10

Simultaneous Equation Techniques[1]

The preceding chapter dealt with multiple correlation analysis. The results of this kind of analysis can be expressed in graphic form in a series of charts, one for each independent variable, showing the net influence of each variable on the dependent variable. Or they can be shown in the form of a mathematical equation, with the dependent variable represented by a term on the left side of the equality sign, and a constant on the right hand side, followed by a series of terms, one for each independent variable, showing the net influence of each variable on the dependent variable in numerical form.

As we have seen, agricultural price patterns evolve through mutual adjustment among a profusion of conditions and economic impulses. Any single relationship or equation is just one strand drawn from the whole tissue of economic interaction. When other closely related processes are assumed to be fixed and frozen, the one hypothetical relation under study may seriously misrepresent the joint processes of which it is only one part. This is, unfortunately, true even if the relationship taken by itself is sensible and verifiable, and even though the single equation contains many variables. Furthermore, if there occurs some basic "structural" change in these closely related processes, the estimates based on past observations may be grossly in error when applied to the new, changed situation.[2]

---

[1] This chapter owes a great deal to K. A. Fox, *Econometric Analysis and Public Policy,* Iowa State Univ. Press, 1958, especially Chaps. 1, 2, 3, and 7; and to M. Ezekiel and K. A. Fox, *Methods of Correlation and Regression Analysis,* Wiley, 1959, Chap. 24, which gives a useful summary.

[2] The classic statement of this last point is in J. Marschak, "Economic Measurement for Policy and Prediction," Chap. 1, pp. 1–26 in *Studies in Econometric Method,* W. C. Hood and T. C. Koopmans, editors, Wiley, 1953.

That is to say, the methods worked out in the previous chapters make it possible to estimate relationships between one dependent variable and one or more independent variables. If, however, there are two or more jointly dependent variables to be explained by other independent variables, this "jointness" or simultaneity in the world may need to be expressed by several simultaneous equations, each of which expresses one of the interrelated processes.

This single equation method is appropriate where the line of causation is clear, and one variable (the dependent variable) is unilaterally determined by the others (the independent variables). This condition is met in many agricultural price analyses—for example, in the simple analysis of the effects of corn production and disposable income on the price of corn from December to May. It is clear here that the price of corn is determined by the size of the corn crop harvested in the preceding November, not vice versa; and it is also clear that the price of corn is determined by disposable income, not vice versa. It is obvious which of the variables is dependent and which ones are independent. The single equation approach is appropriate here, although the results of the study need interpretation if changes in demand and supply are correlated, as shown in the preceding chapter.

WHY SIMULTANEOUS EQUATIONS MAY BE NEEDED

In many cases, the line of causation is not unilateral; it does not go only one way. For example, the price considered may be the price of eggs during the production season. In that case, the price may affect the production, as well as the production affect the price. If the production is affected by the price, but not vice versa, a scatter diagram would yield a supply curve. If the price is affected by the production, but not vice versa, this would yield a demand curve. But if the two variables are jointly determined — if the production is affected by the price, and the price is also affected by the production — a scatter diagram yields neither a supply curve nor a demand curve, but a mixture of both.

It is, in fact, impossible to get a demand curve and a supply curve out of a single equation. When two or more variables are jointly determined, it is impossible to get even one curve—supply curve or demand curve — out of a single equation. It is possible, however, to get both curves out of two equations solved simultaneously.

