Howard Jackson
Prices

ANALYTICAL TOOLS for Studying Demand and Price Structures

Agriculture Handbook No. 146

UNITED STATES DEPARTMENT OF AGRICULTURE
Agricultural Marketing Service
Washington, D. C.

For sale by the Superintendent of Documents, U. S. Government Printing Office
Washington 25, D. C. - Price $1.00

August 1958

ANALYTICAL TOOLS for Studying Demand and Price Structures

by Richard J. Foote
Head, Price and Trade Research Section,
Agricultural Marketing Service

Agriculture Handbook No. 146

UNITED STATES DEPARTMENT OF AGRICULTURE

Washington, D. C.

PREFACE

Considerable progress has been made in recent years in developing methods to be used in analyzing the factors that affect prices and consumption of individual commodities and in studying their demand and price structures. This handbook discusses certain methods which appear to be of value for this purpose. Some of them are relatively standardized; others were developed only recently. These latter have been applied in only a few cases. In most instances, examples are included which indicate specific ways in which these techniques can be used.

The handbook is designed mainly to acquaint research workers in agricultural economics and related subjects with some of the recent developments in the field. No attempt is made to cover all new developments that apply, although many of the more important elements in analysis of demand are touched upon. Use of the handbook presumes a general knowledge of the theories of price and demand and of the standard techniques of regression analysis. Some of the sections also presume a knowledge of college algebra, matrix algebra, and calculus. But the conclusions are presented in nonmathematical terms, so that, except for certain developmental or explanatory sections, the handbook as a whole can be used by those not acquainted with higher mathematics.

Some of this material was included in Agriculture Handbook No. 64, entitled "Analytical Tools for Measuring Demand" by Richard J. Foote and Karl A. Fox, published in 1954. This handbook supersedes Handbook 64. Sources for previously published material and sections prepared by staff members are indicated in the text or by footnote in each section. Helpful suggestions were received from various members of the staff.

CONTENTS

CONTENTS

ANALYTICAL TOOLS FOR STUDYING DEMAND AND PRICE STRUCTURES

By

Richard J. Foote, head, Price and Trade Research Section,
Agricultural Economics Division
Agricultural Marketing Service

The measurement of demand is a complicated subject. Competent analysis requires three things. First, the economist must have a thorough knowledge of the economic factors that affect the commodity and obtain adequate data on which to base the analyses. Second, he must understand economic theory in general. Third, he must be able to use modern techniques of analysis.

This report is mainly concerned with techniques of analysis. But the first few sections discuss ways to make a preliminary survey of the economic relationships to be expected. Here is where the researcher brings to bear his knowledge of the commodity and his understanding of theory. A number of alternative ways to formulate and fit statistical relationships then are discussed in detail. The report next considers some of the techniques that can be used to test the assumed model and to make quantitative estimates, or forecasts. Specific examples in each case are cited when needed to indicate how a particular technique should be used.

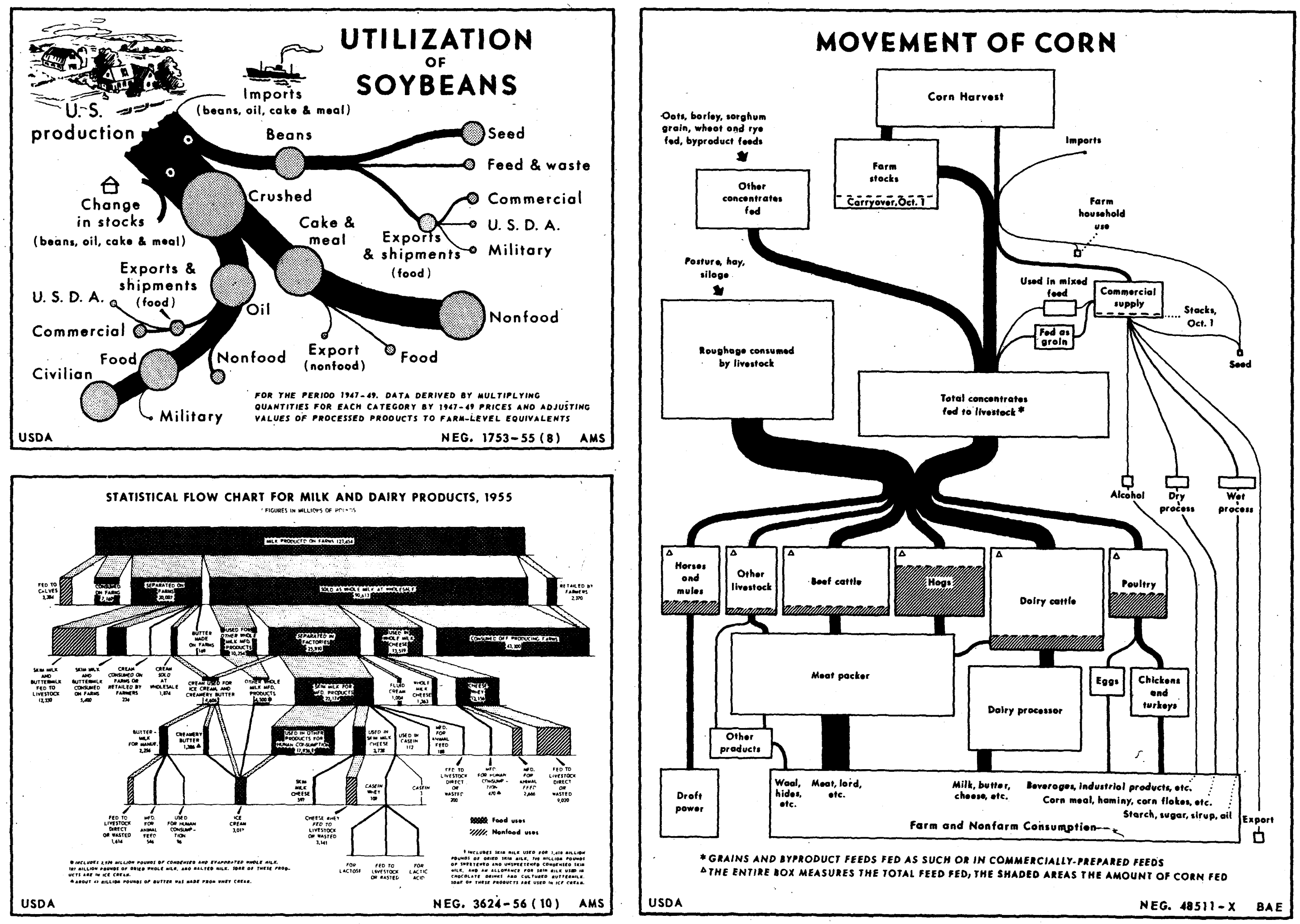
DIAGRAMS OF SUPPLY-DEMAND-PRICE STRUCTURES

Diagrams that show the flow of commodities from producer to consumer in terms either of physical products or marketing channels, or both, have been in use for many years. Such diagrams are useful in showing the relative importance of specified sources of supply or kinds of outlets and, in some instances, the sequence of marketing channels or processing operations. Figure 1 shows three such diagrams taken from Gerra (42) 1/, Rojko (81), and Foote, Klein, and Clough (30).

Similar diagrams that show the economic forces or relationships that affect a given commodity or group of commodities have been developed during the last several years by staff members of the Agricultural Marketing Service. Figure 2 shows four diagrams of this sort taken from Fox (33), Meinken, Rojko, and King (72), Hermie (47), and Armore (5). Similar charts have been prepared for many other commodities. Such diagrams usually are most helpful when they are relatively simple, though, at times, complex diagrams serve a useful purpose. Diagrams of this kind may be used in the following ways: (1) To help the analyst think through basic factors and relationships involved, (2) to aid in the preparation of a logical writeup of the economic structure of the industry, and (3) to assist the reader in following fairly complex relationships and discussions. Any statistical analyses that are run should be consistent with relationships indicated in the diagram though, as with any research work, modifications may be made as the study develops.

1/ Underlined numbers in parentheses refer to Literature Cited, page 196.

DIAGRAMS THAT SHOW COMMODITY FLOWS



U. S. DEPARTMENT OF AGRICULTURE NEG. 4428-57(8) AGRICULTURAL MARKETING SERVICE

Figure 1.--Diagrams that show the flow of commodities from producer to consumer are useful in illustrating the relative importance of specified sources of supply or outlets and in showing marketing or processing sequences.

The two diagrams in the upper part of figure 2 indicate the kind of statistical questions that can be discussed in terms of the diagrams. The first shows the supply-demand structure for a certain type of perishable crop, for example, early-season fresh asparagus. If (1) practically none of the crop normally is unharvested for economic reasons, as suggested by the narrow dotted lines around this factor in the diagram, (2) all of the production moves directly into a single outlet, and (3) no close substitute or complementary commodities exist, then the elasticity of demand in retail markets can be measured directly from certain coefficients in a single equation fitted by least squares. 2/ In the diagram this point is brought out by the fact that no arrows are double headed and, if the dotted arrows are omitted, no circular chains are shown. Elasticity of demand at the farm or local market level can be derived from that at the retail level by appropriate allowance for the assumed nature of the marketing structure (see page 100).

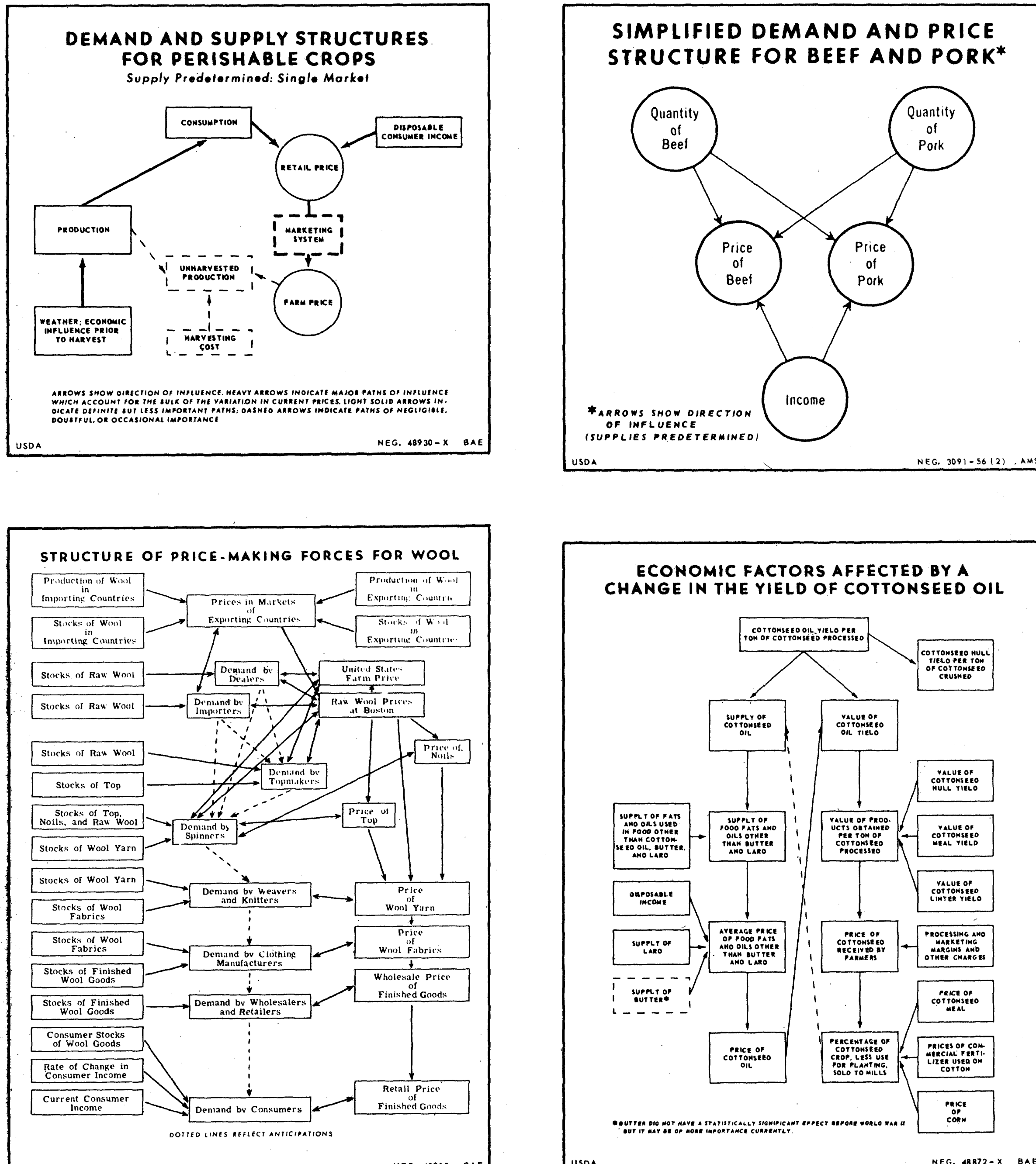
The second diagram assumes that conditions (1) and (2) referred to in the preceding paragraph apply approximately to beef and pork but it shows that the quantity of each affects both its own price and that of its competitor. In the words of Meinken, Rojko, and King (72, pp. 714-715), "a given combination of production of beef and pork results in a unique set of market prices that is simultaneously determined. To obtain estimates of the elasticities of demand that are statistically consistent (see page 58), the parameters in the structural demand equations ... must be estimated by a statistical method that allows for the simultaneity." Methods for doing this are discussed in detail beginning on page 87.

The third diagram is designed to illustrate several features of the demand and price structure for wool in domestic markets. Domestic prices tend to exceed world prices by approximately the amount of our tariff, but prices rise and fall with changes in world supply and demand. Demand for wool at local markets is derived from the combined demands of the many processors and users of wool, but the level of demand at each step in processing is affected by stocks of raw, semifinished, and finished products. A system of equations that allows for the many aspects of simultaneity is needed to describe adequately the domestic wool economy and to measure elasticities at the several marketing levels but, unfortunately, data are not available to fit such a system.

The fourth diagram shows the two channels through which the effects of an increase in the yield of cottonseed oil per ton crushed on total returns received by farmers take place. One channel affects the quantity of oil obtained per ton of seed crushed and therefore directly affects its value. This channel is illustrated by the arrow toward the top of the diagram running from the box entitled "Cottonseed oil yield per ton of cottonseed processed," to the box entitled "Value of cottonseed oil yield." The second channel affects the price of the oil by increasing its total supply. These effects are shown by the second column of boxes and the arrow running diagonally upward from the last box in this column to the box labeled "Value of cottonseed

2/ The reasoning here is developed in detail beginning on page 53.

DIAGRAMS THAT SHOW ECONOMIC RELATIONSHIPS



U. S. DEPARTMENT OF AGRICULTURE NEG. 4429-57(8) AGRICULTURAL MARKETING SERVICE

Figure 2.--Diagrams that show economic forces or relationships are useful in thinking through complex relationships, in preparing a logical writeup of them, and in helping readers to follow the discussion.

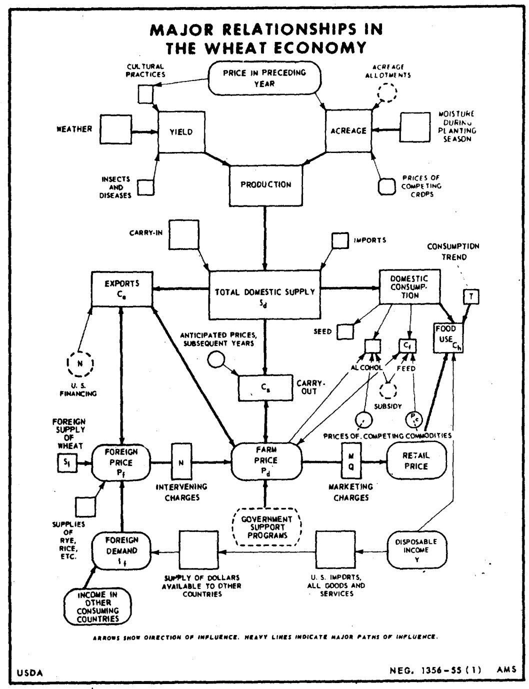
oil yield." Single equations fitted by least squares could be used to measure the relationships suggested by the first and second columns, although this part of the diagram actually is an over-simplification of the fats and oils economy, particularly for the period following World War II. Effects on the total value of cottonseed oil through the first channel can be computed directly by arithmetic. If the dotted line running diagonally upward from the last box in the third column to the box, entitled "Supply of cottonseed oil," was of greater importance, a system of equations would be needed to measure the relationships in the diagram because of the implied circular chain.

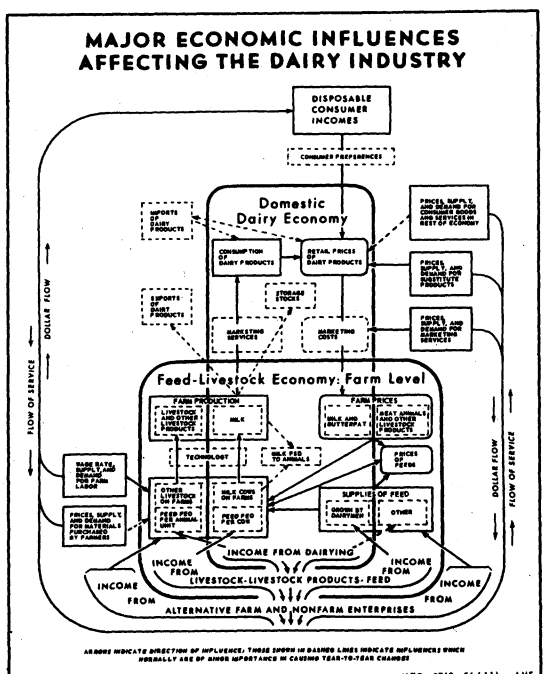
Diagrams shown in figure 3 are designed to illustrate a demarcation which, in the opinion of the author, exists between those that are too complex to be meaningful and complex diagrams that are useful to the analyst or reader. These are taken from Meinken (71), Rojko (81), and Shuffett (84), respectively. Although the author assisted in the development of the first diagram, he now feels that it is too complex to be of much value either to careful or casual readers. The important relationships shown can be discussed more easily and clearly by a direct referral to the system of six equations developed to represent the domestic wheat economy than by trying to follow through the many double-headed arrows and circular chains suggested by the diagram. On the other hand, the diagram developed for the domestic dairy economy, which is at least as complicated, clearly shows that portion of the dairy industry which is included within the overall feed-livestock economy and that part of the feed-livestock economy which is outside of the dairy economy --one of its principal purposes. It also shows certain income and service flows that are discussed in detail in the original text. In this diagram, boxes relating directly to the dairy economy are given in a simplified form, as they relate to aggregates for all milk and dairy products. When relationships among dairy products are studied, a system of equations is required to obtain estimates of the structural coefficients, such as elasticities, which are statistically consistent.

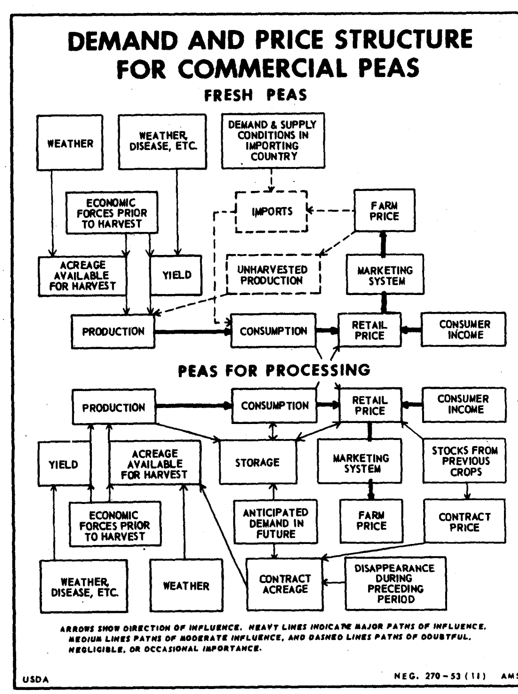
The diagram that relates to the demand and price structure for commercial peas is useful in showing the extent to which fresh peas and peas for processing can be considered separately and the extent to which they must be considered simultaneously. By the same reasoning as developed with respect to the demand structure for beef and pork (see page 3), a system of simultaneous equations is needed to obtain a statistically consistent estimate of the elasticity of demand for either fresh (and frozen) 3/ or canned peas. If the competitive relationships could be ignored, the important relationships indicated by heavy lines in the diagram could be derived from certain coefficients from single equations fitted by least squares.

3/ Shuffett (84, p. 85) shows that by adding frozen consumption to fresh supply, an analysis based on data prior to World War II (when frozen peas were of negligible importance) can be used in the post-War period.

SOME COMPLEX DIAGRAMS







U. S. DEPARTMENT OF AGRICULTURE NEG. 4430-57(8) AGRICULTURAL MARKETING SERVICE

Figure 3.--Some diagrams that show economic relationships are so complex as to materially reduce their value. The diagram for wheat may fall in that category, but those for dairy products and peas are believed to be helpful to readers of the publications in which they were initially published.

Use of heavy lines to show the more important relationships is an aid to readers in following diagrams that are even moderately complex. In the chart for wheat, however, too many relationships are important, so this device is of less value than, for example, in the diagram for peas.

SYSTEMS OF EQUATIONS 4/

In the language that has been developed to consider statistical analysis of economic relations, the process by which a set of economic variables is believed to be generated is called a structure. 5/ The variables whose values are explained by the structure are called endogenous variables, whereas those whose values are determined outside the structure are called exogenous. Lagged values of endogenous variables plus the exogenous variables frequently are grouped together and referred to as predetermined variables. The set of structures that are compatible with the investigator's advance assumptions about the statistical universe from which the data are drawn is called a model. Thus, within a model, we specify which structural relations are assumed to hold exactly and which include an unexplained residual. The former are referred to as identities and the latter, as relations. At times, it is useful to distinguish between economic and statistical models.

The diagrams which have been described are based essentially on economic theory and knowledge of the workings of that part of the economy being studied. The term economic model is applied to the set of structures consistent with the assumptions developed by the investigator from economic theory and knowledge of existing factors that relate to a particular commodity area. Thus an economic model is a set of equations that is consistent with the relationships and identities implied by the diagram.

In the present state of economic and statistical theory, a research worker typically finds it necessary to make additional assumptions for which economic and commodity considerations offer little if any guide. For example, he must specify the algebraic form of the relations and the specific way in which the relations are affected by unobserved influences. Although economic considerations sometimes may exclude certain possibilities, they usually do not provide strong grounds for preferring a particular set of assumptions. The specifications made about these aspects of the structure often are chosen partly to simplify the statistical analysis and are to a considerable extent arbitrary. Here we refer to the set of structures consistent with all the specifications of the investigator (both economic and statistical) as the statistical model.

4/ Material in the first four paragraphs for the most part is adapted from Hildreth and Jarrett (49, pp. 6-7).

5/ Marschak (66, pp. 1-8) gives an easily understood example that illustrates the basic nature of structural relations and parameters. This example is discussed briefly beginning on page 56.

Although the distinction between economic and statistical models is not always a sharp one, it frequently is useful. If the research worker makes explicit the basis for the assumptions underlying the economic model, these can be evaluated by economists and by persons well informed about the particular commodity area. The analyst may wish to experiment with alternative sets of specifications that apply to the statistical model, such as whether to use actual data or logarithms, whether to use first differences, and similar questions. Considerations that help in choosing among the several alternatives are given beginning on page 19.

After preparing diagrams of the sort described in the first section, the analyst will wish, in general, to write down the equations that describe at least the more important economic relations. As mentioned on page 3, two equations are needed to describe the relationships indicated by the chart in the upper left corner of figure 2, and we may wish to use also an identity to remind us of a basic assumption on which this chart is based. In the discussion that follows, we use the following symbols:

Q_p - Production

Q_c - Consumption

P_r - Retail price

P_f - Local market or farm price

D - Disposable income

M - Factors associated with marketing costs

A convenient way to summarize the relations implied by the economic model is as follows:

$$P_r : Q_c, D \quad \text{(price-consumption relation)} \qquad (1)$$

$$P_r, P_f ; M \quad \text{(price level relation)} \qquad (2)$$

$$Q_c = Q_p. \quad \text{(production-consumption identity)} \qquad (3)$$

A colon may be read "depends on;" a comma may be read "and;" and a semicolon may be read "appear in a relation with." The variables to the left of the colon or semicolon are current endogenous variables and those to the right are regarded as predetermined within the particular model. A colon is used only when a single variable appears on the left and the relation is believed to be of a causal nature.

If production is a predetermined variable and consumption and production are assumed to be identical for practical purposes, as implied by equation (3), consumption is assumed to be a predetermined variable. In this model, disposable income also is taken as predetermined. We then are interested in finding the retail price that is consistent, in an economic sense, with this consumption and consumer income. We do this by statistically fitting equation (1). Once we have that price, we can find the local market price from the relationship implied by equation (2).

An advantage in using colons and semicolons rather than a more conventional equation form is that we need not, at this stage, become involved with the essentially statistical problems of whether to use actual data or logarithms or similar matters for which a decision might be implied by a conventional equation. Analysts sometimes have used the form

$$P_r = f_1(Q_c, D) \tag{1.1}$$

$$P_f = f_2(P_r, M) \tag{2.1}$$

but this fails to distinguish between endogenous and predetermined variables and may suggest with respect to equation (2.1) that a third equation with P_r dependent is needed to go from P_r to P_f, whereas our model assumes that the relation is nearly enough a functional one so that only a single statistical relation is needed. If relation (2) is fitted by a simultaneous equations approach, rather than by least squares, results that are algebraic equivalents are obtained regardless of which endogenous variable is treated as the one having a coefficient of one.

Let us now go through a similar formulation for the economic model implied by the diagram in the upper right corner of figure 2. The diagram does not show production of beef and pork, but identities like equation (3) are implied for each item. We omit these identities in what follows. This diagram also omits local market prices and, in the following, we omit the two implied price level equations. Thus the diagram implies the following equations, where the subscript b relates to beef and the subscript p relates to pork:

$$P_b : Q_b, Q_p, D \quad \text{(Price relation for beef)} \tag{4}$$

$$P_p : Q_b, Q_p, D. \quad \text{(Price relation for pork)} \tag{5}$$

As discussed beginning on page 87, if we wish to derive the direct- and cross-elasticities of demand for beef and pork, as normally defined in economic literature, we need to write these equations in another way; but if we think of them chiefly as price determining relations, then the form shown here is correct.

A complete model is one that contains one equation for each endogenous variable. In general, complete models are required if we wish to derive from them equations to be used for analytical purposes or prediction. If we are interested only in ascertaining the probable magnitude of certain coefficients then, at times, a complete model is not required. The author has found by experience that it is generally advisable to at least formulate a complete economic model in terms of the sort of relations given here. Balancing of the total number of endogenous variables with the total number of relations frequently helps in the basic formulation of the model. After the complete model has been written down in symbolic form, a decision may be reached at a later stage to fit statistically only part of the equations.

Our first example might be thought of as a 1-, 2-, or 3-equation model, depending on whether we consider as endogenous variables (1) only P_r, (2) P_r and P_f, or (3) P_r, P_f, and Q_c. In models of this sort, Q_p always is considered as predetermined, as no supply equation is shown. Factors which tend to make production for many agricultural commodities essentially a predetermined variable are discussed beginning on page 44. Since consumption is assumed to equal production, it normally is considered directly as a predetermined variable; but if equation (3) is considered as an integral part of the model rather than a definitional identity, then Q_c would be considered as endogenous. The analyst should always regard the second model as at least a 2-equation one, with both P_b and P_p endogenous; the reasons for this are given on page 3, in the discussion of the related diagram. The equations shown can be used to forecast prices from a given production of beef or pork, but other equations are needed to show how consumption of each depends on the respective prices.

Each of these models is on a national aggregate basis and each assumes production, marketing cost, and disposable income to be predetermined variables and production and consumption to be nearly identical. In some cases, prices at the end of or during the season may decline to a point where only part of the crop is harvested. If this is frequently true, then a harvesting equation must be introduced to show the economic factors that determine the amount of unharvested production. An example of this sort for watermelons is discussed in the next section. For some commodities, marketing charges may depend in part on the quantity moving through the market; if this is so, an additional equation showing factors that affect marketing charges must be added.

ECONOMIC MODELS OF INCREASING COMPLEXITY

The best way to gain facility in formulating economic models is to examine some that have been constructed by other investigators. The examples given here are designed to demonstrate how the number of relations tends to increase as certain simplifying assumptions are relaxed. Only enough detail is given to indicate the basic model-building principles that are involved; the reader is referred to the original sources for complete details.

We indicated on page 3 that if (1) production can be assumed to be determined in advance of the marketing season, (2) essentially all of the production moves directly into a single outlet, and (3) no close substitute or complementary commodities exist, then our model can be thought of as containing only a single structural demand or price equation. But, as noted on page 3, if close substitutes exist, then we must have as a minimum one equation for each of the several substitutes. If in some years part of the crop is unharvested for economic reasons, then we must add to our model an equation for each such item showing the economic factors that determine the quantity harvested.

Suits (88) made use of an equation of the last-named sort in developing an econometric model of the watermelon market. We translate his notation into that used here for our first example. He points out that in no event can the harvest of melons exceed the crop. Thus the harvest equation has the general form

$$\left.\begin{array}{ll} & Q_c = a + b_1 \frac{P}{W} + b_2 Q_p \\ \text{or} & Q_c = Q_p \end{array}\right\} \quad \text{(harvest relation)} \qquad (6)$$

whichever gives the smaller value of Q_c, where Q_p is production, Q_c is the number of watermelons harvested, W is an index of southern wage rates on farms, and P is price. In addition, we have a demand equation similar to that indicated by relation (1) except that it contains two endogenous variables. 6/ Thus, if no close substitutes or complements exist, we have at a minimum a 2-equation model, one equation of which is made up of two alternative forms. The two endogenous variables are, of course, Q_c and P. The composite relation (6) is part of a statistical rather than an economic model, but in this instance the pure economic model does not bring out the nature of the assumed relationship.

We next turn to a commodity for which production is assumed to be a predetermined variable and for which no close substitutes are assumed to exist, but which moves into four price-determined outlets. The commodity is wheat, and we make use of the system of equations developed by Meinken (71). A simplifying assumption is that a single price for wheat can be used in each of the demand equations. The model permits us simultaneously to estimate domestic and world prices of wheat and utilization in each of the four outlets. The following symbols are used: C_h, use for food products; C_f, use for feed; C_e, use for export; C_s, use for commercial storage; P_w, price in world markets; P_d, price in domestic markets; Q_d, domestic supply less use in certain outlets not affected by price; Q_w, world supply; and Z_i, a predetermined variable other than supply. The following structural relations are involved:

$$P_w : Q_w, Z_1 \quad \text{(world price relation)} \qquad (7)$$

$$C_h, P_d ; \text{some } Z\text{'s} \quad \text{(domestic demand for food)} \qquad (8)$$

$$C_f, P_d ; \text{some } Z\text{'s} \quad \text{(domestic demand for feed)} \qquad (9)$$

6/ Suits also gives a supply relation showing the economic factors that affect Q_p but, as only prices in a prior year are used, we need not concern ourselves with this aspect.

C_e, P_d, P_w ; some Z's (domestic net export relation) (10)

C_s, P_d ; some Z's (domestic storage relation) (11)

$C_h + C_f + C_e + C_s = Q_d.$ (domestic supply-utilization identity) (12)

The reader will note that the above system of equations contains no domestic price relation, but no such relation is required because domestic prices must be at a level such that the identity implied by relation (12) holds exactly. Formulation of the complete model verifies this fact, as we have the same number of economic relations as endogenous variables. 7/

This example shows another advantage of writing our economic model in this form; at this stage we need not specify which Z's come into each equation. The number of Z's affects the amount of work involved in fitting and to some extent has an effect on the statistical reliability of the results, but the number included in each relation has no effect on the basic structure of the model unless an identification problem is involved. 8/ As we see later, some equations may have no variables on the right of the semicolon.

We next turn to a model for dairy products developed by Rojko (81) which is similar to that for wheat except that a price relation is used for each of the several dairy products. The retail price of butter, for example, differs from the retail price of milk or ice cream for a variety of reasons, but the value of milk used for making butter at the farm on the average must be equivalently priced with that used for making any other dairy product or no milk will be channeled into the lower-priced outlet. This economic fact is allowed for in the model by showing the price in each outlet as a function of the same set of predetermined variables. Using a notation similar to that in other examples, we have the following symbols in a model that relates to a period following World War II: Q_f, use for fluid milk and cream; Q_b, use for butter; Q_c, use for American cheese; Q_o, use for other manufactured dairy products; Q_m, quantity of margarine sold; P_f, price of fluid milk and cream; P_b, price of butter; P_c, price of American cheese; P_o, price of other manufactured dairy products; P_m, price of margarine; P_s, price of substitutes for cheese; Q, total quantity of milk available for consumption; and D, disposable consumer income. The price of margarine, P_m, is assumed to be determined by the fats and oils economy and thus to be exogenous to this model, but the quantity of margarine sold is determined in part by the dairy economy. P_s is assumed to

7/ The model could have been written to include a domestic price relation but, had this been done, one of the other equations would have been omitted.

8/ Most equations that relate to supply-demand structures for individual commodities are highly overidentified (see page 63). For certain other kinds of models, problems of identification are important, and we may need to specify in the economic model which Z's appear in which equations.

be exogenous to this model, as is D. Q is assumed to be a predetermined variable, although a research project currently underway is attempting to ascertain whether this is justified. The model thus consists of 4 endogenous quantities and 4 endogenous prices of individual groups of dairy products, and the endogenous quantity of margarine sold--or 9 endogenous variables in all. The following structural relations are involved:

$Q_f, P_f ; D$	(demand for fluid milk and cream)	(13)
$Q_b, P_b ; D, P_m$	(demand for butter)	(14)
$Q_c, P_c ; D, P_s$	(demand for American cheese)	(15)
$Q_o, P_o ; D$	(demand for other dairy products)	(16)
$Q_m, P_b ; D, P_m$	(demand for margarine)	(17)
$P_f : D, P_m, Q$	(price relation for fluid milk and cream)	(18)
$P_b : D, P_m, Q$	(price relation for butter)	(19)
$P_c : D, P_m, Q$	(price relation for American cheese)	(20)
$P_o : D, P_m, Q.$	(price relation for other dairy products)	(21)

Since we have 9 equations corresponding to our 9 endogenous variables, this is a complete model. However, the following identity always holds:

$$Q_f + Q_b + Q_c + Q_o = Q. \quad \text{(production-utilization identity)} \qquad (22)$$

An alternative way to formulate a model of this type is as follows. We retain the 5 demand relations, but in place of the 4 price-relations, we use a single price relation for all milk, and then relate each of the individual prices to this overall price through a price level relation similar to that used in relation (2). This gives us 10 relations in all, but our balance is retained as this model has an extra endogenous variable, namely the overall price of milk. Estimates of prices from the alternative formulation in general differ somewhat from those obtained from the original formulation because a different sort of market structure is assumed. Rojko (81) presents results from both formulations for several models that relate to dairy products.

In the models discussed so far, production is assumed to be determined chiefly by economic or other factors that exert their influence before the start of the marketing year. Allowance is made in the watermelon example, however, for partial harvesting of the crop for economic reasons. For a number of livestock products that are produced on a continuous basis, production can be modified during the marketing period if this period is taken to be as long as a year, as is frequently done in statistical analysis. This is true for eggs, because annual production can be affected by current price through both (1) varying the culling rate and (2) changing the number of replacement chicks raised. Our next model, which relates to eggs, allows for such factors by providing relations that permit supply to be partially endogenous. This example was formulated and fitted by Gerra. 9/

The model consists basically of two endogenous supply relations, three identities that relate to supply, a demand equation, and a farm-retail price relation. Four more relations are required, however, to explain endogenous variables that appear in other relations. In a notation similar to that used for other examples in this handbook, the following endogenous variables are involved: Q_p, production of eggs; Q_c and Q_c', consumption of eggs; P_r and P_r', retail price of eggs; P_f and P_f', local market price of eggs; N_a, average number of layers during the year; N_p, number of replacements started during January-June; N_c, number of layers culled during the year; and S', eggs placed in storage during spring. A symbol with a prime (') indicates that the variable relates to the period January-June; lack of a prime indicates that the variable relates to the calendar year. A prime is not used on N_p because, although this variable relates to replacements started in spring, the replacements do not enter the flock until fall. We have 11 endogenous variables. The following predetermined variables need to be specified to indicate the nature of the economic model: R, rate of lay per hen in each year; Q_m and Q_m', miscellaneous uses for eggs other than as food by civilians, including net change in stocks as of January 1 and net exports; N_b, number of layers on hand at the start of the year; N_d, number of layers dying during the year; P_g and P_g', price of feed grains, which is assumed exogenous for this model; and Q_p', production of eggs during January to June, which is assumed to be determined by economic factors in operation prior to the start of the marketing year. The symbol Z_i is used for other predetermined variables. R is taken as predetermined because price relations normally are of such a nature that it pays to feed hens as much as they will eat, so that the rate of lay depends chiefly on developments relating to breeding, methods of feeding and housing, and so forth, rather than on the level of economic factors during the marketing period.

The model consists of the following 11 relations: 10/

$N_p,\ P_f'\ ;\ P_g'$ (replacement relation) (23)

$N_c,\ P_f\ ;\ P_g$ (culling relation) (24)

9/ Gerra, Martin J. The Supply, Demand, and Price Structure for Eggs. Unpublished manuscript. 1957.

10/ The model as used by Gerra differs slightly from this in that Q_m is defined so as to use a plus sign rather than a minus sign in relation (27) and Q_m' is omitted due to a lack of data.

$N_a = N_b + N_p - N_c - N_d$ (average number of layers identity) (25)

$Q_p = R\ N_a$ (production identity) (26)

$Q_c = Q_p - Q_m$ (consumption identity) (27)

$Q_c,\ P_r$; some Z's (price-consumption relation) (28)

$P_f,\ P_r$; some Z's. (price level relation) (29)

The following relations apply to the January-June period:

$S',\ \frac{Q_p'}{Q_p}$; some Z's (storage in spring relation) (30)

$P_r',\ Q_c',\ S'$; some Z's (price-consumption relation) (31)

$P_f',\ P_r'$; some Z's (price level relation) (32)

$Q_c' = Q_p' - Q_m' - S'$. (consumption identity) (33)

The economic reasoning behind these relations should be clear to the reader except for the following: (1) In the statistical model for relation (30), storage is assumed to depend on (a) the ratio between production in spring (assumed to be predetermined) and production for the entire year (assumed to be partially endogenous) and (b) profits made from the storing operation in the preceding year. (2) Q_m and Q_m' are in fact at least partially endogenous, but in most years these net uses for export or storage represent only a small percentage of production or consumption and, when they have represented more than this, they have essentially reflected Governmental activity of an institutional nature. (3) The last four relations are needed in the model to get P_f' for use in relation (23).

The last model that we consider in this section is one relating to asparagus developed by Carstensen. 11/ Here production is assumed to be a predetermined variable, but we consider three producing and consuming regions for fresh asparagus and also allow for production and consumption in terms of national aggregates for frozen and canned asparagus. The regions referred to are (1) the West, (2) the central part of the country, and (3) the East. In terms of volume of production, California is the chief State in the West, Michigan and Illinois in the Midwest, and New Jersey in the East. As data on retail prices are not readily available, market prices are taken at wholesale and for fresh asparagus relate to prices, respectively, in (1) San Francisco and Los Angeles, (2) Chicago, and (3) New York. A national average price must be used for canned and frozen asparagus, as prices by areas are not available

11/ Carstensen, Hans. The Demand and Price Structure for Asparagus. Unpublished manuscript. 1958.

in published form. No allowance is made in the model for foreign trade in the processed items. In effect, net trade is assumed to be zero. This simplification reflects in part a lack of knowledge as to the nature of the economic relations involved. The model relates to the late spring season of heavy production, and each region is assumed to produce enough asparagus at least to take care of its needs for fresh use. The surplus in each area is used for processing.

The notation used for other examples in this handbook is not satisfactory here because of the need to distinguish between the three regions. The symbol Q is retained for production, but the letters A, B, C, and so forth, are used for other variables in the order in which they are defined. The subscript w is used for the West, c for the central part of the country, and e for the East. The following endogenous variables are included in the model: A_w, A_c, and A_e, consumption of fresh asparagus in each area; B_w, B_c, and B_e, wholesale price of fresh asparagus in each area; B, a weighted average national aggregate price for fresh asparagus; C, consumption of canned asparagus; D, wholesale price of canned asparagus; E, consumption of frozen asparagus; F, wholesale price of frozen asparagus; G_w, G_c, and G_e, local market price for fresh use; H_w, H_c, and H_e, local market price for processing; I_w, I_c, and I_e, use for processing in each area. We have 20 endogenous variables in all. The following predetermined variables need to be specified to illustrate the nature of the economic model: Q_w, Q_c, and Q_e, production of asparagus in each area; J_w, J_c, and J_e, economic factors that affect the relative cost of preparing asparagus for sale for fresh use or for processing in each area; K_w, K_c, and K_e, economic factors that affect marketing costs for fresh asparagus; L_w, L_c, and L_e, economic factors that affect marketing costs, including transportation, for canned asparagus; M_w, M_c, and M_e, economic factors that affect marketing costs, including transportation, for frozen asparagus, N, year-to-year net increase in stocks of canned and frozen asparagus. As in other examples, the symbol Z_i is used to represent other predetermined variables.

The model consists of the following 20 relations:

A_w, B_w, D, F ; some Z's	(demand for fresh asparagus in the West)	(34)
A_c, B_c, D, F ; some Z's	(demand for fresh asparagus in the Midwest)	(35)
A_e, B_e, D, F ; some Z's	(demand for fresh asparagus in the East)	(36)
C, D, B, F ; some Z's	(demand for canned asparagus)	(37)
E, F, B, D ; some Z's	(demand for frozen asparagus)	(38)
$Q_w = A_w + I_w$	(production-utilization identity in the West)	(39)

$Q_c = A_c + I_c$	(production-utilization identity in the Midwest)	(40)
$Q_e = A_e + I_e$	(production-utilization identity in the East)	(41)
$G_w, H_w ; J_w$	(price relation between use for fresh and for processing in the West)	(42)
$G_c, H_c ; J_c$	(price relation between use for fresh and for processing in the Midwest)	(43)
$G_e, H_e ; J_e$	(price relation between use for fresh and for processing in the East)	(44)
$G_w, B_w ; K_w$	(price relation between local market and wholesale price of fresh asparagus in the West)	(45)
$G_c, B_c ; K_c$	(price relation between local market and wholesale price of fresh asparagus in the Midwest)	(46)
$G_e, B_e ; K_e$	(price relation between local market and wholesale price of fresh asparagus in the East)	(47)
$D, H_w ; L_w$	(price relation between local market price for processing in the West and wholesale price of canned asparagus)	(48)
$D, H_c ; L_c$	(price relation between local market price for processing in the Midwest and wholesale price of canned asparagus)	(49)
$D, H_e ; L_e$	(price relation between local market price for processing in the East and wholesale price of canned asparagus)	(50)
$F, H_w ; M_w$	(price relation between local market price for processing in the West and wholesale price of frozen asparagus)	(51)

$F, H_c ; M_c$ (price relation between local market price for processing in the Midwest and wholesale price of frozen asparagus) (52)

$F, H_e ; M_e$. (price relation between local market price for processing in the East and wholesale price of frozen asparagus) (53)

The following identities are implied by the model:

$$B = \frac{A_w B_w + A_c B_c + A_e B_e}{A_w + A_c + A_e}$$ (weighted average national price for fresh asparagus) (54)

$$C + E = I_w + I_c + I_e - N.$$ (national aggregate processing identity) (55)

As these examples illustrate, decisions frequently must be made by the analyst regarding the complexity of the model to be used and the degree of aggregation. Klein (58, pp. 185-200) has a useful suggestion here in connection with his discussion of sector models. He suggests that a master model for the entire national economy be formulated; he and others have made some progress in formulating models of this type. 12/ Models for individual industries, individual commodities, or individual regions then can be formulated and fitted as separate entities in such a way that they can be "grafted" onto the master model. In the examples discussed so far, variables that relate to the national economy, such as disposable income, have been treated as exogenous, but we did not mean to imply that this was necessarily an approved method. With Klein's approach, either of the following methods may be used for variables of this sort: (1) The computed value of disposable income for each observation, based on the master model, is used as a predetermined variable in fitting equations for a particular sector or (2) disposable income is treated as an endogenous variable and the predetermined variables on which it depends are brought in as predetermined variables in the system for the sector.

The same general approach can be used in narrower fields. Hildreth and Jarrett (49) developed a model that relates to the feed-livestock economy but for which all livestock and livestock products are aggregated. We have considered in this section a model for eggs for which the price of feed grains is considered as determined outside the egg economy. Instead, we could have used as a price for feed grains the calculated price from the Hildreth-Jarrett model. In later research studies, feed-livestock models might be fitted for separate regions, with separate relations for each major type of livestock,

---

12/ See Klein and Goldberger (60) and Suits and Goldberger (89). For criticisms of these models, see Christ (14) and Fox (37).

and these regional models then could be grafted onto a master feed-livestock economy model, which in turn could be grafted onto a master model of the entire economy. Thus, over time, a series of studies could be built up, each of which would be manageable as a research unit, but all of which would tie together to make a united whole. 13/

CONSIDERATIONS IN FORMULATING STATISTICAL MODELS

After writing down the economic model, we are faced with the difficult task of locating or developing the necessary data and of deciding upon the exact nature of the variables and relations to be used. As noted on page 7, decisions in this area frequently must be based chiefly on experimentation or judgment. In many instances, alternative methods exist, but methodological experts sometimes do not agree as to the best procedure to follow. In certain cases, particular conditions that are known to prevail with respect to the commodity area to be studied indicate a preference for one method over another. In this section, the more important decisions that must be made are listed, and some of the considerations that indicate a preference for one method over another are discussed.

Choice of the Time Unit 14/

Most analyses of factors that affect the price or consumption of a given commodity are based on annual data for either calendar or crop years. This is satisfactory if conditions within the period are sufficiently homogeneous. For products that are produced continuously throughout the year, such as dairy products or eggs, available published data frequently relate to a calendar year and this is the most convenient time unit to use. Data for crops, on the other hand, normally are published by marketing years, so that this is the most convenient time unit. For some items, however, the economic structure differs considerably in one part of the year from that in other parts. For asparagus, for example, production in early spring is concentrated almost entirely in California and practically all of the crop is shipped for use in fresh form. Thus the economic model that relates to early spring differs materially from the one that relates to late spring--described on page 15--and two economic and statistical models for asparagus are required, one relating to early spring and one relating to late spring. A similar situation exists for corn. During the period November to May, the supply variable that affects price in years for which price is determined chiefly by free-market influences is the total supply of feed concentrates for the October-September marketing

---

13/ A start in this direction, for major sectors of agriculture in relation to the general economy, has been made by Cromarty, William A., Economic Structure In American Agriculture. Ph.D. thesis, Michigan State University, 1957.

14/ Decisions with respect to the time unit are based more on economic considerations than on statistical ones and frequently must be made before formulating the economic model.

year. During the period June to September, however, allowance must be made also for the effects on price of new-crop supplies of oats and barley and for the expected size of the new corn crop. [See Foote (26).]

For some crops, the marketing season covers only a few weeks or months. Here, naturally, the several variables used in the analysis must relate to this period. At times, interest is centered on the prediction of sales or prices during specific weeks or months. A study of this sort for apples was made by Pubols (80). He considered factors that affect price in each month of the fall-winter marketing season from September through May, using as variables in each, items that are known in advance of the month for which predictions are made. Foytik (39) made an analysis of the weekly price-demand structure for California plums, and Harrington and Gislason (46) have in process a study of factors that affect daily sales of fruit from retail stores. Shorter periods tend to be more homogenous than longer ones but the effect of this on the analysis may be offset, at least in part, by the fact that irregular or nonmeasurable factors become more important. The period chosen should be long enough to average out the effect of irregular or nonmeasurable factors and short enough to insure that a relatively homogeneous set of factors are operating.

Years To Be Included

Now that enough data have accumulated since World War II to permit running analyses for the postwar years, we generally run analyses for (1) the years between World War I and World War II, (2) the years following World War II, and (3) the entire period. If the differences between the coefficients for the three analyses are not statistically significant (see page 180), results for the entire period are used as the best predictor for the future, since this analysis contains the largest number of observations. Otherwise, the second analysis normally is used as the best predictor for the future. For some commodities, neither price ceilings nor rationing were in effect during World War II and, unless the war is believed to have had other abnormal effects on the demand and price structure, the war years can be retained in the analysis.

At times, other years may be so abnormal, owing to special conditions, that these also should be omitted from the analysis. For example, a linear relation may hold except for unusually large or unusually small crops. If, because of considerations to be discussed later, a linear relation is desired in the fitting process, data for these extreme years may be omitted from the analysis or the actual data may be replaced by figures that are consistent with a linear relation. 15/ Except in extreme cases, the analyst should decide which years to include before he runs the analysis. Calculated values for these years should be obtained and checked to see that they are in line with expectations.

