variable but of two jointly dependent ones x, p: this is a generalization of the least squares principle. In *Econometrica* 1943, Haavelmo pointed out the joint dependency of variables in economic models, and M a n n and W a l d studied the properties of maximum likelihood estimates of structural parameters. At the Cowles Commission the method has been developed by Koopmans and Rubin for the estimation of all structural parameters simultaneously. When one or more structural equations are to be estimated separately, more flexible though less efficient procedures are applied, developed at the Cowles Commission by T. W. Anderson, Girshick and Rubin for the case of a single equation, and by Rubin for the case of several equations. All these methods include cases where the jointly dependent variables occur with time lags.

The economist faces the combined disadvantages of the meteorologist and the engineer. At least until artificial rain was produced a few weeks ago, meteorologists were not able to experiment with weather; nor did they need to, as they did not consider it their task to change weather. Engineers, on the other hand, are required to predict the results of their interventions with nature; but in this task they are helped by experiments. The economist is deprived of experiments that change the structure, yet has to predict results of policies consisting in structural change.

This picture is incomplete. Some economic policies consist, not in changing the structure but merely in fixing the values of exogenous variables. Also, structural estimation may call for different methods in economics, where observations cannot be duplicated; and, say, in astronomy where they can. The principles as well as the techniques of structural estimation require much further study.