We can see what is involved here if we begin with elementary

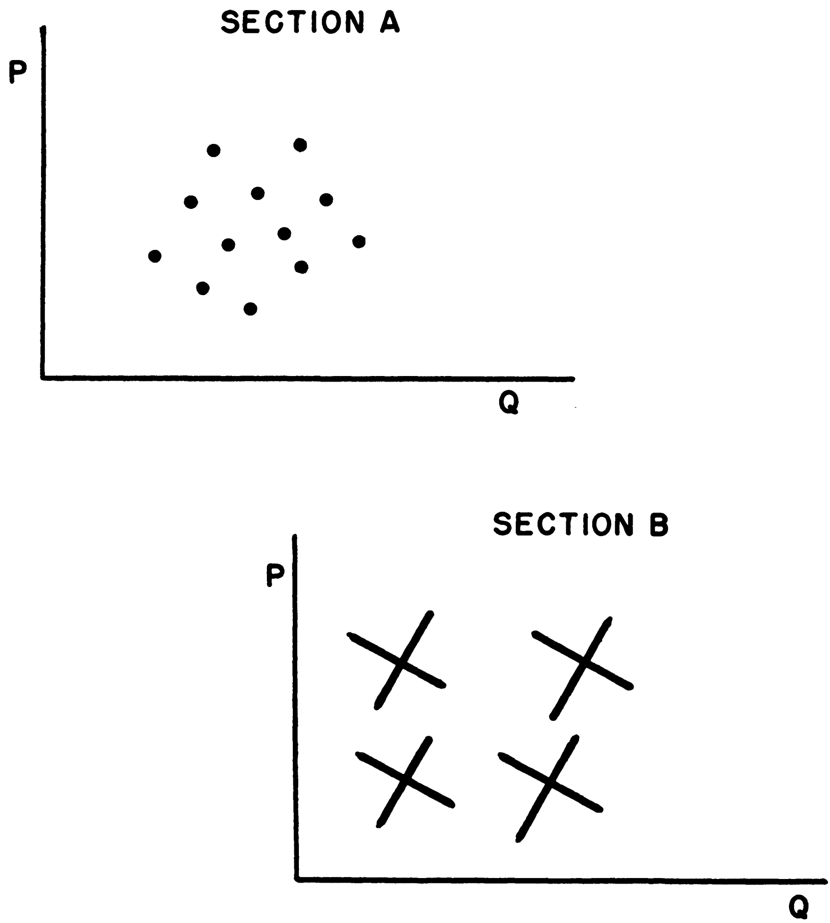

Fig. 10.1 — Hypothetical price and production data plotted in scatter diagrams. Demand and supply both unstable.

concepts and proceed to show when and why simultaneous equation techniques are needed.[3]

Figures 10.1, 10.2, and 10.3 bring together in summary form the elementary concepts developed in the preceding chapter. The raw price and production data for a typical farm product, plotted in a scatter diagram, may look something like Section A of Figure 10.1.

[3] The rest of this section draws on parts of a paper by R. J. Foote, "A Comparison of Single and Simultaneous Equation Techniques," *Journal of Farm Economics,* Vol. 37, No. 5, Dec., 1955, p. 975.