15/ This procedure was used by Meinken (71, pp. 24-25) in deriving figures on use for feed in his 6-equation model of the wheat economy. A scatter diagram between wheat fed and the price spread between wheat and corn indicated

For some studies, the basic structure of the analysis may have changed from one period to another. Under these circumstances, the analysis might be confined to the latest homogeneous period or, if interest exists in the nature and extent of the structural change, "before" and "after" analyses can be run. The model for dairy products described on page 12 relates to the period following World War II when margarine had become a major competitor of butter. In this instance, analyses also were run for the pre-World War II (and pre-margarine) period to show how the increasing competition from margarine has affected the entire demand and price structure for the dairy industry. [See Rojko (81).] If sufficient data were available for the period during which frozen asparagus has been an important factor in total asparagus consumption, only the model described on page 15 might have been fitted. As data on canned asparagus are available for a considerably longer period, a model which allows only for competition between fresh and total processed asparagus was fitted also. [See Carstensen, op. cit.]

Sometimes the nature of the structural change is such that it can be allowed for by basing the analysis on one homogeneous period and then algebraically modifying the model to take account of the structural change. This was done by Meinken (71) in the case of wheat. The model was fitted for the marketing years beginning in 1921-29 and 1931-38. These were years for which average prices and utilization for the entire marketing year are believed not to have been affected significantly by Government action other than processing taxes, tariffs, and export subsidies which are allowed for directly in the model. In years for which domestic prices are maintained above their free-market level by a Government price-support program, the system can be used to estimate the use in each outlet and the probable accumulation of stocks by the Government in the following way: (1) The domestic price, which otherwise would be predicted from the system, is set at the expected season-average level under the support program. (2) This price is used to determine the free-market utilization in each of the four price-determined outlets. (3) When these amounts are subtracted from supplies available for use in these outlets, the remainder is expected to show up as end-of-year stocks held by the Government under the loan program. 16/

---

that, for the years used in the study, wheat fed in 3 years was higher than indicated by a linear regression. The apparently linear relationship for the other years was extended to cover these years, and net exports were increased by the difference between the actual amount of wheat fed and that read from the linear regression. The value read from the line was substituted for the actual amount fed in fitting the 6-equation model and the adjusted data on exports were used.

Exports were picked as the outlet on which a compensating adjustment should be made because the elasticity with respect to price for exports was believed to be of approximately the same magnitude as the elasticity with respect to price for quantity fed. Subsequent analysis confirmed this belief.

16/ Other ways of algebraically modifying Meinken's wheat model to allow for specified changes in structure are discussed in Meinken (71, pp. 49-50, 89-93) and Foote and Weingarten (31). These techniques are summarized beginning on page 189.

We referred earlier (footnote 3) to an analysis for peas by Shuffett (84) in which a similar procedure was used. The statistical analysis was based on data for 1921-41, when frozen peas were of negligible importance. By 1948-52, consumption of frozen peas on a shelled-weight equivalent basis exceeded that of fresh peas on the same basis and, during the last few years, fresh peas have declined to a negligible part of the total consumption. Prices estimated for 1942-52, based on the 1921-41 relationships, differed from actual prices by more than did the estimates in the earlier period and, from 1946 to 1952 (the last year for which data were available at the time his study was published), actual prices were below the estimated price in all years, reflecting increased competition from frozen peas during these years. The equivalent consumption of frozen peas was added to the supply variable for 1937-52, and new estimates of price were computed. Prior to 1944, when consumption of frozen peas per capita first exceeded 0.5 pound, residuals for the two sets of estimates were nearly equal on the average; for all years from 1946 to 1952, estimates based on the supply of fresh peas plus consumption of frozen peas are closer to the actual price than are estimates based on supply of fresh peas alone. From this, Shuffett concluded that fresh and frozen peas are close substitutes and that the analysis based on prewar data can be used in the postwar period by adding frozen consumption to the fresh supply before obtaining the calculated price.

Sometimes analysts combine data from periods that are believed to be non-homogeneous into a single analysis by using the so-called 0-1 variable. Such a variable has a value of 0 in one period and a value of 1 in another period. In a least squares analysis, the regression coefficient on the 0-1 variable indicates the extent to which the dependent variable is larger or smaller in the second period than in the first, after allowing for the net effect of all of the factors specified in the analysis. This approach is satisfactory if the only effect of the change in structure is to affect the level of the dependent variable and if the entire adjustment occurs within a single year. If the change in structure affects the magnitude of the coefficients or the basic nature of the relationships, or the change in structure occurs gradually over time, use of a 0-1 variable is unsatisfactory. Changes that occur gradually over time sometimes can be allowed for by use of a time trend as described on pages 39 to 43.

In some analyses, deviations from one average for certain variables appear relevant for some years and deviations from a different average are relevant for other years. This appeared to be true in an analysis of factors that affect mill consumption of cotton run by Lowenstein and Simon (65). One of the independent variables represented the departure from normal in the ratio of stocks to unfilled orders. The average level of this ratio shifted from about 1 in the period 1926-40 to about one-third in 1947-52. Deviations from each of these averages were used to represent deviations from normal in the respective periods, but these deviations from the two periods then were combined for use in a single analysis.

In working with least squares analyses, we have a method that we can use to test whether a structural change has taken place. If only two variables are involved, or the change is believed to affect the relationship only between the dependent variable and one of the independent variables, say X_2, we follow this method: (1) Create a new variable X_2'. X_2 takes on actual values for the first period and zeros for the second period. X_2' takes on zero values for the first period and actual values of X_2 for the second period. (2) Run an analysis using both X_2 and X_2' and the other variables for the entire period. (3) Test whether the regression coefficients on X_2 and X_2' differ by a statistically significant amount by a one-tail test of the type described on page 123. If the coefficients differ significantly, a change in structure is assumed to have taken place, and analyses should be run separately for each period. Presumably, a similar procedure could be used if structural changes were believed to have taken place for the relationship of the dependent variable to more than one of the independent variable. If this were done, a multiple test on the regression coefficients would be made.

At times, the analyst may know that many changes in structure have taken place but still wish to use a long series of years in the study. This was true for coffee in an analysis run by Hopp and Foote (50). They say, "Because of known inaccuracies in the data on supply and disappearance and the many extraneous factors that have affected the coffee economy, many years are required to obtain statistically-significant results for the more important causal factors." The analysis was based on marketing years from 1882 through 1949, omitting 1890-91, 1914-17, and 1940-46. The equation was constructed by expressing each of the factors relating to supply as a ratio to an appropriate factor relating to demand. For example, the variable relating to available world stocks at the start of the marketing year was divided by average world imports or deliveries for the preceding 5 years to allow for the long-term gradual increase in coffee consumption. They comment, "Thus the analysis became basically a relationship between two types of variables--(1) price (dependent) and (2) supply deflated by demand--but because there is more than one supply factor, a multiple regression equation is required. A time factor also was included and found significant." Coefficients on two of the supply factors differed from zero by a statistically significant amount, and the analysis, when adjusted for certain cyclical effects (see page 117), explained prices satisfactorily during 1949-54.

Use of Prices at the Local Market, Wholesale, or Retail Level in Equations that Relate to Demand

In the section on economic models, we suggest that equations designed to measure the elasticity of demand for consumer goods should be based preferably on prices at the retail level or, if these are not available, then at the wholesale rather than the local market level. In measuring domestic demand for items used to a large extent by industry, a wholesale price may be preferred to a retail price. In measuring the demand for livestock feeds or fertilizer, a local market price is preferred. If a price at the desired level is not available, it is not, in general, possible to measure the elasticity of demand at this level, but, as shown in the section beginning on page 100, it is possible, under relatively unrestrictive assumptions, to set a lower or upper *limit* to the elasticity of demand at this level. 17/

If the available price relates to the local market or wholesale level and an elasticity at retail is desired, then a lower limit to the elasticity of demand at retail is given. If the available price relates to the wholesale or retail level and an elasticity at the local market level is desired, then an upper limit to the local market elasticity is given. In the discussion that follows, we consider problems involved in obtaining a lower limit for the elasticity at retail when only a wholesale or local market price is available. The reader should remember that an upper limit for the elasticity at the local market level is given if we consider a derived measurement at this level based on retail or wholesale prices.

When marketing margins do not depend on the quantity moving through the market.--If a wholesale or local market price is used in an equation designed to set lower limits to the income and price elasticities of demand at retail, we will generally, but not necessarily, wish to include a marketing cost factor as a variable in the equation. The reason for this can be seen by considering the following model. Let Q_c be domestic consumption; D, disposable income; P_r, price at retail; P_w, price at wholesale; and M, a factor relating to marketing costs, such as industrial wage rates, and assume that the following consumer demand and price level relations hold:

$$Q_c = a_1 + b_{11}P_r + b_{12}D \qquad (56)$$

$$P_r = a_2 + b_{21}P_w + b_{22}M. \qquad (57)$$

If data for P_r are not available, we can substitute equation (57) for P_r in equation (56) and obtain the following *partially-reduced form* or *derived demand* equation: 18/

$$Q_c = a_3 + b_{31}P_w + b_{32}M + b_{33}D. \qquad (58)$$

Equation (58) replaces equations (56) and (57) in the statistical model. Inasmuch as M appears in equation (57), it must, if no further equations are specified, appear in equation (58).

An estimate of b_{31} can be used to set a lower limit to the value of b_{11} and, hence, to the elasticity of demand with respect to price at the retail

---

17/ Where the elasticity of demand is defined as a positive quantity.
18/ A partially-reduced form equation is any equation that results from a substitution of this sort. In this case, it is related to what Marshall (67) calls a derived demand curve.

level. 19/ However, when the quantity passing through the marketing sector does not influence the difference between wholesale and retail prices, b_{33} equals b_{12} and we can obtain an exact estimate of the elasticity of demand with respect to income at retail from the corresponding coefficient in equation (58). The same reasoning applies to both coefficients if only local market prices are available except that P_w is replaced by P_f.

In the case considered above it can be shown, under unrestrictive assumptions, that the elasticity of demand with respect to the wholesale or local market price is a lower limit to the elasticity of demand with respect to the retail price; that is, the elasticity obtained from the coefficient b_{31} in equation (58) is a lower limit to the elasticity that would have been obtained from b_{11} in equation (56) had data been available to estimate that equation. This proposition is intuitively obvious, since the existence of marketing groups (processors, dealers, and so forth) between the producer and the consumer should tend to speed up adjustments in the market. This notion is further elaborated and a proof given in the section beginning on page 103.

When marketing margins depend on the quantity moving through the market. --If marketing margins themselves depend in part on the quantity moving through the market, as is true for meats 20/ and may be true for many other commodities, 21/ derivation of the reduced-form or derived demand relationship becomes more complicated. Here, in addition to equations (56) and (57), we may have an equation of the form

$$M = a_4 + b_{41}Q_c + b_{42}W \tag{59}$$

where W represents a factor similar to that represented by M in equation (57). W might be industrial wage rates, or it might be a weighted average of wage rates and per unit transportation and container costs. If data are not available on retail prices, then our partially-reduced form equation must be derived through the following steps:

(1) Substitute equation (59) for M in equation (57) to get

$$P_r = a_5 + b_{51}P_w + b_{52}Q_c + b_{53}W. \tag{60}$$

(2) Substitute equation (60) for P_r in equation (56) to get

$$Q_c = a_6 + b_{61}P_w + b_{62}Q_c + b_{63}W + b_{64}D. \tag{61}$$

19/ The reader may easily verify that it is impossible to obtain an estimate of b_{11} in equation (56) from b_{31}, b_{32}, or b_{33} of equation (58). Hence, it is impossible to estimate exactly the elasticity of demand with respect to price at retail from estimates of the coefficients of equation (58).

20/ See Breimyer (11, pp. 691-693).

21/ See page 107.

(3) Transpose equation (61) to give

$$(1 - b_{62})Q_c = a_6 + b_{61}P_w + b_{63}W + b_{64}D. \qquad (61.1)$$

(4) Rewrite equation (61.1) as

$$Q_c = a_7 + b_{71}P_w + b_{72}W + b_{73}D. \qquad (61.2)$$

Comparison of equation (61.2) with equation (58) indicates that they are of exactly the same form, as W in equation (61.2) is the same variable as M in equation (58). Although equations (61.2) and (58) are of the same form, the parameters in equation (61.2) are combinations of the parameters in equations (56), (57), and (59), whereas the parameters in equation (58) are combinations of those in equations (56) and (57) only; hence, the meaning of the two equations (61.2) and (58) is different. 22/ Estimates of the coefficients in equation (61.2) cannot be used to derive exactly the elasticities of demand with respect to price and income at retail, although they can be used to set appropriate lower limits.

In contrast to the previous case, the income elasticity obtained from the coefficient on income in equation (61.2) is not the same that would be obtained from the corresponding coefficient in equation (56) if data were available to fit this equation. Instead, only a lower limit to this coefficient is given. A proof that the coefficient obtained is a lower limit is given in the section beginning on page 106.

If data are available on prices at the several marketing levels, we may wish to include in our model an equation like (59). If b_{41} does not differ significantly from zero, then we may wish to assume that M in equation (57) can be replaced by a predetermined variable W. If b_{41} does differ significantly from zero and we adopt the framework developed above, our structural model should contain equations like (56), (57), and (59). Under such circumstances, equation (57) becomes an identity. Exact estimates of the elasticities of demand at the several market levels can, of course, be obtained.

A frequently used model for certain agricultural commodities is one that contains an equation at the retail level that relates to consumption and an equation that would be primarily applicable at the farm level that relates to use for livestock feed. If data are not available to permit use of the respective price series in the respective equations, a partially-reduced form equation can be used for one of the two consumption levels. Availability of data on prices in general will determine which consumption equation is expressed in a partially-reduced form. If price data at a wholesale level are used, a partially-reduced form equation might be needed in each case, with a variable added in each equation that relates to marketing costs to and from the wholesale level, respectively.

22/ In the section beginning on page 100 we develop a simpler model, which does not require such involved substitutions to reach the reduced form equation.

If different sorts of structural equations are assumed and partially-reduced form equations are to be used, the research worker should make an algebraic analysis of the sort given in preceding paragraphs to check on the exact form of the partially-reduced form equation and the variables involved.

A discussion of methods that can be used to obtain a derived elasticity of demand for use in particular outlets at the local market or farm level or of the elasticity of total demand in all outlets at the farm level is given in the section beginning on page 108.

Choice and Handling of Variables That Cause Shifts in the Demand Curve

In order to isolate price-quantity relationships or demand curves by statistical means, variables that cause such curves to shift back and forth must be included in the analysis. In equations that relate to the measurement of consumer demand from time series data, it is convenient to divide such variables into four categories: Those that relate to (1) consumer income or other measures of the general level of demand on a national basis, (2) the general price level, (3) supplies or prices of competing products, and (4) population. Each of these is mentioned briefly in the paragraphs that follow. In certain demand equations, other types of variables are included. Comments on some of these are given at the end of the section.

Variables that relate to the general level of demand.--Personal disposable income commonly is used to represent the general level of demand in equations that relate to consumer goods. For products that are used chiefly as industrial raw materials, the Federal Reserve Board index of industrial production might be used instead. Friedman (41) suggests that instead of responding to their current income, consumers respond to their long-run expectations of income. He shows a derived series on a national aggregate basis which he feels represents consumer expectations of income; experiments currently are underway within the Agricultural Marketing Service to learn whether use of this series results in improved analyses.

Allowing for effects of changes in the general price level.--If we have reason to believe that a doubling of all price and income variables has no effect on consumption, effects of the general price level should be allowed for by deflation, that is, by dividing each price, income, or marketing margin variable by a variable such as the Bureau of Labor Statistics Consumers' Price Index. This seems a reasonable assumption with respect to most perishable items and perhaps some of semidurable or durable goods. Changes in the value of money may affect the nature of the demand for certain items that involve large initial expenditures, such as automobiles or refrigerators, owing to the effect of such changes on the relative value of fixed assets or liabilities. In such cases we might (1) allow for changes in the price level by including the Consumers' Price Index as a separate variable or (2) deflate and include variables that relate to fixed net assets in the equation. In the section beginning on page 111, a discussion is given of the use of distributed lags to study demand for items that require substantial initial expenditures by consumers.

If our price variables relate to a level other than at retail, some research workers might feel that the Consumers' Price Index is not a satisfactory series to use as an indicator of the general price level when dealing with analyses of consumer demand. However, if a variable relating to marketing costs is included in the analysis, then we can show that the Consumers' Price Index is an appropriate series to use to represent the general price level by a technique similar to that used on page 24.

When we deflate by an index number that contains the price of the commodity with which the analysis is concerned, we tend statistically to bias the regression coefficients downward. A similar tendency holds when such an index number is included as a separate variable. This has led some analysts to construct special index numbers that eliminate the price of the commodity or commodities under study and to use these to deflate the price variables that appear in each analysis. This is theoretically desirable; whether the extent of the bias is sufficient to warrant the work involved is for the analyst to decide. The unadjusted measure of the general price level, however, should be used to deflate variables that relate to consumer income, and for all variables when studying equations that relate to derived demand.

Allowing for competing or complementary products.--Methods for fitting demand equations that involve competing or complementary products are discussed in detail beginning on page 87. Systems of equations frequently are required to measure the direct and cross elasticities in such cases.

Use of per capita data.--To avoid confusing the time trend for population with one that might reflect other effects, per capita data probably should be used whenever applicable.

As the proportion of people of various ages in the population differs from time to time, specially weighted population aggregates may be needed for certain items. For cigarettes, for example, consumption varies by age groups and differently for people in urban or rural areas, and for men or women. [See Sackrin and Conover (83).] In analyzing time series data, a carefully constructed population aggregate is required to avoid confusing trends in consumption with trends in the proportion of men and women in rural and urban areas and in specified age groups. A similar, though possibly less serious, problem exists for textile fibers, hide and leather products, and certain foods. Research currently underway in the Agricultural Marketing Service deals with the development of a population series adjusted for age composition by sex, to be known as clothing expenditure units, based on survey information covering 16 age breakdowns. If consumption differs greatly among income groups, weighted population aggregates that allow for shifts in the proportion of people in the various income groups may improve results from the analysis of time series data.

Measuring changes in demand for export or storage.--Attempts to develop satisfactory shift variables for use in equations that relate to storage or export have been less successful than others, though some limited progress has been made. An equation relating to exports is included in the model for wheat described on page 11, and equations that relate to storage are included in that model and in the one for eggs described on page 14.

Brandow (10, pp. 7-8), in a study relating to the demand for apples, comments, "Export demand could not be satisfactorily represented by actual exports because normally they reflect both foreign countries' willingness to buy and prices and production in the United States. To isolate the former for the years before the war, a regression of fresh apple exports on production was computed and differences between actual and 'expected' exports were taken as a measure of changes in export demand. For the postwar period, when exports were much lower and in some years considerably influenced by shortages of dollar exchange or by U. S. government export subsidies, simple deviations from average net exports were used."

In some analyses, measures of consumer income or the general price level in one or more foreign countries may be used to measure export demand, just as similar variables are used with respect to the domestic market; at other times the level of stocks or supply in the foreign market may be used as a factor causing the demand for domestic products to shift.

Use of Actual Data or of First Differences

When variables are expressed in terms of the change from the preceding year, we say that we are working with first differences. If we are primarily interested in the percentage change from one year to the next, we may express the change in this way or, as is more common, use is made of the first differences of logarithms. First difference analyses also may be run in terms of the actual year-to-year change in the variables. In some analyses, certain variables are expressed as first differences and other variables are used in an actual, or non-first difference, form.

First-difference equations have been used to some extent by price analysts for many years, but they have come into prominence only recently. In equations designed primarily for use in forecasting, first differences are used when emphasis is placed on measuring the factors that affect year-to-year change rather than on deviations from a long-term average. The recent interest in the use of first-difference equations has arisen partly because of the obvious inapplicability of a pre-World War II average to the postwar period. This reason generally is of less importance when we work with deflated data than when actual data are used. First differences may be hard to use in connection with long-range forecasts and in some types of analyses of relative benefits to be obtained from specified Government programs, as the forecast is in terms of a change in level from a preceding year. In certain cases, first differences might be used to derive statistical coefficients that are assumed to apply in future to non-first difference equations; such would be the case if the statistical conditions, described in the next paragraph, in the period for which data are available are consistent with the use of first differences, but in the period for which forecasts are made are believed to be more nearly consistent with the use of actual data.

From a statistical standpoint, first differences should be used in preference to actual data when the successive unexplained residuals from single-equation analyses based on actual data are almost perfectly serially correlated with a positive sign. A transformation to first differences will eliminate most of the serial correlation in the residuals if the analysis is rerun in terms of the transformed variables. By serial or auto-correlation we mean essentially the correlation between a series of observations and the same series lagged by one or more units of time. If the unexplained residual in one year on the average equals a fixed proportion of the unexplained residual in the preceding year plus a random variable, resulting in some degree of positive serial correlation, then a transformation to first differences may remove some of the serial correlation in the residuals. If the serial correlation is less than 0.5 or negative, a conversion to first differences tends to make the degree of serial correlation in the residuals greater in the transformed than in the original analysis, and first differences should not be used.

Many analyses that relate to economic data tend to give serial correlations in the unexplained residuals that are positive when the analysis is based on actual data. The following considerations have a bearing on the extent to which this is likely to be true. The unexplained residuals may represent essentially random errors in the data 23/; if so, we would expect the serial correlation in the residuals to be close to zero. Unexplained residuals also represent the influence of variables excluded from the equation because (1) we have no data with which they can be measured or (2) the influence of each, on the average, is believed to be too small to warrant their inclusion. If we have an idea as to the nature of these excluded variables, we may be able to make assumptions as to whether, on the average, their combined effect on the dependent variable is likely to be similar from one time period to the next. If it is, we would expect positive serial correlation in the residuals; if not, we would expect the serial correlation in the residuals to be small. Thus, our final appraisal as to the likelihood of positive serial correlation in the residuals depends on (1) the extent to which the residuals reflect errors in the data and the extent to which they reflect omitted variables and (2) the nature of these omitted variables. Another possible cause of serial correlation in the residuals is an incorrect specification of the form of the relation, particularly if one or more of the independent variables tends to follow a time trend.

23/ Errors in the data in most statistical analyses, if allowed for at all, are assumed to relate only to the dependent variable. See page 143.

If, when working with actual data, we find positive serial correlation in the residuals and our appraisal suggests the likelihood of some positive serial correlation but not a serial correlation coefficient of one when based on a time lag of a single year, we might find it desirable to make a first difference transformation. We can then determine by means of the Durbin-Watson test 24/ whether we have reduced the serial correlation to a point where it does not differ from zero by a statistically significant amount. If so, we might assume that the transformation is satisfactory and that the usual tests of significance apply to the regression coefficients obtained.

Our interest in serial correlation stems from the fact that most of the statistical methods described in this handbook are based on the assumption of mutual independence among the successive unexplained residuals. As discussed in detail in the section beginning on page 57, if we are interested in estimating the coefficients of a structural equation that contains only a single endogenous variable, so that the least squares approach is the recommended fitting procedure, then the least squares estimates of the coefficients are both statistically unbiased and consistent. 25/ This will be the case even if the independent variables and the true residuals of the equation are serially correlated [see Wold (103) and Wold and Jureen (106, pp. 208-213)]. However, the usual tests of significance of the regression coefficients do not apply, and other types of estimators might give coefficients with a smaller variance [see Watson (97), Watson and Hannan (98), and Gurland (44)] if we knew how to compute them. Wold (103, pp. 283-284) gives a large sample test of significance for the least squares estimates, but we rarely deal with a large sample in economic research.

Cochrane and Orcutt (15, pp. 54-55) suggest that the use of first differences, rather than the original data, tends to eliminate most of the serial correlation in the residuals of economic relationships. If this is the case, ordinary least squares estimates of the coefficients based on the use of first differences are statistically efficient (see page 58) and the usual tests of significance apply. However, the validity of the Cochrane-Orcutt approach rests on very strict assumptions concerning the form of the serial correlation in the residuals (see page 167). Wold (103) and Gurland (44) point out that only rarely are these assumptions likely to apply. Relaxation of the assumptions even a little leads to great computational complexity. Thus statisticians wish to assume mutual independence in the residuals for two reasons: (1) If they don't, they must make specific assumptions about the form of the nonindependence in order to obtain valid standard errors and estimates of the

---

24/ This test for serial correlation in the calculated residuals of a least squares regression is described on page 173. As pointed out by Cochrane and Orcutt (15, p. 45), a test of this sort, rather than a conventional test of whether the (serial) correlation coefficient differs significantly from zero, is required because the calculated residuals of any least squares regression tend to be biased towards randomness.

25/ These terms are defined on pages 57-58.

coefficients that are statistically efficient, and this is difficult in applied research; and (2) even with the simplest forms of nonindependence, computations to obtain such coefficients and their standard errors become relatively complex.

Based on an experimental study by Orcutt and Cochrane (78), the situation appears to be even more complex for systems of equations. Here we are concerned not only with the degree of serial correlation in the residuals in a single equation but also the degree of correlation among the residuals in the several structural equations. If there is no serial correlation in the residuals, correlation among the residuals in the several equations presents no problem, as the method of fitting used with respect to such equations automatically allows for this. However, when both types of correlation prevail, a shift to first differences appears to be of less value with respect to systems of equations than when we deal with a single equation. Orcutt and Cochrane (78, p. 356), who also consider the necessity of working with relatively small samples in applied research, conclude, "Unless it is possible to specify something about the intercorrelation of the error terms in a set of relations and to choose approximately the correct autoregressive transformations, a certain amount of scepticism is justified concerning the possibility of estimating structural parameters from aggregative time series of only twenty observations."

We now turn to some considerations of the use of first differences that relate more directly to applied research. In essence, however, they are special applications of the statistical considerations discussed in preceding paragraphs.

In some cases, strong trend factors tend to overshadow the effect of economic variables, but the trend factors cannot be allowed for directly in the analysis due to a lack of data or knowledge. If the trend factors can be allowed for by a time trend (see page 39), this might be preferred to a shift to first differences, but if this does not appear feasible, a study based on year-to-year changes may give meaningful results whereas one based on actual data may result in multiple correlations close to zero. This has been found true, for example, in certain studies that relate to factors that affect milk production per cow and to consumption of tobacco products.

Use of first differences has been advocated in the past to avoid obtaining an unduly high correlation between two or more variables when each of them is more closely correlated with an unrealized third factor than with each other. Here it is true that, in certain cases, improved results are obtained by the use of first differences, but more reliable results are obtained when the third variable is included in the analysis and its effects are allowed for by statistical means, either directly or by its use as a deflator.

At times, available data may not be strictly comparable from one year to the next, but the comparability is greater from year to year than over longer periods of time. This is true of some statistics that relate to commercial vegetables, for example. Here the number of States included in the total has been increased gradually over time. Under such circumstances, use of first differences generally improves the reliability of the results.

Measuring Relationships Among Prices Under Alternative Formulations

Armore (5, pp. 52-55) was concerned with measuring price relationships among certain fats and oils or, more correctly, with demonstrating from market data that the price of cottonseed oil is more closely related to the price of certain fats and oils used in food products than to the price of lard, and that it bears little statistical relationship to the price of butter. As indicated in his bulletin, he had strong theoretical and empirical reasons for believing this to be true.

Results from a number of alternative analyses of these data are given in table 1, but only those based on first differences of undeflated data were included in the bulletin by Armore. In each case, the analysis is based on data for the years 1922-40. Correlations shown in the first three columns are those that are directly relevant to the price relationships among the several fats and oils. Within each row, the expected ranking is found, but the magnitude of the coefficients differs greatly from row to row. Correlations shown in the third row are those that would have been used had he relied on first differences to remove the common effect of the general price level on prices of cottonseed oil and the related item.

Lower correlations for butter are obtained when the effects of the general price level on each series are removed statistically by the use of partial relationships, regardless of whether first differences or actual data are used. (See second and fourth rows of table 1.) For lard and the miscellaneous group of fats and oils, however, the correlation is higher when actual data are used and the effects of the general price level are removed by the use of partial relationships than when first differences alone are used, and much higher than when the effects of the general price level are removed by the use of partial relationships in the analysis based on first differences. Simple correlations for the analyses based on deflated data (see last two rows) in each case are higher than the partial correlations for the corresponding analyses based on undeflated data, and in each case are higher when based on actual data than when based on first differences.

Statistical problems involved in attempting to decide whether to deflate by the general price level or to use the general price level as a separate variable differ for analyses of this type from those involved in reaching a similar decision when working with demand equations. Our economic theory with respect to demand equations applies only in terms of relative prices, whereas we may be interested in price relationships that prevail among the undeflated variables when considering relations among prices per se. In order to determine whether a real correlation exists among prices of the individual commodities, we must correct each in some manner for the effects of the general price

Table 1.--Cottonseed oil: Relation of wholesale price to that of specified other items, 1922-40 1/

Correlation between cottonseed oil and related item	Related item			
	Butter 2/	Lard 3/	Other fats and oils used in food 4/	All commodities 5/
Based on actual data:				
Simple	0.83	0.93	0.97	0.88
Partial 6/	7/.15	.75	.88	---
Based on first differences:				
Simple	7/.34	.66	.80	.46
Partial 6/	7/ 8/	7/.41	.66	---
Based on deflated data when using--				
Actual values	.52	.88	.94	---
First differences	.47	.83	.88	---

1/ When all prices are in cents per pound. Cottonseed oil is crude, tanks, southeastern mills. Data on which these analyses are based are given in Armore (5, p. 54).
2/ 92-score, creamery, New York.
3/ Prime steam, loose, Chicago.
4/ Average price of coconut, corn, oleo, palm, peanut and soybean oils, oleostearine, and edible tallow weighted by their average domestic disappearance in 1931-40.
5/ Bureau of Labor Statistics index, 1935-39 = 100.
6/ After allowing for the effect of prices of all commodities.
7/ Does not differ from zero by a statistically significant amount.
8/ Less than 0.005.

level, but we cannot say, without some thought, whether this should be done by deflation or by the inclusion of the general price level as a separate variable.

Kuh and Meyer (62) show that if each of the variables to be deflated is a linear homogeneous function of the deflating variable, then (1) the simple correlation between the deflated series equals the partial correlation between the undeflated series with the deflator held constant and (2) if we study relationships among prices of only two commodities, the simple correlation between the deflated variables equals the multiple correlation based on the deflator and the two undeflated prices.

When we consider linear relationships between two variables, a linear homogeneous function is one for which one variable is a constant proportion of the second variable except perhaps for a random error term, that is, the regression line passes through the zero intercept. In relation to the data shown in table 1, the findings of Kuh and Meyer suggest that if the price of cottonseed oil and the price of butter, for example, each are linear homogeneous functions of the general price level, then we would expect the correlation coefficient in the fifth row of the first column to equal that in the second row, and the correlation in the sixth row to equal that in the fourth row. These coefficients obviously are not equal, a fact which suggests that a linear homogeneous function does not prevail between each of the individual prices and the general price level. Furthermore, we have no reason to expect that this function is one of constant proportionality. In this case it appears likely that, if we are interested in relationships among the undeflated variables, we should obtain the partial correlation coefficients rather than using the computationally-simpler technique of getting the simple correlation between the deflated variables. This is what Armore did in his study. On the other hand, if we are working with a system of equations that involves deflated price variables in certain demand equations and wish to measure relations among these deflated variables, we would use the deflated variables directly in our analysis. Considerations involved in deciding whether to run the analysis in first differences or based on actual data are those discussed on pages 29-33.

Kuh and Meyer give another criterion that may be used in deciding whether to deflate or to include the deflator as a separate variable in the analysis in cases of this sort. This depends on whether the unexplained residuals are more nearly uniform over the range of the independent variable when working with deflated or undeflated data. All of the statistical methods considered in this handbook are designed to be used with data for which the variance of the unexplained residuals about the regression line is uniform over this range. When this variance is uniform, the residuals are said to be homoscedastic. Sometimes the degree of homoscedasticity in the residuals can be increased by deflation. If this appears to be true, and if each of the variables to be deflated appears to be a linear homogeneous function of the deflating variable, then we should deflate. If the variables that might be deflated are not linear homogeneous functions of the deflating variable, and the residuals are expected to be heteroscedastic unless we deflate, we might use a logarithmic or some other transformation to increase the degree of homoscedasticity. If the residuals are expected to be homoscedastic when the variables are used in their undeflated form, then the deflator should be included as a separate variable in the analysis. If interest is centered primarily on the regression rather than the correlation coefficients, Kuh and Meyer suggest that, even if the zero-intercept (or linear homogeneous function) concept is not met, deflation or some similar transformation may often be appropriate. They conclude (62, p. 413) as follows: "The discussion and the cited illustrations point up the fact that each application is likely to have its own special characteristics and that careful consideration must be given to a wide range of alternatives if appropriate estimates are to be obtained."

Deriving Estimates of Quantity or Price From Value Aggregates 26/

A somewhat different problem, though one often identified with that of deflation, concerns the isolation of changes in quantity from data representing value through division of the latter by an appropriate price series. 27/ Obviously, for any individual commodity, the division of value by price should exactly represent quantity. This generally is not true when we divide an aggregate value series by a price index with fixed weights, the usual procedure. But we may be able to regard changes in the derived series as reasonably approximating changes in quantities. For example, consumer expenditures for clothing in constant prices are often derived by dividing dollar expenditures by the Consumers' Price Index for Apparel Products (1947-49 = 100). The result presumably shows the expenditure for clothing in terms of average 1947-49 prices. With prices thus held constant, changes in the derived series are assumed to represent changes in quantities purchased. The same procedure frequently is used to estimate real or unit changes in income, retail sales, wholesale inventories, and so forth.

The reliability of the estimate of changes in quantities obtained by the foregoing procedure depends primarily on the composition and internal stability of the value series in relation to the particular price index employed. Changes in value may reflect shifts between the various products included in the value aggregate and not a change in total quantity or price. If such shifts are important, serious errors may arise from using a price index incorporating weights that remain fixed over relatively long periods of time. In addition, the weights used in the price index may not necessarily be related to the relative importance of the products in the value aggregate even in the base period. To the extent that they differ, the estimate of quantity differs from the actual quantity. However, if the relative importance of the products in the value aggregate remains relatively stable over time, changes in the derived quantities should approximate changes in actual quantities. A somewhat similar consideration applies if the prices in the two series differ.

In some instances, the value aggregate may be divided into several parts. Given appropriate price indexes, estimates of changes in quantity may be obtained for each component. If combination of the breakdowns from a quantitative standpoint is feasible, that is, if the commodities to be combined can be expressed in common quantitative terms, summing the estimates for the subgroups will probably give a better estimate of changes in total quantity than deriving the total estimate directly. In any event, examination of the results at less aggregative levels, where possible, should assist the analyst in evaluating the total estimate.

26/ This section was prepared by Martin S. Simon and Carroll Downey, Agricultural Economic Statisticians, Agricultural Marketing Service.

27/ Or, conversely, estimating changes in price from value data through division by a quantity series.

Use of Arithmetic or Logarithmic Variables and Other Transformations that Relate to Functional Form

Linear (arithmetic) and logarithmic equations are the principal functional forms used in economic analysis. Linear equations give results which, when translated into total value-supply curves, make more economic sense at the extremes than do the results obtained from logarithmic equations. This is true because the total value drops to zero when supply is zero and when supplies are very large. In most cases, however, no data are available for these extreme values; therefore we have little interest in them. In many analyses that have been run within the Agricultural Marketing Service based on undeflated data, logarithmic equations appear to fit the data better than do arithmetic equations. When projections are made from a pre-World War II period into a post-World War II period this is particularly evident. Whether this would hold if the analyses were based on deflated data is not known. Logarithmic equations have the mechanical advantage of yielding curves that have a constant elasticity, but this is not a valid criterion for deciding on their use.

From a statistical viewpoint, logarithmic equations should be used when (1) the relationships between the variables are believed to be multiplicative rather than additive, (2) the relations are believed to be more stable in percentage than in absolute terms, and (3) the unexplained residuals are believed to be more uniform over the range of the independent variables when expressed in percentage rather than absolute terms. To some extent, these items are different aspects of the same thing. The last two conditions are more likely to hold for analyses based on undeflated data than for those based on deflated data, although they might hold in either instance.

Ths use of logarithmic equations to handle relationships that are believed to be multiplicative can be illustrated by the following example. Assume that we use a logarithmic relationship of the following form:

$$\log X = \log a + b \log Y + c \log Z. \qquad (62)$$

When this is translated into natural numbers, the equation becomes:

$$X = aY^bZ^c. \qquad (62.1)$$

Thus, whenever we use an equation in which all of the variables are converted to logarithms, we implicitly assume that the relationship among the variables when expressed in natural numbers is of the form shown in equation (62.1). Frequently, in economic problems, a careful consideration of the variables involved in a particular equation will give strong reasons for expecting the relationships to be either additive or multiplicative. Particularly in those cases where the underlying relationships are believed to be additive, use of logarithmic curves to yield the mechanical convenience of constant elasticity should be avoided. 28/ Methods for computing the elasticity for each point on a curve are discussed in the section beginning on page 78.

In certain analyses, the effects of some variables on the dependent variable are additive, whereas the effects of other variables are multiplicative. A study of factors that affect prices of corn from June to September by Foote (26, pp. 30-41) is of this nature. Four components of supply were included in the initial formulation; these should be treated as additive variables, but the nature of the variables was such as to require that their respective weights be determined by regression analysis. However, the combined supply and the two demand shifters included in the analysis were believed to affect the dependent variable simultaneously, in a multiplicative way. A method by which the several regression coefficients can be estimated through an iterative approach in analyses of this sort is discussed in the section beginning on page 122. In other studies, the variables in each of several demand equations may be related to utilization in a multiplicative way but the several utilizations may be involved in an additive identity. Such is the case with the economic model for dairy products described on page 12. This situation can be handled by using a semi-logarithmic statistical equation; that is, by expressing utilization in actual numbers, and the other variables in logarithms. Relations that have multiplicative aspects can be obtained, of course, in ways other than by use of logarithms, as by the use of cross-product terms as additional variables. An equation of this form might be the following:

$$X = a_1 + b_{12}Y + b_{13}Z + b_{14}YZ. \qquad (63)$$

If Y and Z are each predetermined variables, the variable YZ is computed for each observation, designated as an additional variable, and the entire analysis is treated as a linear regression with three independent variables. In this equation, the effect of Y and Z as such each enters the equation in an additive way but their "joint" or multiplicative effect is brought in as an additional variable. If either Y or Z is endogenous within a system of equations, then the method described in the section beginning on page 71 for handling the multiplicative term is preferred.

28/ Wold and Jureen (106, pp. 258-259) point out, "the demand for food has an income elasticity that presents a clear tendency to decrease as income increases. Such tendencies must always be kept in mind when dealing with demand elasticities that are estimated on the hypothesis of constancy. ... Constant elasticities should in general be interpreted as average values, and in principle they will be valid only for the range of income covered by the data employed. Constant elasticities should, accordingly, not be given without indicating the range to which they refer, and it is necessary in any case to indicate the average income level in the sample." A similar situation probably exists, in some instances, when constant elasticity curves are used to measure the elasticity of demand with respect to price from time series data.

Sometimes economic relations are of such character as to require curves of a very different sort. In a study of smoking habits based on survey data by Sackrin (82, pp. 2-6), the quantity of cigarettes consumed was believed to be related to age and income, but the age relationship was of such a nature that a peak was reached at middle age. A satisfactory equation was obtained by using a relationship of the following form, where Q represents consumption, D represents income, A represents age, and $\bar{A}$ represents the age at which smoking is believed to reach a peak:

$$Q = a_1 + b_{12}D + b_{13}(A - \bar{A}) + b_{14}(A - \bar{A})^2. \quad (64)$$

When b_{13} is zero or positive and b_{14} is negative, the desired inverted parabola on age is obtained.

In general, the type of functional form to be used should be decided before the analysis is run. Except in a few special cases, usually involving thousands of observations, the form of the equation cannot be determined from the data. [See Foote (28, pp. 1-2).]

Problems Related to the Use of Time as a Variable

Time frequently is introduced as a variable into an analysis as a measure of sources of continuous systematic variation for which no data are available. If such sources are believed to be important before the analysis is run and if the time effect is believed to be linear, or curvilinear to only a moderate degree, then a time variable probably should be included in the initial analysis. If its partial regression coefficient fails to differ significantly from zero, it may be omitted. However, if the time effect is strongly curvilinear --for example, U-shaped--then a nonsignificant coefficient may be obtained which is, in a sense, misleading. Thus, if while formulating the model the research worker is uncertain as to the importance of unmeasured factors that change systematically over time, or of the nature of their effect, time probably should be omitted from the initial analysis, but the unexplained residuals should be plotted against time to determine whether they exhibit a nonrandom pattern. If they do, an attempt should be made to discover an economic cause for the pattern.

Sometimes an additional variable can be found that is related to the unexplained residuals. The closeness of the relation can be tested by plotting the residuals graphically against the new variable. Unless the new variable is correlated to a considerable degree with the independent variables included in the original analysis, the graphic chart will give a good indication of the partial correlation on the new variable. If the new variable appears to be a promising one, the analysis probably should be rerun with it included so as to obtain the true multiple and partial relationships. If no such variable can be found, but an explanation of the time trend in terms of technological or institutional developments is available, then a time variable should be introduced into the analysis through the use of an appropriate graphic or mathematical relationship.

Sometimes the unexplained residuals deviate from a random pattern for only part of the period. If a satisfactory economic reason can be found for the deviation, the analysis can be adjusted by use of charts that indicate the partial relationships between the dependent variable and each of the independent variables. (Methods for constructing such charts are given in the section beginning on page 205 and an example of their use is shown in the section beginning on page 174.) The following steps are used: (1) The unexplained residuals for the years involved are plotted against time, and a free-hand trend is drawn through them that coincides with the assumed effect of the explanatory cause. (2) Deviations from this trend are substituted for the original deviations on each of the partial charts. (3) If a change in the slope or general shape of the relation is suggested, and this appears to be in line with expectations based on economic theory and knowledge about the item, a graphic adjustment in the relation is made. (4) An appropriate adjustment is made in the various mathematical coefficients obtained from the original analysis.

Handling of time in logarithmic equations.--In analyses for which the other variables are converted to logarithms, time (t) frequently is used either in a converted or an unconverted form. When it is converted, the first year should be assigned the number 1, as the logarithm of zero is undefined. The following equations illustrate the effect of these alternative treatments:

(1) Use of converted data:

$$\log Y = \log a + b \log X + c \log t \tag{65}$$

$$Y = aX^b t^c. \tag{65.1}$$

Here c normally is close to zero and may be positive or negative. As t increases, we raise a progressively larger number to a certain power.

(2) Use of unconverted data:

$$\log Y = \log a + b \log X + (\log c)\, t \tag{66}$$

$$Y = aX^b c^t. \tag{66.1}$$

Here c (the antilog of log c) normally is close to 1 and always is positive, but log c may be positive or negative and normally is close to zero. As t increases, we raise c (a constant) to a progressively larger power. Results of the two approaches are shown in table 2 and in figure 4.

In effect, these alternative approaches give (1) a power function in t and (2) an exponential function in t. The power function increases or decreases more rapidly for small values of t than does the exponential function, but it soon becomes nearly flat if c is less than unity. The exponential function never declines below zero, and hence for negative time trends becomes flatter as t becomes very large; but when it is used for an increasing time trend, it increases more rapidly as t increases. The typical time trend is

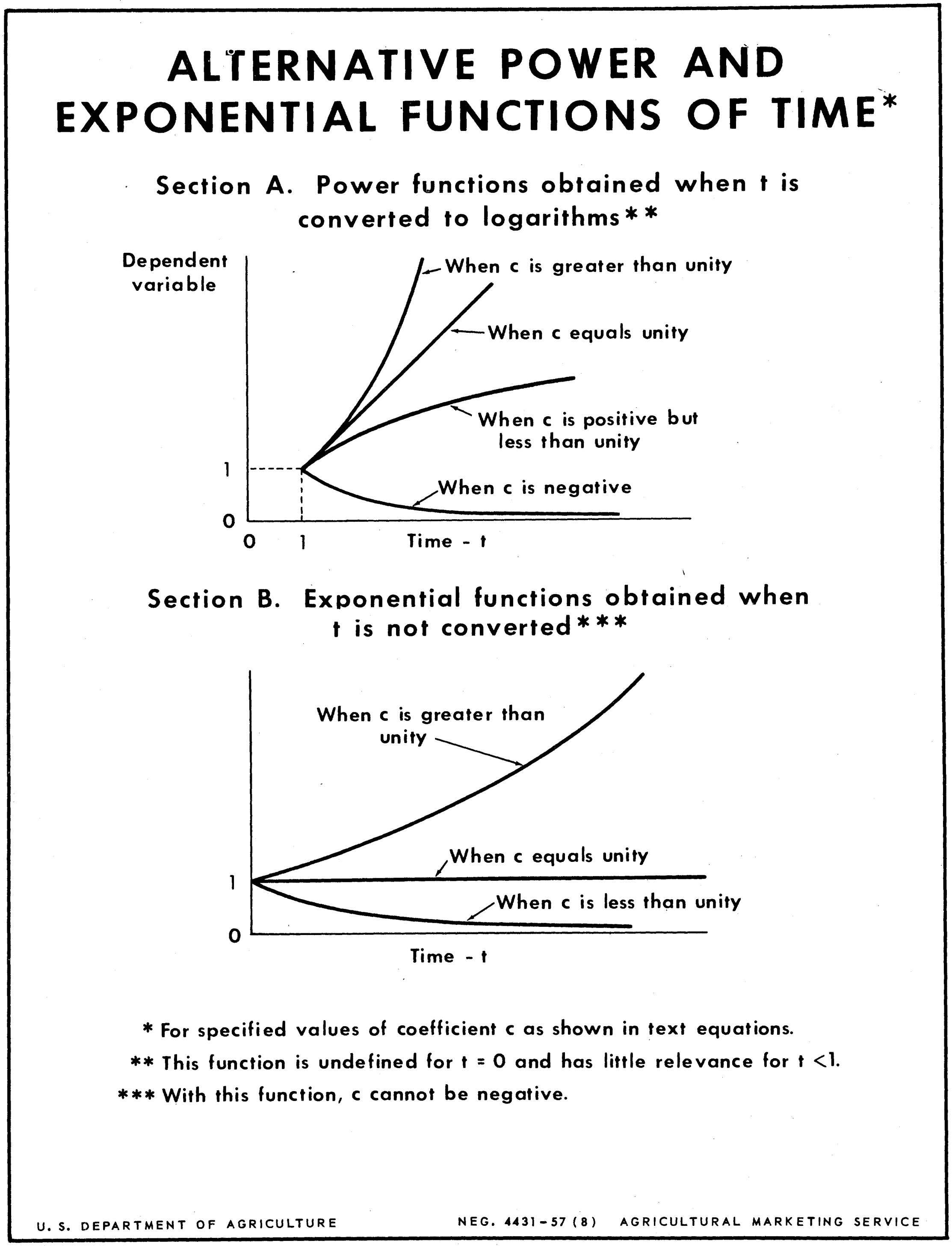


Figure 4.--When analyses are run in terms of logarithms and time is included as a variable, time may be used in a logarithmic or non-logarithmic form. Conversion of time to logarithms gives a greater variety of possibilities, as shown in section A, and is recommended in most cases.

one that increases rapidly during a period of growth and then tends to flatten out. Thus the power function, which requires that t be converted to logarithms, appears more logical for most problems. The wide variety of alternatives permitted by its use is a further advantage.

Data in table 2 illustrate an important aspect of the use of time as a variable. The coefficients have been chosen so that the two curves are similar for t in the neighborhood of 1 to 15 at the lower and higher values of c respectively. But for extrapolations beyond that, the two curves differ markedly. Extrapolation always is dangerous--with time as a variable it is particularly so. And it is most dangerous when a polynomial involving t is used, or when an exponential curve is involved for which the c coefficient is greater than 1.

Table 2.--Results of using time converted to logarithms or not converted in analyses for which the other variables are converted to logarithms

t	Contribution of the time trend to the estimated value of the dependent variable when t is--			
	Converted to logarithms and c equals--		Not converted to logarithms and c equals--	
	-0.1	0.1	0.98	1.02
0	---	---	1.00	1.00
1	1.00	1.00	.98	1.02
2	.93	1.07	.96	1.04
5	.85	1.17	.90	1.10
10	.79	1.26	.82	1.22
15	.76	1.31	.74	1.35
20	.74	1.35	.67	1.49
30	.71	1.41	.55	1.81
40	.69	1.45	.45	2.21
50	.68	1.48	.36	2.69
75	.65	1.54	.22	4.42
100	.63	1.59	.13	7.25

When t is not converted, we can get the value of c by obtaining the antilogarithm of the coefficient attached to t in equation (66). 29/ Each additional year has the effect of raising this coefficient to one higher power.

29/ Mechanically, this can be done most easily by getting 100c by taking the antilogarithm of the coefficient plus 2.

Thus, in the cases used as examples in table 2, we can say that time in effect either raises or lowers the value of Y by roughly 2 percent each year. When the logarithm of t is used, however, no such statement can be made because the percentage effect of a 1-unit increase in t becomes less as t becomes larger. This fact has led some analysts to use t in an unconverted form, but they may have failed to consider the effect of this on the general shape of the curve.

Time trends in relation to first difference analyses.--If first differences are used, the constant value in the equation represents the linear effect of time as, if different from zero, it implies that some change in the dependent variable would occur from the preceding year even if there were no changes in the independent variables. In non-first difference analyses, the constant term has little economic meaning and hence a standard error for it is seldom obtained. When working with first differences, however, the analyst in general will wish to determine whether this coefficient differs significantly from zero. A formula for the standard error of a function for single-equation analyses based on any number of variables is given in Friedman and Foote (40, pp. 17-19), and in certain statistical textbooks. The standard error of the constant term can be obtained from this formula by carrying out the computations when all the independent variables are set equal to zero. This test may be biased if the residuals from the analysis are serially correlated. However, as discussed on page 30, first differences in general are used only when it is believed that the residuals will not be serially correlated.

When all variables in a first difference analysis are converted to logarithms and the constant term differs significantly from zero, we can find the percentage effect of time in each year most easily by taking the antilogarithm of the constant term plus 2. For example, if this term in logarithms was 0.015, the antilogarithm of 2.015 is 103.5, so we could say that the time trend alone tended to increase the dependent variable by 3.5 percent per year. If this term in logarithms was -0.015, the antilogarithm of 1.985 is 96.6, so we could say that the time trend alone tended to decrease the dependent variable by 3.4 percent per year. In such analyses, time has an effect similar to that obtained from the exponential analyses for studies based on actual data, discussed on page 40.

Other effects of time.--Time also may affect an analysis in several other ways. Ways of measuring possible changes in basic structural relationships over time are discussed in the section entitled "Years to be Included" (see page 21). If the coefficients as such have changed in a systematic way over time, this can be measured by the method discussed in the section beginning on page 110. In some studies, certain variables may enter into the equation with a time lag. Such variables can be incorporated easily into the equation if the nature of the lag can be assumed in advance. Use of the theory of distributed lags at times is helpful, particularly if we wish to measure differences between long-run and short-run elasticities of demand or supply. Methods of analysis that appear useful in this area are discussed in the section beginning on page 111. Other ways of incorporating lagged variables into the equations are discussed in the section beginning on page 116.

Deciding Whether Certain Variables are Endogenous or Predetermined

As discussed in a later section, if we wish to estimate certain structural coefficients, such as elasticities, the recommended method of fitting individual equations depends on the number of endogenous variables contained in them and the degree of identification (see page 61). Hence we must decide whether each variable is to be classified as endogenous or predetermined. If we are concerned only with equations designed for forecasting a single variable, the method of least squares may be preferred and this distinction is less important than when structural coefficients are to be estimated. Here we use as the dependent variable that item for which a forecast is desired and as independent variables all relevant items that are expected to be known at the time of forecast (see page 128). A distinction between endogenous and predetermined variables must be made in formulating a system of equations designed for forecasting or analytical purposes, since we must have one equation for each endogenous variable in the system.

An endogenous variable is technically defined as one that is correlated with the unexplained residuals in the structural equation in which it appears. A predetermined variable is independent of the unexplained residuals in the structural equation in which it appears. Predetermined variables are generally defined to include exogenous variables, or those determined outside the particular economic sector under consideration, and lagged values of endogenous variables.

At times, the nature of the economic model itself determines whether a particular variable is assumed to be endogenous or predetermined, as when we decide to take the price of feed as given, in a study relating to the supply-demand structure for eggs, but to let it be endogenous in an analysis of the feed-livestock economy. In making this choice the research analyst is guided by two conflicting aims: (1) To allow for as many interdependent relationships as possible and (2) to keep his models simple enough for them to be fitted statistically without undue effort. Common sense, supplemented by the tests and considerations outlined here, is perhaps the best means for bringing about a reconciliation. Three types of variables for which the classification frequently is not clear are those relating to (1) production, (2) consumption, and (3) the general economy, such as disposable income or prices of all commodities.

Production.--For many industrial products, production is determined to a large degree by current price in relation to certain cost factors. Unexplained residuals that affect the demand equation may well also affect the supply equation, so that such variables probably should be treated as endogenous in a study of the supply or demand structure. In such cases, one or more supply equations are included in the complete economic model. For some livestock products produced on farms this is true, but to a lesser extent than for industrial products. Insofar as production is affected by changes in feeding rates in response to changes in current commodity-feed price ratios or by culling or within-year replacement of producing animals, it may tend to be endogenous, and supply equations may be needed in the economic model. To the extent that production is affected by numbers of animals on hand at the start of the year or that were produced as a result of decisions made before the start of the year, it is independent of the unexplained residuals in the demand equation and hence should be classified as predetermined. For many crops, production is determined chiefly by weather conditions during the growing season and by economic factors before the start of the marketing (and frequently the planting) season. Here production is almost completely a predetermined variable although, as noted on page 11, if only part of the crop is harvested for economic reasons and we refer to that part as production, then to that extent production may tend to be partially endogenous.

A detailed analysis for hogs given by Fox (33, pp. 28-31) is useful in illustrating the sort of considerations that enter into a decision as to whether production should be classified as a predetermined variable or may be at least partially endogenous. We reproduce his discussion in full in the paragraphs that follow.

> "Production of pork obviously is a direct function of the number of hogs slaughtered, their average weight, and the percentage yield of pork per hog. Year-to-year variations in production of pork result mainly from changes in the number of hogs slaughtered.
>
> "The number slaughtered in any given year is determined mainly by the number of sows bred in the preceding year. For example, hogs marketed from September to March were born 6 to 9 months previously, from sows bred 10 to 13 months previously. About June 22, when the size of the spring pig crop is known, a forecast can be made of the number of hogs that will be slaughtered from September to March. Similarly, about December 22, when the size of the fall pig crop is known, the approximate number of hogs that will be slaughtered from April to August of the following year can be forecast.
>
> "The average age at which hogs are marketed can be varied by a few weeks according to how much they are forced during raising and feeding, and by perhaps a week or two according to the exact time chosen for marketing. For example, economic influences current toward the end of a marketing season may determine whether more spring pigs than usual will be carried over into the period for marketing fall pigs, or whether more fall pigs will be marketed early, along with spring pigs. Variations in average marketing dates are directly related to the average weight per hog slaughtered, as late marketings mean heavy weights, early marketings light weights. Variations in the number of gilts saved for breeding mean opposite variations in the number slaughtered currently. These factors influence production of pork relatively little in most years.

"The nature of available official data means that calendar-year estimates of pork consumption must be used in deriving a consumer demand equation. This unit splits the marketing season for spring pigs. However, the logical basis for considering that calendar-year production of pork is predetermined, or nearly so, rests on the 10 to 13 months required for gestation and feeding to market weight, plus a decision-making interval before actual breeding.

"The relevant statistical question in this connection is, What proportion of the variation in calendar-year production of pork is associated with factors known or determined before January 1 and with noneconomic variables operating during the current year? For this purpose, production of pork can be considered to be determined by some or all of the following variables:

Spring pig crop, previous year;
Fall pig crop, previous year;
Breeding intentions for current spring pig crop--that is, number of sows to farrow (reported in previous December)--multiplied by actual number of pigs saved per litter, which depends mainly on natural conditions, including weather, at farrowing;
Supply of corn, previous year;
Hog-corn price ratio, preceding September-December;
Production of corn, current year; and
Short-term expectations regarding price trends which could affect age and weights at marketing.

"The first three variables accounted for more than 93 percent of the variation in production of pork during 1924-41. The report of breeding intentions reflects the influence of other variables such as supplies of feed grain and relative prices of hogs and corn, current and anticipated . Supplies of feed on January 1 relative to numbers of livestock also affect production of pork because of their influence on average slaughter weight and yield of pork per hog. The current year's production of feed grains, which depends primarily on weather, could be introduced as an additional factor which may influence the weights at which hogs are marketed during the latter part of the calendar year and the number of gilts saved for breeding purposes after January 1.

"From this analysis it appears that in 1924-41 variations in calendar-year production of pork were about 95 percent predetermined. The explanation of production is not significantly increased by including the current price of hogs or pork."

In a similar analysis for beef, Fox concludes that about 85 percent of the variation in production can be explained by variables measured or existing at the beginning of the calendar year or before, supplemented by noneconomic factors that operate during the calendar year. Under such circumstances, it can be presumed that production is correlated only slightly with the unexplained residuals in the demand equation.

Rather than base conclusions on a test of this sort, the analyst may prefer to follow the path chosen by Gerra (see page 14) and design a structural analysis to test the extent to which production is in part endogenous. The amount of work required by the Gerra approach, which necessitates the simultaneous fitting of a number of equations, versus the Fox approach, which permits the use of a single least squares analysis to estimate the coefficients in the demand equation, provided production (and consumption) is found to be nearly predetermined, is 10 to 20 times as great in a model comparable in size to that used by Gerra. Therefore, in circumstances of this sort, a system of equations should be used only when it is important to determine the extent to which production is endogenous. It is conceivable that a study like the one Fox made could show that production is only about 10 percent endogenous and yet coefficients from a structural analysis that differ from zero by a statistically significant amount could be obtained on the variables that relate to endogenous supply.

<u>Consumption</u>.--If production is considered to be endogenous, then consumption also normally is considered endogenous, but if production is found to be nearly enough predetermined to so treat it in the analysis, then a separate decision must be made with respect to consumption. Here Fox (<u>33</u>, pp. 12-13) again outlines some relevant considerations. His comments in full follow:

> "Suppose the supply of a given commodity entering the marketing system is not affected by the current market price. Suppose further that the marketing system passes on this supply in a routine way, so that, except for normal wastes and losses in the marketing process, the supply that reaches consumers is exactly equal to that marketed by farmers. In this case, consumption is not determined by prices during the marketing period; it can be used as a predetermined variable.
>
> "If consumption is not exactly equal to farm supply, because of change in stocks or because of existence of foreign trade, cases that are more complicated will arise. Theoretically, the existence of imports suggests one or more supply curves for producers or dealers in the countries from which imports are obtained; the existence of exports denotes the presence in a complete model of the demand curves of one or more foreign countries. If domestic stocks change significantly, in addition to the demand curve for final consumption, a demand curve for storage holdings exists. Thus, in a strict sense, it is clear that if foreign trade or changes in stocks are important, a multiple-equation model is required.

"However, suppose the changes in stocks and in net trade are both small relative to observed changes in consumption and that domestic consumption, accumulation of stocks, and net exports move in the same direction in response to changes in supply. In such cases, changes in domestic consumption can be estimated with considerable accuracy on the basis of changes in supply. If supply is predetermined, consumption also can probably be treated as predetermined under such conditions."

Fox (33, p. 31) then applies these considerations to pork. He says,

"Exports and changes in stocks of pork and pork products (excluding lard) normally are small. In terms of year-to-year changes, 93 percent of the variation in consumption of pork during the calendar-years 1922-41 was associated with variation in the quantity of pork produced. When both variables were expressed in millions of pounds dressed weight, the regression equation was as follows:

$$Q = 50.2 + \underset{(0.05)}{0.75}\,S \qquad (67)$$

where Q is consumption, S is production, and the number in parentheses is the standard error of the regression coefficient. Thus, a 1-million-pound change in production normally was associated with an 0.75-million-pound change in consumption.

"As 95 percent of the variation in production of pork during the interwar period was apparently predetermined, it appears that at least 88 percent (0.95 times 0.93 times 100) of the variation in consumption of pork was predetermined. Alternatively, consumption of pork could have been expressed directly as a function of the variables used to explain production of pork. When this was done, 90 percent of the variation in consumption of pork was associated with the known predetermined factors affecting production. In this instance, the bias that may result from treating consumption as a predetermined variable and using the single-equation approach is probably small."

To increase our knowledge of the effects of changes in production on the several alternative outlets, a set of first difference analyses can be run, with production as the independent variable and with each of the components of the corresponding disposition as dependent variables in a series of simple regression analyses. Year-to-year changes in production are exactly equal to the sum of year-to-year changes in (1) domestic consumption, (2) net foreign trade, and (3) the net change in inventories. These utilization groups could be further subdivided if the data permitted and the problem required it. The technique consists of calculating the simple least squares regression of each distribution category separately upon the production variable, using first differences of actual data in each case. When this is done, the sum of the slopes ("b" values) of these regression equations equals 1.0 and the sum of the constant terms ("a" values) equals 0. 30/ Such a set of analyses indicates the change in domestic consumption, net exports, and stocks, respectively, that would be associated on the average with a given change in production.

If more than one important domestic outlet is available for the commodity, then separate equations within an equation system are needed to measure the structural coefficients that relate to demand in each outlet, and the consumption of each should be considered as endogenous. For some crops, however, varieties grown for processing and those grown for fresh use differ. In such cases, each variety can be considered as a separate commodity, and consumption may be considered as predetermined, provided the other considerations that have been noted permit such an assumption.

As noted previously, if two or more competing commodities are involved, then a system of equations is required to obtain their respective direct- and cross-elasticities of demand even though the consumption of each is thought to be essentially a predetermined variable. However, the nature of the system differs considerably, depending on whether all of the variables relating to consumption can be considered as predetermined or whether some of them are considered as endogenous. The formulation and statistical fitting of such systems is discussed in detail beginning on page 87.

Variables that relate to the general economy.--In arriving at a decision on whether to classify variables that relate to the general economy as endogenous or predetermined, the following kinds of considerations are involved. If we deal with a perishable item, then factors that affect its consumption but are not included in our equation may be correlated only slightly, if at all, with the variables that relate to the entire economy; in such cases, we probably can assume that variables that relate to the general economy are basically predetermined. On the other hand, if we deal with a commodity that can be stored easily by consumers and that frequently is stored by them, then factors that affect the level of current demand (for storage and consumption) but are not included in our equation, are likely to affect the general economy to some extent. In such case a correlation may prevail between the variables that relate to the general economy and the unexplained disturbances in the demand equation.

Consider, for example, the effect of a disturbance like the Korean conflict on purchases by consumers of sugar. Purchases might well increase by more than normally would be expected due to the associated increase in consumer incomes. Thus the unexplained residual for a period immediately following the disturbance probably would be positive and hence partially correlated

---

30/ This approach was developed by Fox and first published in Foote and Fox (29, p. 8). A proof that the sum of the slopes equals 1 and the sum of the constant terms equals 0 is given in Foote and Fox, pp. 79-81.

with the level of consumer income, which also would tend to increase as a result of the Korean conflict. In this instance, we might merely omit one year of data from the analysis, but if disturbances of this sort occurred frequently, and a variable explaining them could not be introduced into the analysis, then income should be treated as endogenous.

If data were available to correct for changes in stocks held by consumers, so that the demand equation reflected only the demand for immediate consumption, then the situation would be similar to that for a perishable commodity. In some cases, use of the year-to-year change in income in addition to the level of income may eliminate the correlation between the unexplained residuals and the general variables. Use of distributed lags, as described in the section beginning on page 111, also would tend to eliminate this correlation. In any of these cases, the general variables can be taken as predetermined.

A similar situation prevails if we consider the supply-demand structure for livestock or livestock products. Farmers essentially sell milk or eggs as fast as they are produced; here general economic variables probably can be considered as predetermined. In relation to livestock that can be held in inventory, a change in general business conditions is likely to affect the nature of the relationship. If this effect is not allowed for specifically in the model, it shows up in the unexplained residuals and a correlation is generated between these residuals and the change in business conditions. Here factors that measure these conditions should be considered as endogenous.

Harberger 31/ clarifies and extends these views. He says, "any price set by the competition of many types of buyers and sellers may frequently be viewed as exogenous when the purpose is to analyze the behavior of only one or a few classes of buyers and sellers, and (consumer) income may often be treated as exogenous when we wish to study the demand for products we believe to be insensitive to the general state of expectations in the economy." As he later points out, not only are durable and semidurable goods likely to be sensitive to the "general state of expectations," but so also are commodities that are considered by the consumer as true luxuries, since the user would tend to "hold off purchases of these if he looked forward to a period of financial stringency and insecurity."

These comments suggest that the treatment of disposable income and certain variables relating to wage rates or other costs as predetermined variables in the models that are discussed in the section beginning on page 10 probably is satisfactory.

Some authors have argued that the relevant consideration is the extent to which consumer income, say, is affected by changes in price or consumption of the given commodity. They have shown that for even the most important

---

31/ Harberger, Arnold C. On the Estimation of Economic Parameters. Cowles Commission Discussion Paper: Economics No. 2088, 1953, p. 31. [Processed.]

individual farm products, such as beef, pork, or fluid milk, the retail value is equivalent only to 2 or 3 percent of disposable income. Hence these authors have assumed that disposable income can be treated as a predetermined variable in statistical analyses of demand for farm products, either singly or in moderately large groups. As each of the items cited is a commodity which, at least until recently, has not been stored in quantity by consumers, these authors come to the same conclusion as that given in the previous paragraph, but for a different reason. Their reasoning suggests that disposable income also could be considered as a predetermined variable when dealing with durable or semidurable products, so long as in value these products represent only a small part of total income. But we conclude that special characteristics of the commodity itself may tend to require that disposable income, and similar variables, should be considered as endogenous, unless the equations are modified in such a way as to eliminate the correlation between these variables and the unexplained residuals from the analysis.

Choice of the Dependent Variable in Least Squares Analyses Designed to Estimate Structural Coefficients

When equations are fitted statistically by a simultaneous equations approach, we in effect relate a linear combination of the endogenous variables to the predetermined variables in the equation; hence no choice is required as to which variable to consider as dependent. If only one endogenous variable is involved in the equation, estimates of the structural coefficients that are statistically consistent can be obtained when the equation is fitted by the method of least squares, provided we take as dependent the single endogenous variable (see page 57). Fox (33) has emphasized, as previously noted, that for many agricultural commodities all of the variables in a demand equation except price can be considered as predetermined. Hence, he argued, the correct way to estimate the elasticity of demand in such cases is to use price as the dependent variable and then algebraically transpose the equation to show consumption as a function of price in deriving the elasticity of demand. This is a sound procedure when the basic assumptions hold, that is, when consumption, consumer income, and other variables in the demand equation can be taken as predetermined and there are no close substitutes or complements. 32/

For reasons discussed in the section beginning on page 67, however, we may at times wish to use the method of least squares to estimate structural coefficients when more than one endogenous variable is included in the equation. Here we must make a choice as to which variable to treat as dependent. Hildreth and Jarrett (49, p. 71) say with respect to this situation, "Intuitive considerations ... suggest that least-squares bias might be minimized by treating as independent those current endogenous variables that are most strongly influenced by predetermined variables not appearing in the equation

---

32/ Fox apparently was unaware of the latter requirement, as was the author of this handbook at the time when Foote and Fox (29) published. Its importance is emphasized by Meinken, Rojko, and King (72).

being estimated. ..." This is directly analogous to the arguments given by Fox when he reasons that production (or consumption) can be treated as a predetermined variable if it were chiefly determined by variables which themselves were predetermined and did not enter into the demand equation.

Brandow (10, p. 17) notes that application of this criterion suggests that price should be treated as dependent in demand equations but that utilization should be treated as dependent in supply equations for particular outlets. He then says, "The grounds for making a choice are so uncertain, however, that solutions were obtained with both quantity and price dependent. ... In two cases (the equations for supply of fresh apples and demand for canning apples) coefficients obtained by least squares with quantity dependent were much closer to" those obtained by a simultaneous equations approach "than were coefficients inferred from solutions with price dependent. In the other two cases, differences were small." He concludes that if the coefficients determined by a simultaneous equations approach "are accepted as criteria, least squares results obtained with quantity dependent were generally better than those obtained with price dependent."

Further empirical research appears to be needed adequately to test the approach suggested on intuitive grounds by Hildreth and Jarrett if further comparisons are to be made of results obtained when a least squares versus a simultaneous equations approach is used to estimate structural coefficients for equations that contain more than one endogenous variable. Such empirical research might well include the measurement of least squares relations when alternative variables are treated as dependent.

Particularly when working with price level relations, we may find that the unexplained residuals are extremely small, so that the choice of the dependent variable in a least squares fit is essentially arbitrary. If the correlation coefficient is nearly equal to 1, it makes little difference which variable is chosen as dependent. When the relationships are no closer than for most of the items shown in table 1, however, a simultaneous equations fit may be preferred if we have an interest in the structural coefficients. Such equations always contain two or more endogenous variables, and the Hildreth-Jarrett criterion as applied to least squares analyses is of no value in choosing the one to treat as dependent.

As shown by the experimental results summarized in the section beginning on page 128, if we are interested only in equations designed for forecasting a single variable, then a least squares analysis under certain circumstances may result in smaller errors of estimate than are obtained from equations fitted by methods that allow for the simultaneity implied by the model. If this is the case, we use as dependent that variable for which a forecast is desired.

CONSIDERATIONS IN CHOOSING A STATISTICAL PROCEDURE FOR FITTING THE EQUATIONS WHEN WE WISH TO ESTIMATE THE STRUCTURAL COEFFICIENTS 33/

Nearly all equations relating to demand before 1950 were fitted by the method of least squares. In 1943, Haavelmo (45) emphasized that entire systems of equations had to be considered as a unit in order to understand and measure quantitatively many economic relationships; staff members of the Cowles Commission for Research in Economics spent a considerable part of the next 10 years in developing statistical methods to provide such measurements. As a result of their work, some analysts concluded that the method of least squares was completely outmoded. Others, however, feel that methods that apply to simultaneous equations are so complex and computationally expensive that they should be avoided whenever possible.

In the paragraphs that follow we start with some elementary economic concepts and then proceed step by step in such a way as to show precisely when and why simultaneous equations techniques are needed. In later sections, computational aspects are discussed.

Some Economic Considerations

In 1927, Working (107) discussed what now is called the identification problem in his classic paper, "What Do Statistical 'Demand Curves' Show?" 34/ In this paper, Working pointed out that when a research worker begins a demand study, he is confronted with a set of dots like that shown in section A of figure 5. He knows that each can be thought of as the intersection of a demand and a supply curve, as in section B, but, without further information, neither curve can be determined from the data. Working then noted that if the demand curve has shifted over time but the supply curve has remained relatively stable, as in section C, the dots trace out a supply curve; conversely, if the supply curve has shifted but the demand curve has remained stable, as in section D, the dots trace out a demand curve. If shifts for each curve have taken place, as in section E, the dots trace out what may look like a structural demand or supply curve, but the slope will be too flat or too steep.

In many analyses of the demand for agricultural products, 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 section F. As discussed on pages 44-49, in some analyses we can assume that the quantity supplied is essentially unaffected by current price. When price is plotted on the vertical scale, the supply curve in such a case is a vertical line, and year-to-year shifts in the supply curve trace out a demand

---

33/ Part of the material in the first three subsections of this section is adapted from that in the article by Foote (27).
34/ A similar line of reasoning is followed by Koopmans (61).

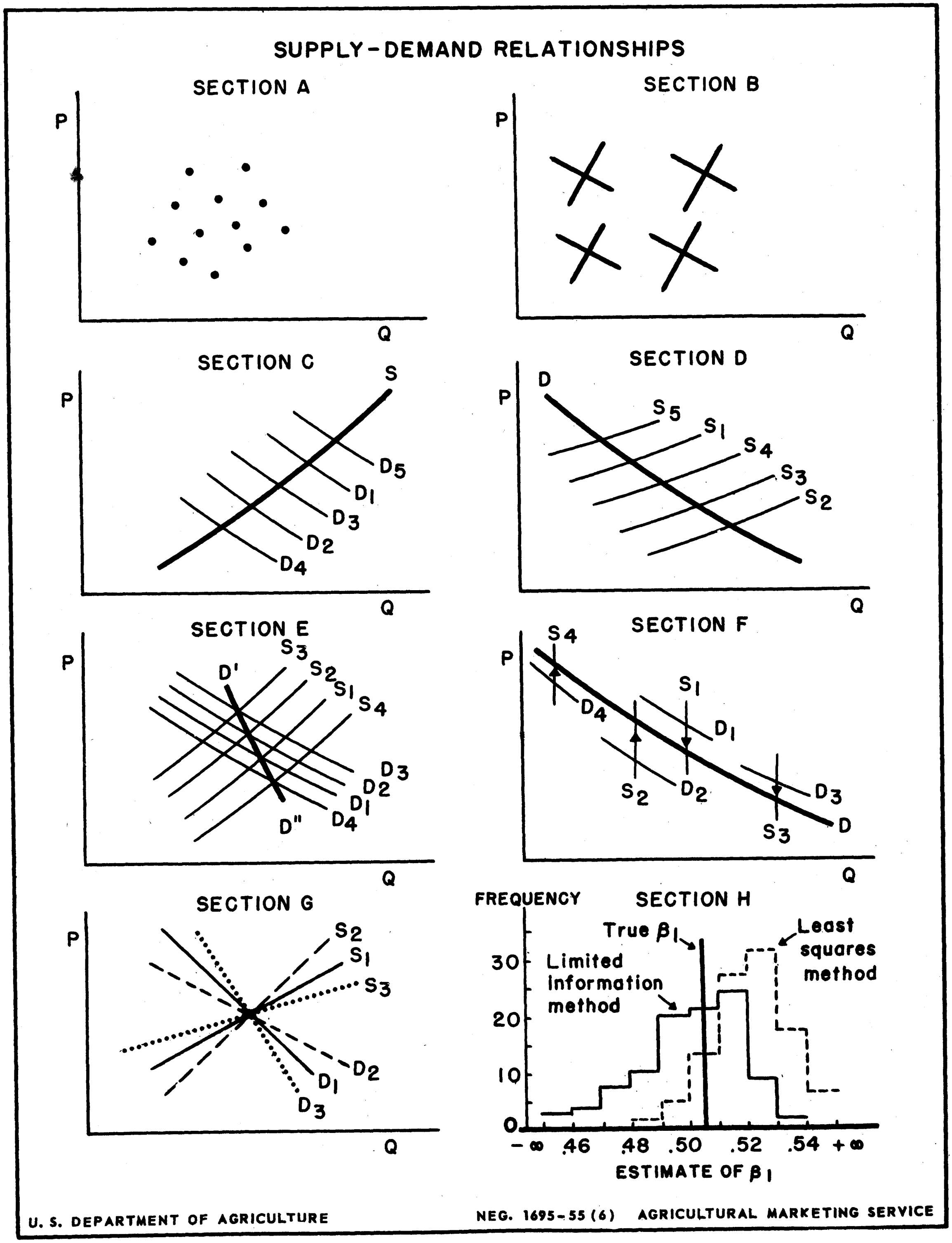


Figure 5.--The relationships shown here bear upon the conditions under which elasticities of demand can be measured by using a single-equation analysis. They are discussed in detail in the text.

curve, just as they did in section D. 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. This point was noted by Working (107, p. 223), emphasized by Ezekiel in a paper published in 1928 (19), and reconsidered in 1953 by Fox (33) in the light of modern simultaneous equations theory. For a number of 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--to use the least squares technique, the supply curve for a given marketing year must be a vertical line. (2) An algebraic transformation must be made after the equation has been fitted to derive the appropriate coefficient of elasticity; the definition as normally used in economic literature is in terms of the percentage change in quantity associated with a given percentage change in price. Other circumstances under which least squares equations can be used to derive coefficients of elasticity are discussed in the section beginning on page 67.

What happens if we have a supply curve that is not a vertical line? If we consider any single point, as in 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 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 involved in the simultaneous equations approach.

Suppose, however, that the analyst has no interest in the true demand and supply curves but only wants a method that will assist him in studying probable future trends in prices. Working had some suggestions on this point, too. He said, "It does not follow from the foregoing analysis that, when conditions are such that shifts of the supply and demand curves are correlated, an attempt to construct a demand curve will give a result that will be useless. Even though shifts of the supply and demand curves are correlated, a curve which is fitted to the points of intersection will be useful for purposes of price forecasting, provided no new factors are introduced which did not affect the price during the period of study. Thus, so long as the shifts of the supply and demand curves remain correlated in the same way, and so long as they shift through approximately the same range, the curve of regression of price upon quantity can be used as a means of estimating price from quantity" (107, p. 227). The experiment described in the section beginning on page 128 tends to verify Working's conclusions, although this analysis deals with a different aspect of simultaneity. However, this experiment suggests that more accurate forecasts can be obtained by including as independent variables in the forecasting equation all of the variables in the system that are expected to be known at the time of forecast. Other aspects of this experiment that relate to forecasting are summarized in the section beginning on page 141.

Our experimental model suggests that, even when no change in structure takes place, least squares equations under certain circumstances give poor estimates of the structural coefficients. Under other circumstances, least squares estimates of the structural coefficients are nearly identical with those obtained by a method that allows for the simultaneity implied by the model, and are close to the true coefficients used to generate the data. The relevant circumstances are discussed in the section beginning on page 67. Effects of certain changes in structure on the forecasting merits of the two sets of equations are discussed in the section beginning on page 139.

Marschak (66) gives an interesting example of the importance of changes in structure on the need for using a complete system of equations. He considers the old problem of taxation of a monopoly. He points out, "Knowledge is useful if it helps to make the best decisions." He considers, among other things, the kinds of knowledge that are useful to guide the firm in its choice of the most profitable output level. If the tax rate has not changed in the past and is not expected to change, the firm can fit an empirical curve to observed data on output and profits and immediately derive the point of maximum revenue. If the tax rate has not changed in the past but is expected to change in the future, the firm could, if it so desired, vary its output and profits under the new tax structure and derive a new empirical relation. But this takes time, and substantial losses might occur during the experimental period. If the firm had taken the trouble to derive the structural demand and net revenue curves, it could immediately determine its most profitable output under the new tax structure. If the tax rate had varied during the initial period, an empirical regression of net revenue on output and the tax rate could have been fitted and used to find the most profitable output under the new tax structure. In the example cited, however, this relation is a quadratic one and, particularly if based on a small sample, the analyst might not be aware of this, whereas if the structural relations had been obtained, this fact would be revealed by algebraic manipulation.

In many real-life situations, changes in structure are frequent. Hence Marschak concludes: "... a theory may appear unnecessary for policy decisions until a certain structural change is expected or intended. It becomes necessary then. Since it is difficult to specify in advance what structural changes may be visualized later, it is almost certain that a broad analysis of economic structure, later to be filled out in detail according to needs, is not a wasted effort" (66, p. 26).

This argument in no way invalidates the use of a single equation to estimate elasticities of demand in those cases for which the equation contains only a single endogenous variable and no substitute or complementary products are involved. In such cases, we may obtain estimates of the structural parameters that can be used in the same way as any other statistically-valid estimates. Instead, Marschak is arguing that only rarely should the economic analyst be satisfied with a purely empirical fit if he can obtain structural relationships with some additional work.

Some Statistical Considerations

We now turn to some statistical considerations that have a bearing on the extent to which we can use least squares to estimate the coefficients in a set of structural equations. In the relationships illustrated in figure 5, we assumed, more or less implicitly, that the points lie exactly on the demand or supply curve. In actual statistical analyses this is never true, since some variables that cause the curves to shift always are omitted and the precise shape of the curves to be fitted are not known. Thus we normally deal with stochastic rather than functional relations. A stochastic relation basically is one that includes a set of unexplained residuals or error terms whose direction and magnitude are usually not known exactly for any particular set of calculations, but whose behavior on the average over repeated samples can be described or assumed.

To take a concrete example, let us consider a single equation from a structural set:

$$Y = a + b_1Z_1 + b_2Z_2 + u. \tag{68}$$

Here Y is a variable for which an estimate is desired, the Z's are two variables which are known to affect Y, and u is an error term. We also have an interest in estimating a and the b's in a way that will meet certain desirable statistical criteria. We assume that for a number of periods we know the value of Y and the Z's. We do not know the value of u but can estimate it in a rough way for any given period as the difference between the value of Y computed from the equation and its actual value.

We know that estimates of the regression coefficients will differ for different sets of observations. We would like to estimate them in such a way that the average value for a large number of periods or samples equals, or nearly equals, the value that would be obtained from a similar calculation based on the combined evidence of all possible samples. Estimates for which these two values are exactly equal are known statistically as unbiased estimates. 35/ We also would like the variation of the estimates about their average or true value to be as small as possible since, under this circumstance, we would have more confidence in any single estimate than if the variation about the average is large. Estimating procedures that give the smallest possible variance within a given set of estimating procedures are known as best estimates. Despite their name, such estimates possess no more desirable properties than many alternative estimates. Although the choice of

---

35/ As Mood (74, p. 149) points out, "it is obviously of some advantage to construct an unbiased estimator, but this is not a very crucial requirement. If the mean of an estimator differs but little from the parameter value relative to the standard deviation of the estimator, the estimator may be quite satisfactory." Unfortunately, it frequently is difficult to ascertain the size of the difference to be expected when we work with estimators that are not known to be unbiased.

terminology is unfortunate, the name has become common in statistical literature. In certain circumstances, we may be unable to obtain best unbiased estimates but able to obtain estimates that are efficient and consistent. A consistent estimate is one that is unbiased when we work with all possible data; it may or may not be biased in small samples. 36/ In actual practice, of course, we never have all possible data, but estimation procedures that give consistent estimates presumably are better even with small samples than are those that are known to be biased even with an infinitely large sample. Efficient estimates are similar to "best" estimates, except that they are known to give the smallest possible variance only when we work with all possible data. 37/

If we use the method of least squares to estimate the coefficients in equation (68), we obtain best unbiased estimates if the u's and Z's meet certain rather rigid specifications. These specifications are given by the well-known Markoff theorem. Some of the specifications that relate to the u's are difficult to state precisely in nonmathematical terms. Essentially, they require that the u's follow some (not necessarily a normal) probability distribution, that their average or expected value be zero, that their variance be finite and independent of the particular values of the Z's (that is, that they be homoscedastic) and, finally, that they shall be serially independent. When working with economic data, we usually assume that these specifications hold, but we may test at least the one regarding serial independence of the residuals after we have run the analysis. As discussed on pages 30-32 and 37-38, in some cases we transform the data to make the variance of u less dependent upon the Z's or to reduce the serial correlation in the u's.

An additional Markoff specification regarding the Z's is easily stated but economists frequently disregard it. To be certain that the least squares approach gives best unbiased estimates, each Z must be a set of known numbers, in contrast to a random variable. When attempting to obtain elasticities of demand, this specification holds only in rare instances. One case is that for which prices are arbitrarily set at certain levels, as in a retail store experiment, and the quantities bought by consumers at these prices are recorded. We know of only three experiments that have been conducted in this way--those

---

36/ The following comments from Mood (74, p. 150) may help to clarify the meaning of the terms unbiased and consistent. For consistent estimates, "the estimate becomes near the true parameter value with probability approaching one as the sample size increases without limit. ... A consistent estimator is obviously unbiased in the limit (that is, as the sample size approaches infinity), but for finite sample sizes it may be biased, though in such a way that the bias approaches zero as n becomes large. An unbiased estimator may or may not be consistent depending on whether or not its distribution becomes concentrated near its mean as the sample size increases."

37/ These definitions have been adopted by mathematical statisticians in part because they simplify certain mathematical proofs and theorems. They are not necessarily considered ideal definitions from the standpoint of applied users of statistical methods.

by Berry, et al (9), Godwin (43) and Jasper (52). The least squares method was developed for use in connection with experiments in the physical sciences, where the independent variables frequently are sets of known numbers, or for the study of relationships between variables, such as heights of fathers and heights of sons, where no "structural" coefficients are involved.

Econometricians have shown that the least squares approach gives estimates of the structural coefficients that are statistically efficient and consistent--but not necessarily best and unbiased--provided the u's meet approximately the same requirements as for the previous case, if the Z's are predetermined variables as defined on page 58. A simplified proof of this is given by Klein (58, pp. 80-85), though a knowledge of calculus and the principles of probability given in chapter 2 of his book are required to follow his development.

We now can reconsider an example cited earlier. In section F we showed a diagrammatic representation of a situation for which the least squares method could be used to estimate the slope of the demand curve. We now know that this estimate will be statistically efficient and consistent only if the quantity consumed and the demand shifters each can be classified as an exogenous or lagged endogenous variable and the other statistical requirements that have been noted, such as lack of serial correlation in the residuals, are met. As noted previously, Fox (33) has argued that the least squares method can be used to estimate the coefficients in the structural demand equations for a considerable number of agricultural products, including meat, poultry and eggs, feed grains, and several fresh fruits and vegetables. 38/ Under the assumptions of Fox, market price is used as the dependent variable and the independent variables are production (which is assumed to be highly correlated with consumption) and some relevant demand shifters.

Another situation under which we can use the method of least squares to estimate elasticities of demand is that for which data are available on purchases or consumption of individual consumers, as prices that confront consumers are determined chiefly by factors other than those that affect their purchases. In this case, consumption is taken as the dependent variable and retail prices of the various items, family income, and perhaps other household characteristics, are taken as independent variables.

As noted previously, least squares equations may yield satisfactory forecasts--and, at times, better forecasts than any other method of fit, depending on the particular circumstances that apply (see page 141).

One further problem needs to be mentioned before leaving the statistical aspects of this subject. In all cases given previously, we have assumed implicitly that the variables are known without error. Any analyst who has been

---

38/ Research currently underway within the Agricultural Marketing Service is designed to measure the extent to which production is a predetermined variable for certain livestock products.

connected with the compilation of data from original sources knows that errors of one sort or another always creep in. These can result from memory bias on the part of respondents, inability to find all of the people in a complete census, errors of sampling, and a host of other reasons beyond the control of the most careful investigator. Whenever we work with economic series, non-negligible errors in the data are known to exist. A reasonably complete consideration of this subject at this point would break the continuity of our discussion. This topic is covered in the section beginning on page 143.

Some Econometric Considerations

Discussion in the preceding section suggests, and econometricians have shown, that we obtain estimates that are statistically biased if we use the least squares approach to estimate the coefficients in an equation that contains current values of two or more endogenous variables, where an endogenous variable is defined as one that is correlated with the unexplained residual. The mathematical nature of the bias has been shown by a number of authors in a supposedly popular way, 39/ but none of these explanations is completely satisfactory for a nonmathematician. We do, however, have some experimental evidence of the kind of statistical bias that results when the method of least squares is applied in such cases. Methods which have been developed to estimate the coefficients in equations that contain two or more endogenous variables are known to give results that are statistically consistent, but methods are not now available that are known to be statistically unbiased. Thus, we cannot say for sure what happens when we work with the small samples that usually are involved in economic research.

An experiment was designed by Wagner (96) to measure the kind of statistical bias that arises when we apply these methods to small samples. As a byproduct of this experiment, we have some concrete evidence of the kind of bias that may arise when we use the method of least squares instead.

In this experiment, a simple 3-equation model was formulated with known coefficients. Variables generated by the model were obtained and random error terms added to them. Two thousand observations were obtained in this way, and they were then divided into 100 samples of 20 observations each. The first equation contains two endogenous variables. Coefficients were obtained for this equation for each sample by the limited information approach, which is discussed on page 63, and by the method of least squares. Since there were 100 samples, 100 separate estimates of the single regression coefficient involved were obtained by each approach. Frequency distributions of these estimates are shown in section H of figure 5, together with the true value of the coefficient. Each method gives estimates that are statistically biased, since the average for the 100 samples differs from the true value, but the 3 highest frequencies for the limited information approach are grouped about the

---

39/ See, for example, Bennion (8), Bronfenbrenner (12), and Meyer and Miller (73).

true value, whereas the 3 highest frequencies for the least squares approach each are to the right of the true value. The average bias of the least squares estimates is almost three times as large as that for the limited information approach. For a second version of this model the two frequency distributions are more nearly similar, but the average bias in the least squares estimates still is almost double that for the limited information estimates. In this study, carried out at Stanford University, a large scale electronic computer was used. Further experiments of similar character, with different sorts of models, currently are underway at Yale University.

The experimental study by the Agricultural Marketing Service, referred to previously and discussed in the section beginning on page 128, also gives evidence of the degree of statistical bias that may result when the least squares approach is used to estimate the coefficients in equations that contain more than a single endogenous variable. As discussed in the section beginning on page 67, it is sometimes possible to estimate in advance the degree of bias that is likely to exist.

In discussing methods that are used to estimate the structural coefficients in equations that involve more than one endogenous variable, it is convenient to introduce a mathematical concept that deals with the degree of identification. We saw earlier that it is sometimes impossible to estimate the coefficients in certain structural equations with the kind of statistical data available (see page 55). Such equations are said to lack identifiability or to be underidentified. In this connection, Durbin (17, p. 354) says, "It is easy to get formal identification by adding extra variables to the model, but if the effect of these variables on the system is small, then for small samples the standard errors of the coefficients will be so large that the situation is almost as bad as if the model were not identified. Unfortunately, nothing can be done to remedy this." Fortunately, most equations within models that relate to the demand and price structure for agricultural commodities are identifiable and, in the discussion that follows, we deal only with identifiable equations. Such equations may be just identified or overidentified. In general, the degree of identification relates to individual equations in a system, not to the entire system.

By algebraic manipulation, we always can write down the equations in a complete system so that the number of equations equals the number of endogenous variables. We then can think of these as n equations in n unknown endogenous variables, and we can always solve the equations so that each endogenous variable is expressed as a function of all of the predetermined variables in the system. These are called reduced form equations. Since each reduced form equation contains only a single endogenous variable, estimates of the coefficients in these equations that are statistically efficient and consistent can be obtained by use of the method of least squares. However, as discussed in the section beginning on page 63, we may at times wish to obtain these coefficients instead by an algebraic derivation from structural coefficients that have been estimated from statistical data.

Just identified equations.--If each equation in the system is just identified, there always exists a unique transformation by which we can go from the coefficients in the reduced form equations to the coefficients in the structural equations. Suppose we have the following structural demand and supply equations, where q is production or consumption, p is price, d is consumer income, w represents important weather factors that affect supply, and each variable is expressed in terms of deviations from its respective mean:

$$p = b_{11}q + b_{12}d \qquad \text{(demand relation)} \qquad (69)$$

$$q = b_{21}p + b_{22}w. \qquad \text{(supply relation)} \qquad (70)$$

A rule of thumb to determine the degree of identification states that each equation is just identified if the number of variables in the system minus the number of variables in each equation equals the number of endogenous variables in the system minus one; more exact rules depend on the rank of certain matrices. As there are 4 variables in the system and 3 variables in each equation and we have two endogenous variables in the system, each equation is just identified.

If we substitute the right hand side of equation (70) for q in equation (69) and the right hand side of equation (69) for p in equation (70) and simplify terms, we obtain the following:

$$p = \frac{b_{11}b_{22}}{1-b_{11}b_{21}} w + \frac{b_{12}}{1-b_{11}b_{21}} d \qquad (71)$$

$$q = \frac{b_{22}}{1-b_{11}b_{21}} w + \frac{b_{12}b_{21}}{1-b_{11}b_{21}} d. \qquad (72)$$

Since the denominator of each term on the right of the equality sign is identical, we can ignore these denominators for the moment. If we divide the coefficient of w in equation (71) by the coefficient of w in equation (72), we obtain an estimate of b_{11}. If we divide the coefficient of d in equation (72) by the coefficient of d in equation (71), we obtain an estimate of b_{21}. Given an estimate of b_{11} and b_{21}, we can estimate b_{12} from the coefficient of d in equation (71) and b_{22} from the coefficient of w in equation (72). This gives the four coefficients needed for our structural equations. Estimates that are uniquely equivalent are obtained by any alternative algebraic manipulation. Since the b's are known to be statistically consistent estimates, provided the necessary criteria with respect to the u's are met (see page 58) and the predetermined variables are known without error, the estimates of the structural coefficients obtained in this way are statistically consistent.

Computationally, we may wish to estimate the coefficients in another way, but the answers obtained are identical with those that would be gotten by an algebraic manipulation of the regression coefficients from the reduced form equations.

Overidentified equations.--Let us now consider an equation that is overidentified. Suppose that equation (70) on supply contains a second predetermined variable, z, that represents lagged values of prices. We now have 5 variables in the system. The supply equation still is just identified, since 5-4 equals the number of endogenous variables in the system minus one. However, equation (69) on demand is overidentified, since 5-3 is greater than 2-1. The new reduced form equations can be obtained by the same general approach as used previously, but the result now looks like this:

$$p = \frac{b_{11}b_{22}}{1-b_{11}b_{21}} w + \frac{b_{11}b_{23}}{1-b_{11}b_{21}} z + \frac{b_{12}}{1-b_{11}b_{21}} d \qquad (73)$$

$$q = \frac{b_{22}}{1-b_{11}b_{21}} w + \frac{b_{23}}{1-b_{11}b_{21}} z + \frac{b_{12}b_{21}}{1-b_{11}b_{21}} d. \qquad (74)$$

With this set of equations, b_{11} could be estimated either by dividing the coefficient of w in equation (73) by the coefficient of w in equation (74) or by dividing the coefficient of z in equation (73) by the coefficient of z in equation (74). Different answers are obtained from the two estimates. It is in this way that overidentified equations differ from just identified ones; for overidentified equations, we have an oversufficiency of information and no direct way to decide which answer to use. In fact, neither answer obtained by the use of reduced form equations is statistically consistent.

It would be possible to estimate the several coefficients in the two structural equations directly by use of a maximum likelihood approach. Maximum likelihood estimates are known to be statistically consistent and efficient. They are used widely in statistical work because the necessary equations always can be derived by performing certain mathematical operations that involve the maximization of the so-called likelihood function. The general approach is the same as for any maximization process by use of calculus, and it is not difficult. For complex systems of equations, however, the mathematics involved in solving the resulting equations is generally complex. That part of Klein (58) referred to on page 59 involved the derivation of maximum likelihood estimates. Methods for obtaining maximum likelihood estimates based on a simultaneous solution for all of the structural equations are discussed by Klein (58) and Chernoff and Divinsky (13) and are called full-information maximum likelihood estimates but, to quote Klein, the computations involved in general are "formidable." Hence this method is seldom used.

Another method, developed by staff members of the Cowles Commission for Research in Economics, is called the single-equation limited-information maximum-likelihood method. In this approach, equations are fitted one at a time and less information is utilized than in the full-information approach. In essence, a least squares fit is obtained between a linear combination of the endogenous variables in the equation on the predetermined variables in the equation, subject to the condition that the predetermined variables in the system but not in the equation have a minimum effect on the combination of endogenous variables. Since, if the model is properly formulated, the latter variables should have no effect on the endogenous variables in the particular equation, this is a reasonable procedure. This method is known to give estimates of the coefficients that are statistically consistent and as efficient as any other method that utilizes the same amount of information, provided the usual criteria with respect to the u's and errors in the data are met. A compromise between this and the full-information method is one called the limited-information subsystem method, in which selected groups of equations are fitted as a unit. Most of the systems of simultaneous equations that have been fitted and that involve overidentified equations have been based on the single-equation limited information approach. This is the approach that generally is meant when we refer to the use of the limited information method.

Alternative Methods

Three alternative methods of fitting equations have been proposed, each of which, if properly used, will yield estimates of the structural coefficients that are statistically consistent, even though more than one endogenous variable is contained in the equation. These methods, however, are not necessarily efficient in a statistical sense. The methods are (1) the recursive approach, which has been advocated particularly by Herman Wold, (2) the method of instrumental variables, and (3) a method proposed by H. Theil.

Recursive systems.--Estimates of the structural coefficients that are statistically consistent are obtained from the recursive approach only when the system of equations has a special form; if it has this form, it is known as a recursive system. A recursive system is made up in the following way: (1) At least one equation contains only a single endogenous variable. As we have pointed out, consistent estimates of the coefficients in such equations can be obtained by fitting them directly by least squares, provided the endogenous variable is treated as dependent. In some systems, several equations contain only a single endogenous variable; each of these is fitted directly by least squares, using the endogenous variable as dependent. (2) At least one other equation must contain only one endogenous variable in addition to those contained in the first set. Consistent estimates of the coefficients in these equations can be obtained if they are fitted directly by least squares, provided calculated values of the endogenous variables included in the equations referred to in item (1) are substituted for actual values before making the computations and the single new endogenous variable is treated as dependent. 40/ (3) The recursive system as a whole must be of such a nature that by

---

40/ Reasons for using calculated values rather than actual values are the following: (1) The unexplained residuals in the several equations within an equation system are assumed to be correlated one with another and, by definition, an endogenous variable in a particular equation is assumed to be correlated with the unexplained residuals in that equation. (2) Calculated values for a given variable in a given equation are known to be uncorrelated with the successive steps each of the equations can be transformed into one that contains only a single endogenous variable other than those which have been treated as dependent in prior analyses. Wold believes that many economic systems are of this nature, whereas certain other econometricians believe that systems of simultaneous relations of a nonrecursive type are more typical. These points of view are interestingly expressed in Wold (104) and the discussion which follows his paper.

Instrumental variables.--The method of instrumental variables is described by Klein (58, pp. 122-125). He says, "Choose as many exogenous variables in the system as there are unknown parameter coefficients of endogenous variables in the given equation. Multiply each variable in the equation by one of the exogenous variables, and sum the equation over all sample observations. Repeat this procedure for each exogenous variable selected from the group outside the given equation and for each exogenous variable in the equation. This will provide as many linear equations in the unknown parameters as there are unknown parameters. Solve for the estimates of the parameters." The consistency property of the estimates remains if lagged endogenous variables as well as exogenous variables are used.

Klein goes on to say, "Obviously, the objection to this technique is that the set of instrumental variables is not unique" except when each of the equations is just identified, in which case the method of reduced forms is equally easy. "In general ... the estimates obtained for the parameters will depend upon the particular set of instrumental variables selected, and this arbitrariness lessens the attractiveness of the method. Some principles can be indicated for the selection of the instrumental set from the total number of exogenous variables in the system, and these principles may help to reduce the degree of arbitrariness involved. One should avoid choosing a set of instrumental variables which is highly intercorrelated and should try to choose that set which shows the greatest correlation with the endogenous variables in the given equation. By the same reasoning, one should choose instrumental variables which are the least correlated with the predetermined variables already present in the equation being studied. ... The use of the method of instrumental variables would seem to be desirable when quick results are wanted in exploratory studies. It should be especially useful in screening multiple possibilities among a group of admissible hypotheses."