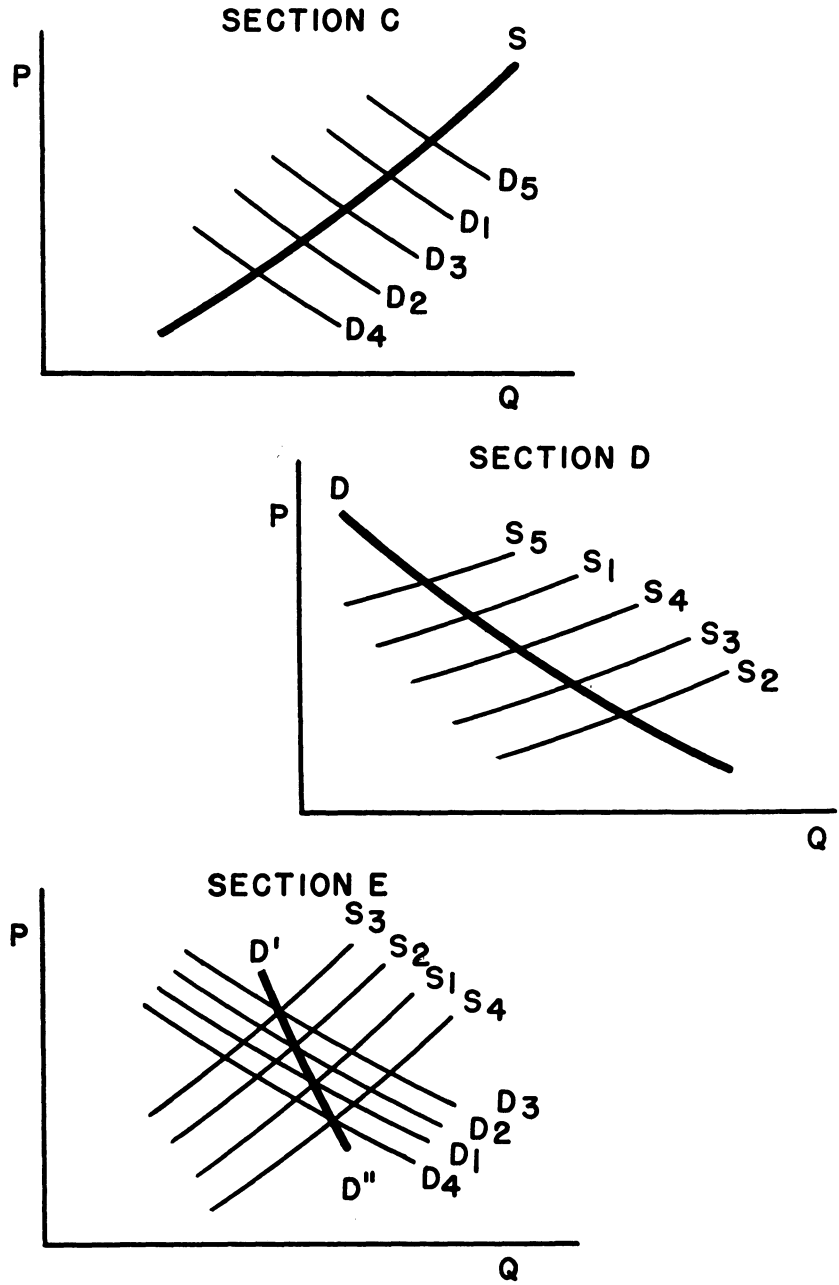

Fig. 10.2 — Hypothetical price and production data plotted in scatter diagrams. Section **C** shows unstable demand and stable supply. Section **D** shows stable demand and unstable supply. Section **E** shows unstable demand and supply, negatively correlated.

Each dot may be thought of as the intersection of a demand and a supply curve, as in Section B; but the elasticities of the curves shown are purely hypothetical, for without further information, neither curve can be determined from the data.

The demand may be unstable, so that the demand curve shifts back and forth over a wide range, while the supply curve remains relatively stable. This is shown in Section C of Figure 10.2. In that case, if the movements of the supply and demand curves are uncorrelated, the dots trace out a supply curve. Conversely, if the

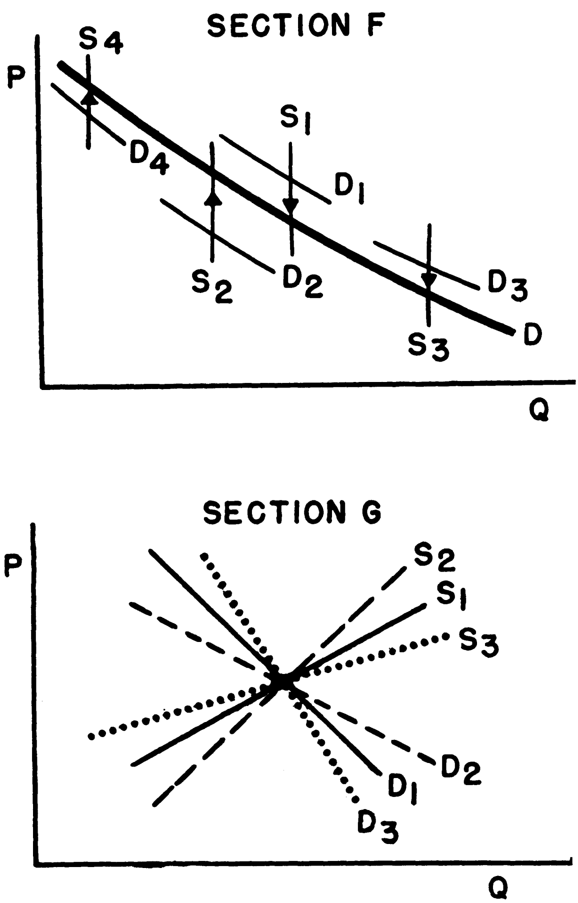

Fig. 10.3 — Hypothetical price and production data plotted in scatter diagrams. Section **F** shows demand and supply both unstable, but demand adjusted to remove instability, and supply completely inelastic. Section **G** shows an intersection point of a demand curve and a supply curve when their elasticities are both unknown.

supply curve is unstable but the demand curve is relatively stable, as in Section D of Figure 10.2, the dots trace out a demand curve.

If the movements of the supply and demand curves are correlated, as in Section E, Figure 10.2, the dots trace out what may look like a demand or supply curve, but the slope will be too flat or too steep.

In many analyses of the demand for agricultural products, the factors that cause the demand curve to shift over time are included as separate variables in a multiple regression equation. In effect, we are then able to derive from our estimating equation an average demand curve. This is indicated in a rough way in Figure 10.3, Section F. In some analyses, we can assume that the quantity supplied is essentially unaffected by current price; in agriculture, a time lag is usually needed before price can affect production. When price is plotted on the vertical scale, the supply curve in such cases is a vertical line, and year-to-year shifts in the supply curve trace out a demand curve, just as they did in Section D of Figure 10.2. Under these circumstances, we may be able to obtain valid estimates of the elasticity of demand by use of a least squares multiple regression analysis for which price is the dependent variable, and supply and some demand shifters are used as independent variables.

For many agricultural products, this set of circumstances permits us to estimate elasticities of demand with respect to price by use of single equation methods. Two points, however, should be kept in mind: (1) price must be used as the dependent variable in order to obtain elasticity estimates that are statistically consistent, since, to use the least squares technique, the supply curve must be a vertical line; and (2) an algebraic transformation must be made after the equation has been fitted to derive the appropriate coefficient of elasticity, since the definition of elasticity is in terms of the percentage change in quantity associated with a given percentage change in price, not the other way around as shown in Section F of Figure 10.3.

What happens if we have a supply curve that is not a vertical line? If we consider any single point, as in Figure 10.3, Section G, we have no way of knowing on which demand and supply curve of a whole family of curves it lies. The basic problem of indeterminateness is similar to that in which correlated shifts in the demand and supply curves take place. What is needed is some hypothesis, adequately tested and proven to be sound, as to the nature of the joint relationships between supply and demand. We should then be able to untangle the two and to obtain a reliable estimate of the slope of each curve. This is essentially what is done by the simultaneous equations approach.

"Simultaneous" refers to the method of algebraically solving or transforming the equations into other equations which can be fitted to the data. It does *not* mean that each equation need not reflect a definite causal relationship; on the contrary, each equation must be "identified," and this usually requires that causal relations be *more* explicitly and boldly stated in simultaneous equations than in multiple regression equations.

Suppose, for example, that X's are dependent variables and Y's are independent variables. The former (single equation) situation may be shown by:

$$X_1 = f\ (Y_1, Y_2, \ldots Y_n \qquad (1)$$

and the resulting regression equation might be:

$$X_1 = a + bY_1 + cY_2 + zY_n + u \ldots . \qquad (2)$$

The latter (simultaneous) situation is shown by:

$$(X_1, X_2 \ldots\ldots X_m) = f\ (Y_1, Y_2, \ldots\ldots Y_n). \qquad (3)$$

The resulting simultaneous regression equations might then appear as:

$$\begin{aligned} X_1 &= a + bX_2 + cY_1 + u \qquad (4) \\ X_2 &= a' + b'\,Y_1 + c'\,Y_2 + u' \\ X_m &= \delta + \beta Y_n + u^2. \end{aligned}$$

The most familiar case of this involves price and quantity as the joint outcome of a supply and demand equilibrium. A separate equation stands for each curve, and in equilibrium the P and X values are identical for both equations; this corresponds to solving the equations simultaneously for the values of P and X.

In some respects this method can best be seen as an extension of the methods of the previous chapter, although it does involve some additional algebraic skills and some new terms. More basically, it stresses the need to set forth clear, logical, and theoretically sound relationships to be tested. This is in contrast to the ever-present temptation with multiple correlation to shop around for variables to explain the dependent one, no matter what the result may mean in theoretical terms. Certain disadvantages may accompany this technique, both in computing effort and in possible error, and its usefulness will depend on the nature of the particular case.

There is no distinct consensus yet on (1) exactly which problems need to be formulated in terms of simultaneous equations, or (2) which of several methods of fitting the equations, once they are formulated, should be used. Since each research project has its own aims and requirements, no general verdict would be sensible. Instead the best choice or combination of techniques needs to be worked out for each case. Before summarizing current discussions on these questions, the general method with a simple supply and demand example will be illustrated.

An Illustration

Suppose that we wish to estimate both a supply and a demand curve for a particular commodity using data on its past prices and quantities. These data look like Figure 10.1 when made into a scatter diagram.

Apparently each point shows the equilibrium of supply and demand for one period, and both curves have been shifting randomly, in response to other influences, in about the same degree. If only one curve had been shifting, the other could easily be estimated, but this has not happened. Suppose that shifts in each curve have been independent of shifts in the other.

At each of these equilibrium points, price and quantity are mutually determined; there is no single direction of cause and effect between them which can be logically *identified,* one way or the other. Since P and X are, in fact, jointly dependent variables, which can be jointly "explained" by other variables, a model using simultaneous equations may be best for estimating either the supply or the demand curve, or both together.

A logical form for the supply and demand curves might be:

$$\text{Demand curve: } X = a_1 + b_1P + u \qquad (5)$$

$$\text{Supply curve: } X = a_2 + b_2P + u \qquad (6)$$

That is, the amount demanded depends on the price, in a way described by a straight line (on numerical or log graph paper) which cuts the quantity axis at some level of $X = a_1$ and which has a slope of b_1 (which may be negative). This line (or curve) shifts in the short run about its long-run position in response to numerous random (or "stochastic") disturbances, which are lumped into the term u. Whatever factors u reflects, they are not correlated with levels or changes of P.