Theil's two-rounds estimates.--Descriptions of his method given by Theil (90, 91) are not likely to be understood by most readers, but Klein (59) discusses the method in terms of the matrices used in connection with limited information estimates. The discussion of Theil's method given here is adapted from that in Klein.

---

unexplained residuals in that equation because the residuals are ignored in the computations. (3) Hence, calculated values for an endogenous variable obtained from one equation are uncorrelated with the unexplained residuals in another equation within the same system, and the calculated series becomes in effect a predetermined variable.

In obtaining the coefficients in an individual equation that contains more than one endogenous variable, we first write the equation with one endogenous variable expressed as a function of the other variables in the equation. We then take each of the predetermined variables in the system in turn and perform computations similar to those used in the method of instrumental variables. If the initial equation is overidentified, this process yields more equations than there are coefficients to be estimated. To make use of all of the information available, Theil runs a least squares relationship between the variable on the left of the equality sign in each of these equations and the several variables on the right of the equality sign, with the value in each equation treated as an observation. This procedure gives estimates of the coefficients on the corresponding variables to the right of the equality sign in the initial equation. These estimates have statistical properties that are equivalent to those of coefficients estimated by the single-equation limited-information method. Klein (59, p. 149) proves that "Theil's method of two-rounds estimation is the same thing as the method of instrumental variables, using a particular linear combination of instrumental variables and leaving no room for arbitrary selection." 41/

The exact procedure can be clarified by a concrete example. Suppose we have N observations for a system of equations that contains 3 endogenous and 4 predetermined variables. Assume that one of the equations in the system contains the following variables, where the y's refer to endogenous and the z's to predetermined variables, each expressed as deviations from their respective means:

$$y_1 = b_{11}y_2 + b_{12}y_3 + b_{13}z_1. \tag{75}$$

This equation is overidentified, since the number of variables in the system minus the number of variables in the equation is greater than the number of endogenous variables in the system less one. If we wish to estimate the coefficients in this equation by the method proposed by Theil, the following steps are used:

(1) Obtain the following equations by multiplying each observation for equation (75) by the indicated predetermined variable and summing over the N observations:

$$\Sigma z_1y_1 = b_{11}\Sigma z_1y_2 + b_{12}\Sigma z_1y_3 + b_{13}\Sigma z_1^2 \tag{76}$$

$$\Sigma z_2y_1 = b_{11}\Sigma z_2y_2 + b_{12}\Sigma z_2y_3 + b_{13}\Sigma z_2z_1 \tag{77}$$

$$\Sigma z_3y_1 = b_{11}\Sigma z_3y_2 + b_{12}\Sigma z_3y_3 + b_{13}\Sigma z_3z_1 \tag{78}$$

$$\Sigma z_4y_1 = b_{11}\Sigma z_4y_2 + b_{12}\Sigma z_4y_3 + b_{13}\Sigma z_4z_1. \tag{79}$$

41/ This point was proved independently by Basmann (7).

(2) Fit the following equation by least squares, using as observations the corresponding four values from equations (76) through (79):

$$\Sigma z_i y_1 = b_{11} \Sigma z_i y_2 + b_{12} \Sigma z_i y_3 + b_{13} \Sigma z_i z_1 . \quad (i = 1,2,3,4) \qquad (80)$$

The estimates of b_{11}, b_{12}, and b_{13} from the least squares fit of equation (80) are the desired ones for the corresponding coefficients in equation (75).

Radner and Bobkoski 42/ compare the time required to use this approach with that of the limited information method for two examples. For one example, Theil's method required 58 percent as long and for the other example, 92 percent. They feel that in neither example does the full advantage of Theil's method become apparent.

The chief disadvantage of Theil's method, in relation to the method of limited information, is that different results are obtained depending on which variable is written to the left of the equality sign.

Reasons Why a Least Squares Fit May Be Satisfactory Even Though the Estimates of the Structural Coefficients are Subject to Some Statistical Bias

Before Christ (14) had published an article, our staff had decided to abandon the practice of publishing structural coefficients obtained by fitting both by least squares and by a simultaneous equations method, such as that of limited information. Instead, we had expected to publish only the results obtained by the latter approach. However, Christ (14, pp. 397-398) says, "... the least-squares method (of estimating the coefficients) in a system of (structural) equations can be likened to a shotgun that scatters its shot fairly close together, but not centered on the bullseye. ... the limited-information method can be likened to a shotgun that scatters its shot less close together than the least-squares shotgun does, and not centered right on the bullseye either, but becoming better centered and approaching perfect centering as the sample size approached infinity. Thus the question of which method to use for any finite sample size is still open, for we do not know how to tell whether the bias of the limited-information method at a given sample size is smaller than that of the least-squares method by enough to compensate for its bigger variance." Christ recommends a continuation of the practice of publishing coefficients obtained by both methods in the hope that empirical evidence eventually will provide a guide to the one to be preferred under specified circumstances. We concur in his recommendation.

Wold and Faxer (105) indicate why in certain instances results of the two approaches are so nearly similar. Suppose the only error which the analyst makes in specifying the nature of the statistical model is one of assuming that

---

42/ Radner, R., and Bobkoski, F. Some Recent Work of H. Theil on Estimation in Systems of Simultaneous Equations. Cowles Commission Discussion Paper: Statistics No. 385. 1954. 10 pp. [Processed.]

a certain variable is predetermined when in fact it is endogenous. This is equivalent to assuming that an independent variable in a least squares analysis is uncorrelated with the unexplained residual when in fact a non-zero correlation exists. In such a case, Wold and Faxer show that the degree of specification error, as contrasted with the sampling error which applies in any case, in a 2-variable analysis is equal to the degree of correlation between the independent variable and the unexplained residual times the unexplained variance as a ratio to the total variance in the dependent variable. Although the exact formula is slightly more complex, a similar situation holds when more than two variables are involved in the analysis. Thus, if the analysis explains a substantial part of the variation in the dependent variable, the degree of specification error must of necessity be relatively small. This explains why the least squares approach frequently is satisfactory to estimate structural coefficients when used for price level relations even though each of the price variables clearly is endogenous. Also, if the degree of correlation between the assumed predetermined but actually endogenous variable and the unexplained residual is small, the degree of specification error also must be relatively small. This justifies the argument advanced by Fox (33, p. 8) when he said, "... there are certain cases, particularly in analysis of agricultural prices, in which simultaneity is of limited importance. In such cases it is doubtful whether the elaborate procedures of the Cowles Commission will improve or even change the results of the single-equation approach within the limits of sampling error."

Results from an experiment.--A Monte Carlo experiment was designed by us to determine the relative merits of equations fitted by (1) least squares and (2) limited information for use in forecasting under specified conditions. This experiment is described in detail in the section beginning on page Some of the results obtained from it have a direct bearing on the question being considered in this section, namely the probable size of the statistical bias in estimating the coefficients in an equation by least squares versus limited information.

The model used in this experiment is similar to that for wheat described in the section beginning on page 11, except that it involves fewer equations and variables. The initial model was formulated in such a way as to bring about a squared correlation coefficient (or coefficient of determination) of about 0.3 among the unexplained residuals in the three stochastic equations. We anticipated that this degree of correlation would induce a fairly high correlation between the endogenous variable to be treated as independent in the least squares fit and the unexplained residuals in each of the three equations. However, the latter correlation coefficient squared in no case exceeded 0.05. As the multiple coefficient of determination for each of these structural equations exceeded 0.9, differences between the coefficients estimated by (1) least squares and (2) limited information were negligible. This confirms empirically the theorem of Wold and Faxer.

The model was reformulated to bring about a squared correlation coefficient of more than 0.9 between the unexplained residuals in the three equations and to reduce the multiple coefficient of determination for the structural equations moderately. These changes were designed to assure that different coefficients would be obtained by the two approaches. This change in the model resulted in increasing the squared correlation coefficients between the endogenous variable written on the right of the equality sign and the unexplained residuals in each structural equation to 0.7 or more. Based on the theorem of Wold and Faxer, a substantial degree of statistical bias would be expected if the coefficients in these equations were estimated by the method of least squares and, in fact, the coefficients estimated by least squares did differ considerably from the structural coefficients and by more than would be expected based on their standard errors (see page 131). Results when these and related equations were used for forecasting are discussed in the section beginning on page 128.

This experiment suggests that a high degree of correlation between a so-called endogenous variable and the unexplained residuals in a particular structural equation will exist only when a high degree of correlation exists among the unexplained residuals in the several equations within the system. Thus, unless the latter correlations are high, coefficients estimated by least squares may be nearly the same as those obtained by a method that allows for the simultaneity implied by the system, such as limited information. If this is true, the computational simplicity of the method of least squares suggests that it could be used to provide approximate estimates of the structural coefficients, except when a high degree of correlation among the unexplained residuals in the several equations is anticipated.

Results from empirical research.--This hypothesis, if subsequently verified, would explain why some systems of equations given coefficients that are nearly identical when fitted by least squares versus limited information, as was true for the 6-equation model for wheat used by Meinken (71), and why other systems give quite different answers by the two methods, as was true for the 9-equation dairy models used by Rojko (81). The four utilization equations in the model for wheat relate to food, feed, export, and storage, respectively, and negligible correlations between the unexplained residuals in the several equations would be anticipated. Factors that affect the consumption of the several dairy products but are not included in the equations are more homogeneous, so that a fairly high degree of correlation would be expected among the unexplained residuals in the several equations. Systems of equations for certain other commodities have been fitted by these alternative approaches, and results obtained also are consistent with those that would be anticipated from an application of this hypothesis. Further mathematical, experimental, or empirical evidence will be required for complete verification.

AN EXAMPLE THAT ILLUSTRATES SOME FITTING PROCEDURES 43/

Beginning on page 14, we discuss an 11-equation model of the egg economy. At an earlier time, this model contained 9 equations, but the economic basis for it is similar to that for the 11-equation model. In this section, we

---

43/ This section is based chiefly on material prepared by Martin J. Gerra, agricultural economic statistician, Agricultural Marketing Service. A knowledge of the notation and procedures given in Friedman and Foote (40) is assumed.

discuss the fitting procedure for the 9-equation model, giving only such economic aspects as are required in connection with the fitting process. Methods for fitting systems of equations by the limited information approach, or a slight modification of it for equations that are just identified, are given in detail in Friedman and Foote (40). The discussion included here supplements that given by these authors, chiefly in two ways: (1) More detail is shown with respect to obtaining linear approximations for non-linear variables and (2) steps that can be run efficiently on certain electronic computers are delineated. The latter, of course, depend on the nature of the data processing equipment available and the computational programs that have been written.

Equations and Variables in the Statistical Model

The following variables are taken as endogenous in this model: Q_E, P_R and P_R', L_F, C_F, L_C, P_F and P_F', S'. Their exact economic meaning need not be specified. The remaining variables are assumed to be predetermined.

Structural form.--The following structural equations are involved in a system representing the supply and demand for eggs. The nine equations shown correspond to the nine endogenous variables in the system. Therefore, this is a complete system.

$$\frac{Q_E}{H} = a_1 + b_{12} P_R + c_{11} \frac{I}{H} + c_{12} P_M + c_{13} P_C + c_{14} P_B + c_{15} P_t + c_{16} P_o + u_1 \qquad (81)$$

$$Q_E = (Q_A L_F) - A \qquad (82)$$

$$L_F = L_J + C_F - M - L_C \qquad (83)$$

$$C_F = a_4 + b_{42} \frac{P_F'}{P_G'} + u_4 \qquad (84)$$

$$L_C = a_5 + b_{52} \frac{P_F}{P_G} + u_5 \qquad (85)$$

$$P_F = a_6 + b_{62} P_R + c_{61} W + u_6 \qquad (86)$$

$$P_R' = a_7 + b_{72} \frac{Q_F'}{H} + b_{73} \frac{S'}{H} + c_{71} \frac{I}{H} + c_{72} P_M' + c_{73} P_C' + c_{74} P_B' + c_{75} P_t' + c_{76} P_o' + u_7 \qquad (87)$$

$$P_F' = a_8 + b_{82} P_R' + c_{81} W + u_8 \qquad (88)$$

$$\frac{S'}{H} = a_9 + b_{92} \frac{Q_F'}{(Q_E + A)} + c_{91} F + u_9 \qquad (89)$$

In this notation, the a's represent the constant term in each equation, the b's and c's represent structural coefficients that apply to the endogenous and predetermined variables, respectively, and the u's represent random error terms. No coefficients or random error terms are involved in equations (82) and (83) because they are identities; they need not be fitted by statistical means.

Linearized form.--Because the equations in the egg model are stated in linear terms, the endogenous variables in the model used in nonlinear combinations (for example, $\frac{Q_E}{H}$, $Q_A \cdot L_F$, $\frac{P_F}{P_G}$, and so forth) were transformed into linear approximations by use of formulas given by Klein (58, pp. 120-121). The linear approximations are then substituted for the original variables. Combinations of variables that are assumed to be entirely predetermined, however, are treated as a single composite without linearization as, within the analysis, they are assumed to be given.

The formulas given by Klein are:

$$XY \cong \bar{Y}X + \bar{X}Y - \bar{X}\bar{Y} \tag{90}$$

$$X/Y \cong \bar{X}/\bar{Y} + X/\bar{Y} - (\bar{X}/\bar{Y}^2)\,Y \tag{91}$$

where $\bar{X}$ and $\bar{Y}$ are the means of X and Y, respectively. If either the product or the quotient is multiplied by a constant, then each term in the transformation is multiplied by the constant.

Rewriting the structural equations in linearized form, we obtain:

$$\frac{\bar{Q}_E}{\bar{H}} + \frac{1}{\bar{H}} Q_E - \frac{\bar{Q}_E}{\bar{H}^2} H = a_1 + b_{12}P_R + c_{11}\frac{I}{H} + c_{12}P_M + c_{13}P_C + c_{14}P_B + c_{15}P_t + c_{16}P_o + u_1 \tag{81.1}$$

$$Q_E = (\bar{Q}_A \cdot L_F - \bar{Q}_A \cdot \bar{L}_F) + (\bar{L}_F \cdot Q_A) - A \tag{82.1}$$

$$L_F = L_J + C_F - M - L_C \tag{83}$$

$$C_F = a_4 + b_{42}\left[\frac{\bar{P}_F'}{\bar{P}_G'} + \frac{1}{\bar{P}_G'} P_F' - \frac{\bar{P}_F'}{\bar{P}_G'^2} P_G'\right] + u_4 \tag{84.1}$$

$$L_C = a_5 + b_{52}\left[\frac{\bar{P}_F}{\bar{P}_G} + \frac{1}{\bar{P}_G} P_F - \frac{\bar{P}_F}{\bar{P}_G^2} P_G\right] + u_5 \tag{85.1}$$

$$P_F = a_6 + b_{62}P_R + c_{61}W + u_6 \tag{86}$$

$$P_R' = a_7 + b_{72} \left[\frac{\bar{Q}_F'}{\bar{H}} + \frac{1}{\bar{H}} Q_F' - \frac{\bar{Q}_F'}{\bar{H}^2} H \right] +$$

$$b_{73} \left[\frac{\bar{S}'}{\bar{H}} + \frac{1}{\bar{H}} S' - \frac{\bar{S}'}{\bar{H}^2} H \right] + c_{71} \frac{I}{H} + c_{72} P_M' +$$

$$c_{73} P_C' + c_{74} P_B' + c_{75} P_t' + c_{76} P_o' + u_7 \qquad (87.1)$$

$$P_F' = a_8 + b_{82} P_R' + c_{81} W + u_8 \qquad (88)$$

$$\frac{\bar{S}'}{\bar{H}} + \frac{1}{\bar{H}} S' - \frac{\bar{S}'}{\bar{H}^2} H = a_9 + b_{92} \left[\frac{\bar{Q}_F'}{(Q_E + A)} - \frac{\bar{Q}_F'}{(Q_E + A)^2} (Q_E + A) + \frac{1}{(Q_E + A)} Q_F' \right] + c_{91} F + u_9 \qquad (89.1)$$

Fitting the Linearized Model

In fitting the model of the egg economy, equations (82.1) and (83) are excluded because they have no statistical coefficients. However, in using the model for analytical purposes or forecasting, these equations are used together with the seven equations that involve statistical coefficients. The seven equations each contain more than one endogenous variable and are each overidentified. Therefore, they are fitted by the limited information method.

The analysis is based on first difference values for the years 1931-41 and 1946-54. Conditions during these years are assumed to have been sufficiently homogeneous to permit their inclusion in a single analysis. The war years 1942-45 are excluded because of abnormal circumstances that influenced supply and demand in that period. First differences are used because of the several trend factors that have affected supply and demand.

Variables used in the M_{zz} and M_{zy} matrices.--The M_{zz} matrix usually contains moments for all the predetermined variables in the system. In some cases, however, the number of predetermined variables in the system exceeds the number of available observations and modifications must be made. To prevent the obtaining of indeterminate results, we must have at least one more year of data than there are predetermined variables in the M_{zz} matrix.

In the model of the egg economy, there are 21 predetermined variables and 18 observations for each variable; at least 4 predetermined variables must be omitted. For reasons given in subsequent paragraphs, 6 variables are excluded from the matrix of predetermined variables M_{zz}, leaving 15 variables in this matrix.

Moments for the 5 predetermined variables--P_M', P_C', P_B', P_t' and P_o'--that relate to prices of competing or complementary goods during the first half of the year, are excluded from the matrix of predetermined variables for the entire system, because these variables are highly correlated with their respective annual averages--P_M, P_C, P_B, P_t, and P_o. Moments for these variables, however, are included in certain other matrices when equation (87.1) is fitted individually. With the method of limited information, moments for some predetermined variables can be dropped from the matrix of predetermined variables for the entire system if this appears desirable, provided sufficient predetermined variables are used to provide identification, although the estimates of the coefficients tend to be less efficient in a statistical sense than those in which moments for all the predetermined variables in the system are used.

Moments for the predetermined variable M, which relates to mortality, are omitted from the M_{zz} matrix because reliable data are not available. 44/ Experience in fitting models of this sort indicates that improved results are obtained when variables that poorly represent the economic factor that should be in the structural model are omitted from the M_{zz} matrix. Moments for such variables should be used when fitting the particular equation in which they are found, but, in this case, the omitted variable is in an identity, hence no statistical fitting is involved.

The linearized values of each of the predetermined variables used in combination with an endogenous variable in the structural equations are used only once in computing moments for the M_{zz} matrix. For example, Q_E/H, when transformed into linearized form, is:

$$\frac{\bar{Q}_E}{\bar{H}} + \frac{1}{\bar{H}} Q_E - \frac{\bar{Q}_E}{\bar{H}^2} H.$$

Similar last terms are involved for Q_F'/H and S'/H. As the respective last terms in each are perfectly correlated, only one, namely - $(\bar{Q}_E/\bar{H}^2)$ H, is used in computing moments for the M_{zz} matrix. As an alternative, moments for H as such might have been used.

In computing moments for the endogenous variables, however, the entire linearized value of the combination of endogenous and predetermined variables is used in place of the original variable. For example, for Q_E/H, the following linearized form is included in computing the moments for the M_{zy} matrix:

44/ Mortality may be, in fact, an endogenous variable but to so treat it would have required an additional equation to keep the model complete. If it depends on variables that are included in the M_{zz} matrix, then no loss in efficiency is engendered by omitting it from this matrix.

Table 3.--Variables used in obtaining moments for specified matrices and equations

Variables used in--	
M_{zz} matrix	M_{zy} matrix
$-\frac{\bar{Q}_E}{\bar{H}^2} H = z_1$	$\frac{\bar{Q}_E}{\bar{H}} + \frac{1}{\bar{H}} Q_E - \frac{\bar{Q}_E}{\bar{H}^2} H = y_1$
$I/H = z_2$	$P_R = y_2$
$P_M = z_3$	$C_F = y_3$
$P_C = z_4$	$\frac{\bar{P}'_F}{\bar{P}'_G} + \frac{1}{\bar{P}'_G} P'_F - \frac{\bar{P}'_F}{\bar{P}'^2_G} P'_G = y_4$
$P_B = z_5$	$L_C = y_5$
$P_t = z_6$	$\frac{\bar{P}_F}{\bar{P}_G} + \frac{1}{\bar{P}_G} P_F - \frac{\bar{P}_F}{\bar{P}_G^2} P_G = y_6$
$P_O = z_7$	$P_F = y_7$
$\bar{L}_F Q_A = z_8$	$P'_R = y_8$
$A = z_9$	$\frac{\bar{Q}'_F}{\bar{H}} + \frac{1}{\bar{H}} Q'_F - \frac{\bar{Q}'_F}{\bar{H}^2} H = y_9$
$L_J = z_{10}$	$\frac{\bar{S}}{\bar{H}} + \frac{1}{\bar{H}} S' - \frac{\bar{S}'}{\bar{H}^2} H = y_{10}$
$-\frac{\bar{P}'_F}{\bar{P}'^2_G} P'_G = z_{11}$	$P'_F = y_{11}$
$-\frac{\bar{P}'_F}{\bar{P}_G^2} P_G = z_{12}$	$\frac{\bar{Q}'_F}{(\bar{Q}_E + A)} - \frac{\bar{Q}'_F}{(\bar{Q}_E + A)^2} (Q_E + A) + \frac{1}{(\bar{Q}_E + A)} Q'_F = y_{12}$
$\frac{1}{\bar{Q}_E + A} Q'_F = z_{13}$	
$F = z_{14}$	
$W = z_{15}$	

Predetermined variables excluded from M_{zz} matrix but used in individual equations

$P'_M = z_{16}$	$P'_t = z_{19}$
$P'_C = z_{17}$	$P'_O = z_{20}$
$P'_B = z_{18}$	

$$\frac{\overline{Q}_E}{\overline{H}} + \frac{1}{\overline{H}} Q_E - \frac{\overline{Q}_E}{\overline{H}^2} H,$$

and for S'/H, we include:

$$\frac{\overline{S}'}{\overline{H}} + \frac{1}{\overline{H}} S' - \frac{\overline{S}'}{\overline{H}^2} H.$$

This is done because a single structural coefficient applies to the entire combination.

If both the numerator and the denominator are endogenous variables, then no terms from the linearized variable are involved in the M_{zz} matrix. As in the previous case, the entire linearized variable is treated as a unit in obtaining moments for the matrices that involve endogenous variables.

The next step in the fitting process is to assign numbered y's and z's to all of the variables, making use of the linearized forms. The letter z is used if the variable is predetermined and the letter y if it is endogenous. In table 3, the variables are designated as being included in either the M_{zz} matrix or the M_{zy} matrix, or as being excluded from the M_{zz} matrix but used in fitting the individual equations. The y's included in the identities are omitted from the M_{zy} matrix as these equations need not be statistically fitted.

Variables used in fitting the individual equations.--Listed in table 4 are the variables used in fitting each of the seven equations--they are designated by their structural symbols. The endogenous variables are classified under Y* and the predetermined variables under Z*. The corresponding y's and z's are indicated in parentheses.

Considerable care must be used in computing the linearized values of the variables, particularly when first differences are used, to assure that enough decimals are carried to make certain that the final differences shall contain several significant digits. The linearized values in all cases are obtained before the first differences are computed.

In deriving the reduced form forecasting equations after the model has been fitted, the term in the linearized value containing the endogenous variable is treated as an unknown value while the remaining terms in the linearized value are treated as known values. For example, equation (84.1), after values have been obtained for a_4 and b_{42}, is rewritten as:

$$C_F - \frac{b_{42}P_F'}{\overline{P}_G'} = a_4 + b_{42}\frac{\overline{P}_F'}{\overline{P}_G'} - b_{42}\frac{\overline{P}_F'}{\overline{P}_G'^2} P_G'.$$

If first differences are used in fitting the model, the term $b_{42} \frac{\bar{P}_F}{\bar{P}'_G}$ in the forecasting equation equals zero because the first difference of a constant is zero.

Table 4.--Variables used in fitting specified equations

Equation	Y*	Z*
(81)	Q_E/H, P_R (y_1, y_2)	I/H, P_M, P_C, P_B, P_t, P_0 (z_2, z_3, z_4, z_5, z_6, z_7)
(84)	C_F, P'_F/P'_G (y_3, y_4)	---
(85)	L_C, P_F/P_G (y_5, y_6)	---
(86)	P_F, P_R (y_7, y_2)	W (z_{15})
(87)	P'_R, Q'_F/H, S'/H (y_8, y_9, y_{10})	I/H, P'_M, P'_C, P'_B, P'_t, P'_0 (z_2, z_{16}, z_{17}, z_{18}, z_{19}, z_{20})
(88)	P'_F, P'_R (y_{11}, y_8)	W (z_{15})
(89)	S/H, $Q'_F/(Q_E + A)$ (y_{10}, y_{12})	F (z_{14})

For certain economic variables that normally change slowly from month to month, an annual average is used in conjunction with other more volatile variables that relate to the first half of the year. For example, annual disposable income was used in the January-June equation (87.1). This is done to simplify the fitting process.

Computations on an Electronic Computer

This section outlines the procedure for estimating on an electronic computer (SEAC) 45/ the structural coefficients and their standard errors for this system of simultaneous equations. In using the limited information approach, we estimate the coefficients of one equation at a time, but two preliminary computations are made which are applicable to all equations in the system. These are (1) computing adjusted augmented moments for all the

---

45/ The National Bureau of Standards, Washington, D. C., designated the computer in use when these computations were made as Standards' Electronic Automatic Computer, or by the sobriquet SEAC. The entire fitting process could have been programmed for SEAC. However, costs of programming were so variables in the system; and (2) obtaining the matrix product $M_{yz} M_{zz}^{-1} M_{zy}$ which, when the entire operation is done on desk calculators, is obtained by performing a forward Doolittle solution, using all the variables listed above as being included in the M_{zz} and M_{zy} matrices. 46/

Computing adjusted augmented moments.--As the first step in computing the adjusted augmented moments, the first differences and means for the 27 variables in M_{zz} and M_{zy}, and a check sum, are obtained on a desk calculator. In setting up the worksheet for this step, the Z values should be listed first in the sequence in which they occur in the system, followed by the Y values, also listed in sequential order, with the check sum at the end. The uncorrected sums of squares and cross-products then are obtained on SEAC by use of an A'A routine. 47/ The uncorrected sums of squares and cross-products are corrected, augmented, and adjusted, in the usual manner 48/ on a desk calculator.

Because the predetermined variables z_{16}, z_{17}, ... z_{20} are excluded from the matrix of predetermined variables (see page 73), it is necessary to carry out an additional computation to obtain their adjusted augmented moments. These predetermined variables (z_{16}, ... z_{20}) are to be used in fitting equation (87.1), and consequently the corrected sums of squares and cross products of z_{16}, ... z_{20}, the endogenous variables appearing in equation (87.1), namely y_8, y_9, and y_{10}, and the predetermined variable z_2 are obtained directly on SEAC and then are augmented and adjusted in the usual way on desk calculators.

---

large that, instead, certain operations were done on SEAC and others, on desk calculators. Slightly different procedures probably would be involved in using a different electronic computer, but the basic principles illustrated here would be applicable.

46/ If this operation is done on desk calculators, only that part of the product matrix actually needed is computed. See Friedman and Foote (40, p. 67).

47/ A program for obtaining corrected sums of squares and cross-products is available on SEAC, but the machine does not have sufficient internal memory to handle a problem of this magnitude. Uncorrected values can be obtained by a matrix multiplication process for which the machine has a larger capacity. A program to obtain adjusted augmented moments for up to 28 variables has been written by the Agricultural Marketing Service for the IBM 650.

48/ See Friedman and Foote (40, pp. 3-7). Particularly in a problem of this size, it is desirable to combine tables 1, 2, and the upper part of table 3, as shown in their publication. The k_i should be entered as a third row in the first section of table 1. Computations shown in the second and third row of each section other than the first of their table 1 can be retained in the machine and the difference recorded directly by use of negative multiplication. Two additional rows then should be added to each such section of table 1, the first of which contains the $k_i k_j$ and the second, the adjusted augmented moments. When this is done, the adjusted augmented moments are entered directly in the first row of each section other than the first in table 3 and the first section of table 3 is omitted. This procedure has been adopted as a standard one for any problem in our central computing unit.

Obtaining $M_{zy} M_{zz}^{-1} M_{yz}$.--In this step the matrices M_{zz} and M_{zy} are set up. The M_{zz} matrix includes the adjusted augmented moments of the predetermined variables $z_1, z_2, \ldots z_{15}$, and is, therefore, a 15 x 15 symmetric matrix. The M_{zy} matrix includes the adjusted augmented moments of the z's ($z_1, z_2, \ldots z_{15}$) on the y's in the seven equations to be fitted by limited information ($y_1, y_2, \ldots y_{12}$), and is a 15 x 12 matrix. The matrix product is obtained directly on SEAC, using standard routines for matrix inversion and multiplication. The latter can be carried out by making use of an A'BA routine, where A is the transpose of M_{zy} and B is the inverse of M_{zz}.

Additional steps in the limited information fit.--Values for the seven submatrices $M_{y*z} M_{zz}^{-1} M_{zy*}$ are obtained directly from the complete matrix $M_{yz} M_{zz}^{-1} M_{zy}$.

The remaining steps to obtain the structural coefficients and their standard errors in a limited information fit for each of the seven equations follow the procedure outlined by Friedman and Foote (40). With two exceptions, these computations are performed on desk calculators. Because equations (81.1) and (87.1) contain 6 predetermined variables each, it would be a laborious process to compute M^{-1}_{z*z*}, P', and $M_{y*z*}P'$ on a desk calculator. These values for equations (81.1) and (87.1) are, therefore, obtained on SEAC, using standard matrix routines. 49/

SPECIAL ECONOMETRIC TOPICS

In this section we consider points that would have led us too far astray from our main theme had we taken them up earlier.

Elasticity of Demand 50/

Most research analysts are familiar with the term, "elasticity of demand," which was popularized, if not invented, by Alfred Marshall. Elasticity of demand is the ratio of the percentage change in consumption of a commodity to the associated percentage change in its price. In mathematical notation it is written as

---

49/ By noting the matrix operations that are involved in getting $M_{y*z*}P'$, it will be seen that

$$M_{y*z*}P' = M_{y*z*}M^{-1}_{z*z*}M_{z*y*}.$$

If a matrix routine is available on the computer to get A'BA, then this triple product can be obtained directly once M^{-1}_{z*z*} has been obtained. However, if this is done,

$$P' = M^{-1}_{z*z*}M_{z*y*}$$

must be obtained as a separate operation, using an ordinary AB routine.

50/ Some of the material in this section is adapted from that in Waugh (100, pp. 56-57) and Foote and Fox (29, pp. 36-38). Similar material is given in many textbooks on economics.

$$\eta = \frac{dq}{dp} \cdot \frac{p}{q} \tag{92}$$

where dq/dp is the derivative of the demand curve at the point (p, q) when quantity is plotted on the vertical scale. The magnitude of dq/dp can be measured empirically by finding the change in quantity associated with a small change in price at the particular point on the demand curve.

Many economists have trouble with coefficients of elasticity. Figure 6 shows how this coefficient can be measured graphically. The curved line on the diagram represents an assumed demand curve for eggs. The scales for consumption and prices would not need to be shown. They are unimportant, because the coefficient of elasticity is invariant to changes in scale provided that the axes start at the origin. Suppose we want the coefficient of elasticity at the point (p=a, q=c). We draw the indicated straight line tangent to the demand curve at that point. The elasticity in question is -a/b, as shown in the following paragraph. For this example, this equals -35.5 divided by 64.5 based on the scales shown. In terms of small squares on the grid, this equals -17.75 divided by 32.25. Either computation indicates an elasticity of -0.55.

This piece of graphics comes from Marshall (67, pp. 102-103). It derives from the definition of elasticity given in equation (92). In this chart, dq/dp = - (c+d)/(a+b) and, by similar triangles, (c+d)/(a+b) = c/b. Also p = a and q = c. So

$$\frac{dq}{dp} \cdot \frac{p}{q} = \frac{-c}{b} \cdot \frac{a}{c} = -a/b. \tag{93}$$

Some economists have found the concept of elasticity so difficult that they have used "arc elasticity," or the "average elasticity of a curve." If the graphic approach to elasticity is used, there is little need for such concepts. The elasticity coefficient shown here is identical to the mathematical point elasticity and is easy to compute.

For most demand curves, including linear ones, the elasticity differs at every point on the curve. If a linear relation is given when both quantity and price are expressed in logarithms, then the elasticity is the same at every point on the curve and equals the slope of the curve. A proof of this can be derived as follows. For such a curve,

$$\log q = a + b \log p. \tag{94}$$

Applying standard principles of differentiation from calculus, we obtain

$$\frac{1}{q}\,\frac{dq}{dp} = \frac{b}{p} \tag{95}$$

or

$$\frac{dq}{dp} \cdot \frac{p}{q} = b. \tag{95.1}$$

A comparison with equation (92) shows that, for this curve, $b = \eta$.

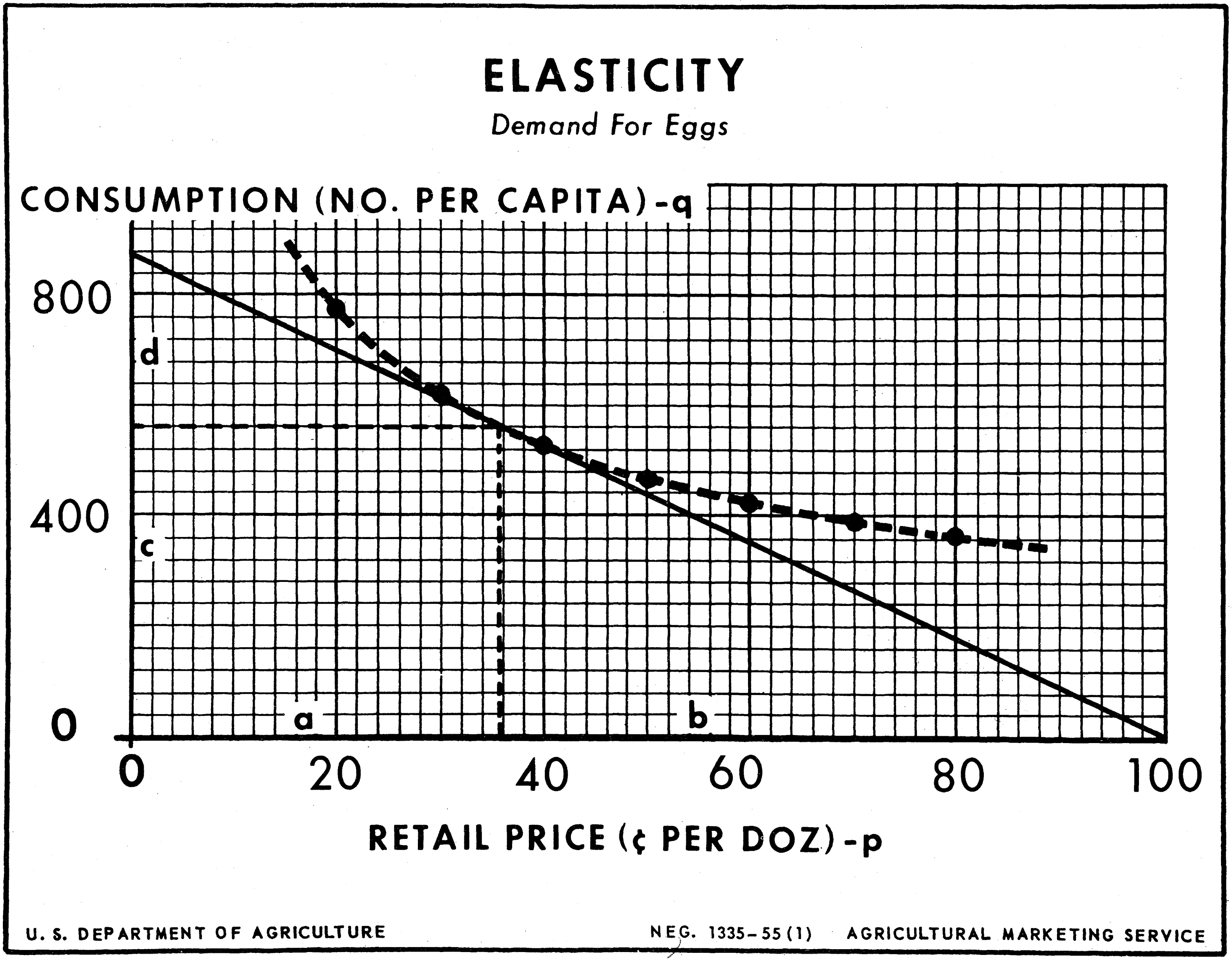


Figure 6.--The curved line in this diagram represents an assumed demand curve for eggs. A graphic method for measuring the elasticity of demand at a particular point on any demand curve is illustrated.

The cross elasticity of demand is defined as the percentage change in consumption of one commodity that is associated with a 1 percent change in the price of a second commodity. If the direct elasticity and the cross elasticity have the same sign, so that an increase in price of the second commodity results in a decrease in consumption of the first commodity, we say that they are complementary products. If the direct and cross elasticities are of opposite sign, so that an increase in price of the second commodity results in an increase in consumption of the first commodity, we say that they are competing products. If the direct and cross elasticities are of the same magnitude when the respective quantities and prices are expressed in comparable terms, we say that the two commodities are perfect complements or substitutes. If the cross elasticity equals zero, or nearly so, we assume that the two products are independent in demand. Special techniques for studying relations among complementary and substitute commodities are given in the section beginning on page 87.

If the elasticity, when defined as a positive number, is less than unity, so that a given percentage change in price results in a smaller percentage change in consumption, we say that demand is inelastic. For such commodities, as supply increases the total value tends to decrease. Demand for most farm products, at least at the local market level, tends to be inelastic. If the elasticity is larger than unity, so that a given percentage change in price results in a larger percentage change in consumption, we say that demand is elastic. For such commodities, as supply increases the total value tends to increase. If the elasticity equals unity, then the total value remains unchanged regardless of the quantity.

Henry L. Moore, a pioneer in the field of price analysis, was interested in the reciprocal of demand elasticity, which he called "price flexibility." Price flexibility is the ratio of a percentage change in the price of the commodity to the associated percentage change in its consumption. But the term is also applied to the ratio of percentage change in price to the associated percentage change in whatever supply variable is used in the analysis (consumption, production, or total supply). Obviously, there is no logical reason why the percentage relationship between price and production or price and supply should be the exact reciprocal of the ratio between changes in consumption and changes in price.

Reasons for different elasticities of demand.--The immediate object of a statistical demand analysis is the measurement of relationships rather than an explanation of the particular values obtained. But the analyst feels under pressure to rationalize the numerical results on either a commonsense or a theoretical basis. It is often said, for example, that the demand for an item that takes up an almost infinitesimal fraction of total income is likely to be highly inelastic. But this is not the relevant factor that tends to make for inelasticity; rather it is that the commodity has no close substitutes. A commodity such as potatoes or onions, on this score, tends to have a less elastic demand than a commodity such as beef, pork, or chicken which has several fairly close substitutes. Whether two commodities are effective economic substitutes depends on consumer attitudes toward them.

Problems involved in interpretation.--Many possible complications arise in interpreting coefficients that are intended as elasticities of demand. For example, moderate changes in relative prices might not cause measureable changes in relative consumption of two commodities. But sharp increases in the price of one or a continued wide price differential might lead to a substantial and possibly cumulative or irreversible shift from this commodity to the other. Although the immediate cause was physical shortage of butter rather than its high price, the change in relative consumption and in consumer attitudes toward butter and margarine during World War II is one of the most dramatic cases on record.

Some economists run an analysis in which price is made a function of the production or supply of a commodity. They refer to the reciprocal of this relationship as the elasticity of demand for the commodity. If the commodity has more than one outlet (including storage), the elasticity of demand in any one of these outlets need not equal the reciprocal of "price flexibility with respect to production." Changes in commercial stocks and in net exports of meat, for example, tend to cushion the effects of a change in production of meat upon its retail price. The reciprocal of the price-production relationship suggests a unit "elasticity of demand" for meat. But the regression of consumption of meat upon its retail price gives an elasticity of demand of around -0.6. Year-to-year changes in consumption normally are highly correlated with changes in production, but during 1921-41 they were only 70 percent as large. Hence the elasticity of consumer demand is only 70 percent as large as the reciprocal of the price-production flexibility.

An elasticity of demand for "meat to store" and one for exports of meat from this country could be calculated. An elasticity of supply for imports of meat (perhaps a separate one for each major country from which meat is imported) also would be involved. Ordinarily, elasticity of demand means the elasticity of domestic consumption with respect to price. It is likely that the elasticities of demand for exports of meat and meat to store are greater than the elasticity of demand for consumption.

The three major categories of utilization for wheat have very different demand curves and elasticities. For example, the price elasticity of demand for wheat for domestic food use is less than -0.1 (see table 5). The demand for wheat as a livestock feed is inelastic as long as the price of wheat is considerably above the price of feed grains. But if the price of wheat falls to or slightly below the price of corn on a pound-for-pound basis, the use of wheat for feed is likely to increase tremendously. In other words, the demand for wheat in the range of (say) 20 cents a bushel below the price of corn to 5 or 10 cents above is highly elastic. The elasticity of demand for exports of wheat has varied in the last 30 years. In the late 1920's, when exports of wheat from this country amounted to about a sixth of total world exports and exports to Europe amounted to 5 to 7 percent of European production, the elasticity of demand for our wheat exports may have been substantially greater than one. Dollar rationing as such--that is, the setting aside by other countries of a specific dollar amount to be spent for our wheat--would imply a unit elasticity of demand with respect to price. An elasticity of demand for wheat for export under conditions of extensive import and exchange regulations may have little meaning.

The implication of this is that research workers should indicate specifically the particular utilization, or set of utilizations, to which a given coefficient of elasticity refers. The ratio of a percentage change in total utilization of wheat to a percentage change in the price of wheat should be a weighted average of the elasticity in each utilization group.

A further problem in interpretation is that, except for double logarithmic curves, the elasticity differs at every point on the curve and, in a statistical analysis, also differs, depending on the particular values assumed by other variables in the analysis. This problem frequently is ignored, in a sense, by computing the elasticities when all variables are at their average values. But, in comparing results from one analysis with those from another, this practice is undesirable because the averages depend on the particular years on which the analysis is based. More reliable comparisons could be obtained from elasticities derived for a uniform year or period of years. In such computations, use should be made of calculated values for the dependent variable rather than actual values.

Table 5, adapted from Meinken (71, p. 43), shows the effect on several elasticity and price flexibility coefficients of using values of the variables for specified time periods. The analysis was based on data for 1921-29 and 1931-38, and results of the elasticity computations are shown based on averages for these years and also for 1931 and 1953, respectively, to contrast levels in years of depression and prosperity. Publication of tables of this sort to illustrate the range in elasticities that are suggested by an analysis under specified conditions appears desirable.

The value of statistical regression coefficients depends upon the extent to which they improve our ability to act intelligently and appropriately in specific situations. An application that is frequently encountered is the question of whether compensatory payments or purchase and diversion programs would be less costly in supporting the price of perishable commodities. So far as cost to the public treasury is concerned, the answer to this question turns largely on the elasticity of demand for the commodity. If the elasticity of demand is unity, this tends to make the costs to Government of the two methods of price support identical. If demand is highly inelastic, purchase and diversion is less expensive to the Government; if demand is more than unit elastic, compensatory payments presumably are less expensive to Government as well as more satisfactory to consumers. The reasons are discussed in detail by Fox (32). The accuracy of statistical regression coefficients and their validity in the particular context--time, place, and duration of time with which a projected program is concerned--bear upon the quality of administrative actions and the overall effectiveness of Government programs.

Table 5.--Wheat: Coefficients obtained by a simultaneous-equations approach expressed in percentage terms based on values of the variables for specified periods

Coefficient	Determined at values for years beginning July-- Average 1921-29 and 1931-38	1931	1953
Demand elasticity:			
Use for food with respect to Kansas City price	-0.04	-0.02	-0.11
Use for net export with respect to spread between world and Kansas City price	.18	.23	.39
Use for feed with respect to spread between price of wheat and corn	-.40	-.02	-1.75
Use for storage with respect to deflated Kansas City price of wheat	-2.73	-.75	-4.76
Income elasticity of use for food	.20	.16	.76
Price flexibility (world price) with respect to--			
World supply	-1.43	-2.48	-.74
World price level	1.18	1.52	1.48

Adapted from Meinken (71, p. 43).

Problems of Aggregation 51/

Most applied economists in the supply, demand, and prices area are interested chiefly in the behavior of market aggregates. This interest stems from their desire: (1) to forecast such economic variables as prices, production, and utilization, and (2) to analyze probable effects on these variables of changes in institutional factors such as price supports, freight rates, tariffs, import quotas, and so forth.

51/ This section is based chiefly on an unpublished paper by Richard J. Foote and Marc Nerlove, Agricultural Marketing Service, entitled "Why Applied Research Workers are Interested in Problems of Aggregation." These problems are discussed in detail by Theil (92) and, more briefly, Allen (2, pp. 694-724).

As is well known, however, economic theory with respect to supply and demand is couched largely in terms of the individual. More important, we have a great mass of data, some aggregative, some not. Hypotheses consistent with most or all of our data may or may not be useful for forecasting or for assessing the effects of alternative policies; but there is surely some presumption that such hypotheses should be better for both purposes than those consistent with only a small part of the available data. Thus, applied economists are interested in a body of theory that, in many specific situations, will help to develop hypotheses, both qualitative and quantitative, consistent with available data that relate both to aggregates and individuals, or relatively small groups, and in methods by which these hypotheses can be tested against available data.

Although much remains to be done in this area, some progress has been made. In this section, we mention briefly some research studies that have been published. Our discussion can be shortened by using the terms "microtheory," "microvariables," "aggregates," and "macrotheory." Microtheory is that body of economic theory that is concerned with the behavior of individuals, either consumers or production or marketing firms. Microvariables are those that relate to such individuals. Aggregates, as their name implies, are totals or index numbers of the microvariables. Macrotheory deals with economic relationships that should hold between aggregative variables.

The consistency approach.--Consider the traditional trichotomy: microtheory, aggregates, macrotheory. The consistency approach may be characterized as follows: Given any two of the foregoing, determine the third in such a way that it is consistent with the other two. Thus the consistency approach has three variants: (1) Given the microtheory and certain properties of the macrotheory, construct aggregates of the microvariables consistent with the microtheory and the given properties of the macrotheory. This approach was taken by Klein (56). (2) Given the microtheory and certain aggregates of the microvariables, find a macrotheory which holds between the aggregates and which is consistent with the microtheory. This approach was taken by May (68, 69). (3) Given a macrotheory which holds between certain aggregates, find which microtheories are consistent with the macrotheory and under what assumptions we have consistency. This approach was taken by Klein (57) in a second study in this area.

Theil (92, p. 5) gives a concrete example that relates to studies of this sort. He says, "Suppose, for instance, that a microtheory tells us that each family's consumption is an increasing function of its income. Suppose also that all family incomes move up and down simultaneously. Then total consumption must increase whenever total personal income does, so that a macrotheory in terms of these aggregates which tells us that total consumption is a decreasing function of total income is certainly not consistent with the microtheory."

The analogy approach.--Holte 52/ and Theil (92) are included in what we call the analogy approach.

Theil's method may be characterized as follows: Suppose we are given a microtheory and some simple aggregates of the microvariables. We assume a macromodel, analogous to the micromodel, which relates the aggregates. Now we estimate the macromodel statistically and ask what the meaning of the estimated parameters is in terms of the parameters of the microtheory. In particular we ask how the estimated parameters differ from simple combinations of the microparameters, such as sums or averages. Allen (2, p. 709) summarizes some of Theil's results as follows:

"In the simple case where individual demands for a commodity are aggregated into a single relation between total demand and aggregate income, the results can be expressed and interpreted:

"(i) an exact total demand relation is obtained by aggregating incomes with weights proportional to individual marginal propensities to consume (or individual income elasticities);

"(ii) with simple aggregation, the total demand relation is determined statistically and its parameters generally depend on the movement of incomes over the time period considered;

"(iii) with simple aggregation, again, the only situation in which there is never contradiction in prediction between individual and total demand relations is when all individual marginal propensities to consume are equal;

"(iv) the only system of fixed weighting in aggregate income which leads to no contradiction in prediction is that where weights are proportional to individual marginal propensities to consume, as in (i)."

Holte's approach is similar to Theil's in aim, although it differs in method. In describing his approach, Holte 53/ writes "Another approach is to specify a micromodel, postulate a macromodel and state that the macroparameters are such simple functions of the microparameters as implicitly assumed by many economists. We will then generally get a macromodel which is wrong (if the micromodel is assumed to be true). By investigating what determines 'the degree of wrongness' in the macromodel we may perhaps obtain knowledge which makes it possible to describe some types of economic situations in which the analogy approach is a sound one, and other types in which it is dangerous."

The formal approach.--The "formal" approach, given that title for lack of a better one, is characterized as follows: Given a microtheory and some distributional assumptions concerning some or all of the relevant variables which

---

52/ Holte, Fritz C. A New Approach to the Aggregation Problem. Cowles Foundation Discussion Paper No. 21. 1956. 26 pp. (Processed.)