The supply equation has an exactly similar meaning, though its slope (b_2) will presumably be positive. These are "structural equations," relating price and quantity in ways which are sensible and defensible in theory. The coefficients a_1 and a_2 and b_1 and b_2

are called *structural parameters;* it is the values of these which we wish to estimate. A single equation, multiple regression approach does not give us estimates of these structural parameters, but it requires less rigid and less hazardous assertions about cause-and-effect than do simultaneous equations.

Though equations (5) and (6) may be logically correct, one can see intuitively that they cannot be fitted to the roundish scatter of $P : X$ dots in a scatter diagram to give a good estimate of either b_1 or b_2 separately. We cannot *identify* whether the supply curve alone determines price and quantity, or if the demand curve does so. To put it in statistical terms, we cannot fit either equation using available data to give unique estimates of the structural parameters b_1 or b_2.[4] But, if two changes are made in the equations, it may be possible to estimate both equations together.

First add to each equation a "predetermined" variable. These correspond to "independent variables" in multiple regression. Such a variable may be either truly exogenous to (or "outside") the model; that is, it may represent any physical, social, or economic factor which unilaterally influences demand or supply, but is not in turn influenced by them — weather, for example, or GNP. Or it may be simply the level of one of the already-present variables (in this case P or X) at an earlier period; that is, a "lagged endogenous variable" such as P_{t-1}, X_{t-2}, etc. A logical choice for the demand equation might be consumer income; although for a corn demand equation, one might use number of beef cattle. For the supply equation some earlier supply measure, such as previous plantings or number of hogs six months previously (i.e., lagged by six months), might be chosen. If such lagged endogenous variables are used, one must be sure that they influence P and X but not the other way around. To be precise, they must be *recursive.*

In selecting the predetermined variables (exogenous or lagged endogenous), we are drawing, out of the grab-bag random u and v terms, the most likely explanatory variables. Just as we add, one

[4] If for example price at time one determines quantity at time two, this unilateral causation satisfies identification requirements, and

$$X_2 = a + bP_1 + u$$

can be uniquely estimated for b. This one-way causal relationship between time periods, with no reverse influence from period two on period one, is called a *recursive* relationship. Recursiveness may be required in structural equations as well as in single equation methods; for example, factor Z in equation (8) will probably be recursively related to both X and P. On recursiveness see H. Wold and L. Jureen, *Demand Analysis,* Wiley, 1953, especially pp. 48–71, 202–04. On identification, see T. C. Koopmans, "Identification Problems in Economic Model Construction," Chap. 2 in Hood and Koopmans, *op. cit.;* and Fox, *op. cit.,* pp. 26–29.

by one, only the most reasonable independent variables to multiple regression equations, in this instance too we will select only the most logical variables. Suppose that an income variable Y is chosen for the demand equation and that some factor Z based on weather, or previous plantings, or previous prices, is added to the supply equation. The result of this first step is the two structural equations:

$$\text{Demand equation: } X = a_1 + b_1P + c_1Y + u' \qquad (7)$$

$$\text{Supply equation: } X = a_2 + b_2P + c_2Z + v' \qquad (8)$$

Since the stochastic terms u and v no longer include Y and Z, they are given as u' and v'. The equations are in fact logical hypotheses about demand and supply.

The *second* step uses algebra to transform or solve these equations for P and X, taking P and X as dependent upon the predetermined variables Y and Z and on the shift factors or disturbances u' and v'. This gives the following two equations which are called *reduced-form* equations:[5]

$$P = A_1 + B_1Y + C_1Z + d_1 \qquad (9)$$

$$X = A_2 + B_2Y + C_2Z + d_2. \qquad (10)$$

These equations differ in form from the structural equations, and they sometimes have no inherent logical significance of their own as they stand. But, like the structural equations, they have parameters or coefficients (A, B, C) and these reduced-form parameters can be transformed back algebraically to derive the structural parameters. For instance, in this example d_1 and d_2 include the disturbances u' and v' and the structural parameters b_1 and b_2.
And

$$b_1 = C_2/C_1;\ b_2 = B_2/B_1;\ C_1 = B_1\ [\ -\ (b_1 - b_2\ ];$$
$$C_2 = C_1\ (b_1 - b_2).$$

[5] Note that the system of structural equations is *complete,* as well as that each separate equation is *identified.* This is because the number of endogenous variables equals the number of equations. This allows us to solve to get these two reduced-form equations in which each endogenous variable is expressed as a function of (i.e., is dependent on) all the predetermined variables in the system. If the system were incomplete — with more endogenous variables than equations — such reduced-form equations could not be derived for each endogenous variable. If, on the other hand, the system of equations included more equations than endogenous variables (this is usually called an over-identified system), the system could not be uniquely estimated. More than one version of some of the reduced forms would be possible, leading to indeterminacy of the estimates of both the reduced forms and the structural parameters. This is the case with equations (5) and (6) above.

Both completeness of the system and identifiability of each single equation are necessary conditions for solving for reduced forms and estimating the structural parameters. For more detailed discussion on this point see Fox, *op. cit.,* Chap. 1. Solution of simultaneous equations above follows customary algebraic methods. A step-by-step solution of these two equations can be found in Ezekiel and Fox, *op. cit.,* Chap. 24.

And so on. If we can estimate statistically the reduced-form parameters, then we can work out estimates of the structural parameters, including a_1, a_2, b_1 and b_2. These parameters define the demand and supply curves themselves.

It is possible to fit the reduced-form equations statistically, since each has one "dependent" variable, plus "independent" variables, in the manner of familiar least-squares regression equations. The further statistical requirement that d_1 and d_2, the random residuals of the regression, be independent of Y and Z is also met. If there is any doubt of this, it can be checked after the estimation is done by seeing if the residuals seem to have any systematic pattern. Each equation may now be fitted by itself, using the methods in the previous chapter; that is, by either graphic or, more usually, least-squares estimation, or possibly using the maximum-likelihood methods discussed later in this chapter. Note, however, that a single value estimate of each parameter (A, B, and C) is required, so straight lines only can be fitted; though, of course, logs could be used to provide for some curvilinearity.

If the resulting correlation for each reduced-form equation is satisfactory (in terms of R^2 standard error, and confidence levels; or by a visual check of the scatter) then the algebraic transformation of the reduced-form parameters to the structural parameters will be worth doing. Since this transformation is mathematically precise, it faithfully transmits back into the structural estimates both the accuracy and the errors present in the reduced-form estimates. Similarly the degree of goodness of fit for the structural parameters (in terms of standard errors) can be derived, and the residuals of the structural equations can be analysed for auto-correlation, using ratios of derived values of u^1 and v^1.

Whatever their significance as theoretical propositions, the reduced-form equations may be extremely useful for making predictions or determining policy. This is because they imply a "cause and effect" relation and, when properly fitted, enable one to estimate the degree of change in the dependent variable associated with changes in the independent variables. Since the latter are by definition preknown or preset (possibly under direct policy control) this knowledge may have great practical use.

In partial contrast, structural equations often have a more scholarly role, estimation of their parameters being more generally devoted to hypothesis testing and measuring. Reduced-form estimation might be used in predicting output or prices in the future, and in "predicting" the effects on output or price (in the future or in the past) of given policy changes. To estimate demand or supply elasticities, structural equations are necessary.

This distinction between reduced-form and structural equations is blurred because structural parameters are often crucial for policy, and are, in any event, implicit in the reduced-form parameters. Moreover, the usefulness of the reduced-form parameters for prediction purposes depends on the constancy of the structural parameters which are implied in them. If there are structural changes in the relationships involved — in this case if the demand or supply curves permanently shift (involving a change in an a or a b, or both) — it is clear that the prediction coefficients must be revised. So an understanding of the past and likely future behavior of the structural parameters should underlie any use of the reduced-form equations for prediction.

CHOOSING EQUATIONS AND METHODS

After two decades or more, our knowledge about the relative merits of multiple correlation and structural equation techniques, and of alternative methods of estimating them, is at present still in the formative stage. Although certain types of problems have been explored, some of them at great length, there is general agreement that much more testing is needed, and that no general choice between them is either possible or desirable.

The initial elegance and mystique of the simultaneous equations approach has been somewhat dimmed in the face of evidence on several points that it does not solve all problems, nor does it always do the best job even on those for which it is best suited. Despite this, interest in this approach has encouraged a substantial advance in sophistication and care in framing hypotheses and stating questions, and more recently additional practical arguments favoring simultaneous systems have been advanced. Nonetheless, the single equation multiple regression method has turned out to be difficult to defeat, and it may continue to be satisfactory or at least useful for many, if not most, research projects in agricultural price analyses.

Single or Simultaneous Equations?

Karl Fox has concluded on the basis of a number of empirical demand studies that many agricultural situations lend themselves to single equations least-squares estimations as well as or better than to formulation into two or more simultaneous structural equations.[6] For these situations a "uni-equational complete model" gives satisfactory estimates and is economical of computing effort.