53/ Ibid., p. 2.

serve to differentiate the individual units, derive both the appropriate aggregates and the macrotheory. Although this approach is an old one, we mention only a few recent examples: (1) Tobin (93) used the assumption of a random distribution of tastes to derive aggregate demand functions from individual demand functions under rationing. (2) Farrell (21, 22) followed Tobin's lead and applied the results to combine cross-section data on incomes and ownership of automobiles with time series data on income, prices, and stocks of automobiles of various ages. While Farrell's empirical results are not precise, his approach shows clearly the relation between problems of aggregation and the construction of hypotheses consistent with different types of data. (3) In an interesting paper, Houthakker (51) showed how to aggregate the technological possibilities of individual producing units into a production function for a group of units such as an industry. Houthakker showed that, if we assume a linear programming model for each producing unit and a Pareto distribution of fixed resources among producing units, then we can derive a Cobb-Douglas production function for the industry as a whole. Houthakker's approach is especially interesting, because it points out a way in which technological and engineering data may be combined with time-series or cross-section data on inputs and outputs. Its major defect is that it deals only with firms and industries producing a single product. (4) Friedman (41) also used the "formal" approach. He divides consumption and income into two components, permanent and transitory. He shows that the relation between consumption and income in a cross-section depends on certain characteristics of the joint distribution of permanent and transitory components of consumption and income among individuals and on the macro- and micro-relations between the permanent components. Similarly the relation between consumption and income over time depends on certain characteristics of the joint distribution among time periods. Although Friedman does not explicitly investigate the direct combination of cross-section and time-series data, his approach lends itself to such investigations.

Each of these approaches to problems of aggregation undoubtedly has something to offer the applied research worker, although the "formal" approach may prove the most fruitful of the three in generating hypotheses that are consistent with several types of data. Further research by econometricians appears to be needed before practical techniques for applied research workers become available. Future developments in this area should be watched closely.

Equation Systems for Competing Commodities

Meinken, Rojko, and King (72, pp. 711-712) point out, "Since the days of Walras, at least, economists have recognized the importance of substitution in demand. Theoretically, the consumption of a commodity depends not only upon its own price, but also upon prices of all other commodities. The Walrasian formulation of the problem is too unwieldy for purposes of statistical measurement. Many statisticians, however, have measured the interrelationships among the demands for two or three commodities--neglecting the minor effects of other commodities and services. Thus, they have measured competition between beef and pork, without worrying too much about the undoubted fact that beef consumption may be slightly affected by the price of gasoline, shoes, and movie tickets.

"Most of the statistical work done on this problem has been based either upon demand and cross elasticities, or upon the relation of consumption ratios to price ratios. Recently Waugh (99) has derived a partial indifference surface from market data. ... Since the demand functions for two commodities, the relation of the consumption ratios to the price ratios, and the indifference function all involve the same variables--prices, quantities, and income--and provide measures that are designed to indicate the competitive relationship between two commodities, it appears, on an intuitive basis at least, that it would be possible by appropriate mathematical transformations to go directly from any given approach to the other two. We shall, in fact, show that it is possible to go directly from the demand functions to the ratios, or to the indifference surface, but it is not possible to go the other way."

These authors next consider in detail three approaches that have been commonly used to study relationships among competing products. They show that one of these typically has been used in an incorrect way and that a second frequently yields misleading conclusions.

They say, "Demand theory conventionally specifies that, for an individual consumer, the quantity of beef or pork consumed depends upon prices of beef, pork, and all other commodities, the individual consumer's income, and factors that reflect changes in tastes and preferences." If time series data that relate to purchases by individual consumers are available, a least squares regression of consumption of either product on their respective prices and other relevant variables provides coefficients that can be used to derive estimates of the direct and cross elasticities that are statistically consistent.

These authors then say, "Market demand, which is the summation of these individual demands, may be defined as follows:

$$Q_b = f(P_b, P_p, Y, u_1) \tag{96}$$

$$Q_p = f(P_b, P_p, Y, u_2) \tag{97}$$

where the Q's represent the aggregate consumption of beef (Q_b) and pork (Q_p), the P's represent market prices of beef (P_b) and pork (P_p); Y represents aggregate consumer income; and the u's represent random disturbances that affect consumption of beef and pork. As no separate allowance is made for substitute commodities such as other meats and fish, the u's also include the effect of changes in the price or supply of these. Other meats and fish are believed to be relatively unimportant in affecting the quantity of beef and pork consumed.

"If time series data on prices, quantities and incomes are given, the method used to estimate the coefficients in these demand relations depends on assumptions that are made regarding the type of functional relation that generates the observed data. For beef and pork, as for many other agricultural commodities, production and consumption in any given period are essentially predetermined, that is, the supply curve is completely or almost completely inelastic. ... When consumption of two competing agricultural commodities can be assumed for all practical purposes to be independent of prices in the current period, the procedure normally has been to estimate the coefficients for the relations that express the price of each good as a function of the two quantities and income. This procedure gives (statistically) biased results when applied to two or more competing commodities because ... prices for each commodity are simultaneously determined by the interaction of demand factors and the supply of each. Thus a given combination of production of beef and pork results in a unique set of market prices that is simultaneously determined. To obtain estimates of the elasticities of demand that are statistically consistent, the parameters in the structural demand equations (96) and (97) must be estimated by a statistical method that allows for this simultaneity. Equations of the sort discussed here always are just identified. Hence, the reduced-form method of fitting simultaneous equations can be used to estimate the coefficients" (72, pp. 713-715). This method is discussed in detail in the section beginning at the bottom of this page.

With respect to the third method, these authors (72, p. 717, 726, 732-733) say, "Various research workers have derived statistical estimates of the elasticity of substitution (E_s) by a short-cut method of relating price ratios and consumption ratios. ... Two conclusions with respect to the elasticity of substitution can be drawn. First, an empirical estimate of the true E_s can be obtained by relating price ratios and consumption ratios only under very restrictive conditions, ... When price ratios and quantity ratios are related, a poor fit may be indicated for direct substitutes and a good fit for independent commodities. Moreover, the regression coefficient will not necessarily tell us whether goods are competing, independent, or complementary, let alone the 'ease of substitution.' Second, it is demonstrated that even knowing that two goods are competing, the numerical value of E_s $[0 - \infty]$ tells us little about the 'ease of substitution' or the degree of competitiveness between the goods, its designed purpose. This follows since the measure E_s is a combination of direct price elasticities and cross price elasticities and income elasticities." Hence they conclude "that the statistical estimation of E_s is of dubious value, and its estimation by the price-ratio consumption-ratio method, with no other check as to the nature of the demand interrelationship, is meaningless."

Estimation of direct and cross elasticities from market data when consumption of each commodity can be taken as a predetermined variable.--In this section, we consider methods that can be used to estimate direct and cross elasticities of demand, as commonly defined in economic literature, for three commodities that are believed to be competing, when consumption of each is believed to be a predetermined variable (see page 47). The use of three commodities, rather than two as done frequently in the past, illustrates certain aspects of the method, namely, that of providing an analytical framework which readily can be extended to more competing commodities. To simplify the notation and equations, we assume that all variables that depend in part on (1) the size of the population are expressed in per capita terms and (2) the general price level are deflated. Consumer income is considered as a predetermined variable and consumption of each item is assumed to be a linear function of retail prices of the several competing items and consumer income. Most of these assumptions must be made implicitly in order for equations (96) and (97) to be consistent with the general principles developed so far in this handbook.

We thus imply that our model consists of the following structural equations. The subscripts a, b, and c are used respectively for the three commodities. The variables are assumed to be expressed as deviations from their respective means so that the constant terms can be omitted.

$$q_a = b_{11}p_a + b_{12}p_b + b_{13}p_c + b_{14}y \qquad (98)$$

$$q_b = b_{21}p_a + b_{22}p_b + b_{23}p_c + b_{24}y \qquad (99)$$

$$q_c = b_{31}p_a + b_{32}p_b + b_{33}p_c + b_{34}y. \qquad (100)$$

It is now clear why such equations are often just identified. To be just identified, the number of variables in the system minus the number of variables in each equation must equal the number of endogenous variables in the system minus one. Stated another way, the number of variables omitted from each equation must equal the number of endogenous variables less one. If we have n commodities and hence n endogenous prices, we always omit consumption (a predetermined variable) of n-1 commodities from each equation; hence each equation is just identified. This condition holds only if each predetermined variable (any number) in the system, but excluding the consumption variables, appears in each equation. For example, equations (98) to (100) remain just identified if we allow for the nonlinear per capita and deflated variables by the method described in the section beginning on page 71, as we add two predetermined variables to the system but incorporate each into each equation, so that the counting rule is unaffected.

However, if any variable besides prices of the n commodities is assumed to be endogenous, as for example consumer income, the system of equations as given is incomplete as there are more endogenous variables than equations. If consumer income is the additional endogenous variable, balance may be restored by adding an income equation which includes special factors that explain income behavior and that can be treated as predetermined variables. The demand equations then are overidentified unless only one special factor is used to explain income, and the income equation always is overidentified. In the remainder of this discussion, we consider all variables other than price to be predetermined.

As discussed in the section beginning on page 62, in the method of reduced forms each endogenous variable is expressed as a function of all of the predetermined variables in the system. These equations are known as reduced form equations; estimates of the coefficients that are statistically consistent and efficient can be obtained if they are fitted directly by least squares as each contains only a single endogenous variable. In this instance, the three reduced form equations are as follows:

$$p_a = B_{11}q_a + B_{12}q_b + B_{13}q_c + B_{14}y \tag{101}$$

$$p_b = B_{21}q_a + B_{22}q_b + B_{23}q_c + B_{24}y \tag{102}$$

$$p_c = B_{31}q_a + B_{32}q_b + B_{33}q_c + B_{34}y. \tag{103}$$

By the same kind of algebraic manipulations as shown on page 62 the relations between the coefficients in the reduced form and the structural equations can be derived. Such relations are given by Meinken, Rojko, and King (72, p. 734) for a system of equations for two commodities. They point out that the reciprocal of the "price flexibility" (see page 81) for beef equals the direct price elasticity for beef only if the cross elasticities of beef on pork and pork on beef each are zero. They say further, "The algebraic relation of the coefficients indicates clearly that the reciprocal of the cross price flexibility does not give the cross elasticity of demand. The same holds for the income coefficient."

Algebraic relationships of this kind become complex when we deal with more than two commodities, but the relationships can be stated simply by use of a matrix notation. It is suggested that readers who are unacquainted with this notation skip the next three paragraphs. Meinken, Rojko, and King introduce time as a variable into each of their equations, and in the example given here we do likewise so as to include two predetermined variables in each equation.

In order to use a matrix notation, it is convenient to write our equations making use of somewhat different symbols. The reduced form equations with which we deal in this section are as follows:

$$p_1 = B_{11}q_1 + B_{12}q_2 + B_{13}q_3 + C_{11}y + C_{12}t \tag{104}$$

$$p_2 = B_{21}q_1 + B_{22}q_2 + B_{23}q_3 + C_{21}y + C_{22}t \tag{105}$$

$$p_3 = B_{31}q_1 + B_{32}q_2 + B_{33}q_3 + C_{31}y + C_{32}t. \tag{106}$$

In connection with systems of equations, the predetermined variables frequently are designated by the letter z. Here we wish to treat the predetermined variables that relate to consumption in one way and the other predetermined variables in another way. So we include y and t in a vector (Z), the three quantities in a vector (Q), and the three prices in a vector (P). Primes are used on the vectors in equations (110) and (110.1) to indicate that they are to be written as column vectors. We also work with a matrix (B) of the B_{ij} coefficients and a matrix (C) of the C_{ij} coefficients.

Each of the reduced form equations is fitted by least squares, using the single endogenous variable as dependent. After obtaining estimates of the B_{ij} and C_{ij} coefficients, we wish to transform these algebraically into the structural coefficients--the b_{ij} and c_{ij}--that are used in the following structural equations:

$$q_1 = b_{11}p_1 + b_{12}p_2 + b_{13}p_3 + c_{11}y + c_{12}t \qquad (107)$$

$$q_2 = b_{21}p_1 + b_{22}p_2 + b_{23}p_3 + c_{21}y + c_{22}t \qquad (108)$$

$$q_3 = b_{31}p_1 + b_{32}p_2 + b_{33}p_3 + c_{31}y + c_{32}t. \qquad (109)$$

Equations (104) to (106), treated as a unit, can be written in the following matrix notation:

$$P' = BQ' + CZ'. \qquad (110)$$

If we multiply each term of equation (110) by B^{-1} and rearrange the resultant terms, we obtain:

$$Q' = B^{-1}P' - B^{-1}CZ'. \qquad (110.1)$$

In this form, we have a matrix equation that is equivalent to our structural equations, where the b_{ij} coefficients are found in matrix B^{-1} and the c_{ij} coefficients in the product matrix $- B^{-1}C$. By carrying out the matrix operations indicated, we obtain the structural coefficients. The matrix equation applies directly to any number of variables.

We now summarize the above for readers who are not acquainted with matrix notation although we shall, in fact, make use of such notation. Derivation of the structural coefficients involves the following steps:

(1) We fit the reduced form equations (104) to (106) by the method of least squares. In doing so, advantage should be taken of the fact that the same independent or predetermined variables are used in each equation [see Friedman and Foote (40, pp. 69-76)].

(2) We take the B coefficients obtained in this way and write them in the following form:

$$\begin{bmatrix} B_{11} & B_{12} & B_{13} \\ B_{21} & B_{22} & B_{23} \\ B_{31} & B_{32} & B_{33} \end{bmatrix}$$

This is in effect a matrix and must be inverted according to standard rules for a nonsymmetric matrix. Any mathematician and most statisticians know how to do this, and programs for inverting matrices are available for any electronic computer. 54/ The answer is in the same form as this, but different numbers appear in the cells.

(3) The b coefficients for the structural equations (107) to (109) are found in the following locations within the inverse of the matrix of B coefficients:

$$\begin{bmatrix} b_{11} & b_{12} & b_{13} \\ b_{21} & b_{22} & b_{23} \\ b_{31} & b_{32} & b_{33} \end{bmatrix}$$

(4) We form a matrix of the C coefficients obtained by least squares of the following form:

$$\begin{bmatrix} C_{11} & C_{12} \\ C_{21} & C_{22} \\ C_{31} & C_{32} \end{bmatrix}$$

and premultiply it by the inverse of the matrix of the B coefficients. This again is an operation that can be performed by any mathematician and most statisticians or can be done electronically. 55/ Upon completion of this operation, a new set of numbers are formed, within which the c coefficients for the structural equations are found at the following locations:

$$\begin{bmatrix} c_{11} & c_{12} \\ c_{21} & c_{22} \\ c_{31} & c_{32} \end{bmatrix}$$

By making use of these simple principles and calling upon the aid of mathematicians to perform what for them are simple mechanical operations, we can go from the coefficients of the reduced form equations to the coefficients of the structural equations with greater ease than by carrying out algebraic manipulations of the sort described on page 62 even for two competing commodities, and we can handle as many commodities as are involved in the analysis.

54/ An outline for carrying out such computations on desk calculators is given by Friedman and Foote (40, pp. 98-100), and elsewhere.

55/ Matrix multiplication on desk calculators is described by Friedman and Foote (40, pp. 23-26), and elsewhere.

Unfortunately, it is difficult to obtain standard errors for the structural coefficients even for two competing commodities, although methods to derive these standard errors are available. 56/

Estimation of direct and cross elasticities from market data when consumption of one or more commodities is at least partially endogenous.--When we drop the assumption that consumption is a predetermined variable, we face a variety of equation systems. However, as some of the equations normally are overidentified, the aspects of simultaneity that cause complications when we fit by the method of reduced forms are taken care of automatically, so that the derivation of direct and cross elasticities from such systems frequently is a simple matter. This point can be illustrated by reference to two equation systems referred to previously.

In the model for dairy products shown on page 13, three equations involve cross elasticities, namely equation (14) for butter, (15) for cheese, and (17) for margarine. In the economic model, two endogenous variables are shown on the left of the semicolon in each case, namely the quantity of the product in question and the price of either that commodity or its competitor. In writing down the equations in the statistical model, however, the quantity of the product normally is written to the left of the equality sign in systems of this type, and the direct and cross elasticities can be obtained directly from the coefficients shown. Standard errors for the structural coefficients are given directly by the limited information method. A similar situation prevails with respect to the model for asparagus shown on page 16. Although four endogenous variables are shown to the left of the semicolon, only a single endogenous quantity is involved and this normally is shown on the left of the equality sign in the statistical model. Again, the direct and cross elasticities can be obtained directly from the structural coefficients when the equations are written in this way.

The author, in cooperation with Olman Hee, Anthony S. Rojko, and Will Simmons of the Agricultural Marketing Service, recently developed a model for potatoes during late winter and spring that illustrates the kind of algebraic manipulations that are required when consumption is at least partially endogenous and when the structural equations contain more than a single quantity variable. As a unique set of manipulations in general is required for each such model, only the general approach is outlined here. In this model, we consider the demand for old and new potatoes separately in each of two periods --late winter (January-April) and spring (May-June). Price relationships between new crop potatoes in the two periods also are measured. The supply of old crop potatoes to be consumed during both periods is determined by stocks on January 1, but the quantity used in each period is determined simultaneously by the economic variables that enter into the model. Production of new

---

56/ The general approach is described by Klein (58, pp. 258-259), but certain partial derivatives must be obtained which for these equations are algebraically complex.

crop potatoes within each period is assumed to be a predetermined variable. Demand for new potatoes in period II is assumed to depend in part on the quantity of new potatoes available in period I.

The following variables enter into the economic model. As in previous models discussed in this section, we assume that quantitative variables are expressed in per capita terms and that variables affected by the price level are deflated. As prices at retail are not available in the detail required, a marketing margin variable is included in each of the demand equations (see pages 24-26).

P_{O_I} - Local market price of old potatoes in period I

$P_{O_{II}}$ - Local market price of old potatoes in period II

P_{N_I} - Local market price of new potatoes in period I

$P_{N_{II}}$ - Local market price of new potatoes in period II

S_I - Stocks of old potatoes used for consumption in period I

S_{II} - Stocks of old potatoes used for consumption in period II

S - Stocks of old potatoes on hand January 1

Q_I - Production of new potatoes in period I
Q_{II} - Production of new potatoes in period II
} In each case, most of these are used directly for food

Y_I - Disposable consumer income in period I

Y_{II} - Disposable consumer income in period II

M - Factors that affect marketing costs.

The first 6 variables are assumed to be endogenous.

The following equations are involved in the economic model:

$S_I, P_{O_I}, P_{N_I} ; Y_I, M$	(Demand for old potatoes in period I)	(111)
$P_{O_I}, P_{N_I} ; Q_I, Y_I, M$	(Demand for new potatoes in period I)	(112)
$S_{II}, P_{O_{II}}, P_{N_{II}} ; Y_{II}, M$	(Demand for old potatoes in period II)	(113)
$P_{O_{II}}, P_{N_{II}} ; Q_{II}, Q_I, Y_{II}, M$	(Demand for new potatoes in period II)	(114)

$P_{N_I}, P_{N_{II}} ; Q_I, Q_{II}$ (Normal relation between prices of new potatoes in periods I and II) (115)

$S_I + S_{II} = S.$ (Identity with respect to stocks) (116)

Equation (114) contains more than one variable that relates to quantity, whereas, to determine direct and cross elasticities, as normally defined in economic literature, we need a single quantitative variable as a function of the several prices and other variables used in the model. An equation in the needed form can be obtained by the following algebraic operation:

Substitute equation (115) into equation (114) to eliminate Q_I and get

$$Q_{II} = f(P_{O_{II}}, P_{N_I}, P_{N_{II}}, Y_{II}, M). \tag{117}$$

Any person who has a good grasp of college algebra can make substitutions of this sort.

Standard errors of the coefficients involved in the modified equation can be determined by the method given by Klein (58, pp. 258-259). The amount of work involved in using this method depends on the nature of the algebraic transformations. In some cases, the task is complex; for this equation, the task is easy, given a knowledge of calculus.

A partial indifference surface.--In this section, we present a method developed by Waugh (99) to derive a partial indifference surface from market data, although the discussion follows the summary of his approach given by Meinken, Rojko, and King (72, pp. 726-732). They say, "The partial indifference function used by Waugh is

$$Z = A \frac{[1 - (c - g)]}{[1 + (b - f)]} Q_b^{1+(b-f)} + Q_p^{1-(c-g)} \tag{118}$$

where Z represents the elevation of a surface. For each value of Z, there exists a contour line (partial indifference curve) that represents substitution possibilities of beef and pork for which consumers are indifferent. In contrast to a production function, in which Z can be taken as a direct measure of output, the Z in (118) is not intended to give a direct measure of the level of utility. Instead, Z may be considered a monotonic [or steadily increasing or decreasing] function of utility from which it is possible to obtain contours (indifference curves) for each value of Z."

If P_r equals the ratio of beef prices to pork prices at retail, this partial indifference function implies that

$$P_r = AQ_b^{(b-f)}Q_p^{(c-g)} \quad 57/ \tag{119}$$

57/ The reasoning involved is given by Meinken, Rojko, and King (72, p. 727) and is summarized here. If Z is a single valued, monotonic function of utility, say $Z = \phi(U)$, the marginal rate of substitution (or the slope of a

In logarithmic form, this equals

$$\log P_r = \log A + (b - f) \log Q_b + (c - g) \log Q_p. \qquad (119.1)$$

Estimates of the coefficients in equation (118) can be obtained algebraically from the coefficients in equation (119) or (119.1) which, in turn, can be derived from an equation, fitted by least squares, of the form

$$\log P_r = \log {}^{a}/_{e} + (b - f) \log Q_b + (c - g) \log Q_p + (d - h) \log Y, \qquad (120)$$

which includes income as a variable. For any given level of income, if we substitute the value of Y in equation (120) and transform to the original non-logarithmic variables, we obtain an equation like (119), from which we can obtain the coefficients in equation (118). In so doing, the value (d - h) log Y must be combined with log ${}^{a}/_{e}$ to get log A from which the value of A is derived.

Waugh's analysis implies that for each level of money income, there is a definite indifference surface for beef and pork. This means that the constant (A) in equation (119) may vary with income. This also is true for the constant in equation (118). In a sense, equation (118) is that of a three-dimensional surface, and may be thought of as a cross section of a four-dimensional surface, holding income constant at some specified level.

Meinken, Rojko, and King then apply this method, with some modification, to their data on Canadian beef and pork consumption and prices, and state, "Figure 7 shows, for each year, the contour line (partial curve) associated with the quantities of beef and pork consumed had per capita real income been

---

contour line) is

$$\frac{\partial Z}{\partial Q_b} \Big/ \frac{\partial Z}{\partial Q_p} = \frac{\partial[\phi(U)]}{\partial Q_b} \Big/ \frac{\partial[\phi(U)]}{\partial Q_p}$$

where $\partial[\phi(U)]/\partial Q_b$ equals the marginal utility of beef and $\partial[\phi(U)]/\partial Q_p$ equals the marginal utility of pork. In competitive equilibrium, the marginal rate of substitution between beef and pork must equal the ratio of their prices or

$$\frac{\partial Z}{\partial Q_b} \Big/ \frac{\partial Z}{\partial Q_p} = P_r.$$

But the slope of the contours of Z in equation (118) at any point is

$$\frac{\partial Z/Q_b}{\partial Z/Q_p} = AQ_b{}^{(b-f)}Q_p{}^{(c-g)},$$

which also equals the price ratio, that is,

$$P_r = AQ_b{}^{(b-f)}Q_p{}^{(c-g)}.$$

\$648.53 in that year. ... Figure 7 also shows the adjusted price ratio (P_r') associated with the quantities of beef and pork consumed during that year had per capita real income remained constant at \$648.53. ... If conditions of competitive equilibrium are met and if the contour function represents the true substitution possibilities between beef and pork, we would expect the adjusted price ratio to be tangent to the contour line for the particular quantities of beef and pork consumed for that year. 58/ Inspection of figure 7 tends to substantiate this formulation." A chart of this sort also is presented in the original article by Waugh (99).

Measuring Relationships Among Complementary Products 59/

Some commodities, like edible fats and oils and sugar, are used almost exclusively as ingredients in combinations of foods in which they often account for a relatively small part of the total cost. At any given time, consumption of such commodities tends to be related to consumption of other foods by a set of technical coefficients--number of teaspoonfuls of sugar per cup of coffee, ounces of butter-plus-margarine per pound of bread, ounces of salad oil per unit of salad vegetables, and so forth. If the demand for sugar-using foods increases with consumer income, the demand for sugar also will appear to increase with income, even though no consumer actively values sugar for its own sake.

Consider the following two demand equations for sugar:

$$q_s = a_1 + b_1 p_s + c_1 y \tag{121}$$

and

$$q_s = a_2 + b_2 p_s + c_2 y + d \sum_{i=1}^{m} w_i q_i , \tag{122}$$

58/ The shape of the indifference curve depends on the degree of substitution between a pair of commodities and becomes a straight line for goods that are perfect substitutes. The shape of the empirical indifference curve presented here depends directly on a particular combination of elasticities of demand for beef and pork. As implied in the discussion of the elasticity of substitution, results obtained from a method of estimation that is based on some combination of direct and cross price elasticities do not provide sufficient information (1) to conclude that a pair of commodities are substitutes and (2) to determine the precise degree of substitution when they are known to be substitutes. Thus, the empirical indifference curve given here also is subject to these limitations. We know that beef and pork are substitutes, but the relative flatness of the curve does not tell us the precise degree of substitution between these commodities.

59/ This material was developed initially by Fox and was first published by Foote and Fox (29, p. 18). A brief discussion along similar lines is given by Fox (33, pp. 16-17).

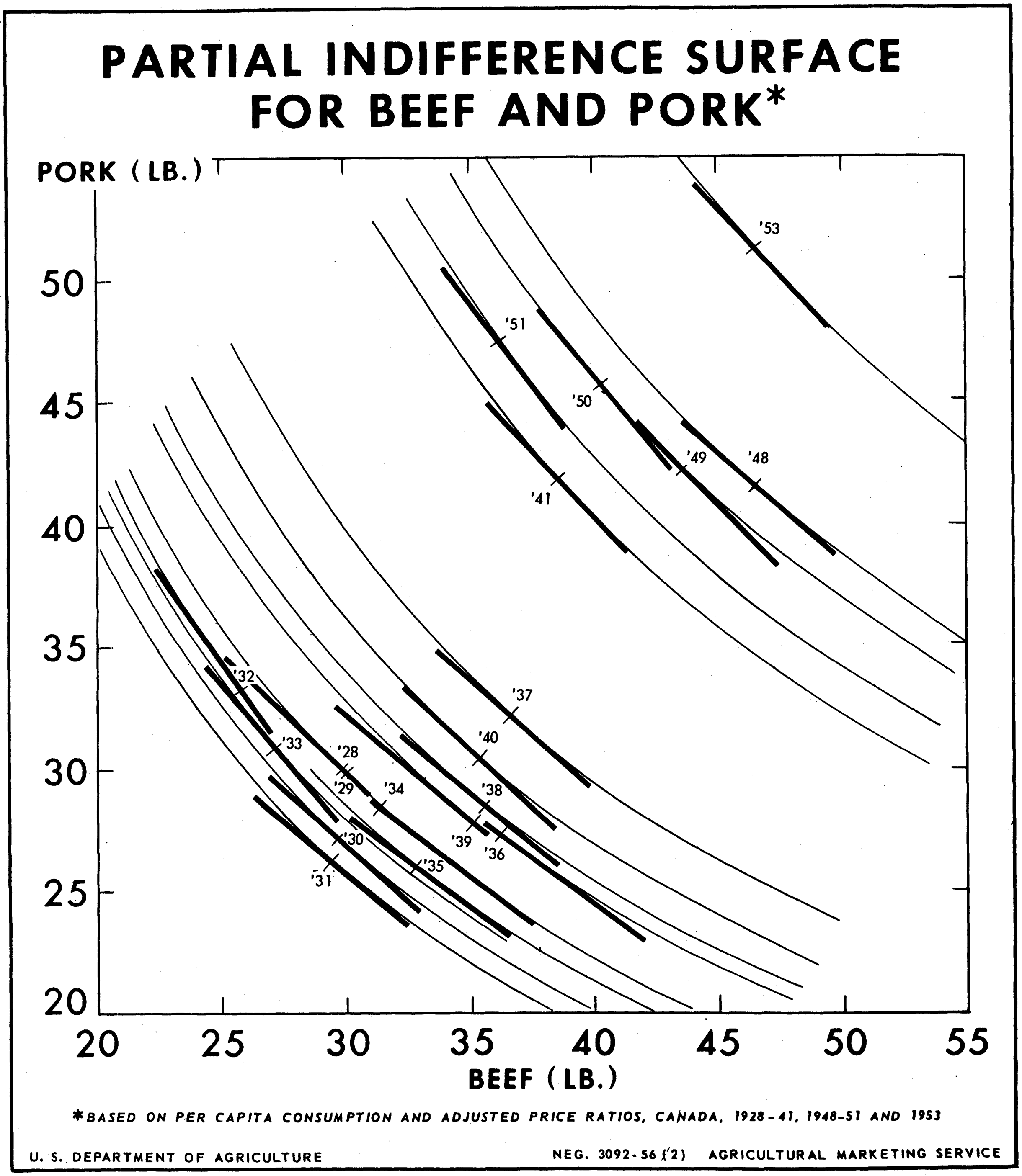


Figure 7.--In the method developed by Waugh (99), the extent to which the adjusted price ratio approximates the tangent to the contour curve for the particular quantities of beef and pork consumed for that year provides confirmation of the belief that these contour curves represent true indifference surfaces for a given level of consumer income.

in which the last term of equation (122) is an index number of consumption of sugar-using foods weighted by the quantity of sugar (w_i) customarily used with a unit of each food. Given a certain consumption level for sugar-using foods, both price and income elasticities of demand for sugar are believed to be small. Hence, in general, b_2 would be considerably smaller than b_1 and c_2 would be considerably smaller than c_1.

Other examples of completing goods may be found in rayon and cotton in the very short run (a few weeks or months). If the proportion of rayon in a rayon-cotton blend is inflexible, an increased price for cotton, passed through into the finished product, may curtail consumption of both cotton and rayon. During a longer period and for the textile market as a whole, the competitive relation between rayon and cotton would predominate. Completing relationships also may exist among different types of tobacco used in a standard blend.

Derived Demand Equations, Partially Reduced Form Equations, and the Estimation of Demand at Different Market Levels 60/

In this section we consider a commodity that is generally sold by producers to intermediaries, who may be processors, wholesalers, retailers, and the like. We think of a market as divided into three sectors or groups of individuals: (1) consumers, (2) the processing and marketing group, and (3) producers. For given values of other relevant variables, we may think of the quantity produced, the quantity consumed, the price paid to producers, and the price paid by consumers (that is, the retail price) as being determined by four relations: (1) the producers' supply relation, (2) the marketing group's demand relation, (3) the marketing group's supply relation, and (4) the consumers' demand relation. Let

Q_c = quantity consumed

Q_p = quantity produced

P_r = retail price

P_w = price paid to producers

D = disposable income

Z_1 = other factors affecting consumer demand

60/ This section was prepared by Marc Nerlove, agricultural economic statistician, Agricultural Marketing Service. It is based in part on Hildreth and Jarrett (49, pp. 107-112) and in part on Nerlove, Marc: The Predictive Test as a Tool for Research: The Demand for Meat in the United States, M. A. thesis, Johns Hopkins University, 1955, especially pp. 5-18.

Z_2 = other factors affecting the marketing group's supply and demand

Z_3 = other factors affecting producer supply

The four relations can be written in the following functional form, where the Greek letter indicates that we have an equation of some sort that involves the variables shown in the parentheses on the left of the equality sign:

$\xi(Q_c, P_r, D, Z_1) = 0$ (consumer demand) (123)

$\delta(Q_c, P_r, P_w, Z_2) = 0$ (marketing group's supply) (124)

$\sigma(Q_p, P_r, P_w, Z_2) = 0.$ (marketing group's demand) 61/ (125)

$\mu(Q_p, P_w, Z_3) = 0.$ (producer supply) (126)

In this model, we neglect problems raised when a large part of the quantity produced is exported or stored or when a large part of the quantity consumed is imported or comes at times from stocks. If the commodity is highly perishable, or if storage is sufficiently expensive, we may assume that the quantity sold to consumers is the same as that sold to the marketing group. In this case $Q_c = Q_p$, when the two variables are measured in comparable units.

If equations (123) and (125) have an appropriate form--for example, linear or linear in the logarithms of the variables--the marketing group's supply and demand relation may be represented by a single equation, which can be called the marketing group's behavior relation:

$\eta(Q_c, P_r, P_w, Z_2) = 0.$ (marketing group's behavior) 62/ (127)

Equations (123), (126), and (127) constitute a simple model; more complicated forms are likely to be encountered in applied work. But the model shown is useful to illustrate several fundamental points. Equation (127) is a reduced form equation, although it is not, strictly speaking, a derived demand or supply equation.

Derived demand and partially reduced form equations.--If equations (123) and (127) are of an appropriate form, we can eliminate the retail price from each. This follows from the equilibrium condition that the quantity supplied

---

61/ σ is a demand function for a factor of production, namely, the raw material; actually our system contains demand functions for other factors, but we neglect them here.

62/ This equation corresponds to equations (57) or (60) on pages 24 and 25, respectively. This equation is interpreted as a supply function for marketing services in subsequent paragraphs. It can be obtained from equations (123-126) in a variety of ways; in each, exactly the same equation is given.

by the marketing group equals the quantity demanded by the consumer group. If this is done, we obtain a relationship among the variables Q_c, P_w, D, Z_1, and Z_2 which might be called a derived demand relation:

$$\theta\ (Q_c,\ P_w,\ D,\ Z_1,\ Z_2) = 0. \qquad \text{(derived demand relation) } \underline{63}/ \qquad (128)$$

Alternatively, under the conditions specified, we can eliminate the price paid to producers. In this way, we obtain a relationship among the variables Q_c, P_r, Z_2, and Z_3:

$$\tau\ (Q_c,\ P_r,\ Z_2,\ Z_3) = 0. \qquad \text{(partially reduced form supply) } \underline{64}/ \qquad (129)$$

An important aspect of equations (128) and (129) is that each is a partially reduced form equation derived from a structural equation by the elimination of a price through the use of an equilibrium condition.

Hildreth and Jarrett (49, p. 108) make the following interesting point concerning partially reduced form equations: "Equations obtained by simultaneously eliminating one or more equations and one or more endogenous variables from a model have been called partially reduced form equations in various discussions. In a certain fundamental sense, all equations we are likely to deal with may be regarded as partially reduced form relations. It is always possible to imagine a more fundamental explanation of the phenomena that we observe, involving more equations and more endogenous variables. If the model we use is a reasonable one, it should, in principle, be possible to derive it, either exactly or approximately, from the more fundamental model by successive elimination of variables."

Thus, additional equations and equilibrium conditions might be used to eliminate more variables than P_r or P_w from the system defined by the equations (123-126). For example, if Z_2 and Z_3 are prices of inputs or other outputs of the marketing or producer group, we could obtain the demand or supply schedules for the factors or products to which these prices apply from a knowledge of the appropriate production functions. With these equations and the appropriate supply equations for factors or demand equations for products, we could, by using the equations already in the model, eliminate Z_2 and Z_3. 65/

63/ This equation corresponds to equations (58) or (61.2) on pages 24 and 26, respectively.

64/ This type of relationship was used by Nerlove, op. cit., in his study of the demand for meat.

65/ Conceptually, this procedure is analogous to that used in showing the relationship between the Walrasian general equilibrium framework and the Marshallian partial equilibrium framework.

The basis for the statement made on page 24, that M cannot be eliminated from equations (57) or (58) unless further equations are specified, follows from the discussion in the text.

If data are available on all the variables, we may estimate the parameters in equations (123), (126), and (127) jointly. If, however, retail prices or prices paid to producers are not available, we must use a model which includes a partially reduced form equation; that is, we must use the model consisting of equations (123) and (129) if prices paid to producers are not available, or the model consisting of equations (126) and (128) if retail prices are not available. Unless additional equations and equilibrium conditions are specified, the variables Z_2 and/or Z_3 cannot be eliminated from the system; hence, the procedure of estimating the elasticity of demand at the producer level of the market by using an equation which does not contain Z_2, such as

$$\theta' \, (Q_c, P_w, D, Z_1) = 0 \tag{130}$$

is incorrect, unless we assume that no variables other than Q_c, P_r and P_w enter the marketing group's behavior relation.

Thus, if retail prices are not available, we cannot simply substitute prices at some other level of the market; additional variables relating to the behavior of the group of individuals between the final consumer and the level of the market to which the price variable refers should, almost always, be included in the demand equation. For example, if wholesale rather than retail prices are used, variables that relate to the behavior of retailers should be included in the demand function, since the latter is really a partially reduced form equation or a derived demand equation and not a consumer demand equation. If the coefficients of these additional variables do not differ from zero by a statistically significant amount, they may be dropped from the analysis on the assumption that they do not affect the behavior of any group of individuals standing between the final consumer and the level of the market at which demand is measured.

Elasticities of demand and supply from partially reduced form equations in relation to those from structural equations.--Since, in many instances, we may be forced to use a partially reduced form equation in our models rather than estimating equations (123), (126), and (127) jointly, the relation between the elasticities of demand or supply derived from the partially reduced forms, equations (128) or (129), and those that would be derived from the original structural equations, (123) or (127), is of some interest. Hildreth and Jarrett (49, pp. 108-112) obtain such a relation when a derived demand equation is estimated rather than a true consumer demand equation. Nerlove (op. cit.) obtained such a relation when a partially reduced form supply equation is estimated rather than a true producer supply equation. Here we state their results without proof. Beginning on page 106, a simple proof of one of the Hildreth-Jarrett results is given.

Let

$E_{Q_c P_r}$ = elasticity of demand at retail with respect to price

E_{Q_cD} = elasticity of demand at retail with respect to disposable income

$E_{Q_cP_w}$ = elasticity of demand with respect to the price paid producers derived from estimates of the coefficients in equation (128), the derived demand relation

$E_{Q_c'D}$ = elasticity of demand with respect to disposable income derived from estimates of the coefficients in equation (128), the derived demand relation

All these elasticities are defined so as to be positive. What is the relation between $E_{Q_cP_r}$ and $E_{Q_cP_w}$ and between E_{Q_cD} and $E_{Q_c'D}$?

Let $E_{P_rP_w}$ be the elasticity of the retail price with respect to the price paid producers. This elasticity could, if data were available, be derived from estimates of the coefficients in equation (127). It might be called the "elasticity of price transmission." Hildreth and Jarrett show that

$$E_{Q_cP_w} = \frac{E_{Q_cP_r} \cdot E_{P_rP_w}}{1 - A \cdot B} , \tag{131}$$

where A is the change in retail price associated with a unit change in the quantity consumed in relation (127), that is,

$$A = \left(\frac{\partial P_r}{\partial Q_c}\right)_\eta , \tag{132}$$

and B is the change in the quantity demanded by consumers associated with a unit change in retail price, that is,

$$B = \left(\frac{\partial Q_c}{\partial P_r}\right)_\xi = -E_{Q_cP_r} \cdot \frac{Q_c}{P_r} . \tag{133}$$

The minus sign is introduced on the right because $E_{Q_cP_r}$ is defined as a positive quantity whereas $\left(\frac{\partial Q_c}{\partial P_r}\right)_\xi$ is, of course, typically negative. Similarly, Hildreth and Jarrett show that

$$E_{Q_c'D} = \frac{E_{Q_cD}}{1 - A \cdot B} , \tag{134}$$

where A and B are defined as before.

Two important groups of relations follow almost immediately from equations (131) and (134):

(1) If the quantity passing through the marketing sector does not influence the retail price, so that A = 0, the income elasticity of derived demand equals the income elasticity of consumer demand at retail. 66/ Beginning on page 107, however, a discussion is given which suggests that, in general, A is a positive number different from zero. Since B normally is negative, it then follows from equation (134) that the income elasticity of derived demand tends to be less than the income elasticity of consumer demand at retail. Thus, in general,

$$E_{Q_c'D} \leq E_{Q_cD}. \tag{135}$$

Hence, if retail prices are unavailable, the best we can do is to set a lower limit to the income elasticity of consumer demand at retail from the coefficients in the partially reduced form equation.

(2) From equation (131) we see that, even if A = 0, the price elasticity of derived demand differs from the price elasticity of consumer demand at retail by an amount depending on the elasticity of price transmission. Unless we can specify whether $E_{P_rP_w}$ is greater than, less than, or equal to one, we cannot say whether $E_{Q_cP_w}$ is a lower or upper limit to $E_{Q_cP_r}$. For a large class of cases, however, we expect the elasticity of price transmission to be less than or equal to one. This class includes such cases as (a) a constant dollar marketing margin, (b) a constant percentage marketing margin, and (c) a marketing margin that is any increasing monotonic function of the quantity passing through the marketing system. 67/ If $E_{P_rP_w}$ is less than or equal to one, it follows from equation (131) that

$$E_{Q_cP_w} \leq E_{Q_cP_r}. \tag{136}$$

Thus, in a wide variety of cases, the price elasticity of derived demand is a lower limit to the price elasticity of consumer demand at retail.

We next state Nerlove's result for the relation between the elasticity of supply derived from the partially reduced form supply function and the true producer supply elasticity. 68/

66/ The reader should remember that equation (127) shows essentially the relation between retail and wholesale prices. If A = 0, then the marketing margin is unaffected by the quantity moving through the marketing sector.
67/ Proofs of these statements are given in the Appendix, pages 203-205.
68/ This result is given and proved by Nerlove (*op. cit.*, pp. 8-10). The proof there follows the same general approach as that used by Hildreth and Jarrett (49, pp. 108-111).

Let

$\tilde{E}_{Q_cP_w}$ = elasticity of producer supply with respect to the price paid to producers, that is, as derived from estimates of the parameters of equation (126)

$\tilde{E}_{Q_cP_r}$ = elasticity of supply at retail, that is, as derived from estimates of the parameters of equation (129), the partially reduced form supply relation

Both $\tilde{E}_{Q_cP_w}$ and $\tilde{E}_{Q_cP_r}$ tend, of course, to be positive. Nerlove shows that

$$\tilde{E}_{Q_cP_r} = \frac{\tilde{E}_{Q_cP_w} \quad E_{P_rP_w}}{1 - A \cdot C} , \tag{137}$$

where C is the effect of a unit change in the price paid to producers on the quantity supplied, that is,

$$C = \left(\frac{\partial Q_c}{\partial P_w}\right)_\mu = \tilde{E}_{Q_cP_w} \cdot \frac{Q_c}{P_w} . \tag{138}$$

The result which follows from equation (137) is not as clear cut as the results which follow from equations (131) and (134). If the quantity passing through the marketing system has no effect on the retail price, A = 0. In this case, by the arguments given above, the elasticity of supply at retail tends to be less than the elasticity of supply at the producer level, since $E_{P_rP_w}$ tends to be less than unity. On the other hand, since C is positive, if the quantity passing through the marketing system does affect the retail price strongly, 1 - A · C may be much less than one and tend to offset the effect of the elasticity of price transmission. It may thus happen that the elasticity of supply at retail exceeds the elasticity of supply at the producer level. For commodities like eggs, meat, or milk, there may be both a significant producer supply response to current price and a significant effect of the quantity passing through the marketing sector on the marketing margin. In these cases, the elasticity of supply at retail may well exceed the elasticity of supply at the producer level. In such instances, results from the use of least squares procedures to estimate the consumer demand function at retail may be more seriously biased, in a statistical sense, than consideration of the elasticity of producer supply would lead one to expect.

<u>Proof that the price elasticity of derived demand is typically less than the price elasticity of demand at retail</u>.--A simple proof of equation (136) may be obtained from a particular interpretation of the marketing group's behavior relation--equation (127). It is primarily because of an interest in this interpretation that we give the proof here.

Equation (127) states that there is a relation between the quantity passing through the marketing sector, the retail price, the price paid to producers, and other variables related to the marketing sector. In order to write this relation, we assume that the quantity supplied by producers is approximately the same as the quantity used by consumers. The marketing group may be viewed as adding certain services to the commodity under consideration as it passes through the marketing sector which, in total, are roughly proportional to the quantity passing through. Consequently, equation (127) may be interpreted as a particular kind of supply function, namely, the supply function for marketing services.

A supply function for a commodity or service usually contains the price of that commodity or service, but equation (127) contains two prices, the retail price and the price paid to producers, and neither is the price of marketing services. If the function η is of the appropriate form, we can remedy this situation. 69/ Let the difference between the retail price and the price paid to producers be P_s, so that

$$P_s = P_r - P_w \, . \tag{139}$$

P_s is the dollar amount per unit of the commodity passing through the marketing system which goes to individuals in the marketing group. If the services which they supply are roughly proportional to the quantity passing through, P_s may be interpreted as the price of marketing services; hence, if η is of an appropriate form, we may derive the following equation from equations (127) and (139) and this can be interpreted as the supply of marketing services relation:

$$\lambda\,(Q_c, P_s, Z_2) = 0. \qquad \text{(supply of marketing services)} \tag{140}$$

From equation (140) we see that the supply of marketing services has a well-defined price elasticity

$$E_{Q_cP_s} = \left(\frac{\partial Q_c}{\partial P_s}\right)_\lambda \frac{P_s}{Q_c} \, . \tag{141}$$

Under conditions of competition in the marketing sector, $E_{Q_cP_s}$ tends to be positive.

We are now ready to show that, in general, the price elasticity of consumer demand at retail is greater than the price elasticity of derived demand. If we differentiate the identity (139) with respect to Q_c and multiply both sides of the result by Q_c/P_r, we have

$$\frac{\partial P_s}{\partial Q_c}\frac{Q_c}{P_r} = \frac{\partial P_r}{\partial Q_c}\frac{Q_c}{P_r} - \frac{\partial P_w}{\partial Q_c}\frac{Q_c}{P_r} \, . \tag{142}$$

69/ Whatever the form of η , this always is possible if we allow approximations by a Taylor series.

With some manipulation, we have

$$\frac{\partial P_s}{\partial Q_c} \frac{Q_c}{P_s} \frac{P_s}{P_r} = \frac{\partial P_r}{\partial Q_c} \frac{Q_c}{P_r} - \frac{\partial P_w}{\partial Q_c} \frac{Q_c}{P_w} \frac{P_w}{P_r} . \tag{143}$$

Recalling that elasticities of demand are defined to be positive, we have, from equation (143),

$$\frac{1}{E_{Q_cP_s}} \cdot \frac{P_s}{P_r} = - \frac{1}{E_{Q_cP_r}} + \frac{1}{E_{Q_cP_w}} \cdot \frac{P_w}{P_r} . \tag{144}$$

By use of equation (139), equation (144) can be rewritten

$$\frac{1}{E_{Q_cP_s}} \left(1 - \frac{P_w}{P_r}\right) = - \frac{1}{E_{Q_cP_r}} + \frac{1}{E_{Q_cP_w}} \frac{P_w}{P_r} . \tag{145}$$

Adding and subtracting $\frac{1}{E_{Q_cP_r}} \cdot \frac{P_w}{P_r}$ from the right hand side of equation (145) and performing some algebraic manipulation, we find

$$\frac{1}{E_{Q_cP_w}} - \frac{1}{E_{Q_cP_r}} = \left(\frac{1}{E_{Q_cP_s}} + \frac{1}{E_{Q_cP_r}}\right) \left(\frac{P_r}{P_w} - 1\right) . \tag{146}$$

Since the retail price always exceeds the price paid to producers, and since $E_{Q_cP_s}$ and $E_{Q_cP_r}$ are defined so as to be positive, we see that the right hand side of equation (146) is positive. Thus

$$\frac{1}{E_{Q_cP_w}} \geq \frac{1}{E_{Q_cP_r}} . \tag{147}$$

But equation (147) implies that

$$E_{Q_cP_w} \leq E_{Q_cP_r} , \tag{136}$$

that is, that the price elasticity of derived demand is a lower limit to the price elasticity of consumer demand at retail, which is what we set out to prove.

In passing, we should point out that it is incorrect to include the marketing margin, P_s, in any partially reduced form equation derived from equations (123), (126), and (127). In particular P_s should not be used in place of the variable Z_2 in the derived demand equation, since the identity (139) must be included in any system containing equation (140); when it is, it becomes clear that P_s is not an independent variable, but a derived one.

Problems of estimation.--We now turn to some problems of estimation. These are: (1) If we know the price elasticity of consumer demand at retail, how can we estimate the price elasticity of derived demand? (2) If we know the price elasticity of derived demand, how can we estimate the elasticity of demand at retail? Unless we can assume that the quantity passing through the marketing system has no effect on the marketing margin and unless we know the elasticity of price transmission, it is clear that the second question cannot be answered.

If the system containing equations (123), (126), and (127) has been estimated jointly, 70/ then we have estimates of $E_{Q_c P_r}$, $E_{P_r P_c}$, A, and B; consequently, the elasticity of derived demand, that is, the price elasticity of demand at the producer level of the market, may be estimated by applying equation (131). If the quantity passing through the marketing system does not affect the marketing margin, A = 0 and the price elasticity of derived demand equals the price elasticity of consumer demand at retail times the elasticity of price transmission. 71/ As indicated, the elasticity of derived demand is less than the elasticity of demand at retail, because $E_{P_r P_w}$ typically is less than one.