Fox cites the recursiveness of many production and marketing processes as one explanation for the fact that in many previous studies both methods have given virtually identical estimates of structural parameters. In these cases, which have included both just-identified and over-identified systems, no departure from the simpler single equation least-squares method seems to be called for. Fox also concluded that structural changes in demand for agricultural products have been more gradual since prewar years than some analysts have thought. Therefore, the need to recognize explicitly the possible changes in these parameters would have been less than has been argued by some.

Fox and others conclude that generally the importance of the bias of least-squares methods, which may stem from their neglect of simultaneity in real world processes, will be less than has been feared. This is partly because simultaneity may not in fact be so prevalent; partly because other problems such as auto-correlation and limited dependent variables may also deserve care; and partly because other feasible methods may not on balance be much superior to the single-equation multiple-regression method.

Methods of Estimation: Least-Squares or What?

On a more technical level, there has been extended discussion and some testing of alternative statistical methods of estimating reduced-form and structural parameters, once they have been decided upon; on this there is "no verdict yet."[7] Several alternatives to ordinary least-squares are current, all of them involving simultaneity; namely, "two-stage least-squares;" "limited-information maximum-likelihood;" and "full-information maximum-likelihood" methods. These methods, some aspects of which are still being developed, differ in complexity and ease of use, and an explanation of them would go well beyond this discussion.

There are three ways to evaluate which methods are best for given situations: mathematical theorems, real world studies, and controlled artificial experiments. The first has not been fruitful because it can deal only with infinite samples, and it is precisely for small samples (from say 20 to 60) that comparisons are needed. Real world studies have tended to show similarity among the re-

[6] See Fox, *op. cit.*, Part I, pp. 1–150.
[7] "A Symposium on Simultaneous Equations Estimation," *Econometrica*, Vol. 28, No. 4, Oct., 1960, pp. 835–71.

sults given by different methods, with somewhat better results from least-squares than had eariler been expected.

Controlled experiments (so-called Monte Carlo tests) whereby true values are derived for an artificial model programmed in a computer, and then several small-sample estimates are made using alternative techniques, has tended to strengthen confidence in the simultaneous methods, especially for over-identified systems. However, most of the testing models used so far are not of the sort relevant to agricultural price analysis.

Klein has argued recently that even though the alternative statistical methods may give nearly identical estimates of *structural* parameters, these slight differences may be seriously magnified in the transformation to *reduced-form* parameters.[8] For example a .06 difference in estimating the marginal propensity to consume may become a .68 difference in the income multiplier which is related to it. So even if least-squares estimates of structural parameters are only slightly biased from the true values, an estimation or prediction using the reduced-form equations may contain substantial error.

The degree of this magnification depends wholly on the nature of the structural and reduced-form systems; bias may be enhanced, left unchanged, or even diminished by the transformation. In agricultural price analysis instances of increased error in the reduced-forms parameters may be unlikely. Also, if estimation of structural parameters is the main object of the study, then the problem of magnified reduced-form error naturally fades.

The present situation for equations and statistical methods may be summed up briefly, in somewhat more rigorous terms. Single equation least-squares is, in general, likely to give biased estimates, because it ignores possible simultaneity. On the other hand, for small sample estimation it is generally most suitable, since the other methods are known to be unbiased only asymptotically; that is, for very large samples. Simultaneity may also be a mixed blessing, for if some structural equations are incorrectly specified, or if their variables are correct by displaying great variance, simultaneous methods may spread error into estimates for the other equations.

This suggests first that it may not be possible with structural equations simply to set them up on *a priori* grounds and then run the test; some shopping around and exchanging of variables and equation forms may be necessary to get "correct" structural equa-

[8] L. R. Klein, "The Efficiency of Estimation in Econometric Models," Cowles Foundation Paper No. 157; also in *Essays in Economics and Econometrics*, Chapel Hill, 1960, pp. 216–32.

tions. Second, as for statistical techniques, using least-squares is the best safeguard against using an "incorrect" model, but if you are sure of the model, then such methods as limited-information maximum-likelihood may be be somewhat better. The latter will usually have somewhat larger variance; that is, their estimates will be more accurate (centered on the "true" value) but less reliable (more widely spread). If the predetermined variables are correlated among themselves ("multicollinearity") or are auto-correlated (i.e. a variable with regular waves may correlate highly and spuriously with itself), the least-squares results will be liable to error, but less so generally than the other methods. In such situations other precautions will be needed in any case.