In order to derive the price elasticity of consumer demand at retail from a derived demand function, we must also estimate the marketing group's behavior relation or the supply of marketing services relation. Since retail prices are typically not available when we estimate a derived demand relation directly from the data, it usually is not possible to estimate the elasticity of price transmission by estimating the marketing behavior relation. If, however, we can obtain an elasticity of price transmission on the basis of outside knowledge or incomplete data, we can obtain the price elasticity of consumer demand at retail by simply dividing the price elasticity of derived demand by the elasticity of price transmission. This procedure rests, of course, on the assumption that the quantity passing through the marketing sector does not affect the marketing margin.

Variables to be included in the marketing group's behavior relation.--In general we should include in the marketing group's behavior relation (1) the quantity consumed, (2) either (a) the retail price and the price paid to producers or (b) the difference between these prices and (3) other variables, referred to collectively as Z_2, that affect the marketing margin.

Interesting problems arise when we consider the additional variables represented by Z_2. Price series for many factors of production used in the marketing sector are rarely available; furthermore the marketing sectors for many commodities are characterized by rapid technological change. Wage rates in industries connected with the marketing sectors for many farm commodities, however, are frequently available. One way to allow for the marketing margin

---

70/ If the quantity supplied does not depend on current price, equation (126) may be omitted and quantity taken as independent.

71/ The procedure is in fact the one suggested by Fox in Foote and Fox (29, p. 40). Illustrative computations were given in the publication cited, but we have not reproduced these results here because the price elasticities of consumer demand at retail, from which the derived demand elasticities were computed, were obtained from analyses based on undeflated data.

to depend on (1) the amount of commodity flowing through the marketing sector, (2) the level of technology in the marketing sector, and (3) the one factor price available (namely, wage rates in the marketing industry) is to assume that the marketing margin is simply a function of unit labor cost in marketing. 72/

Allowing for Systematic Shifts in the Regression Coefficients Over Time

Foytik (39), in a study of the demand for plums, uses an interesting technique to determine whether the regression coefficients on the several variables in the demand equation change in a systematic way as the season progresses. He states that two alternative methods might be used. In the first, data for each week are considered as a separate set of observations, and a demand curve for each week is derived. The regression coefficients obtained are listed in sequence. Unless they seem to follow a systematic pattern over time, the differences between weeks are assumed to be statistically nonsignificant. The second approach, which was the one used by Foytik, considers all the weekly observations as an entirety, and tests whether systematic changes in the regression coefficients are statistically significant. This is done by use of the following equation, where $i = W =$ week of season, $P =$ price, $Q =$ quantity sold, and $D =$ consumer income:

$$P_i = a + (b + b'W)Q_i + (c + c'W)D_i + (d + d'W)Q_{i-1} + (e + e'W)W. \qquad (148)$$

If the coefficients on the cross-product terms b', c', d', and e' differ from zero by a statistically significant amount, the coefficients within the basic demand equation can be assumed to change in a systematic way over time. Equations of this type restrict the changes in the net regression coefficients to a well-defined, smooth pattern, although the use of second and higher degree terms of W in the parentheses would permit the rates at which the coefficients change over the season to themselves increase or decrease gradually. Foytik (39, p. 445) states, "It should be pointed out that the more complex the function is made, by the use of such additional terms, the more difficult it is to make a suitable economic justification even though the statistical fit is improved."

Foytik fitted an equation of this type to 270 observations covering weekly prices and sales at the New York auction market for 1928-48, excluding the years during World War II. His final equation retained all of the coefficients shown in equation (148) except b' and d'. Thus his study suggests that, for plums in the New York market, the relationships between price and (1) current and (2) lagged quantity remain essentially unchanged throughout the season, but the relationship between price and consumer income changes in a systematic way, with changes in income having a less important effect on price as the season progresses. This result appears reasonable.

72/ This approach was taken by Gerra in the model referred to on page 14 and by Nerlove, *op. cit.*

Use of Distributed Lags in Demand Analysis 73/

Irving Fisher (24) was the first to use and discuss the concept of a distributed lag. Although the idea that one economic variable depends on another variable lagged in time is an old one, the use of lagged variables in empirical research has been restricted primarily to areas other than demand analysis. In this section we discuss briefly some of the analytical problems involved in attempting to introduce variables with a distributed lag into a demand study and then illustrate the general approach with an application to measuring the long-run demand for automobiles.

In economics, a cause often produces its effect only after a lapse of time. For example, a drop in the price of potatoes in the fall cannot affect potato acreage until the following spring, nor can it decrease potato production until the following fall. The lapse of time between a cause and its effect is called a lag. The lag may be a specific time, say three months, or one year. But in many cases, the effects of an economic cause are spread over many months, or even many years. In such cases, we have a distributed lag.

Distributed lags with respect to variables that affect consumption may arise for the following reasons:

(1) Psychological. Under this category we include forces of habit and assumptions on the part of the consumer that changes may be only temporary.

(2) Technological. These include factors such as, in a general case, lack of knowledge about possible substitutes or, in a specific case, the inability to increase greatly the use of frozen foods without first acquiring adequate freezer storage space.

(3) Institutional. This category includes (a) situations in which certain contractual items of expenditure or savings may need to be adjusted before shifts can be made in consumption patterns, and (b) situations resulting from the fact that some markets, particularly for durable goods, are imperfect in an economic sense.

Analytical approaches.--One way to measure the degree of lag with respect to a particular variable is to find by statistical analysis that distribution of lag which maximizes the effect of the causal factor. Empirical analyses based on this approach have been run which (1) make no assumption as to the form of the distribution of lag, (2) are designed to estimate certain characteristics of an assumed general form for the distribution, or (3) derive and statistically fit a model based on the fundamental cause of the distributed lag but which yields a specific distribution of lag only incidentally.

One way to formulate models for generating distributed lags is to assume that the lags arise chiefly because of technological and institutional

---

73/ Material in this section is condensed from Nerlove (75, 76).

rigidities. The traditional distinction between long- and short-run elasticities rests largely on causes of this sort. Although observable points each lie on a short-run curve, the coefficients of the long-run curve can be estimated if the model is properly formulated. Relations between long- and short-run demand curves depend chiefly on the path that observed consumption would follow if it moved directly toward its long-run equilibrium level following an initial change in a causal variable. The shape and form of such paths is determined by the type of institutional and technological rigidities that exist.

Another way to formulate models that generate distributed lags is to assume that technological and institutional rigidities are absent but that uncertainty about the future exists and that habit is a powerful force. Under conditions of uncertainty, a change in price or income may be thought of by consumers as divided into two components--one permanent and one transitory. The permanent part changes expectations in all relevant future years, whereas the transitory part changes expectations in only some future years, or perhaps in no future year. The average level about which future prices or incomes are expected to fluctuate is called the "expected normal;" this level is affected only by that part of the initial change in a causal factor that is considered permanent. If forces of habit are strong, the effects of a change in current price or income on consumer behavior are slight compared with the effects of a change in the expected normal. Hence, models need to be formulated so as to emphasize the effects of changes in the expected normal.

Hicks (48, p. 205) defines "the elasticity of a particular person's expectations of the price of a commodity x as the ratio of the proportional rise in expected future prices of x to the proportional rise in its current price." The elasticity has been found by empirical research to normally range between 0 and 1. Many factors affect the elasticity of expectations; some affect all commodities equally, whereas others affect different commodities differently. Among the latter is the typical variance of prices of the commodity. For any given commodity the elasticity of expectations need not be stable over time, but models are simpler if we assume that it is stable. Frequently an equation that contains one or more distributed lags can be reduced by algebraic manipulation to an equation that does not contain such lags. The reduced equation then can be fitted statistically and the results used to obtain certain characteristics of the lag. A difficulty with this approach is that the number of variables added in the reduced equation is greater than or equal to the square of the number of variables with distributed lags that enter the initial equation. Thus the method is not feasible when the initial equation contains more than two or three variables with distributed lags.

In working with commodities that substitute for or complement each other, we generally are interested in a system of equations rather than a single equation. Each equation frequently contains values of the same variables. The multiple equation method of reduction of a demand equation which involves distributed lags takes advantage of the existence of all the interrelated equations. By way of contrast with the single equation method of reduction described in the preceding paragraph, if each original demand equation contains n expected normal prices and expected normal income, each reduced demand equation in the multiple set includes, in addition to n + 1 current values of prices and income, n + 1 lagged values of the quantities demanded and aggregate consumption of all commodities. Thus, whereas the single equation method leads to one reduced demand equation containing $(n + 1) + (n + 1)^2$ independent variables, the multiple equation method leads to n reduced demand equations each containing $2(n + 1)$ independent variables. For statistical purposes, the latter are clearly preferable when n exceeds 1 or 2. Some systems of demand equations are of such a nature as to be "separable" in a certain mathematical sense. An example is a system that contains several demand equations and a consumption function; here we may (1) reduce the whole system, including the consumption function or (2) solve the consumption function for expected income, substitute this into the demand equations, and then reduce only the demand equations. This is the sense in which the term "separable" is used.

Equations that contain distributed lags that are due only to technological or institutional rigidities can be reduced easily. Reduction becomes more complex when the lags are due to uncertainty about the future, but a considerable degree of simplification may be obtained by the multiple equation method of reduction. When the lags are due both to uncertainty and rigidities, in general, no simplification of reduction is possible. If, however, the distributions of lag are of a special form, simple reduction is possible.

Statistical estimation of the coefficients in demand equations that contain distributed lags can be done theoretically in two ways: (1) by dealing directly with the equation that involves the distributed lags or (2) by using one or more reduced equations. Maximum likelihood procedures under the first approach are available. This approach requires a large number of repeated steps; for this reason, it is called the "iterative" method of estimation. In order to use the iterative method, the unexplained residuals in each equation that contains a distributed lag must be normally and independently distributed. If the lags arise solely because of technological and institutional rigidities, this will be true only under special conditions. If the lags are caused by uncertainty, problems of serial correlation in the residuals do not arise but the iterative procedure is computationally feasible only in the simplest cases. Thus the iterative approach can be used only under special circumstances.

If a distributed lag is due only to rigidities of a technological or institutional nature, the coefficients can be estimated easily by using a single reduced equation fitted by least squares. Estimation becomes more complicated if the lags are caused by uncertainty about the future. If three or more variables with distributed lags are involved and we use the single-equation method of reduction, the number of variables in the reduced equation becomes so large as to make statistical fitting virtually impossible with time series of normal length. Complications from serial correlation of the residuals also enter. Only in the simplest of cases should the noniterative method be based on a single reduced equation. In using the multiple reduced equation approach, we must assume that the distributions of lag for the same variable in each of the equations is the same. If the multiple equation method can be used, it is computationally much simpler than the single equation method.

In analyzing the demand for commodities in general, if anticipated distributed lags result from uncertainty about the future, we should specify a system of equations before proceeding with any estimation. If the distributed lags are believed to result only from technological or institutional rigidities, only a single demand equation for the particular commodity need be specified.

An empirical example.--The example described here relates to the measurement of the long-run demand for automobiles. The same general approach, however, may prove useful for other consumer durables or near-durables, such as certain items of clothing and other textile products, household appliances, and so forth. Only enough detail with respect to the specific variables used is given to indicate the economic basis for the equations in the model.

In this system of equations, we assume that the quantity of automobiles demanded by consumers can be represented by the total stock of cars adjusted in some manner for age, make, and model. As cars are known to depreciate rapidly the first few years and more slowly thereafter, we can adjust for the age factor approximately by applying a constant percentage rate of depreciation. Use of such a depreciation rate permits us to derive the total stock of automobiles, adjusted for age, from an index of past purchases of new cars which takes into account differences due to make and model. Let s_t be the stock of automobiles during period t, d be the percentage rate of depreciation and x_t be new car purchases during period t. Then we may write

$$s_t = x_t + (1-d)x_{t-1} + (1-d)^2 x_{t-2} + \dots \qquad (149)$$

$$= x_t + (1-d)s_{t-1}. \qquad (149.1)$$

A specific value for s_t is not needed in fitting the model, but a measure of x_t is required.

The demand for the total stock of automobiles depends on the usual variables that enter into a demand equation, namely relative price, consumer income, and technological factors such as miles of highways, degree of urbanization, and so forth. However, consumers may be unable or unwilling to adjust stocks immediately to the equilibrium level suggested by current income and prices. For example, consumers may prefer to pay off certain installment debts before purchasing a new car, or they may believe that a current change in income is not permanent and hence defer purchases of expensive items until they feel that it is permanent. Each of such factors tends to cause a lag in adjustment of purchases to current economic stimuli.

Since these economic stimuli are constantly changing, consumers never reach an equilibrium position--they only move toward it. Thus we cannot expect a direct relation between observed stocks of cars, or new purchases, and current values of prices or consumer income. Instead, purchases of new cars are related to current <u>and</u> <u>prior</u> values of the economic stimuli. As we show in the paragraphs that follow, we can <u>deduce</u> the long-run equilibrium relations from observed changes in new car purchases, price, income, and other variables by making certain assumptions as to how consumers adjust toward an equilibrium position.

Let s_t^* be the long-run equilibrium stock of cars desired by consumers, p_t be current relative price, y_t be current real income, and z_t be other variables that affect long-run demand. Then

$$s_t^* = a + b_1 p_t + b_2 y_t + b_3 z_t. \tag{150}$$

It seems reasonable to suppose that the closer consumers are to equilibrium, the less incentive they have to overcome the costs and frictions of adjustment. When current stocks are far out of line with equilibrium stocks, on the other hand, consumers have strong incentives to make necessary adjustments. Thus it seems reasonable to assume that the <u>rate</u> of adjustment is proportional to the amount of imbalance or that

$$s_t - s_{t-1} = \beta\,(s_t^* - s_{t-1}). \tag{151}$$

We now use these basic equations to derive an equation that can be used for statistical estimation. This is done as follows:

(1) Rewrite equation (151) to get

$$s_t = \beta\, s_t^* + (1-\beta)s_{t-1}. \tag{151.1}$$

(2) Lag equation (151.1) by one period to get

$$s_{t-1} = \beta\, s_{t-1}^* + (1-\beta)s_{t-2}. \tag{151.2}$$

(3) Multiply equation (151.2) by (1 - d) to get

$$(1 - d)s_{t-1} = \beta\,(1 - d)s_{t-1}^* + (1 - \beta)(1 - d)s_{t-2}. \tag{152}$$

(4) Subtract equation (152) from equation (151.1) to get

$$s_t - (1 - d)s_{t-1} = \beta\,[s_t^* - (1 - d)s_{t-1}^*] + (1 - \beta)[s_{t-1} - (1 - d)s_{t-2}]. \tag{153}$$

(5) From equation (149.1)

$$x_t = s_t - (1 - d)s_{t-1} \tag{149.2}$$

$$x_{t-1} = s_{t-1} - (1 - d)s_{t-2}. \tag{149.3}$$

(6) By substitution in equation (153) we get

$$x_t = \beta\,[s_t^* - (1 - d)s_{t-1}^*] + (1 - \beta)x_{t-1}. \tag{153.1}$$

(7) Using equation (150) to substitute for s_t^* and s_{t-1}^*, we get

$$x_t = \beta[a + b_1p_t + b_2y_t + b_3z_t - (1-d)(a + b_1p_{t-1} + b_2y_{t-1} + b_3z_{t-1})] + (1 - \beta)x_{t-1} \quad (153.2)$$

$$= A + B_1p_t + B_2p_{t-1} + B_3y_t + B_4y_{t-1} + B_5z_t + B_6z_{t-1} + b_7x_{t-1}. \quad (153.3)$$

If we fit equation (153.3) by least squares, we can estimate the coefficients in equations (149.1), (150), and (151) as follows:

$$\beta = 1 - B_7 \quad (154)$$

$$b_1 = B_1/\beta \quad (155)$$

$$b_2 = B_3/\beta \quad (156)$$

$$b_3 = B_5/\beta \quad (157)$$

$$a = A/(\beta + d - 1). \quad (158)$$

But

$$d = (B_2 + b_1)/b_1 \quad (159)$$

$$= (B_4 + b_2)/b_2 \quad (160)$$

$$= (B_6 + b_3)/b_3. \quad (161)$$

The estimate of d in equation (153.3) is overdetermined, the alternative answers for d obtained from equations (159) - (161) are not the same. Nerlove suggests using that B_i which relates to the variable that is measured most accurately by available data.

Results of an analysis of this sort for 1922-41 and 1948-53 are given in Nerlove (75, pp. 62-63). By relating new car purchases to current and lagged price, current and lagged consumer income, and lagged new car purchases, he obtains a multiple coefficient of determination of 0.9, a value for β of 0.7 and, when d is estimated from the coefficient on lagged income, a value of 0.5. The long-run elasticity on relative price is estimated to be about unity, and on real income, about 4.

Measuring Cyclical Factors that Affect Demand

Sometimes the residuals from a demand analysis follow a pattern that suggests a cyclical shift in the demand relations. Three examples are cited in this section.

Fats and oils.--In a study of factors that affect wholesale prices of fats and oils used in food products, Armore (5, pp. 56-58) states, "When the final analysis was completed, the residuals appeared to be correlated with changes in the general price level and with changes in prices of fats and oils. As discussed in the body of this bulletin, accumulation and reduction of inventories by members of the fats and oils trade during periods of changing prices would be expected to have such an effect. To throw more light on this point, year-to-year changes in the dependent variable were added as a fourth independent variable. The multiple coefficient of determination was raised from 0.92 to 0.96, so that the unexplained variation was reduced from 8 to 4 percent, and a statistically significant partial coefficient of determination of 0.52 was obtained for the new variable. The addition of this variable has little longer-term forecasting value but it does confirm the importance of allowing for the effects of changes in inventories in making short-term forecasts."

Coffee.--In a similar study of factors that affect prices of coffee, Hopp and Foote (50, pp. 433-434) write,

"When the unexplained residuals were plotted against time as a check on the degree of serial correlation, it was found that they followed a definite cyclical pattern. ... For coffee, however, use of the change in price as an additional independent variable had practically no effect on the regression equation or the residuals. ... Further study of the unexplained residuals ... indicated that in general they are positive when prices are rising or remain relatively high and they are negative when prices are declining or remain relatively low. If these are designated as inflationary and deflationary periods, respectively, 18 out of 20 residuals for the inflationary years were positive and 25 out of 34 residuals for the deflationary years were negative. On the average, prices were 21 percent higher during inflationary periods than would have been expected from the regression equation and 11 percent lower during deflationary periods. ... When the computed prices were adjusted in this way, the percentage of variation explained by the analysis was increased from 70 to 84 percent. This improvement by stratification of the variation into these 2 classes was highly significant statistically.

"The economic explanation of this appears to be similar to that for fats and oils. When supplies are declining, efforts are made to maintain coffee inventories, and prices tend to be higher than would be expected from the level of supply in relation to current consumption. When supplies are increasing, inventories can be reduced; hence prices tend to be lower than would be expected based on relative supplies.

"The only notable exception to this cyclical effect was during 1931-34. In this deflationary period, prices declined, but they did not fall as low as would have been expected based on the original regression equation. In the extreme situation that existed at that time, when values fell exceptionally low as supply reached an all-time peak in relation to consumption, some market resistance apparently developed to further price declines."

Allowance for cyclic position significantly improved forecasts made from this analysis for the years 1949-53.

Mill demand for cotton.--Lowenstein and Simon (65, pp. 106-110), in a study of factors that affect mill consumption of cotton, use a different approach. They write,

"In any given period mill consumption of cotton may be out of balance with consumer purchases of cotton products because of changes in inventories at various levels of fabrication and distribution. For example, when inventories of cotton products are being built up at any level of marketing, the increments represent an increase in demand for cotton fabrics, and hence for cotton, over and above current consumption.

"Of importance among factors that affect inventories are changes in sales or expectations thereof. Merchants may try to keep inventories in a fixed--or relatively fixed--ratio to their rate of sales. If consumer purchases of cotton goods decline, merchants would try, other factors being the same, to adjust inventories to a level commensurate with the changed conditions of demand. Current demand would be satisfied temporarily from stocks of goods produced in the past. Orders for cotton products thus would tend to decline by an amount greater than that of the decrease in retail sales. As the decline in demand spreads to preceding stages of distribution and manufacture, it would grow in intensity to the extent that inventories on these levels also are reduced proportionately. Ultimately, the magnified reduction in consumer demand is reflected back to the mill level. A result of the adjustments in inventories along the line is a rate of cotton consumption less than that indicated by the decrease in consumer demand. Although inventory changes may bear some relationship to changes in consumer income, the latter measure could not be expected to account fully for the effect of the former on mill consumption of cotton.

"Other factors that affect demand for goods for inventory, given the marketing and technological structure of the industry, include fears of shortages, expectations concerning price changes, and other facets of the economic outlook. Prices themselves are not immune to inventory changes. For example, the decision to acquire additional stocks, perhaps in line with a rise in consumer purchases, would add to the upward pressure on prices for textiles. The price rise, in turn, induces, for speculative and precautionary reasons, further increases in stocks and an inventory-price-inflation spiral may develop. Thus actions taken with respect to stocks may affect and be affected by prices.

"In addition to changes in inventory, the size of the total inventory in the cotton textile system is to be noted. Obviously, the larger this inventory, the deeper will be the effect of, and the longer the adjustment to, a decline in consumer demand. Conversely, if demand were to increase suddenly and sharply, with inventories overly low, an industry-wide speculative movement could be generated because of the prevailing tight supply condition. Stocks of cotton goods thus affect demand schedules for cloth and hence mill demand for cotton.

"A change in demand for cotton goods is translated at the mill level into a change in volume of new business both for immediate and future delivery. The reaction of output to a change in demand, however, is usually not instantaneous. It takes time for production to adjust to a new level of sales. Influential among reasons for the relatively slow response of production are the momentum of the manufacturing process, uncertainty concerning the lasting nature of the change, the time it takes to obtain additional materials or to cancel orders, and cost and time considerations relating to removing shifts and shutting down looms or to adding shifts and starting up idle equipment. Initially, mill stocks of cotton textiles would tend to bear the brunt of a change in demand, probably varying inversely to it. Theoretically, adjustment in output, when it comes, would account for the involuntary change in stocks plus any tendency for mill inventories of textiles to be brought into line with the new level of demand. Hence cotton consumption would be expected to reflect both the lag in response of output to a change in demand and the resulting adjustment in level of stocks.

"Adjustment in production is often carried too far; that is, output is found to be forthcoming at a rate too high or too low when compared with the level of demand. Hence it may more than compensate for the earlier change in stocks. If output were maintained at a level above that of demand, textile stocks would tend to accumulate. But if output were cut back and maintained below the level of demand, stocks would tend to decline. If, concurrently, demand were to shift in the opposite direction the imbalance could be magnified. Ultimately production would have to be adjusted and, if carried too far again, could affect stocks similarly but in an opposite direction.

"Prices of cotton textiles, actual or expected, have been omitted from the preceding discussion but actually they are interwoven into the dynamics of the industry. The apparent tendency of price to respond almost immediately to changes in the demand for cotton textiles is _prima facie_ evidence of the lagged output response. If the change in demand is sudden, following a period of stock accumulation or liquidation, the effect on prices can be extreme until the necessary adjustment is made in stocks and output. If the change in price initiates a further change in inventory demand and possibly a price-inventory spiral up the line, the shift in demand would be greater and the supply adjustment required by the industry magnified. When prices are thought to be fully discounted and production curtailment is proceeding apace, the desire to cover forward at low prices or the incentive to acquire stocks in anticipation of higher prices may initiate a buying wave. At this point risks connected with stock acquirement may be low compared with those associated with the continued deferment of needs. The low textile prices also may lead to an increase in retail sales with all its back ramifications on the demand for textiles. On the other side, expectations of lower prices may cause a general falling off in demand.

"Largely as a consequence of these varied forces and their interrelationships, mill product stocks tend to change in a cyclical fashion, frequently being out of line with demand. The tendency for output to be kept at relatively high levels in the short run despite unfavorable economic conditions apparently is characteristic of the cotton textile industry. Clearly the tendency on the part of the industry not to respond readily or properly to changes in demand can affect the timing and extent of mill consumption of cotton considerably and, if possible, should be accounted for in the mathematical formulation of mill demand for cotton.

"Recently the American Cotton Manufacturers Institute, Inc.--henceforth designated as the Institute--made available to the United States Department of Agriculture for research purposes data on production, stocks, and unfilled orders of cotton broad woven goods in physical units at the mill level. ... Stock and unfilled orders data represent the mills' position as of the end of a reporting period, generally at or near the end of a calendar month. Data on production cover the intervening period; hence they coincide approximately with a calendar month. ...

"Because mills participate in the Institute's statistical program on a voluntary basis, the percentage of the industry covered by the reports tends to vary. ... Because of ... the lack of strict comparability over time, it was found advisable to use the data in a ratio form to adjust roughly for changes in the reporting sample.

"The ratio of mill stocks of cotton cloth to unfilled orders was computed as of the end of each month for the full period covered by the data. ... The ratio reflects the degree of imbalance between stocks, output, and demand at the mill level. When the ratio is relatively high, unless an increase in demand is forthcoming, a downward adjustment in output to reduce stocks in indicated. Conversely, a relatively low ratio suggests the likelihood of a higher output rate in the near future. The ratio indicates also the cyclical character of changes in inventories of mill products.

"Obviously some inventory is necessary if a business is to function properly and efficiently. The amount of inventory not considered excessive may vary directly with the volume of business, so that a relatively constant ratio between the two is sought. Whether mill stocks of cotton cloth are too high or too low at a given time probably depends more on the amount of business expected in the near future--reasonably approximated by the level of unfilled orders--than on past volume.

"Some 'normal' ratio of stocks to unfilled orders thus may be postulated about which the actual ratio would fluctuate and toward which it would tend. Departure from normal--indicative of imbalance in the industry--would be expected to lead to changes in mill consumption of cotton. For want of information, it was decided to use the average of the ratios as normal. ...

"The residuals from Analysis II in logarithms were found to be fairly closely correlated with actual deviations from normal of the stock-unfilled order ratio for cotton cloth. The best result--a coefficient of correlation of -0.85--was obtained when the new variable led the residuals series by 5 months. This lead is consistent with the 3 to 6 months' lead that would have been expected from *a priori* considerations.

"To throw more light on the importance of the effect of imbalance in mill inventories of cotton cloth on mill consumption of cotton, the deviations of the stock-unfilled order ratio from normal were added as a fifth independent variable in Analysis III. As the new variable is believed to affect mill consumption of cotton in an additive fashion, actual deviations from normal were used. The other variables ... were kept in logarithms when the analysis was run. Analysis III is based on 1927-32, 1935-40, and 1948-52, the only full years for which data on stocks and unfilled orders are available when a lead of 5 months is employed. The results ... are surprisingly good. ... All of the coefficients are statistically significant. ... Changes in actual deviations from normal of the stock-unfilled order ratio, on the average, account for a larger percentage of the variation in mill consumption of cotton than does the price of cotton or the consumption of synthetic fibers, after allowing for the effects of the other independent variables. Changes in disposable personal income and changes in the ratio of the current to the preceding year's income, in that order, are still more important in this respect."

Whenever the residuals from an analysis, plotted against time, appear to follow a pattern, an attempt should be made to ascertain whether cyclical effects could be anticipated for the industry being studied and, if so, appropriate allowance should be made. At times, the analyst may know in advance that such cyclical effects may be important and, if so, a special variable like that used by Lowenstein and Simon might be incorporated into the analysis, or checks like those used by (1) Armore (5) and (2) Hopp and Foote (50), respectively, might be made.

Obtaining Additive Weights for the Components of Variables to be Used in a Logarithmic Analysis 74/

The following method can be used whenever additive weights are desired for the components of one or more variables to be used in a logarithmic analysis. The method has been applied mainly to problems that involve the determination of proper weights for several components of a supply variable. Two research papers have contributed to the statistical problem considered here. Foote (25) demonstrated that a successive approximation method similar to the one outlined here will converge to the least-squares value. Ezekiel (20, pp. 390-403) proposed methods along these lines for studying joint relationships while holding other variables constant.

Steps in the analysis.--The following steps are used when weights for components of a supply variable are desired and price is used as the dependent variables:

74/ Material in this section is adapted from Foote (26, pp. 34-41).

1. The components of the supply factor are combined in some simple way. As a first step in the example for livestock feed, they were converted from bushels to tons and added together. The composite is used as one independent variable in a logarithmic analysis.

2. Values of the dependent variable from this analysis are adjusted for the effects of the independent variables other than supply, and the antilogarithms obtained. In the case of first-difference analyses, some additional calculations are involved. The calculated value plus 2 has for its antilogarithm a figure in terms of percentage of the preceding year. This should be applied to the actual price in the preceding year to get an estimated price. The difference between this estimate and the actual price in the preceding year is the adjusted dependent variable that should be used as the dependent variable in the linear first-difference analysis discussed in step 3.

3. The adjusted dependent variable is used as the dependent variable in a linear analysis in which the separate components of the supply factor are used as independent variables. The partial regression coefficients from this analysis are proportionate to the weights that should be applied to the individual supply components.

4. The following method may be used to test whether these partial regression coefficients differ significantly from each other: (a) Compute the simple linear correlation between the dependent variable used in step 3 and the supply factor based on constant weights for the components. In the case considered here, the latter is the same as the supply variable used in step 1. Compute the unexplained sum of squares for this analysis. If X_0 is the dependent variable, this equals $\Sigma(X_0 - \overline{X}_0)^2\ (1 - r^2_{01})$. (b) Compute the unexplained sum of squares for the linear multiple correlation analysis in step 3. This equals $\Sigma(X_0 - \overline{X}_0)^2\ (1 - R^2_{0.1\ \ldots\ p})$. (c) Take the difference between these two sums of squares and divide it by P-1. This represents the variance owing to the differences between the regression coefficients in the linear analysis. (d) Divide the sum of squares obtained in step b by N-P-1. This represents the error or remainder variance. (e) Compute the ratio between the variance in step c and the variance in step d, and compare this with the tabular values in an F table, using P-1 and N-P-1 degrees of freedom. If the ratio obtained is larger than the tabular value at the 5-percent point, the differences between the regression coefficients are statistically significant. If this is approximately true, the weights determined in step 3 should be used, provided they appear logical.

A method of successive approximation to refine the various coefficients is next considered. This is used only if the F-ratio obtained in step 4e is equal to or larger than that at, say, a 10-percent probability level. If the regression coefficients do not differ significantly, the simple weighting procedure originally used in step 1 is satisfactory, and that analysis becomes the final one.

5. Obtain a new supply composite by multiplying each component by its partial regression coefficient from the linear analysis or by weights that are proportionate to these coefficients. In the case of first-difference analyses, the actual values are multiplied by the weights, the total obtained, and the first-difference logarithms are derived from this total.

6. Use this composite with the other variables originally used in step 1 to obtain a second approximation for the logarithmic analysis.

7. If the second approximations for the partial regression coefficients for the variables other than supply differ from the first approximations obtained in step 1, get a new series for the dependent variable adjusted for the effects of these nonsupply factors and obtain the antilogarithms of these.

8. Use the variable obtained in step 7 as the dependent variable in a linear analysis, with the same components of supply as independent variables as in step 3.

9. If the regression coefficients obtained in step 8 differ from those obtained in step 3, repeat steps 5, 6, 7, and 8 until the results converge to stable values. The final results yield a set of weights to be applied to the components of supply to give a composite factor of supply that will maximize the multiple correlation in the logarithmic analysis, for which price is the dependent variable and supply is one of several independent variables.

Application to an analysis relating to corn.--For the series of analyses discussed here, the basic analysis is a logarithmic one, using first differences, for which the dependent variable is the June-to-September average price received by farmers for corn. The independent variables are: (1) A composite feed-concentrate-supply variable, for which the weights are determined from a separate linear analysis; (2) a composite units of livestock production variable, for which the weights were partially determined from a separate linear analysis; and (3) average June-to-September index numbers of prices received by farmers for livestock and livestock products.

The dependent variable for the linear analysis involving the separate components of the supply factor was determined by adjusting year-to-year changes in the price of corn for the effects of year-to-year changes in units of livestock production and prices of livestock, using the highest-order partial regression coefficients from the logarithmic analysis. Likewise, the dependent variable for the linear analysis involving the separate components of the units-of-livestock-production variable was determined by adjusting year-to-year changes in the price of corn for the effects of year-to-year changes in supply of feed and prices of livestock, using the highest-order partial regression coefficients from the logarithmic analysis. Results for each succeeding approximation were used to improve the results for the following analysis.

In an iterative analysis of this type, values for certain coefficients may be assumed at the start, as the results of successive iterations converge toward the statistically most likely or "true" values. During the process of analysis, certain coefficients may be obtained which are inconsistent with a priori expectations. Particularly if the standard errors of these coefficients are large, consistent coefficients may be assumed or the variable may be omitted from the analysis, in which case a regression coefficient of zero is assumed. Table 6 lists the coefficients obtained or assumed for succeeding approximations for each of the three analyses involved in the over-all analysis of factors that affect the price of corn from June to September.

Results obtained from the second approximation for the analysis dealing with the components of the supply of feed concentrates were tested to determine whether the regression coefficients differed significantly from each other. An F of 3.30 was obtained, compared with a 5-percent probability value of 3.24. As the individual coefficients in general seemed reasonable, we decided to use them. As the coefficient for concentrates other than corn, oats, and barley fed ($b_{04.123}$) was positive and had such a high standard error, this factor was omitted. Year-to-year changes in this item are extremely small.

In the first analysis, data on livestock production units for the July-to-September quarter only were used. In the second approximation, data for the July to September quarter were weighted by 2 and for the October-to-December quarter by 1. The series on prices of corn was then adjusted for the effects of supplies of feed concentrates and prices of livestock. The adjusted price was used as the dependent variable in a linear analysis in which the two livestock production series were used as separate independent variables. These two variables explained only 7 percent of the residual variation in prices of corn. This is consistent with the highest-order partial coefficient of determination between prices of corn and production of livestock from the logarithmic analysis, which was only 0.05 for this approximation. The analysis indicated, however, that, if anything, October-to-December livestock production units (for which the sign of the regression coefficient was consistent with a priori expectations) were more important than July to September livestock production units (for which the regression coefficient had the "wrong" sign). In subsequent logarithmic analyses, these two series were given equal weights.

In the basic analysis of the factors that affect June-to-September prices of corn, the regression coefficient for the units-of-livestock-production factor ($b_{02.13}$) had the wrong sign on the first approximation. This occasionally occurs in analyses of this type and may merely indicate the need for additional approximations. In this instance, the dependent variable, in terms of first-difference logarithms, for the second approximation for the linear analysis involving the components of the supply of feed concentrates was obtained by use of the formula $X_{0.3} = X_0 - b_{03.1} (X_3 - \bar{X}_3)$. In other words, the computations for this approximation were based on results that would have been obtained had the X_2 factor been omitted from the analysis.

Table 6.--Corn: Successive approximations for the three sets of multiple correlation analyses used in connection with the basic analysis of factors that affect June to September prices

ANALYSIS DEALING WITH FACTORS THAT AFFECT THE PRICE OF CORN, JUNE TO SEPTEMBER

Coefficient	Approximation 1st	2d	3d	4th
$R^2_{0.123}$	0.634	0.738	0.770	0.758
$b_{01.23} \pm s_{b_{01.23}}$ ..	-.82 ± 0.36	-1.65 ± 0.39	-1.70 ± 0.34	-1.59 ± 0.34
$b_{02.13} \pm s_{b_{02.13}}$ ..	-1.06 ± .54	.23 ± .69	.76 ± .69	.70 ± .70
$b_{03.12} \pm s_{b_{03.12}}$ ..	1.21 ± .29	1.10 ± .24	.87 ± .23	.84 ± .23
ANALYSIS DEALING WITH COMPONENTS OF THE SUPPLY OF FEED CONCENTRATES				
$R^2_{0.1234}$	---	0.530	0.587	0.656
$b_{01.234} \pm s_{b_{01.234}}$	1/1	-2.69 ± 0.69	-2.88	-3.19
$b_{02.134} \pm s_{b_{02.134}}$	1/1	-.68 ± .40	-.78	-.81
$b_{03.124} \pm s_{b_{03.124}}$	1/1	-.85 ± .87	-.61	-.54
$b_{04.123} \pm s_{b_{04.123}}$	1/1	1.93 ± 3.80	1/0	1/0
ANALYSIS DEALING WITH COMPONENTS OF PRODUCTION OF LIVESTOCK				
$R^2_{0.12}$	---	---	0.071	---
$b_{01.2}$	1/1	1/2	1/2/1	1/1
$b_{02.1}$	1/0	1/1	1/2/1	1/1

1/ Assumed values.
2/ Calculated values for this analysis were as follows:

$b_{01.2} \pm s_{b_{01.2}}$, -1.82 ± 1.62; $b_{02.1} \pm s_{b_{02.1}}$, 0.96 ± 1.05.

In general, the multiple correlation coefficient for the basic analysis should be the same or higher for each succeeding approximation, and the regression coefficients should change in an orderly way. In the analyses discussed here, this takes place through the third approximation but does not hold for the fourth. This is believed to reflect rounding errors. The weights for the components of the feed-concentrate-supply factor were carried to two decimals only, which is equivalent to two significant figures for the smaller coefficients. Thus, there is no sound way of choosing between the third and the fourth approximations. The fourth was chosen mainly because the weights for the supply factor continue to change in an orderly way through that approximation.

Comparison of results for corn, oats, and barley.--Meinken (70) ran similar analyses for oats and barley for July to October when marketings are heaviest for these crops. In each case, the price received by farmers was the dependent variable and a first-difference logarithmic analysis was used covering the years 1922 to 1941. In these analyses, as for corn, the three

Table 7.--Corn, oats, and barley: Relative weights by which specified components of the supply of feed grains should be multiplied to obtain a composite supply factor for use in analyses of factors that affect their respective prices 1/

Item	Corn	Oats	Barley
Months covered by analysis	June-Sept.	July-Oct.	July-Oct.
Weight obtained for specified component of supply:			
Stock of corn, July 1:			
Weight	0.64	0.19	0.11
Standard error 2/	.11	.11	.08
Prospective new crop of corn:			
Weight	.16	.24	.14
Standard error 2/	.06	.06	.04
New-crop supply of--			
Oats:			
Weight	3/.10	.57	.20
Standard error 2/	.13	.17	.18
Barley:			
Weight	3/.10	0	.55
Standard error 2/	.13	---	.42
Sum of weights	1.00	1.00	1.00

1/ When all components are expressed in tons.
2/ Based on standard errors of respective regression coefficients.
3/ Components for oats and barley were combined in the analysis for corn.

variables--(1) a composite of supplies of old- and new-crop corn and new-crop oats and barley, (2) production of livestock from July to December, and (3) prices received by farmers for livestock and livestock products for the months included in the analysis--were used as independent factors. Weights for the components of supply were determined by a linear first-difference analysis through the method discussed in this section.

Table 7 (see page 127) compares the weights obtained from the analyses for oats and barley with those obtained for corn. In each case, the current supply of the commodity as such carries the largest weight, and the weights for other components are consistent with a priori expectations. The regression coefficient for barley in the analysis for oats had the wrong sign but, as it did not differ significantly from zero, its value was assumed to be zero in these computations. Results for the analyses for oats and barley confirm those obtained for corn in indicating that this general approach is useful in analyzing problems of this type.

An Experiment to Test the Relative Merits of Least Squares and Limited Information Coefficients for Forecasting Under Specified Conditions

As discussed in the section beginning on page 57, if an equation within a system of equations contains two or more variables that are endogenous and this equation is fitted statistically by the method of least squares, the estimates of the structural coefficients obtained will tend to be biased in the sense defined on page 57. Under the circumstances described beginning on page 67, the degree of bias may be small, but in general more accurate estimates of the structural coefficients can be obtained by methods that allow for any simultaneity that may exist within a system of equations than by a direct fit by the method of least squares.

But in practical matters of business and politics, the economist is not ordinarily asked to compute a structural coefficient. Rather, he is asked to estimate the expected value of one or more variables. These estimates--or forecasts--frequently are obtained by using coefficients estimated by statistical means and known or assumed values for certain related variables. A farmer may want an estimate of the expected price of apples in late winter. An economic analyst may provide such an estimate based on given data on apple production and consumer income.

With respect to problems of this type, there seems to be less agreement about the relative merits of least squares and of structural analysis. Some econometricians and many applied research workers believe that the best unbiased forecasts are obtained from equations fitted directly by least squares. Others apparently believe that more accurate forecasts can be made from equations derived by limited information. Klein (58, p. 253) implies that better estimates are obtained from "reduced form" equations derived from a structural approach because these take into account certain restrictions within the model.

To attempt to clarify questions with respect to the best method when we are chiefly interested in forecasts for a single variable, the author, in cooperation with Frederick V. Waugh, recently made a Monte Carlo analysis of some constructed data. We believe the results throw some light upon the proper uses of both the least-squares regressions and the structural equations. Possibly these results would have been anticipated by econometricians who have specialized in this area. Each of us must admit that we were surprised by some of the results. We realize also that this one test will not settle forever all arguments about the relative merits of these two approaches in economic research. But we sincerely hope that these results will help to bring about more understanding than now exists.

The model.--Other writers, such as Bennion (8), Bronfenbrenner (12), and Marschak (66), have used either theoretical or Monte Carlo models in an attempt to shed light on the relative merits of these alternative approaches. In general, none of these models are appropriate for the particular tests that we had in mind. The one used by Bennion, for example, yields exact values for any one variable given exact values of the other two. Hence it cannot be used to test the relative merits of alternative methods when we wish to predict one of the endogenous variables given a knowledge of other variables in the model.

Our model is a simplified version of the one used for wheat by Meinken (71) (see page 11). In economic terms, we have three price-determined utilizations (Y_1, Y_2, Y_3) that add to a fixed supply (Z_4), and price (Y_4) must be at a level to equate demand with supply. Each utilization is a function of (1) price and (2) a demand shifter that is unrelated to this system of equations. The disturbances in the three demand equations were computed so as to be correlated. To add variety, we specified that the coefficients on (1) price and (2) the demand shifter for equation (2) are identical. Coefficients on price were chosen so that one was greater than unity, one equaled unity, and one was close to zero. The following equations were used:

$$Y_1 = -0.1\ Y_4 + Z_1 + u_1 \tag{162}$$

$$Y_2 = -(Y_4 - Z_2) + u_2 \tag{163}$$

$$Y_3 = -4\ Y_4 + Z_3 + u_3 \tag{164}$$

$$Y_1 + Y_2 + Y_3 = Z_4 \tag{165}$$

$$u_1 = \tfrac{1}{2}(5v + w_1) \tag{166}$$

$$u_2 = 2(5v + w_2) \tag{167}$$

$$u_3 = 4(5v + w_3). \tag{168}$$

The Z's were drawn from a rectangular distribution, making use of random digits published in the Journal of the American Statistical Association. Each column of five digits was divided into two 2-digit columns and a 1-digit column that was ignored. All numbers of 40 or less in the 2-digit columns were recorded until sufficient numbers were obtained for the initial analyses. For those analyses involving forecasts when the Z's vary over twice the range used in the period of fit, all numbers of 80 or less were recorded. When the Z's fall outside the initial range, numbers of 60 to 80 were recorded.

The v's and w's were obtained in the following way, making use of the same tables of random digits: (1) Each column of 5 digits was treated as a unit. (2) If the last digit was even (or zero), the number was recorded as positive; if the last digit was odd, the number was recorded as negative. Numbers of 49998 or above were ignored. (3) Use was made of a table in Ostle (79, pp. 442-443) entitled "Areas of the Standard Normal Curve." A decimal point was mentally placed at the left of the first digit and the interval that included it in the right hand column of each section of the table was used to select the assigned number. For example, the first random number was 48190. The relevant portion of the table shows

2.09	.48169
2.10	.48214

Hence the first recorded number is 2.095. It is positive because the last digit in the 5-digit number is even. Similar procedures were used to obtain the remaining observations. Thus the v's and w's, and therefore the u's, are approximately normally distributed with mean zero.

The u's and Y's were then calculated from the given values of the Z's, v's, and w's. To avoid possible small sample bias, 100 observations were used in fitting the structural equations by, respectively, least squares and limited information. The results are shown in the tabulation on page 131.

If the first and third equations and the first formulation of the second equation are used, as in the original model, then the coefficients estimated by limited information are considerably closer to the coefficients used to generate the data than are those estimated by least squares and each coefficient is within two standard errors of the "true" value. All of the least squares coefficients, on the other hand, differ from the coefficients used to generate the data by more than three times their respective standard errors, although the standard errors estimated by least squares are, in each case, moderately to materially smaller than those estimated by limited information. It is possible, however, that the expected value, in a statistical sense, of the coefficients estimated by least squares differs from the coefficients used to generate the data and represents instead an average of a large number of equations that best forecast the Y_i ($i = 1, 2, 3$) given Y_4 and the respective Z's.

Neither method of fit gives good results for the second formulation of the second equation. The limited information coefficients differ from the "true" value by more than do the coefficients estimated by least squares, but they are each within two standard errors of the true value whereas the least squares estimates differ from the coefficients used to generate the data by materially more than twice their smaller standard errors.

The following tabulation shows the results obtained by fitting, respectively, by limited information and least squares:

First equation:

Used to generate data	$Y_1 = 0 - 0.1\ Y_4 + 1\ Z_1 + u_1$
Limited information fit 1/	$Y_1 = 1.4 \underset{(0.11)}{- 0.16}\ Y_4 \underset{(0.028)}{+ 0.963}\ Z_1$
Least squares fit 1/	$Y_1 = -1.5 \underset{(0.02)}{+ 0.19}\ Y_4 \underset{(0.013)}{+ 0.958}\ Z_1$
Related coefficients	$R^2 = 0.98 \quad r^2_{Y_4u_1} = 0.70 \quad r^2_{Z_1u_1} = 0.02$

Second equation:

First formulation:

Used to generate data	$Y_2 = 0 - 1\ (Y_4 - Z_2) + u_2$
Limited information fit 1/ ..	$Y_2 = -0.5 \underset{(0.13)}{- 1.14}\ (Y_4 - Z_2) + u_2$
Least squares fit 1/	$Y_2 = 6.6 \underset{(0.08)}{- 0.53}\ (Y_4 - Z_2)$
Related coefficients	$r^2 = 0.31 \quad r^2_{(Y_4-Z_2)u_2} = 0.26$

Second formulation:

Used to generate data	$Y_2 = 0 - 1\ Y_4 + 1\ Z_2 + u_2$
Limited information fit 1/ ..	$Y_2 = 6.0 \underset{(1.6)}{- 2.5}\ Y_4 \underset{(0.4)}{+ 1.4}\ Z_2$
Least squares fit 1/	$Y_2 = -5.5 \underset{(0.07)}{+ 0.27}\ Y_4 \underset{(0.05)}{+ 0.78}\ Z_2$
Related coefficients	$R^2 = 0.79 \quad r^2_{Y_4u_2} = 0.73 \quad r^2_{Z_2u_2} = 0.004$

Third equation:

Used to generate data	$Y_3 = 0 - 4\ Y_4 + 1\ Z_3 + u_3$
Limited information fit 1/	$Y_3 = 6.6 \underset{(1.1)}{- 4.4}\ Y_4 \underset{(0.27)}{+ 0.93}\ Z_3$
Least squares fit 1/	$Y_3 = -11.0 \underset{(0.12)}{- 1.45}\ Y_4 \underset{(0.09)}{+ 0.56}\ Z_3$
Related coefficients	$R^2 = 0.62 \quad r^2_{Y_4u_3} = 0.77 \quad r^2_{Z_3u_3} = 0.003$

1/ Numbers in parentheses are the standard errors of the respective coefficients.

Certain coefficients of multiple or simple determination are shown below each equation. The first is the coefficient of multiple determination estimated by least squares, except for the first formulation of the second equation, where the simple coefficient is relevant. The second shows the simple coefficient between the endogenous variable treated as independent in the method of least squares and the unexplained residual in that equation. The third shows the simple coefficient between the predetermined variable and the unexplained residual in each equation. The latter is assumed by the model to be negligible and in no case exceeds 0.02.

Forecasts from the price-utilization equations.--In this experiment, we first assume that the econometrician is asked to forecast--or estimate--Y_1, Y_2, and Y_3, using equations like (162), (163), and (164). He does not know the parameters in these equations, but must estimate them from sample data. We assume that he knows the value of Y_4 and the Z's in the forecast period from an independent source. For purposes of estimation, he may fit three least-squares regression equations, using Y_1, Y_2, and Y_3, respectively, as dependent variables and Y_4 and the respective Z's as independent variables. 75/ Or he may use the limited information method to obtain estimates of the structural coefficients. Here the several Y's--or simultaneously determined variables--are treated as endogenous and the Z's, as predetermined. A second approach for predicting the first three Y's from the Z's, with or without a knowledge of Y_4, is discussed in the next section.

Four types of test samples were used, each involving paired samples of 25 observations: (1) When the Z's were drawn over the same range as for the observations used in the initial fit and (a) varied from observation to observation or (b) were fixed over the 25 observations, (2) were drawn over twice the range, and (3) were drawn completely outside the range of the Z's used in the initial fit. The latter have some relevance in applied work, as for example when an analysis for a period prior to World War II is used for forecasting in a period following World War II. In each case, the range in the v's and w's was the same as for the initial observations.

By making use of the alternative estimates of the coefficients in the price-utilization equations, as shown in the tabulation on page 131, and known values for Y_4 and the several Z's, estimates of the first three Y's were obtained for each of the 8 test samples. Differences between the actual and computed utilizations for each equation were obtained. For the observations used in the initial fit, these differences add to zero. In a test sample, however, they would not necessarily add to zero and, for some methods and some samples, the arithmetic sum differed considerably from zero. If the symbol d

---

75/ As discussed in the section beginning on page 51, there are several schools of thought concerning the correct method to choose the variable to be treated as dependent in a least-squares analysis. Here we choose as the dependent variable the one that is to be estimated and as independent variables those whose values are given at the time of forecast.

is used to represent these differences, estimates of the mean squared error should be computed by the formula:

$$s_d^2 = \frac{\Sigma d^2}{N - 1} . \qquad (169)$$

This is used rather than the usual formula for the square of the standard deviation which is

$$s_d^2 = \frac{\Sigma d^2 - \bar{d}\,\Sigma d}{N - 1} \qquad (170)$$

because the latter does not take into account errors caused by a faulty estimate of the level of the regression curve.

A comparison of the mean squared errors for the 4 samples for which the Z's were drawn over the same range as in the period of fit indicates that in every case they are smaller for the estimates based on coefficients obtained by the method of least squares than for those obtained by limited information even though these coefficients differ materially from those used in constructing the model. (See table 8.) Ratios of the alternative pairs of mean squared errors range from 1.4 to 14.6. For the two samples for which the Z's were drawn over twice the range, the mean squared errors based on coefficients obtained from the equations fitted by least squares are smaller for 4 comparisons and larger for 2 comparisons. For the two samples in which the Z's lie outside the initial range, the mean squared errors based on least squares are smaller for 2 comparisons and larger for 4 comparisons. In each of the last-named comparisons, however, the ratio of the larger to the smaller mean squared error ranges only between 1.1 and 2.1. Variations between samples within any pair of samples in general are greater when the Z's are drawn outside the initial range.

Mean squared errors are smaller in most cases for the least squares coefficients than for those computed by limited information, even though the latter are closer to the coefficients used to generate the data, because the unexplained residuals in the structural relations are correlated with the values of Y_4. The structural coefficients do not allow for this correlation, whereas the least squares estimates include both the structural relation between Y_i $(i = 1, 2, 3)$ and Y_4 and also the relation between Y_4 and the respective u_i. This is the reason both for (1) the difference between the least squares estimates of the coefficients and the structural coefficients and (2) the better forecasts of Y_i given by the least squares equations <u>given a knowledge of Y_4</u>.

This point is so fundamental that it seems desirable to illustrate it graphically. The samples for which the Z's are fixed are ideal for this purpose, because the scatter between the respective Y_i $(i = 1, 2, 3)$ and Y_4 are

Table 8.--Difference between actual and computed utilizations when price is given and estimates are made from the price-utilization equations: Mean squared error for alternative methods and samples

When the Z's are drawn over--	Sample: First, Least squares	Sample: First, Limited information	Sample: Second, Least squares	Sample: Second, Limited information
The same range as in the period of fit and--				
They vary:				
Y_1	3.3	8.5	2.1	9.4
Y_2	95.4	134.5	85.4	170.4
Y_3	124.3	549.7	209.4	675.9
Are fixed:				
Y_1	3.0	9.8	.8	11.1
Y_2	70.5	150.3	107.3	173.3
Y_3	99.5	603.0	94.4	703.0
Twice the range:				
Y_1	7.7	9.9	7.9	10.8
Y_2	121.2	99.4	165.3	208.3
Y_3	755.3	435.8	515.7	622.0
Outside the range:				
Y_1	14.4	13.1	13.8	24.7
Y_2	221.5	189.7	339.3	158.2
Y_3	931.8	498.2	924.2	952.2

in effect partial relationships for the fixed Z. Figure 8 shows the scatter between Y_1 and Y_4 for the first of these samples. Similar results are given for Y_2 and Y_3. Because a positive correlation exists between Y_4 and u_1, the line of best fit differs in slope from the structural relation between these two variables. The least squares equation based on the initial sample of 100 observations is not a perfect fit for this sample of 25 observations, but it is much closer to a fit than is either the limited information estimate or the structural relation. We believe that this is the point that Bennion (8) had in mind, but this seems a clearer example.

Forecasts based on all of the predetermined variables.--An alternative method of estimating the Y's, which can be used without a knowledge of Y_4, is to make use of the reduced form equations that involve the four given or predetermined variables in the model. To make this test, the following equations first were fitted by least squares, using the respective Y's as dependent variables and the several Z's as independent variables. 76/

76/ The numbers beneath the coefficients in each case are the respective standard errors.

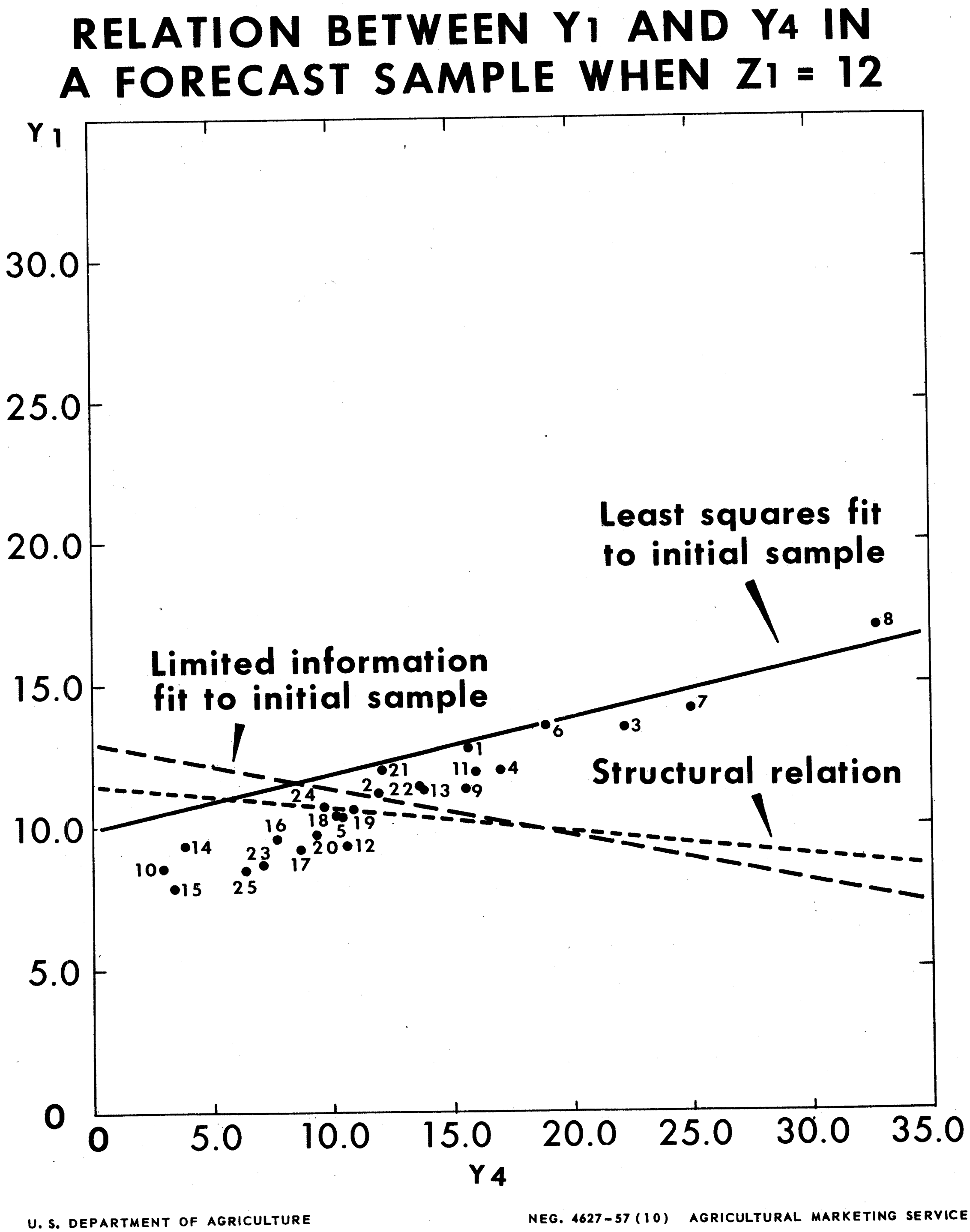


Figure 8.--A regression equation obtained by least squares more nearly fits the data that show the relationship between two endogenous variables than does the true structural relation because the endogenous variable that is treated as independent in the least squares equation is correlated with the unexplained residuals in the structural equation.

$$Y_1 = \underset{}{1.5} + \underset{(0.02)}{0.95}\, Z_1 - \underset{(0.02)}{0.01}\, Z_2 - \underset{(0.02)}{0.05}\, Z_3 - \underset{(0.02)}{0.00}\, Z_4 \qquad R^2 = 0.97$$

$$Y_2 = 3.5 - \underset{(0.04)}{0.27}\, Z_1 + \underset{(0.04)}{0.81}\, Z_2 - \underset{(0.04)}{0.24}\, Z_3 + \underset{(0.04)}{0.16}\, Z_4 \qquad R^2 = 0.87$$

$$Y_3 = -5.0 - \underset{(0.06)}{0.68}\, Z_1 - \underset{(0.05)}{0.80}\, Z_2 + \underset{(0.06)}{0.28}\, Z_3 + \underset{(0.06)}{0.85}\, Z_4 \qquad R^2 = 0.85$$

$$Y_4 = 4.2 + \underset{(0.06)}{0.10}\, Z_1 + \underset{(0.06)}{0.24}\, Z_2 + \underset{(0.06)}{0.14}\, Z_3 - \underset{(0.06)}{0.27}\, Z_4 \qquad R^2 = 0.27$$

The last equation shows price as a function of production and the demand shifters in the respective demand equations. The other equations normally would not be used except in connection with a system of equations.

Next, the following equation for Y_4 was computed by algebra from the coefficients obtained by the limited information approach, using the method described in Friedman and Foote (40, p. 84):

$$Y_4 = -1.33 + 0.17\, Z_1 + 0.20\, Z_2 + 0.16\, Z_3 - 0.18\, Z_4.$$

Once a value for Y_4 is obtained, this value and the known values of the Z's can be substituted in the first and third equations and the first formulation of the second equation shown in the tabulation on page 131, using the equations obtained by the limited information fit, to obtain computed values for the remaining Y's. Results of this approach differ from those obtained from the reduced form equations, as fitted by least squares, because each of the equations in this system is overidentified.

As before, differences between the calculated and actual values for the four Y's were obtained, and related mean squared errors were computed for each of the first six samples used in the preceding test. The results are shown in table 9.

To justify the elaborate fitting procedures involved in the limited information approach so far as forecasting is concerned, better forecasts should be given by the algebraically-derived limited information equations than by the reduced form equations fitted by least squares. In 17 out of 24 comparisons, the mean squared error is less for the limited information than the least squares forecasts, but the average difference differs from zero by a statistically significant amount only if we use a probability level of about 20 percent. 77/ Except for greater variability of results in the latter case,

---

77/ In Cowles Foundation Discussion Paper No. 11, dated April 20, 1956, entitled "The Efficiency of Estimation in Econometric Models," L. R. Klein emphasizes that, "While the squared discrepancies between predicted and actual values of endogenous variables in the model will be smaller over the sample period if calculated from (unrestricted) least squares estimates of the no consistent differences in these relationships were found when the Z's were drawn over (1) the same range as in the period of fit and (2) twice the range.

Table 9.--Difference between actual and computed endogenous variables when estimates are made from all of the predetermined variables in the system: Mean squared error for alternative methods and samples

When the Z's are drawn over--	Sample			
	First		Second	
	Least squares	Limited information	Least squares	Limited information
The same range as in the period of fit and--				
They vary:				
Y_1	4.4	3.8	4.6	3.9
Y_2	18.1	18.9	23.2	25.9
Y_3	38.4	36.6	46.0	45.7
Y_4	47.4	44.0	55.3	54.2
Are fixed:				
Y_1	4.8	4.3	5.3	4.8
Y_2	24.4	22.6	21.0	23.3
Y_3	48.9	44.3	43.8	45.7
Y_4	53.0	48.6	57.1	57.0
Twice the range:				
Y_1	5.5	5.8	5.3	5.2
Y_2	26.7	22.9	25.2	50.9
Y_3	55.1	29.1	50.6	70.5
Y_4	58.0	33.0	59.0	53.1

Comparison of the respective mean squared errors shown in tables 8 and 9 gave results that at first thought were surprising but, on further examination, appear reasonable. In the computations shown in table 8 we assume that price (Y_4) is known and then estimate each of the utilizations based on this price and the exogenous variable in the respective structural equation. In table 9, we ignore this available information with respect to price but make

---

reduced form equations than if calculated from any set of estimates of structural parameters using *a priori* restrictions," the algebraically-derived estimates will tend to give better predictions outside the sample. A proof is given of the proposition that "the more one uses valid information in the form of *a priori* restrictions imposed on the system, the more efficient are the estimates."

use of all of the Z's in the system. Mean squared errors for the limited information estimates in table 9 are in every case smaller than the corresponding mean squared errors in table 8 and in many cases are materially smaller. Moreover, they also are smaller than the corresponding least squares estimate in table 8 in each comparison except for Y_1 when the Z's are drawn over the same range as in the period of fit. Again the differences are substantial in most instances. Mean squared errors for the least squares estimates in table 9 are smaller than the corresponding values in table 8 in the same set of circumstances. The implication of these results is that even if somehow we knew or could estimate Y_4 in advance of making a specific forecast, we would do better to forecast the several utilizations from all of the predetermined variables in the structural system than by making use of the estimate of the endogenous variable Y_4 and the specific Z occurring in equations (162), (163), or (164) in either the limited information or the least squares estimate of the structural equation for the particular utilization. These results have important implications with respect to the kinds of variables that should be included in equations designed basically for forecasting.

Forecasts based on all of the predetermined variables and the known price.--What happens if we make use of the known price in addition to all of the predetermined variables in the system? At least for the initial sample, we know that our least squares estimates can be no worse than if we ignore this information. But, on first thought, we might not expect much improvement since this price itself is based on the several predetermined variables in the model. As degrees of freedom, as well as computational time, are always a problem when we work with multiple regression analyses, it seemed worthwhile to find out how much we gain by including in our estimating equation such endogenous variables as are expected to be known at the time of forecast.

This test was conducted in two stages: (1) We determined the extent to which the multiple coefficient of determination for the initial sample was increased when the least squares estimating equations for Y_1, Y_2, and Y_3, respectively, include Y_4 in addition to the four Z's in the model. The results are shown in table 10. In every case the multiple coefficient of determination is increased substantially by the inclusion of Y_4. (2) For the first and third pairs of test samples, we determined the accuracy of forecasts based on the least squares regression equations when Y_4 is included. The results are shown in table 11. In every case the mean squared errors are smaller than by any of the other methods that were used. Thus it is evident that if we know the value of one of the endogenous variables at the time of forecast, we should fit a least squares equation that includes this as one of the independent variables and use this equation as a forecasting mechanism.

When we work with the algebraically-derived equations, we have no way of making use of an advance knowledge of Y_4 except by using this price directly in the structural equations which, as we have seen, results in poor forecasts. At least for this model, the slightly greater accuracy of prediction which the algebraically-derived reduced form equations have over the least squares reduced form equations disappears completely if we compare the limited-information results as shown in table 9 with those obtained from the least squares equations when Y_4 is included as an independent variable as shown in table 11.

Table 10.--Initial sample: Multiple coefficients of determination for least squares equations that include all of the predetermined variables

When price (Y_4) is--	Dependent variable		
	Y_1	Y_2	Y_3
Excluded	0.97	0.87	0.85
Included	.997	.98	.98

Table 11.--Difference between actual and computed endogenous variables when estimates are made from Y_4 and all of the predetermined variables in the system: Mean squared error for alternative samples when use is made of least squares regressions

When the Z's are drawn over--	Sample					
	First			Second		
	Y_1	Y_2	Y_3	Y_1	Y_2	Y_3
The same range as in the period of fit	0.5	5.06	3.35	0.1	2.56	5.31
Twice the range	.4	5.39	5.39	.3	5.12	5.42

Forecasts under a changed structure.--For the two samples in which the Z's vary from observation to observation but were drawn over the same range as for the observations used in the initial fit, the model was recomputed by omitting the second equation. In a real situation this would be representative of a commodity with three major utilizations in the period of fit but only two in the period of forecast. The results, which are shown in table 12, differ only slightly from those obtained for the basic model. The least squares estimates are far superior to those from the limited information approach when use is made of the price-utilization equations. However, use of the forecasting formulas derived by algebra from the limited information coefficients when the second equation is omitted resulted in errors that are smaller or about the same for Y_1 but much smaller for Y_3 than either set obtained by use of the price-utilization equations.

We first thought that no use could be made of the reduced form equations fitted by least squares under this sort of a change in structure. Subsequently, we decided to try these equations when Z_2 is omitted. An adjustment was, of course, required in the constant term. Results are shown in the table when Y_4 is (1) excluded and (2) included. Particularly when Y_4 is included, the mean squared errors are considerably smaller than when estimates are made from the structural equations. Best estimates for Y_3, however, are obtained from the algebraically-derived reduced form equations. This may result from the particular form of the system of equations, since the error in Y_3 can be no greater than that for Y_1 since these two endogenous variables must add to a fixed sum. For this phase of the experiment, a model that involved four utilizations would have been preferable.

Table 12.--Difference between actual and computed endogenous variables when Y_2 is omitted from the structure: Mean squared error for alternative methods when the Z's are drawn from the same range as in the period of fit

Method	Sample					
	First			Second		
	Y_1	Y_3	Y_4	Y_1	Y_3	Y_4
When Y_4 is given and use is made of price-utilization equations fitted by--						
Least squares	4.5	181.9	---	3.9	316.8	---
Limited information	8.4	544.2	---	8.8	643.3	---
Estimated from Z_1, Z_3 and Z_4 using equations obtained by--						
Least squares when Y_4 is--						
Excluded	4.7	166.4	44.4	5.6	155.3	34.5
Included	2.3	88.8	---	2.3	139.0	---
Algebraic derivation from the limited information coefficients when the second equation is omitted from the model	4.2	4.2	33.2	4.1	4.1	38.9

<u>Other tests that might be made</u>.--The tests discussed here answer most of our questions with respect to how to forecast from a price-utilization model when price is at its equilibrium level. Additional tests would be desirable

to determine (1) how to forecast several interrelated utilizations, either singly or in combination, when price is at a level considerably different from its equilibrium value, as is apt to be the case under Governmental regulation or support, and (2) forecasting techniques when supply is simultaneously determined with demand.

Conclusions.--1. Table 13 summarizes the basic conclusions with respect to the relative merits of the several methods of estimation when applied to samples outside the initial period of fit. Results are compared for five alternative methods of estimating the respective utilizations and for two alternative methods of estimating price. For three of the methods of estimating the utilizations, price is assumed to be known exactly from an outside source. In comparing some of the alternative methods, either (a) no consistent superiority was in evidence or (b) the results, although slightly superior for one method or the other, were not sufficiently different to be of practical significance. In such cases, an identical ranking number is assigned to each method.

In forecasting price, the two methods, each based on all of the predetermined variables in the system, are closely similar, but the algebraically-derived limited information coefficients have a slight advantage, particularly when the Z's are from twice the initial range.

With respect to forecasts of the utilizations, the least squares estimates derived from all of the Z's and the known Y_4 give the best forecasts except when Y_2 is omitted from the structure; in the last named case, the algebraically-derived limited information estimates based on all the Z's are best, but this may result from a particular characteristic of this model. Poorest forecasts are obtained when use is made of the structural equations and the known Y_4; in this phase of the experiment, the coefficients obtained by least squares give better forecasts than those obtained by limited information even though the latter are closer to the true structural values. This results because the unexplained residuals are correlated with Y_4.

2. The importance of classifying variables into (a) endogenous and (b) predetermined--which is equivalent to specifying whether they are or are not correlated with the unexplained residuals in the structural equations--is upheld when we are interested in estimating the structural coefficients, although (as noted on page 67) the statistical bias that results if we use the method of least squares rather than a simultaneous equations approach to estimate structural coefficients tends to be small unless the several unexplained residuals are themselves highly correlated. In connection with forecasts, however, this distinction is not important. We should use as independent all variables that are expected to be known at the time of forecast. As noted in item 1, the forecast equation, if fitted by least squares, should include all predetermined variables in the system plus any endogenous variables that are expected to be known.

3. All of these conclusions apply when price is at its equilibrium level. Further research is needed to determine how forecasts should be made when price is at some other level, as it might be under government controls.

Table 13.--Approximate ranking of alternative methods of forecast in order of minimum mean squared error: Specified conditions 1/

WITH RESPECT TO THE THREE UTILIZATIONS

Method	With no change in structure when the Z's are from the-- Same range as in the period of fit	With no change in structure when the Z's are from the-- Twice the range	With a change in structure due to the omission of Y_2 when the Z's are from the same range as in the period of fit
When Y_4 is given and use is made of the price-utilization equations fitted by--			
Least squares	3	3	4
Limited information	4	3	5
Estimated from all predetermined variables in the system using equations obtained by--			
Least squares when Y_4 is--			
Excluded	2	2	3
Included	1	1	2
Algebraic derivation from the limited information coefficients	2	2	1
WITH RESPECT TO PRICE			
Estimated from all predetermined variables in the system using equations obtained by--			
Least squares	1	2	1
Algebraic derivation from the limited information coefficients	1	1	1

1/ Identical rankings indicate inconclusive or nearly identical results. The best method is identified by the number 1.

Allowance for Errors in the Data

In a statistical analysis, we have no way of distinguishing between that part of the unexplained residual that is due to errors in the data and that part that is due to omitted variables. 78/ For this reason, methods have been developed to handle equations for which the variables are (1) subject to error but in which the true variables are assumed to be related in a functional way --called "error" models--or (2) known without error but assumed to be related in a stochastic way--called "shock" models, but practically no progress has been made in developing methods for handling stochastic relationships when the variables are subject to error--a "shock-error" model. The latter is, of course, the situation which prevails in any empirical analysis based on time series data. All of the statistical methods discussed so far are based on an assumption that the independent or predetermined variables, at least, are known without error.

Adjusting the sums of squares for assumed errors.--Fox suggested a way to modify the method of least squares to correct for errors in the data, provided we are willing to make an informed guess about the average magnitude of the errors. 79/ Such guesses or estimates can be made by carefully reading how the series was compiled, or preferably, by talking to the person in charge of the compilation. For most series, we have a fairly good idea as to whether the error is in terms of, say, a few percent, 10 to 20 percent, or something higher. If we wish, we can experiment with various assumptions about the level of error, compare the effect on our coefficients, and then arrive at a rather good idea as to the effect which likely errors may have on their magnitude.

The nature of this correction can be shown easily in the following way. When we express variables in terms of deviations from their respective means, the least squares regression coefficient between two variables is given by:

$$b_{yx} = \frac{\Sigma yx}{\Sigma x^2} . \tag{171}$$

Suppose that y and x are each subject to error and that the magnitude of the error is given by d and e, respectively. The simple regression coefficient then equals:

$$b^*_{yx} = \frac{\Sigma(y+d)(x+e)}{\Sigma(x+e)^2} . \tag{172}$$

Let us now expand the numerator and denominator. We obtain:

78/ Effects of the omitted variables frequently are referred to as "shocks" or "random shocks" in econometric literature.

79/ This approach and some examples based on it first was published in Foote and Fox (29, pp. 29-35).

$$\Sigma(y+d)(x+e) = \Sigma yx + \Sigma ye + \Sigma dx + \Sigma de \tag{173}$$

$$\Sigma(x+e)^2 = \Sigma x^2 + 2\Sigma xe + \Sigma e^2. \tag{174}$$

If d and e are random and independent, and we have a large sample, any summation term that involves one or both of them as a cross-product equals approximately zero. Summation terms that involve their squares, however, do not equal zero. Applying this principle, when the variables are subject to error, we can write the value of the regression coefficient in the following way:

$$b^*_{yx} \cong \frac{\Sigma yx}{\Sigma x^2 + \Sigma e^2} . \tag{175}$$

This gives a biased estimate of b_{yx} because the denominator is too large. But if we reduce the sum of squares for the independent variable by an amount proportional to the percentage error, then we obtain an estimate of the regression coefficient that is approximately unbiased provided the only source of bias is that due to errors in the data. Such a correction can be made easily if we know the level of error approximately. The reader will note that the bias is independent of any error in y, as the only term involving y is a cross-product.

One way to obtain the simple coefficient of determination, or the square of the correlation coefficient, is from the following formula:

$$r^2_{yx} = \frac{(\Sigma yx)^2}{\Sigma y^2 \; \Sigma x^2} . \tag{176}$$

Although an algebraic derivation comparable to that used in obtaining equation (175) involves too much detail to give here, it is intuitively obvious and easily can be shown that if we reduce the sum of squares for both the dependent and the independent variables by an amount proportional to their respective percentage errors, we obtain an estimate of the coefficient of determination that is approximately unbiased provided the only source of bias is that due to errors in the data.

In working with multiple regression and correlation coefficients, similar reasoning suggests the desirability of reducing the sums of squares for all variables by amounts proportional to their respective errors. Insofar as this general approach is correct, we see no reason why the same sort of correction could not be applied to the sums of squares that are involved in equations to be fitted by the simultaneous equations approach. Some empirical evidence that the whole approach may be in error is presented immediately after the following example which shows one way to use this method of allowing for estimated errors in the data as an analytical tool.

In analyses of the demand and price structure for lettuce based on single equations, Shuffett (84, pp. 21-28) obtained coefficients of multiple determination ranging between 0.06 and 0.19. He cites evidence to support the hypothesis that the poor results may be caused at least in part by (1) simultaneity between demand and supply relationships, since part of the crop frequently is unharvested for economic reasons and (2) errors in the data. As partially unharvested acreage and some totally unharvested quantities are not reported in the statistics on lettuce, he felt that the simultaneous relations could not be fitted with available data. By use of the technique suggested by Fox, however, he did arrive at estimates of possible effects of errors in the data.

He says: "Some of the unexplained price variation for lettuce may have occurred because of the errors in measurement of the variables involved. According to estimates of the former Bureau of Agricultural Economics (94), between 25 and 50 percent of the total supply of lettuce was not included in commercial supplies during the years for which the statistical analyses were fitted. Noncommercial production may be correlated with commercial production to some extent, but the home-garden production would tend to remain relatively more stable from one year to another than commercial production, as production for home use would not be expected to respond much to prices in previous years. Some error also is involved in the price estimates because the point within the marketing system to which the prices apply varies from State to State. As relative production varies from year to year, this factor would cause some distortion in the series relating to season average prices for the country as a whole" (84, p. 29).

He considers an analysis of factors that affect the season average price of lettuce which uses as independent variables annual production and disposable income per capita. All variables were expressed as first differences of logarithms and the equation was fitted for the years 1921-41. A coefficient of multiple determination of 0.08 was obtained when none of the variables were adjusted for possible error, and the absolute value of the net regression coefficient on production was 0.4. An analysis based on the same variables but assuming random measurement errors of 25 percent for production and 10 percent for season average price increased the coefficient of multiple determination to 0.12 and the absolute value of the regression coefficient on production to 0.7. Computations based on a 50-percent random error component in production and 25 percent in disposable income raised the coefficient of multiple determination to 0.41 and the regression coefficient to 2.3.

Shuffett (84, p. 32) concludes, "These computations indicate that an extremely high level of error must be assumed in order to raise the coefficient of multiple determination to a statistically significant level. They also indicate that, although the coefficients for the unadjusted analysis may be biased toward zero to a considerable extent, an extremely high level of error must be assumed to change the conclusion that demand at the farm level is elastic. Satisfactory analyses designed to measure the elasticity of demand for lettuce probably would require both an allowance for a moderate level of error in the data, as in the first adjusted analysis illustrated here, and an allowance for the simultaneous determination of quantities moving into marketing channels and price as discussed in the preceding section."

A Monte Carlo study of the effect of errors of observation.--Ladd (63) presents results from an experimental study of an overidentified 2-equation model for which the only basic assumption (see page 58) which is violated is that the variables are known without error. Thirty samples of 30 observations each were constructed and fitted by both the method of least squares and that of single equation limited information. In one sample, all of the variables were taken without error, whereas in a set of 30 samples, normally distributed random errors of a magnitude such as to give an average relative error variance ranging from 0.08 to 0.265 for each of the endogenous and predetermined variables were used. The errors were computed in such a way as to be distributed serially independently and independently of each other, of the true variables, and of the random shocks. Comparisons were made between the results obtained from the single sample for which the variables were known without error and an average for the 30 samples for which errors were taken in the variables.

Although the results of any experimental study of this sort may reflect in part the particular model used in the test, the findings are helpful in appraising the potential usefulness of methods like that proposed by Fox. The following summary of this study is based essentially on that given by Ladd (63, pp. 294-295):

(1) The presence of errors of observation imparts little statistical bias to either the least squares or the limited information coefficients, but does increase the standard errors of these coefficients. (The first finding directly contradicts the assumption made by Fox.)

(2) The distribution of the limited information estimates approaches the normal distribution quite rapidly.

(3) A standard t-test (see page 178) rejects the null hypothesis that the two methods of estimation give the same average values on five of the six coefficients; on two because of their sizeable least squares bias, on one because of the different impact of the random shocks on the two methods, and on two because of the differential effects of errors of observation on the two methods.

(4) The least squares method can be applied directly to a structural equation having endogenous variables on the right of the equality sign if the covariance between each explanatory endogenous variable and the random shock is small. (This confirms the viewpoint presented on page 67.) Even in cases in which this covariance is quite high, a coefficient may be obtained whose least squares bias is negligible, but we have no way of knowing in advance whether this will be the case.

(5) The estimated limited information standard errors and variances understate the reliability of the corresponding coefficients because of an

inherent characteristic of the method, the presence of errors of observation in the data, or a combination of these two factors.

(6) In an equation with moderate or high covariance between an explanatory endogenous variable and the random shock, the least squares standard errors overestimate the true standard errors. (This contradicts the findings from the experimental model discussed beginning on page 128.)

Comments of Kendall and Wold with respect to this problem.--In part because of the conflict in conclusions with respect to the effect of errors in the data on the coefficients as assumed by Fox and as found by experiment by Ladd, we conclude this section by quotations from two mathematical statisticians who have investigated this area rather carefully.

In a detailed discussion of structural, regression, and functional relationships, Kendall (55, pp. 20-24) includes comments with respect to a number of 2-variable cases in which one or both of the variables is subject to error. If x' and y' are the observed variables that are known to be subject to error, Kendall mentions that Lindley (64) has shown, among other things, that the regression of y' on x' is not linear even if that of y on x (the true variables) is linear unless certain conditions hold. In particular, if the distribution of x is normal, only normality in v (the error in x) will preserve the linearity of y' on x'. Kendall then derives a regression coefficient similar to that given in equation (175) but comments that even if we assume that the distribution of v (the error in x) is known with known variance, the problem of testing it for significance is formidable. He concludes his discussion of this topic as follows: "The errors of observation impair our estimators, vitiate our tests of significance and even bend our regression lines unless we are prepared to postulate normality in the variates, which is asking rather a lot, particularly in economic work. But they are not quite as devastating as they look. A slight departure from linearity will sometimes allow the ordinary theory to be used as an approximation. None the less, it would appear that much more effort is needed to reduce errors of observation than has sometimes been supposed" (55, pp. 23-24).

As a summary to a similar discussion of the effects of errors of observation, Wold (104, p. 44) concludes as follows: Corrections for errors in the explanatory variables "have been worked out on the assumptions that the errors are independent of the error-free variables and that the error variances are known a priori. ... The disadvantage in such correction methods is that in practice we have little or no information about the observational errors, neither of their presence nor of their distributional properties. Hence, if it is felt that the data contain observational errors that are not negligible (negligible relative to the influences upon the effect variable that are due to unspecified causal factors), it is dangerous to employ the data for more than a tentative orientation."

Effects of Serial Correlation in the Error Terms 80/

In this section we summarize recent work on the statistical effects of possible nonindependence in the true error terms of an economic relationship. This detailed summary is included in this handbook for two purposes: (1) The original literature is highly mathematical and is not summarized well in any other place and (2) actually little is known in a general way about the effects of serial correlation in the error terms. By spelling out in detail what is known, the reader can see by default how much is not known in this area.

Nonindependence in the error terms, according to Cochrane and Orcutt (15, pp. 36-38), usually takes the form of positive serial correlation of the error terms for the following reasons: "(1) Systematic errors may arise from a faulty choice of the form of relationship assumed to exist between economic variables. Since the economic variables are positively autocorrelated, 81/ then, in general, errors of this type will be positively autocorrelated. Further the shortness of most available time series makes the statistical results meaningless if very complicated relationships are adopted, so that errors of this type are inevitable.

"(2) Error terms may arise owing to the omission of variables, both economic and noneconomic, from the analysis. Important variables may be omitted either because they are not available or because their importance is not realized. Furthermore, because of the brevity of available time series, it is also frequently necessary to neglect variables which individually have but small influence. Nevertheless, it is evident that the total influence of a number of such variables may be very substantial and highly positively autocorrelated. ...

"(3) The series of data used may not measure exactly what is required for the particular analysis. Insofar as the discrepancy is one of coverage, it seems reasonable to believe that the error term involved will have much the same autoregressive properties as economic series in general. Insofar as the discrepancy is more nearly what might perhaps be called pure error of observation, it would appear more difficult to say anything about whether or not it is autocorrelated. 82/ ..."

80/ This section was prepared by Marc Nerlove, agricultural economic statistician, Agricultural Marketing Service.

81/ Evidence for the tendency of economic variables to be positively autocorrelated is presented in Orcutt (77). Economic variables are, of course, not necessarily positively autocorrelated.

82/ Because of the way in which agricultural series are constructed, it seems likely that such errors in the series as occur will tend to be positively autocorrelated. For a description of the estimating procedures used to construct agricultural series see U. S. Dept. Agr. Mis. Pub. 703 (95); and for a discussion of some of their statistical properties, see Nerlove, Marc, Thus, we not only expect nonindependence in the error terms, but we expect this nonindependence to take the form of positive autocorrelation. In most of the following discussion, however, we do not restrict ourselves to positive autocorrelation but consider temporal nonindependence more generally. In order to understand some of the recent work on the statistical effects of nonindependence, we introduce the notion of a stochastic process.

Stochastic processes 83/.--A stochastic process is simply a generalization of the notion of a random variable. 84/ As is well-known, the probability that a random variable x is less than a particular value u defines the cumulative distribution function of the random variable, F(u). This function has the following properties:

$$\left.\begin{aligned} 0 \leq F(u) &\leq 1 \\ \lim_{u \to \infty} F(u) &= 1 \\ \lim_{u \to -\infty} F(u) &= 0 \\ \frac{dF}{du} &\geq 0. \end{aligned}\right\} \qquad (177)$$

The idea of a random variable may be generalized easily to that of a random vector. In this case the cumulative distribution function has an many arguments as there are components of the random vector. If, for example, the random vector has n components, its cumulative distribution function has n arguments.

In principle, the notion of a random vector and a cumulative distribution function are valid irrespective of the nature of the measurements or of the dimensionality. Specifically, they apply if the random vector is a time series of any finite number of observations. Since the observations of a time series are, in principle, unlimited in number, an extension of the notions of an n-dimensional random vector and cumulative distribution function is required. We are thus led to consider a random vector of the form

---

Estimates of the Elasticities of Supply of Corn, Cotton, and Wheat, Ph.D. thesis, Johns Hopkins University, 1956, pp. 101-138. Statistical problems resulting from errors in the variables are discussed briefly in the preceding section.

83/ This section is based largely on Wold (102, pp. 1-8, 31-66) and Wold and Jureen (106, pp. 149-168). Readers who are unacquainted with calculus and matrix algebra may prefer to skip to the concluding paragraphs of this section, which begin on page 169.

84/ Cramér (16, pp. 57-58) gives a clear but nonrigorous definition of a random variable.

$$x = (\ldots, x_{t-1}, x_t, x_{t+1}, \ldots) \quad (178)$$

where t runs through all the integers from $-\infty$ to $+\infty$. The variables x_t are all random, but they differ from random variables in general in that they are ordered. The random vector x--of infinite dimensionality--is called a stochastic process; it expresses the ordered development (usually, although not necessarily, through time) of a series of events subject to random influences. Any particular sequence, x, is called a realization of the process. For example, the error terms of an economic relationship are a realization of a stochastic process.

Since any m components of the vector x form a random vector of finite dimensions, we have a joint cumulative distribution function for the m random variable components which we may write as

$$F(t, m; u_1, \ldots, u_m). \quad (179)$$

The particular vector to which equation (179) refers is

$$(x_{t+1}, x_{t+2}, \ldots, x_{t+m}). \quad (180)$$

Thus, for a stochastic process, we have a whole set of cumulative distribution functions, rather than a single function as we have when dealing with a random variable or a random vector. This set of distribution functions is generated, for a fixed positive integer m, by letting t range over the integers from $-\infty$ to ∞. These distribution functions must be consistent in the sense that

$$\lim_{u_{m+1}, \ldots, u_n \to \infty} F(t, n; u_1, \ldots, u_n) = F(t, m; u_1, \ldots u_m), \quad (181)$$

where $m < n$ and $-\infty < t < \infty$. A set of distribution functions which satisfy equation (181) may be said to determine the probability density of a particular realization of a stochastic process. 85/ The notion of a stochastic process may be extended to cover that of a stochastic vector process. In this case each component of x is interpreted as a vector.

In our discussion of the effects of nonindependence of the error terms in an economic relationship, it is useful to distinguish between two fundamental types of stochastic processes: stationary processes and evolutionary processes. These two classes are mutually exclusive and exhaust the totality of all stochastic processes. A process is called stationary if and only if the set of distribution functions $\{F\}$, generated from equation (179) by letting t range over all the integers from $-\infty$ to ∞, satisfies the following relation

$$F(t, m; u_1, \ldots, u_m) = F(t+v, m; u_1, \ldots, u_m) \quad (182)$$

85/ See Wold and Jureen (106, p. 153).

for all $m > 0$ and all t and all u_i $(i=1, \ldots, m)$ between $-\infty$ and ∞. The definition of stationarity contained in equation (182) amounts to treating time, or the ordering variable t, in a purely relativistic sense. The probability laws assumed to rule a stationary stochastic process "... depend on time in such a way that if we replace time as measured from a fixed time point by a time variable measured from another time point, the probability laws will remain the same. In other words, if the development in a time series [or stochastic process] is known up to a certain time point, say t, the probability laws ruling the continued development will depend only on the behavior of the time series up to the time point t, not on the actual value of t." Wold (102, p. 4). All stochastic processes which are not stationary are evolutionary in the sense that the properties of their sets of distribution functions do depend on the particular point from which we measure time.

For a stationary stochastic process we define the mean and variance as follows

$$\left.\begin{aligned} m &= E\,x_t = \int_{-\infty}^{\infty} u\, d_u\, F(t, 1; u) \\ \sigma^2 &= E(x_{t-m})^2 = \int_{-\infty}^{\infty} (u-m)^2\, d_u\, F(t, 1; u). \end{aligned}\right\} \qquad (183)$$

The correlation of x_t with x_{t+n} or x_{t-n} (the two are equal if the process is stationary) is called the nth order autocorrelation coefficient and is defined as

$$\rho_n = \frac{E(x_{t-m})\,(x_{t+n} - m)}{\sigma^2}\,. \qquad (184)$$

The set of autocorrelation coefficients generated by letting n range over the integers from $-\infty$ to ∞ is called the correlogram of the stationary process.

The fact that equation (182) holds for a stationary process is crucial in the above definitions, for this equation clearly implies that the mean, variance, and correlogram of the process are independent of the ordering variable t. If this were not the case, m, σ^2, and $\{\rho_n;\ -\infty < n < \infty\}$ all would be functions of time, and therefore not useful as characteristics of the process. Although stationarity of a process implies that it has a mean, variance, and correlogram independent of time, the converse is not generally true. If a process has a mean, variance, and correlogram independent of time, we say that it is stationary to the second order. Stationarity implies stationarity to the second order, but a process stationary to the second order may be evolutionary in the sense that it does not satisfy equation (182).

In most applications of the theory of stochastic processes to economic relationships, the error term is taken to be a stationary process. Intuitively, this implies that the economic structure underlying the relationship has been in its present form for some time past, so that the random influences at work have had sufficient time to settle down to an orderly existence. Not only, however, is it usually difficult to find periods of sufficient length over which we may assume no change in structure within the particular set of relationships under consideration, but it is even more difficult to find a period before which no structural changes have recently occurred. Be this as it may, stationary processes are the most susceptible of precise formulation, and no very useful results have yet been obtained without the assumption of stationarity.

In subsequent discussion it is useful to distinguish between two fundamental types of stationary stochastic processes: (1) a purely random process, and (2) a process of moving summation. A stochastic process

$$x = (\ldots, x_{t-1}, x_t, x_{t+1}, \ldots) \tag{178}$$

is called purely random if, for all t and n, its distribution functions satisfy the following condition:

$$F(t, n; u_1, \ldots, u_n) = \prod_{i=1}^{n} F(t, 1; u_i). \tag{185}$$

Thus, the error terms of a purely random process are assumed to be completely independent of one another. For economic series, this assumption is not likely to be justified. For any purely random process the correlogram reduces to

$$\rho_0 = 1, \rho_1 = \rho_2 = \ldots = 0. \tag{186}$$

Let $e = (\ldots, e_{t-1}, e_t, e_{t+1}, \ldots)$ be a purely random process with zero expectation. For all t between $-\infty$ and ∞, we may form the sum

$$z_t = a_0 e_t + a_1 e_{t-1} + \ldots, \tag{187}$$

where the a's are constants, independent of t, such that $\sum_i a_i^2$ is finite (thus insuring that z_t is a finite number).

The stochastic process defined by

$$z = (\ldots, z_{t-1}, z_t, z_{t+1}, \ldots) \tag{188}$$

is clearly stationary, for the set of distributions $\{F\}$ generated by combining equations (185) and (187) satisfies the condition stated by equation (182), provided, of course, that the a's of equation (187) do not depend on time. For the expectation and the variance of z we have

$$\left.\begin{aligned} m_z &= Ez_t = 0 \\ \sigma_z^2 &= Ez_t^2 = \left(\sum_i a_i^2\right) \sigma_e^2 , \end{aligned}\right\} \tag{189}$$

where σ_e^2 is the variance of the purely random process e. The correlogram of the process z is given by

$$\rho_n = \frac{\sum_i a_i \; a_{n+i}}{\sum_i a_i^2} \tag{190}$$

for n ranging over the integers from $-\infty$ to ∞. The process z is called a <u>process of moving summation</u>.

The process of moving summation has two special cases which are of interest: (1) the process of moving averages, and (2) the autoregressive process. The latter plays a major role in current discussions of the statistical effects of serial correlation. The process of moving averages is defined by equation (187) when the summation is taken over a finite number of terms only. Thus, if

$$z_t = a_0 e_t + a_1 e_{t-1} + \dots + a_h e_{t-h} \tag{191}$$

the process z is one of <u>moving averages</u>. The expectation of the process and its variance are given by equations (189) except that the summations in these formulae are only over a finite number of terms. The correlogram is given by equation (190) except that $\rho_n = 0$ for $n > h$. The name is unfortunate, as this is not an average in the usual sense, but it has become established.

An <u>autoregressive process</u> is a sort of inverse of a process of moving summation. Suppose that for all t between $-\infty$ and ∞ we have

$$z_t + b_1 z_{t-1} + \dots + b_h z_{t-h} = e_t, \tag{192}$$

where the process e is a purely random process. Equation (192) is an h-order difference equation in z_t, and may be solved for z_t as a function of the <u>series</u> e_t when t lies between $-\infty$ and ∞. The solution is of the form

$$z_t = e_t + a_1 e_{t-1} + a_2 e_{t-2} + \dots , \tag{193}$$

where the a's are functions of the b's in equation (192). The process z defined by equation (193) is a special case of the process of moving summation in which $a_0 = 1$. The process defined by an equation such as (192) or (193) is called an <u>h-order autoregressive process</u>, where h is the number of past values of z_t in equation (192); thus,

$$z_t + b_1 z_{t-1} = e_t \tag{194}$$

is a first-order autoregressive process. In order that z_t, as given by equation (193), be a finite number, the a's and hence the b's of equation (192) must satisfy certain conditions. We need not, however, discuss these here. For further details see Wold and Jureen (106, pp. 162-163).

A commonly discussed stochastic process is the Markov process. A stochastic process x is a Markov process if

$$E(x_t \mid x_{t-1}, x_{t-2}, \ldots, x_{t-n}) = E(x_t \mid x_{t-1}) \tag{195}$$

for all n > 2. Thus an autoregressive process is a Markov process if and only if it is first-order (that is, h=1). For an autoregressive process which is Markov, the correlogram has a particularly simple form:

$$\rho_n = (-b_1)^n \qquad (n = 0, 1, 2, \ldots). \tag{196}$$

Ordinary least squares in matrix form 86/.--In a subsequent section, we take up Aitken's method of generalized least squares, a method developed to meet complications introduced when the error terms of an economic relationship cannot be assumed independent. Since Aitken's method cannot conveniently be discussed except in matrix form, we first discuss ordinary least squares in matrix form.

Consider the vectors

$$y = \begin{pmatrix} y_1 \\ y_2 \\ \cdot \\ \cdot \\ \cdot \\ y_t \end{pmatrix}$$

$$x_i = \begin{pmatrix} x_{i1} \\ x_{i2} \\ \cdot \\ \cdot \\ \cdot \\ x_{it} \end{pmatrix} \qquad (i=1, \ldots, n)$$

$$\beta = \begin{pmatrix} \beta_1 \\ \beta_2 \\ \cdot \\ \cdot \\ \cdot \\ \beta_n \end{pmatrix}$$

86/ This section is based on Kempthorne (54, pp. 54-59).

and

$$e = \begin{pmatrix} e_1 \\ e_2 \\ \cdot \\ \cdot \\ \cdot \\ e_t \end{pmatrix} \qquad (197)$$

and the matrix

$$x = [x_1, x_2, \ldots, x_n]. \qquad (198)$$

We suppose that, at any time t, a relationship holds between y_t, the x_{it}'s, and e_t:

$$y_t = \sum_i \beta_i x_{it} + e_t. \qquad (199)$$

The last relationship can be more conveniently written in matrix form as

$$y = X\beta + e. \qquad (199.1)$$

In the theory of ordinary least squares, we generally assume that the matrix X is given, that is, is nonstochastic, and that the error terms e have the same distribution, zero mean, and finite variance. If X is nonstochastic, the distribution of e_t cannot involve any of the x_{it}.

The error terms, or the vector e, may be thought of as part of the realization of a stochastic process. If the process is stationary, the probability laws holding for the particular part represented by e are characteristic of the entire process. In the ordinary theory of least squares, e is assumed to come from a purely random stationary stochastic process. Its correlogram must be that given by equation (186), so that its variance-covariance matrix, Eee', may be written

$$\sigma^2 . I , \qquad (200)$$

where σ^2 is the variance of the process and I is an identity matrix. Under our assumptions we also have that

$$Ee = 0. \qquad (201)$$

The true error terms should be carefully distinguished from the unexplained residuals which might be calculated from equation (199) given y and X and an estimate of β. Let b be an estimate of β and u be the vector of calculated residuals, then by definition we have

$$y = X\beta + u. \qquad (202)$$

The least squares estimate of β, say b, is defined as that estimate, given y and X, which minimizes the sum of squared residuals, u'u. From equation (202) we see that u'u may be written

$$u'u = (y - Xb)'(y - Xb) \tag{203}$$

$$= (y' - b'X')(y - Xb) \tag{203.1}$$

$$= y'y - 2b'X'y + b'X'Xb. \tag{203.2}$$

Equation (203.2) may be differentiated with respect to $b_1, b_2, \ldots, b_n$. These derivatives, when set equal to zero, give n equations which may be written as

$$-2X'y + 2(X'X)b = 0 \tag{204}$$

or

$$X'Xb = X'y. \quad \text{87/} \tag{204.1}$$

Equation (204.1) may be solved for the vector b, giving

$$b = (X'X)^{-1} X'y , \tag{205}$$

and the resulting estimates minimize the sum of the squared residuals, that is, they are least squares estimates of β.

An estimate of β is called linear if it is a linear function of the vector y, say Ay. The estimates b are clearly linear since we may set $A = (X'X)^{-1} X'$. Under the assumptions, the estimate b is also an unbiased estimate of β. To see this, we substitute equation (199.1) in equation (205) and take the expected value:

$$E\,b = E(X'X)^{-1} X'y$$

$$= E(X'X)^{-1} X'(X\beta + e)$$

$$= E[\beta + (X'X)^{-1} X'e]$$

$$= \beta + E(X'X)^{-1} X'e . \tag{206}$$

Since, by assumption, X is nonstochastic and Ee = 0, the final term in equation (206) vanishes; hence, Eb = β, that is, b is an unbiased estimate of β.