For under-identified structural systems, none of the methods satisfactorily estimates the structural parameters, although the reduced forms may be handled best by least-squares. For just-identified systems the methods will be about equally good, both for structural equations and reduced forms, since they are all basically equivalent. For over-identified systems, simultaneous methods are superior though lengthy for estimating structural parameters; from them the reduced forms can be derived without the risk of magnified error which Klein warns against. The advantage of working only with just-identified systems is evident; in this way one in effect forestalls the question of methods. Klein also rightly notes that greater access to electronic computers sharply reduces the advantages of single equation least-squares in the way of computational ease.

CRITERIA FOR CHOICE

The decision to use structural equations and the writing of the model, if the decision to use one has been made, depends on an intimate understanding of the real-world processes to be estimated and the variables which may be used. It is important that the equations adopted reflect reasonable, appropriate, and useful hypotheses about the world, as well as that they satisfy certain technical requirements of identifiability, consistency, and ease of fitting.

In addition to these theoretical and statistical standards, the analyst will often need to make special allowance for the eventual uses of the results, for policy and other purposes. This bears on the choice of statistical techniques for fitting simultaneous equations as discussed above. In some cases bias in the estimates would be especially harmful; in other cases possible bias may matter little compared to other possible weaknesses, such as variance. In general, structural equations techniques, like any other approach, may need careful adjustment to the needs and dangers of the particular case.

Several steps or rules are customary in posing and estimating a structural system, such as a supply-demand estimation:

1. All the variables which may be relevant are listed. These may include such "economic" variables as prices, quantity, incomes, costs, acreage and other inputs, imports, and the like; and such others as perhaps rainfall and temperature. The scope of the variables may vary (i.e. state, regional, national, or by sectors); they may include composite index variables or first differences; and lags of various durations may be specified. All of the variables must then be classified as either (a) endogenous or (b) predetermined (exogeneous or lagged endogeneous).

2. The structural equations (i.e. a "model") are worked out, each one representing as accurately as possible some theoretical or factual relationship. Equations may represent (a) hypothesized economic behavior (including most theoretical relationships); (b) institutional rules; (c) technological laws of transformation (such as production functions); or (d) definitions (in the form of identities).

3. Logical and defensible analysis must govern both the classification of variables and the writing of structural equations. Variables should not be reclassified, or equations rewritten, or extra equations added, to make the system identifiable or easier to estimate.

4. One or more equations should contain at least two endogenous (jointly dependent) variables. This of course reflects the assumed simultaneity.

5. The number of structural equations (of all sorts) will equal the number of endogenous variables, to provide completeness of the model. Then any equations containing only one endogenous variable can be fitted straightaway by least-squares. If a logical *and* identifiable system cannot be written, this may simply reflect actual under-identification in the real-world process. In such cases — of which price determination under oligopoly and duopoly conditions may be an example — any estimation would have to be forced, and might yield misleading results.

6. As in any analysis, Occam's razor should be used to keep the model within reasonable bounds, especially if data are unreliable and the area of research is a new one. Access to an electronic computer will, of course, increase the extent and complexity of systems that may be tried.

7. Solving the structural equations for the endogenous variables provides the reduced-forms, which can then be estimated by least-squares or some other method. In many, if not most cases, least-squares (or even linear graphic analysis) will suffice, but special care should be taken if the reduced-forms are to be used for predictive purposes. Extensive treatment of computational problems is given in Friedman and Foote's 1955 handbook.[9]

The elasticities of demand for various nondurable consumer goods, including foods, have been computed in about 200 different studies by the use of simple equations and simultaneous equations, some over-identified and some just-identified. The results of these studies are brought together in one large 13-page table in "Price Elasticities of Demand for Nondurable Goods, With Emphasis on Food" by Richard J. Foote (USDA, March, 1956). See also G. E. Brandow, "Interrelations among Demands for Farm Products and Implications for Control of Market Supply," Bul. 680, Aug., 1961, Pennsylvania State University, Agr. Exp. Sta., University Park, Pennsylvania.

An excellent discussion and appraisal of simultaneous equations is given in M. J. B. Ezekiel and K. A. Fox, *Methods of Correlation and Regression Analysis,* Wiley, 3rd ed., 1959, Chap. 24. See also R. J. Foote, "Analytical Tools for Studying Demand and Price Structures," Agr. Handbook No. 146, USDA, Aug., 1958.

---

[9] J. Friedman and R. Foote, *"Computational Procedures for Handling Systems of Simultaneous Equations,"* USDA, 1955.

A useful summary of the results of numerous price analyses for the chief farm products in the United States is given in a 131-page mimeographed report by H. E. Buchholz, G. G. Judge, and V. I. West, "A Summary of Selected Estimated Behavior Relationships for Agricultural Products in the United States," USDA Res. Rept. AERR-57, Oct., 1962.

four

SALES, PRICES, COSTS, AND RETURNS