87/ The reader can verify that equation (204.1) gives the usual normal equations and that the process of deriving them by differentiation of equation (203.2) is exactly equivalent to the usual procedure.

The variance-covariance matrix of b is given by

$$E(b - \beta)(b - \beta)' = E[(X'X)^{-1} X'ee'X(X'X)^{-1}]$$
$$= (X'X)^{-1} X' E(ee')X(X'X)^{-1}$$
$$= \sigma^2 (X'X)^{-1} \qquad (207)$$

by equation (200). The diagonal elements of $\sigma^2 (X'X)^{-1}$ are the variances of the estimates b and are clearly positive.

We may now ask the question as to whether the estimates b have minimum variance as compared with other linear unbiased estimates, that is, whether or not they are statistically efficient as compared with other linear unbiased estimates. Consider a linear estimate of β, By. If the estimate By is to be unbiased we must have

$$E(By) = E(BX\beta + Be) = \beta . \qquad (208)$$

Since B and X are nonstochastic and Ee = 0, equation (208) implies that

$$BX = I. \qquad (209)$$

We may arbitrarily write the matrix B as

$$B = [(X'X)^{-1} X' + C]. \qquad (210)$$

By equation (209) we have

$$CX = 0, \qquad (211)$$

where 0 is the zero matrix. Hence,

$$\hat{b} = [(X'X)^{-1} X' + C]y \qquad (212)$$

is an arbitrary linear unbiased estimate of β subject to equation (211). Substituting equation (199.1) in equation (212) and proceeding as before, we see that the variance-covariance matrix of $\hat{b}$ is

$$E(\hat{b} - \beta)(\hat{b} - \beta)' = E \{[(X'X)^{-1} X' + C]ee' [X(X'X)^{-1} + C']$$
$$= \sigma^2 [(X'X)^{-1} + CC'], \qquad (213)$$

since $Eee' = \sigma^2 I$ and CX = 0. Since each diagonal element of CC' is simply the sum of squares of the elements in the corresponding row of C, they must be positive. Hence, unless the matrix C is the zero matrix, the variances of the estimates $\hat{b}$ are greater than those of b. For this reason the least squares estimates, b, are called "best" linear unbiased estimates. In passing, we note that the assumption that $Eee' = \sigma^2 I$, that is, that the error terms are part of the realization of a purely random stationary stochastic process, was essential in the derivation of the fact that the least squares estimates are efficient. This assumption was not necessary in the derivation of the fact that b is an unbiased estimate of β.

Wold (103) also has shown that, under the assumptions, the least squares estimates, b, are consistent estimates of β. This proof likewise does not depend on the assumption that $Eee' = \sigma^2 I$.

Effects of nonindependence on ordinary least squares estimates.--When e is not assumed to form part of the realization of a purely random stationary stochastic process, we no longer can assume that $Eee' = \sigma^2 I$; hence, it no longer follows that the ordinary least squares estimates are "best" in the sense of being most efficient. Watson (97) discusses the loss of efficiency that results when Eee' is incorrectly assumed to be equal to $\sigma^2 I$. He does this quite generally under the assumption that the e form part of the realization of a stationary process or a process stationary to the second order. In a second article, Watson and Hannan (98) apply the results of Watson (97) to the special cases in which e forms part of the realization of an autoregressive or a moving average type of stationary process. Gurland (44) investigates the loss of efficiency in least squares analysis when the error terms form part of an evolutionary stochastic process. Wold (103) discusses least squares regression with autocorrelated variables and error terms, but his results are more properly described when we come to methods for dealing with nonindependence.

Suppose that the error vector, e, forms part of a realization of a stationary stochastic process. As we have seen, there exists for such a process a well-defined mean, variance, and correlogram. The correlogram for such a process may be written

$$\rho_n = \rho_{-n} \quad \text{for } -\infty < n < \infty , \tag{214}$$

since the correlation of the t^{th} component with the t + nth component and the correlation of the t^{th} component with the t - nth component are the same if the process is stationary. Thus

$$Eee' = \Omega , \tag{215}$$

where

$$\Omega = \sigma^2 \begin{pmatrix} 1 & \rho_1 & \rho_2 & \cdots & \rho_n \\ \rho_1 & 1 & \rho_1 & \cdots & \rho_{n-1} \\ \rho_2 & \rho_1 & 1 & \cdots & \rho_{n-2} \\ \vdots & \vdots & \vdots & & \vdots \\ \rho_n & \rho_{n-1} & \rho_{n-2} & \cdots & 1 \end{pmatrix} \tag{216}$$

and σ^2 is the variance of the process. The matrix Ω is not only symmetric but the elements are equal along any northwest-southeast diagonal. 88/ If the process assumed is not at least stationary to the second order, the correlogram of the process is in general undefined 89/ and Ω does not possess the desirable properties just mentioned.

The form of the matrix Ω may be illustrated for several types of stationary stochastic processes:

(1) For a purely random process,

$$\Omega = \sigma^2 I. \tag{217}$$

(2) For a process of moving summation, reference to equations (187) and (190) indicates that

$$\Omega = \sigma^2 \begin{pmatrix} 1 & \sum_i a_i a_{i+1} & \sum_i a_i a_{i+2} & \cdots \\ \vdots & 1 & \sum_i a_i a_{i+1} & \cdots \\ & \vdots & \vdots & \vdots \\ & & \cdots & 1 \end{pmatrix}. \tag{218}$$

(3) The same expression for Ω holds when the process is one of moving averages except that the summations are now over a finite number of terms.

(4) The general autoregressive process also falls into the same class as the process of moving summation, since an autoregressive process of finite order may be transformed into a process of moving summation of infinite order; hence, Ω may also be expressed as in equation (218).

(5) If the autoregressive process is Markov, we have by equation (196):

$$\Omega = \sigma^2 \begin{pmatrix} 1 & b_1^2 & -b_1^3 & \cdots & (-b_1)^n \\ b_1^2 & 1 & b_1^2 & \cdots & (-b_1)^{n-1} \\ -b_1^3 & b_1^2 & 1 & \cdots & (-b_1)^{n-2} \\ \vdots & \vdots & \vdots & & \vdots \\ (-b_1)^n & (-b_1)^{n-1} & (-b_1)^{n-2} & \cdots & 1 \end{pmatrix} \tag{219}$$

88/ Such a matrix is known as a Laurent matrix; see Wold and Jureen (106, p. 156).

89/ Note, however, that it is defined if the process is evolutionary but stationary to the second order.

The particularly simple form of Ω in equation (219) has led to the frequent assumption that the error process is autoregressive and Markov.

If the variance-covariance matrix of the error term is actually Ω but we take it to be $\sigma^2 I$ in order to justify least squares, what will be the result? As indicated in the preceding section, our least squares estimates, b, of β are still unbiased and consistent; they are not, however, statistically efficient nor will the ordinary t- and F-tests apply (see pages 178-183).

Let us consider the relationship to be estimated as being given by equation (199.1). Our least squares estimates of β are given by

$$b = (X'X)^{-1} X'y. \tag{205}$$

We would take the variance-covariance matrix of the b to be

$$V(b) = \sigma^2 (X'X)^{-1}, \tag{220}$$

whereas in fact it is not. Substituting $Eee' = \Omega$ in equation (207), we find that actually

$$V(b) = (X'X)^{-1} X'\Omega X(X'X)^{-1}. \tag{221}$$

Our estimate of the error variance would be

$$s^2 = \frac{1}{t-n} [(y - Xb)' (y - Xb)]. \tag{222}$$

This estimate is biased as:

$$E(s^2) = \frac{1}{t-n} E[(e-X(X'X)^{-1} X'e)' (e-X(X'X)^{-1} X'e)]$$

$$= \frac{1}{t-n} E[e'e - e'X(X'X)^{-1} X'e], \tag{223}$$

where the expression in the brackets in equation (223) is obtained from equation (222) by substituting for y and b from equations (199.1) and (205). The trace of a matrix is the sum of its diagonal elements; if $Eee' = \Omega$, $Ee'e = tr\,\Omega$, where $tr\,\Omega$ means the trace of the matrix Ω. By some manipulation it may also be shown that

$$E[e' X(X'X)^{-1} X'e] = tr[X' \Omega X(X'X)^{-1}], \tag{224}$$

so that

$$E(s^2) = \frac{1}{t-n} [tr\,\Omega - tr\,(X' \Omega X(X'X)^{-1})]. \tag{225}$$

Ω is a t X t matrix; if $\Omega = \sigma^2 I$, $tr\,\Omega = \sigma^2 t$. $X' \Omega X(X'X)^{-1}$ is an n X n matrix; if $\Omega = \sigma^2 I$, $tr(X'\Omega X(X'X)^{-1}) = \sigma^2 n$; hence, equation (225) shows that s^2 is unbiased only if $\Omega = \sigma^2 I$.

Watson (97, p. 329) is able to set limits on the maximum and minimum bias in the estimate s^2. These limits depend on the characteristic roots of the matrix Ω . 90/ Watson is able to derive these limits only for the case in which the independent variables are uncorrelated and form an orthonormal set, that is, only when

$$X'X = I. \tag{226}$$

The limits in this case are as follows:

Max. bias = σ^2 (the mean of the t-n greatest characteristic roots of Ω - the least root of Ω)

Min. bias = σ^2 (the mean of the t-n least characteristic roots of Ω - the greatest root of Ω). (227)

It is clear that the upper limit is always positive and the lower limit always negative. When $\Omega = \sigma^2 I$, both limits are zero. The presence of nonindependent error terms is commonly believed to make the least squares estimate of the variance of the coefficients deceptively small; equation (227) shows that this may not always be the case. In practice, the limits in equation (227) are not very useful for, although we can always transform the independent variables in such a way that they are uncorrelated, that is, so that X'X is diagonal, we cannot always transform them in such a way that X'X = I, as required if they are to be orthonormal.

In the preceding section we showed that the least squares estimates are statistically efficient as compared with any other linear unbiased estimates provided the error terms e form part of the realization of a purely random process. When the error terms do not do so, the ordinary least squares estimates in general are not efficient. Our definition of the efficiency of one estimate relative to another, as given on page 208, is not rigorous; we now need a somewhat sharper concept. The joint efficiency of an estimate b* relative to an estimate b, where b* and b are both vectors, is defined as the ratio of the determinant of the variance-covariance matrix of b* to the determinant of the variance-covariance matrix of b, that is,

$$Eff(b) = \frac{|V(b^*)|}{|V(b)|}, \tag{228}$$

90/ The characteristic roots of a matrix A may be found by solving the polynomial $|A - \lambda I| = 0$ for λ. If the matrix A is nXn, the equation $|A - \lambda I| = 0$ is a polynomial of degree n. It therefore has n possible solutions for λ , if complex solutions are allowed. See Klein (58, pp. 340-341).

where b* is an alternative estimate to b against which we measure the efficiency of b. 91/ If b* is the most efficient linear estimate, then the ratio given in equation (228) becomes closer to one as the linear estimate b becomes more efficient as an estimate of β . Hence, provided we can find the most efficient linear estimate, the ratio (228) can be used to measure the loss in efficiency if b is used instead of b*.

Aitken (1) has shown that when the variance-covariance matrix of the error terms is Ω , the most efficient linear estimates, b*, have a variance-covariance matrix

$$V(b^*) = (X' \Omega^{-1} X)^{-1}. \quad 92/ \tag{229}$$

Consequently by equations (221) and (229) we have

$$\mathrm{Eff}(b) = \frac{|(X' \Omega^{-1} X)^{-1}|}{|(X'X)^{-1} X' \Omega X (X'X)^{-1}|}$$

$$= \frac{|X'X|^2}{|X' \Omega X|\,|X' \Omega^{-1} X|}, \tag{230}$$

since the determinant of an inverse of a matrix is the reciprocal of its determinant and the determinant of the product of two square matrices is the product of their determinants. Watson (97, pp. 330-331) shows that the ratio (230) is indeed less than one, so that the least squares estimates, b, are statistically inefficient. Watson also shows that this ratio tends to one as Ω tends to $\sigma^2 I$ and that it has, in general, a definite lower limit which depends on Ω and on X. His proof does not require the assumption that $X'X = I$. In a second article, Watson and Hannan (98) show that the loss of efficiency may be very great, that is, the lower limit to the ratio (230) may be close to zero, if strong serial correlation of the error terms is present due to the fact that they form part of a realization of an autoregressive or moving average process.

Since the estimate of the error variance, s^2, is biased, the ordinary t- and F-tests do not apply to the least squares estimates if the error terms do not form part of the realization of a purely random process. Watson (97, pp. 331-340) shows that the attributed level of significance at which the ordinary t- and F-tests are carried out is not the true level of significance; he is able, however, to set limits to the true level of significance. These limits depend on the variance-covariance matrix of the error terms and are difficult to compute except for the simplest cases. The limits converge to the usual levels of significance as $\Omega \to \sigma^2 I$.

91/ See Watson (97, p. 330) and Gurland (44, p. 220).
92/ More is said about this in the following section. See equation (237).

Gurland (44) investigates the case in which the error term forms part of the realization of an evolutionary stochastic process but which the process would be autoregressive and Markov given sufficient time. Gurland assumes that the origin of the process is not sufficiently far back in time so that the initial conditions may be neglected. He shows that additional loss of efficiency can result from not specifying the initial conditions of the process. Thus, even if we use the methods suggested in the following section, which are based on the assumption of stationarity, large losses in efficiency may result because the processes are not stationary but evolutionary. As indicated earlier, evolutionary processes appear more appropriate in economics than stationary ones, and the reader should bear this in mind when reading the following section.

Aitken's generalized least squares and other methods for dealing with nonindependence.--Aitken (1), Cochrane and Orcutt (15), Klein (58, pp. 86-89), and Wold (103) all have suggested methods for coping with the problems raised by nonindependence. The methods of Cochrane and Orcutt and Klein each are derived from Aitken's method of generalized least squares; Wold's method is not directly derived from Aitken's method. All methods rest on an assumption of stationarity or stationarity to the second order. All methods except Aitken's involve considerable simplification with regard to the stochastic process underlying the error terms.

Beginning on page 160, we discuss the situation in which $Eee' = \Omega$ but in which we assume that $Eee' = \sigma^2 I$, that is, a special form for Ω. We indicate that (a) a considerable loss of statistical efficiency may result from such incorrect specification, and (b) in such a situation the ordinary t- and F-tests of the regression coefficients do not apply. Aitken's generalized least squares aims at both statistical efficiency and appropriate t- and F-tests; it solves the general problem by transforming the variables y and X between which the regression is taken. Instead of taking the regression as stated in equation (202), Aitken recommends transforming all the variables by premultiplication by a matrix H,

$$Hy = HXb + Hu, \qquad (231)$$

where H is such that

$$H \Omega H' = I \qquad (232)$$

where we set σ^2 equal to one without loss of generality. By equation (199.1) we have the corresponding transformation of the true equation

$$Hy = HX\beta + He. \qquad (233)$$

The variance-covariance matrix of the transformed error terms is

$$E\, Hee'H' = H(Eee')H' = H \Omega H' = I \qquad (234)$$

by equation (232), the definition of H. The transformation procedure, therefore, brings us back to the case in which ordinary least squares is appropriate and efficient, albeit the variables are now Hy and HX rather than y and X. If we substitute Hy for y and HX for X in equation (205), we obtain the least squares estimates appropriate to the case in which Eee' = Ω ,

$$b^* = (X'H'HX)^{-1} X'H'Hy$$

$$= (X'\Omega^{-1}X)^{-1} X'\Omega^{-1}y. \qquad (235)$$

The variance-covariance matrix of b* may be found as follows: Substituting equation (233) for Hy in equation (235) we have

$$b^* = (X'H'HX)^{-1} X'H'HX\beta + (X'H'HX)^{-1} X'H'He$$

$$= \beta + (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}e. \qquad (236)$$

Hence, Eb* = β and

$$E(b^* - \beta)(b^* - \beta)' = V(b^*)$$

$$= (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}ee'\Omega^{-1}X(X'\Omega^{-1}X)^{-1}$$

$$= (X'\Omega^{-1}X)^{-1}. \text{ 93/} \qquad (237)$$

The residual variance, s^2, is estimated by

$$s^2 = \frac{1}{t'-n} [(Hy - HXb^*)' (Hy - HXb^*)], \qquad (238)$$

where t' is the number of rows in Hy and HX. Note that because of the transformation by premultiplication by H, the vectors Hy and HX do not have the same number of rows as y and H; they have, in fact, fewer rows for a finite sample size. It can be shown that $E(s^2) = \sigma^2$, so that the estimate s^2 is unbiased. The ordinary t- and F-tests apply to the coefficients b*.

In order to use Aitken's generalized least squares, we must be able to specify Ω exactly. It is naturally never possible to do so, although reasonable assumptions can sometimes be made. Watson's analysis, discussed in the preceding section, may be applied to the situation in which Ω is specified incorrectly; similar conclusions result. Hence, incorrect guesses about Ω may be costly in terms of the statistical efficiency of the analysis and the accuracy of t- and F-tests.

Aitken's method may be illustrated by an example: Suppose that the stochastic process from which e is derived is autoregressive and Markov, so that

---

93/ Note that the fact that both Ω and $X'\Omega^{-1}X$ are symmetric is used to derive equation (237). Ω is symmetric only on the assumption of stationarity.

$$e_t + b_1 e_{t-1} = \eta_t , \tag{239}$$

where η_t is purely random and hence independently distributed. By equation (219) the matrix Ω may be written

$$\Omega = \begin{pmatrix} 1 & b_1^2 & -b_1^3 & \cdots & (-b_1)^n \\ b_1^2 & 1 & b_1^2 & & \vdots \\ -b_1^3 & b_1^2 & 1 & & \vdots \\ \vdots & & & & \vdots \\ (-b_1)^n & \cdots & \cdots & & 1 \end{pmatrix} , \tag{240}$$

where we again assume that $\sigma^2 = 1$. If b_1 is known or assumed, it is possible to specify the matrix H used to transform the variables, transform them, and estimate by ordinary least squares. When Ω is as in equation (240), the matrix H is a t-1 x t matrix of the form

$$H = \begin{pmatrix} b_1 & 1 & 0 & \cdots & 0 & 0 \\ 0 & b_1 & 1 & \cdots & 0 & 0 \\ 0 & 0 & b_1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & 0 & & b_1 & 1 \end{pmatrix} \text{94/} . \tag{241}$$

Equation (241) may be derived quite simply: Consider one of the equations (199),

$$y_t = \sum_i \beta_i x_{it} + e_t. \tag{199}$$

94/ Actually

$$\Omega^{-1} = \begin{pmatrix} 1 & b_1 & 0 & \cdots & 0 & 0 \\ b_1 & 1+b_1{}^2 & b_1 & \cdots & 0 & 0 \\ 0 & b_1 & 1+b_1{}^2 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & b_1 & 1 \end{pmatrix}$$

whereas, in this case,

Lag equation (199) one period, multiply the result by b_1, and add the result to equation (199). We get

$$y_t + b_1 y_{t-1} = \sum_i \beta_i (X_{it} + b_1 X_{it-1}) + e_t + b_1 e_{t-1}. \tag{242}$$

By equation (239)

$$e_t + b_1 e_{t-1} = \eta_t , \tag{239}$$

where η_t forms part of the realization of a purely random process. When e is not autoregressive and Markov, it is difficult to specify the transformation matrix H; it is always possible, however, to use equations (235) and (237) to estimate b* and V(b*), respectively. This latter procedure, actually, is simpler than transforming the variables, even in the case where the error process is autoregressive and Markov.

Klein (58, pp. 86-89) discusses a simple generalization of Aitken's method for a special case. Klein supposes that the error process is autoregressive and Markov but that the autoregressive coefficient, say b_0, is unknown. Let H be defined by equation (241), except that we substitute b_0 for b_1 in this equation. We may substitute Hy for y and HX for X in equation (203.2) to get

$$u'u = y'H'Hy - 2b'X'H'Hy + b'X'H'Hb, \tag{243}$$

where u'u is the residual sum of squares. In theory we can minimize this residual sum of squares with respect to any number of parameters, that is, the b's and whatever other parameters enter $H'H = \Omega^{-1}$; in practice, however, only the case which Klein has chosen is practical. In Klein's case, only one additional parameter, b_0, is present. Even this case presents computational difficulties of a high order; Klein shows that the solution of the n + 1 equations, obtained by minimizing u'u with respect to b and b_0, requires the solution of a high order polynomial. The order of the polynomial varies with

---

$$H'H = \begin{pmatrix} b_1^2 & b_1 & 0 & \cdots & 0 \\ b_1 & 1+b_1^2 & b_1 & \cdots & 0 \\ 0 & b_1 & 1+b_1^2 & \cdots & 0 \\ \vdots & \vdots & \vdots & & \\ 0 & 0 & 0 & & 1 \end{pmatrix}$$

The discrepancy is only $1-b_1^2$ in the upper left hand corner. The discrepancy is due, of course, to the fact that we have only a finite number of observations, t, and, therefore, that the matrix H cannot be square. It can be shown that this discrepancy does not affect the validity of Aitken's method [see Stone (87, p. 289)].

n, the number of independent variables. For example, when n=2, the polynomial is of order 9; when n=3, it is of order 13; and when n=4, it is of order 17. The reader is referred to Klein (58, pp. 86-89) for further details.

Cochrane and Orcutt (15) propose an iterative procedure based on Aitken's method. Their method can be applied only when the error process is of a relatively simple type; for example, first- or second-order autoregressive. The suggested procedure, in their words, is as follows: "First estimate the desired regression coefficients by ordinary least squares and obtain the resulting series of residuals. Then estimate from those residuals by least squares the autoregressive parameters of a one or two lag difference equation. Use the autoregressive parameters to make an autoregressive transformation of the observed series aimed at randomizing the error term [that is, to determine the matrix H], and reestimate the desired regression coefficients. 95/ Put these revised estimates back in the original equation, obtain the resulting series of residuals and estimate their autoregressive parameters. Use these to make a new autoregressive transformation of the original series [sic!] 96/ and so on until estimates of the desired regression coefficients are obtained which are consistent with estimates of the autoregressive parameters of the residuals in the sense that no further adjustments are necessary."

Wold (103) shows that the autoregressive parameters estimated from the calculated residuals of a least squares regression are consistent estimates of the true autoregressive parameters. In general, however, by the nature of least squares, we expect the estimated autoregressive parameters to be biased towards zero. Consequently, the Cochrane-Orcutt method may stop short of the desired goal. Watson and Hannan (98) show that even slight misspecification of Ω may result in serious loss of statistical efficiency and significance levels for t- and F-tests that are greatly out of line with the true significance levels. Nevertheless, the Cochrane-Orcutt method is probably a good one, in first approximation.

Wold (103) 97/ proposes a method, not derived directly from Aitken's method, which aims only at correct t- and F-tests for the ordinary least squares estimates and not for statistical efficiency as well. Wold discusses the case in which there is autocorrelation both in the independent variables

---

95/ An alternative to Cochrane's and Orcutt's procedure at this point is to use the autoregressive parameters, obtained from the series of calculated residuals, to obtain an estimate of the matrix Ω^{-1}. The latter can then be inserted in equations (235) and (237) in order to obtain b* and V(b*). The estimates so obtained are slightly different from those of Cochrane and Orcutt because of the finite sample size.

96/ The author of this section interprets this sentence to mean that the second-round autoregressive parameters should be used to transform the series already transformed by using the autoregressive parameters estimated in the first round, and so on. As it stands the sentence does not make much sense.

97/ See also Wold and Jureen (106).

and in the error term, that is, when both the independent variables and the error term form part of the realization of an autoregressive process. Wold first proves that the least squares estimates are unbiased and consistent estimates of β. Unbiasedness, as we have already seen, is a consequence of the fact that the proof of the unbiasedness of ordinary least squares estimates does not depend on any assumptions about the error variance-covariance matrix. Unbiasedness and consistency are highly desirable properties, and efficiency may well be disregarded, which is exactly what Wold does. Wold makes use of equation (221). When the error variance-covariance matrix is Ω, but we obtain the ordinary least squares estimates, b, their variance-covariance matrix is given by

$$V(b) = (X'X)^{-1}X'\,\Omega\,X(X'X)^{-1}. \tag{221}$$

Wold shows that the least squares estimate, s^2, of the error variance, σ^2, is statistically consistent. 98/ Consequently, we may replace σ^2 by s^2 in equation (221). Wold also shows that the least squares estimates of the autoregressive parameters, obtained from the calculated residual terms, are consistent estimates of the true autoregressive parameters. We may, therefore, estimate the matrix Ω. Replacing Ω in equation (221) by its estimate we obtain consistent, but not necessarily unbiased, standard errors for the least squares estimates. For small samples, these estimates may be severely biased, probably toward zero. Wold points out that the standard errors so obtained are usually, but not always, greater than the ordinary standard errors. When estimating the correlogram of the error term, simplifying assumptions usually have to be made; typically one cannot handle the assumptions of an autoregressive process of more than first or second order.

In the simple case in which there is only one independent variable, formula (221), with the appropriate substitutions, results, after some manipulation, in

$$\text{var } b_1 = \frac{s^2}{s_1^2} \; \frac{1 + 2\,\hat{\rho}_1\, r_1^{(1)} + 2\,\hat{\rho}_2\, r_2^{(1)} + \cdots}{t - 2}, \tag{244}$$

where s^2 is the estimated residual variance, s_1^2 is the variance of the independent variable x_1, $(1, \hat{\rho}_1, \hat{\rho}_2, \ldots)$ is the estimated correlogram of the error process, and $(1, r_1^{(1)}, r_2^{(1)}, \ldots)$ in the estimated correlogram of the independent variable. 99/

98/ This follows from equation (223), but the proof is involved.

99/ For the simple case presented in equation (244), the standard error of b_1 reduces to the ordinary least squares standard error when $r_1^{(1)} = r_2^{(1)} = r_3^{(1)} = \ldots = 0$, that is, when the independent variable is not serially correlated. Thus, for nonstochastic independent variables, the ordinary least

Conclusions.--We have seen that nonindependence leads, in least squares procedures, to loss of statistical efficiency and incorrect t- and F-tests; it does not lead to statistical bias or inconsistency. Various approximate procedures have been developed for regaining some of the lost efficiency and deriving correct t- and F-tests of the regression coefficients. These procedures are based on the assumption of stationarity, which, as we have indicated, is not particularly appropriate in the economic realm. Moreover, although the procedures possess large-sample validity, it is doubtful that they are good approximate procedures for dealing with the short series which are typical in economics. Nevertheless, if severe serial correlation is present, the methods discussed in this section probably are better than nothing.

We have not discussed the effects of nonindependence in the context of simultaneous equations. Little is known about this subject. We close with two remarks: (1) When a system is just identified, our entire discussion applies to the least squares estimates of the reduced form parameters. We do not know the effects of nonindependence on the limited information estimates of the parameters in an overidentified equation. (2) When lagged endogenous variables are included as predetermined variables, the situation is much worse than that described above. The least squares estimates of the reduced form parameters not only are inefficient in a statistical sense but they also are biased and inconsistent if the reduced form error terms are not independent.

Spatial Equilibrium Analyses 100/

The analyses and techniques discussed so far deal chiefly with studies designed to obtain quantitative measurements of the demand and price structure for a given market area, where this area frequently is the United States as a whole or, at times, the entire world. Spatial equilibrium analyses instead are concerned with studying interrelationships between regions or countries. They deal with geographical price equilibriums under stated conditions or changes in conditions and flows of commodities between several areas.

This type of model is valuable because it is operational and the computations are manageable. Information obtained from solving spatial price equilibrium models permit one to make predictions as to the direction and magnitude in which the variables of the system will change when some change is made

---

squares standard errors are consistent but not unbiased estimates of the true standard errors. This result, however, holds only asymptotically for large samples. Watson (*, pp. 24-26) shows that, for the simple case of one independent variable, this result holds even for small samples. In dealing with economic time series, however, nonserially correlated independent variables are rare indeed.

*Watson, G. S. Serial Correlation in Regression Analysis. Ph.D. thesis, North Carolina State College, 1951.

100/ Most of the material in this section is adapted from Judge (53).

in the data of the problem, for example, the effect of a rise in unit transport costs between regions i and j. By utilizing the results of these equilibrium models, certain insights into the changing character of a particular industry can be obtained.

The conceptual framework used to explain the mechanism which generates the level and location of consumption, the market prices and the geographical flows among regions is a perfectly competitive market in space, form, and time. The necessary condition for a perfectly competitive market that is relevant in this model is a uniform price which differs only by transportation and handling costs.

All firms are assumed to have the objective of maximizing net profits and thus to choose that market which, deducting transportation charges, will yield the largest possible profit. The supply sources and markets within each region are assumed to be at a single point. Transportation cost per unit between each possible pair of regions is assumed to be independent of volume and direction. Commodities are assumed to be homogenous, that is, an item produced in one region is the same as that produced in another. The time period considered usually is one year and the quantity available by regions, population, and disposable income are assumed predetermined variables within this time period. The model employed normally is static, that is, it does not allow for accumulation or depletion of inventories. Therefore, all demands must be met from current production. It is further assumed that the market demand schedules for each region are known; they may or may not be identical.

Given the predetermined geographical supplies and a unit cost of shipping from source i to market j, the problem is to maximize profit (minimize transportation costs) to each source subject to the side conditions given by the market demand curve in each region. In addition, this maximization is also subject to the following conditions involving shipments: (1) all of the supplies from each source must be sold, (2) no negative amounts can be shipped, and (3) shipments are made only to markets where per unit transportation costs do not exceed price. The first assumption can be dropped if this appears desirable, as when purchases by an individual firm or Governmental agency from a number of suppliers are being studied. It usually is retained when studies for an entire industry are made.

Our problem involves the simultaneous maximization of profits by a number of entrepreneurs, but it can be transformed into a single maximization problem. We may visualize a low calibre "mastermind" who is in control of all sources of supply and tries to maximize his total profits but does not realize his decisions affect prices. In that case he will direct his manager at every supply source to ship to those markets and only those markets yielding the greatest per unit profit. The optimum set of flows can be defined as the set chosen by a monopolistic firm which encompasses the entire industry and wishes to minimize its cost of meeting given demands. This problem is formulated in the framework of linear programming; it can be shown, however, that the resulting minimum cost set of flows is the one that would be determined under conditions of a perfect market.

A solution to this problem exists provided the quantities to be disposed of by the various sources are not so large that they cannot be sold at prices covering the transportation costs. The final solution must of course satisfy the conditions of the regional supply and market demand relationships and must be consistent with the decision rules given previously regarding shipments. The solution obtained is unique except for the case when two or more sources find two markets equally profitable. In this case more than one optimum shipment plan exists.

The solution yields a price for each source and market, quantities that will be shipped over each path, and thus the net surplus or deficit for each region. Once the model has been fitted, it can be used to measure the effect on prices and shipments of (1) a change in transportation costs, (2) a change or shift in demand within one or more regions, or (3) changes in production in total or by areas. If the model is an international one, it can be used to measure the impact on prices and foreign trade in the several countries of changes in tariffs, export subsidies, quotas, and the like. Simple examples of this sort are described by Fox (35).

Methods of analysis are discussed by Judge (53) and in the references cited by him. Applied analyses have been made by Judge for eggs and by Fox (34) and Fox and Taeuber (38) for livestock and feed grains. A number of studies that relate to industrial commodities, either for individual firms or entire industries, also have been published.

STATISTICAL CONSIDERATIONS IN INTERPRETING THE COMPLETED ANALYSIS 101/

Considering the Signs and Magnitudes of the Statistical Coefficients

The first check that normally is made on a regression or structural analysis is to determine whether the signs and general magnitude of the coefficients are consistent with expectations. Price-quantity coefficients, for example, always should be negative and income coefficients for most commodities should be positive. At times, advance information is available that strongly suggests whether certain coefficients should be elastic or inelastic and the nature of the signs that should hold for certain cross-elasticities. At other times, a purpose of the analysis is to ascertain the basic nature of these coefficients and, in such circumstances, these criteria cannot be used for checking purposes. "Wrong" signs or magnitudes from statistical analyses may occasionally lead to a revamping of the underlying theory but they are more likely to indicate the need for a different statistical approach in the analysis.

101/ Part of this material was originally published by Foote and Fox (29, pp. 25-29, 35-36) and part is adapted from Foote (28). Other sources are indicated by footnote.

As noted on page 43, the constant value in a first-difference analysis should be consistent in sign and approximate magnitude with expectations based on any unmeasured trends that are believed to affect the dependent variable, provided this coefficient differs from zero by a statistically significant amount.

Considering the Unexplained Residuals

In the 1920's and the early 1930's, major emphasis in regression analysis frequently was placed on obtaining a high multiple correlation coefficient. Dozens of analyses often were run, each differing slightly from the others by the use of (1) slightly different forms of variables of major importance, (2) alternative variables, each of which were assumed to have only a minor effect on the dependent variable, or (3) alternative methods of manipulating the data. If run by the graphic approach, several successive approximations would be obtained for each variant. Multiple correlation coefficients of 0.99 or above were a common occurrence; yet frequently these analyses had little value for forecasting for even a year or two. The trouble lay in the fact that the analyses had been manipulated to such an extent that they gave a false sense of security--an almost perfect fit was obtained for the particular sample but this sample, which seldom contained more than 20 to 30 observations, in a sense no longer represented the entire universe.

Now it is expected that part of the variation in the dependent variable will remain unexplained. Factors that cause unexplained residuals can be divided into three types:

1. Those due to errors in the data. (See page 143 for methods that can be used to estimate the importance of this factor.)

2. Those due to the omission of certain variables. This may be because the analyst fails to think of them, because no data are available, or because in most years they are so minor as not to be worth including in the study. This is the kind of random error which normally is assumed in a least-squares analysis and it is the type of error allowed for in simultaneous-equation "shock" models.

3. Those resulting from the use of wrong types of curves, incorrect lags, and similar factors. Charts that indicate the degree of partial correlation help to measure the relative importance of the nature of the chosen curves in causing residuals (see page 174).

Residuals for years not included in the analysis frequently are larger than those for the years included. The increase in unexplained variation may be caused by the following factors: (1) Extrapolation beyond the range of data included in the analysis. Some tests designed to measure the importance of this are discussed in the section beginning on page 184. (2) A change in the basic structure of the relationships may have occurred. This might reflect a change in the nature of the curves or the increased importance of some omitted factor. The charts described in the section beginning on page 174 help to indicate whether the nature of the curves has changed. (3) The regression curves in the analysis tend to adjust so far as possible to compensate for the types of errors discussed in the preceding paragraphs. Deviations in the later years reflect not only the true variability but also the extent to which the regressions were warped from their true shape to keep the deviations for the years included in the analysis as small as possible. This factor was particularly important in connection with some of the analyses run by the graphic method in the early 1930's.

If analyses are carefully formulated, making use of the basic principles given in this handbook, and manipulations purely for the purpose of raising the multiple correlation coefficient or increasing the degree of fit in some short test period are avoided, the results have a good chance of applying in future periods with as high a degree of accuracy as found for the years included in the analysis. Of course if a basic change in structure has occurred, an adjustment to allow for this is required (see page 21).

Testing for Serial Correlation in the Residuals

Durbin and Watson (18) developed a method by which the unexplained residuals from an equation fitted by least squares can be tested to see if successive values are correlated. Use of the limits shown in their table must be regarded as approximate when this test is applied to residuals from equations fitted by the limited information approach or to equations fitted by least squares that contain a lagged endogenous variable. But no exact test is available for such residuals.

In using this test, we compute the following statistic:

$$d' = \frac{\sum_{t=2}^{N} (d_t - d_{t-1})^2}{\sum_{t=1}^{N} d_t^2}, \qquad (245)$$

where d_t is the unexplained residual for observation t. If gaps occur in the data, as when certain war years are omitted, the number of observations that enter into the numerator are reduced by one for each gap, since these observations consist of first differences. To retain internal consistency in the formula, the number of observations that enter into the denominator also should be reduced by one for each such gap. This can be done by omitting the d_t that immediately follows the gap in obtaining $\sum_{t=1}^{N} d_t^2$.

If the dependent variable in the analysis is converted to logarithms, so that the residuals are in these terms, the Durbin-Watson test should be applied to the residuals in logarithms.

Friedman and Foote (40, pp. 77-78) discuss the use of this test and reproduce the required table from the original Durbin-Watson article. The test as described by them applies when positive or negative serial correlation is believed equally likely. Anderson (3, p. 118) points out that "In most cases, the experimenter desires a test of the null hypothesis against the alternative of positive correlation." (Reasons are given beginning on page 148 of this handbook.) In the notation used by Friedman and Foote, under these circumstances we expect a small value of d' when the null hypothesis is false, and the following testing procedure should be used: If the computed value of d' is less than the tabulated value for d_L, the null hypothesis is rejected. If d' is greater than d_U, the null hypothesis is assumed to be true. If d' lies between d_L and d_U, the test is inconclusive. When use is made of this one-tailed test, the level of significance is half that of the two-tailed test described by Friedman and Foote.

Testing for Nonlinearity in an Equation

If the equation of a straight line is fitted to data that are from a nonlinear relationship, adjacent deviations tend to be of the same sign, indicating that the residuals from the regression line are not distributed randomly. A measure of the relationship between adjacent deviations can be obtained by modifying the statistic developed by Durbin and Watson (18) to test for serial correlation. The modification consists of letting t represent observations arranged in order of magnitude for a specified independent variable. If more than one independent variable is used in the analysis, a separate test must be made for each one. The test is precisely the same as that described in the preceding section. If d' or 4-d' is less than d_L, we assume that the relationship may be curvilinear. If both d' and 4-d' are greater than d_U, we assume that there is no departure from linearity. If neither of the computed values is less than d_L, but one of them lies between d_L and d_U, the test is inconclusive.

Charts that Indicate the Degree of Partial Correlation

A method is described in the Appendix (page 205) which permits the construction of charts that indicate approximately the degree of partial correlation between the dependent variable in a least squares analysis and each of the independent variables. An excellent example of the use of such charts to determine, after running a study initially, whether the correct form of the relationships was assumed is an analysis by Armore and Burtis (6), made in 1950, of factors that affect consumption of fats and oils other than butter. The analysis first was run in linear terms, following which such a set of charts was developed. The chart in the upper left hand corner of figure 9 shows, for example, the relation between consumption and price after adjusting the dependent variable for the effects of the other variables in the analysis. 102/ Linear relations give a fairly good fit for the years prior to

---

102/ As shown by Foote (25), the simple regression between such variables equals the partial regression but the simple correlation equals the part

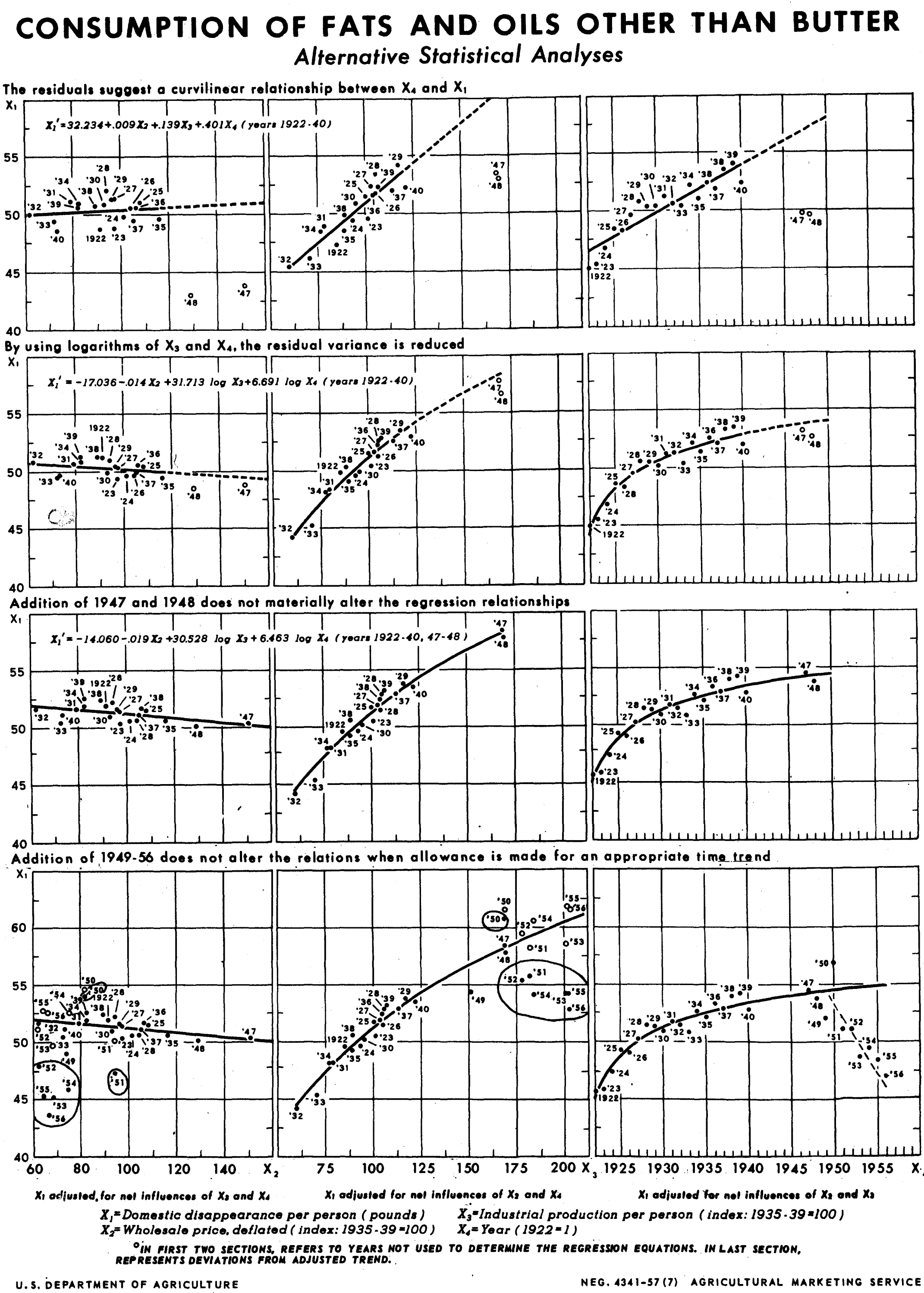


Figure 9.--The first three rows of small charts illustrate how this analysis was modified at the initial fit to take account of ideas obtained from the charts. The last row of small charts indicates further adjustments when it was brought up-to-date through 1956.

World War II on which the analysis is based, although a definite curvilinear relation is suggested for "time" even for these years. Addition of dots that relate to 1947 and 1948, the only post-World War II data available at that time, suggested the need for a curvilinear relation for both X_3 and X_4. Use of a semi-logarithmic relation for these variables (1) improved the fit materially for the years included in the analysis, (2) gave fairly good forecasts for the postwar years, and (3) resulted in a sign on the variable that relates to price that is consistent with economic expectations. Subsequent inclusion of data for 1947 and 1948 resulted in a slightly steeper regression on price and a moderate improvement in fit for these years.

Although some of the variables used in this analysis no longer are available, other variables that are designed to measure the same economic phenomena can be obtained. As a matter of interest, this analysis was brought to date through 1956. The results are shown in the bottom section of figure 9. Technological developments in the fats and oils industry suggested that a U-shaped time trend might be indicated if this analysis were brought to date, but the sharp downturn suggested by the statistical analysis was not foreseen. It is generally known, however, that the demand for fats and oils in this country, even when butter is excluded, has dropped sharply due to increased use of synthetic products, particularly in soaps and cleaners, paints and varnishes, and emulsifiers. This downtrend may well have begun about 1950, as new plants for use of synthetics that had been planned immediately following the end of World War II came into production. Thus the dashed line in the small chart in the lower right hand section of figure 9 may well represent a true measure of the effect of technological developments on the domestic demand for fats and oils during the period 1950-56.

This study furnishes an excellent example of how charts of this sort can be used to revise an analysis graphically. It would be difficult, if not impossible, to obtain a mathematical function that would reproduce the time trend shown for the up-to-date analysis. The graphic curve, however, is all that is needed. Deviations from the dashed line in this chart can be plotted around the original regression curves shown for the other variables to determine whether the new data suggest a change in these relations. These deviations are plotted as circles in the two relevant charts. In neither case is a change in relation suggested, even though the new observations with respect to industrial production are mostly outside the range of the original observations. Thus, our study suggests that the basic relationships between consumption, prices, and industrial production that held during 1922-48 also hold during 1949-56 after making allowance for trends that resulted from technological developments. Other uses for such charts are described on page 172.

---

correlation. However, unless the degree of intercorrelation among the several independent variables is large, the part correlation nearly equals the partial correlation.

The Use and Interpretation of Tests of Significance

Yule and Kendall (108, p. 437) state, "It cannot be over-emphasized that estimates from small samples are of little value in indicating the true value of the parameter which is estimated. ... Nevertheless, circumstances sometimes drive us to base inferences ... on scanty data. In such cases we can rarely, if ever, make any confident attempt at locating the value of a parameter within serviceably narrow limits. For this reason we are usually concerned, in the theory of small samples, not with estimating the actual value of a parameter, but in ascertaining whether observed values can have arisen by sampling fluctuations from some value given in advance." Tests of significance as commonly used are designed to measure whether the observed value differs significantly from zero. This is referred to as "the null hypothesis." In most cases, a test could equally well be made as to whether the observed value differs significantly from some other value.

Tests of significance may be made under any of the following conditions: (1) With no previous knowledge; or (2) applied to a factor that is believed to be unimportant. In these cases, a nonsignificant value for a regression or correlation coefficient would indicate that the factor should be omitted from the analysis. (3) Applied to a factor which for theoretical reasons is believed to be important. A nonsignificant result in this instance does not indicate that the factor is not important. It does indicate a need for further evidence to prove beyond reasonable doubt that the factor is important. Therefore, such a factor might be left in an analysis, particularly if its effect is believed to be small. If repeated samples of 20 observations were drawn from a population for which the true partial correlation was 0.3, nonsignificant results, based on a 5-percent probability standard, would be obtained about three-fourths of the time. Such a factor would not greatly affect the calculated value in most years, but if interest were centered on the structural nature of the relationship, the analyst might wish to include it tentatively pending further evidence.

Requirements for tests to be valid.--Yule and Kendall (108, pp. 437-438) state, "Our results will be strictly true only for the normal universe. ... It appears that, provided the divergence of the parent from normality is not too great, the results which are given below as true for normal universes are true to a large extent for other universes. ... If there is any good reason to suspect that the parent is markedly skew, e.g. U- or J-shaped, the methods ... cannot be applied with any confidence." Nonparametric tests which are independent of the nature of the distribution from which the sample was drawn are available in certain cases [see Siegel (85)]. But in general, the requirement of normality is not very restrictive.

As discussed in the sections beginning on page 143 and 148, respectively, if tests of significance as applied to correlation measurements are to be valid, the independent or predetermined variables must be known without error and the unexplained residuals must be randomly distributed. Certain procedures that can be used if these conditions are not met are discussed in those sections.

<u>Some common tests</u>.--1. Student's t is defined as the difference between any particular value assumed to be the true value in the universe for the statistical measure examined and the value determined from the sample divided by the standard error of this coefficient. In common practice, the assumed value is taken as zero, so that t is given as the sample value divided by its standard error. In such cases, the null hypothesis is tested. If we do not know whether a regression or correlation coefficient should be positive or negative, then we use the conventional or 2-tailed test. If economic theory leads us to expect one sign or the other, then we should use a 1-tailed test, for which the level of significance is half that of the 2-tailed test. The t-test can be used equally well to test for significant deviations from any other assumed value, such as a slope of -1 for a price-quantity regression coefficient. The t-test may be applied with reasonable confidence to sample values that depart somewhat from normal in their distribution, but it should not be used, of course, where other more appropriate tests are available.

2. For samples of 20 or 30 observations, the standard error of a correlation coefficient cannot be estimated with much reliability if the correlation in the universe is high, whether positive or negative. Therefore, when testing hypotheses other than the null hypothesis, Fisher's z transformation should be used. See Ferber (<u>23</u>, pp. 381-386). Snedecor (<u>86</u>, pp. 113, 286) provides convenient tables for testing whether simple, partial, and multiple correlations differ significantly from zero.

3. Chi-square is used to determine whether a series of frequencies differ significantly from a theoretical or expected set of frequencies. Chi-square is defined as

$$\chi^2 = \Sigma \frac{(x-m)^2}{m} , \qquad (246)$$

in which x is the observed and m is the expected frequency. A somewhat different use of chi-square is described beginning on page 184.

4. The F-test is used to determine whether the assembly of data in various classes defined by certain restrictions exhibit on the average greater differences between members of different classes than between members of the same class. If they do, it is concluded that the restrictions result in significant separation of the data into at least two classes; that is, that the means of at least two classes differ. For example, the F-test was used to learn whether the weights differed significantly for the several supply components in the analysis for corn discussed on page 122.

The F-test also can be used in connection with a problem in which the value of certain regression coefficients are assumed in advance. The following steps are used in such cases:

a. Minimize the sum of squares, making no assumption about the coefficients. The degrees of freedom attached to this sum of squares equals the number of observations minus the total number of coefficients.

b. Minimize the sum of squares after assigning values to such coefficients as are to be tested.

c. Find the difference between the first and the second sums of squares. The degrees of freedom attached to this difference equals the number of coefficients to which a value was assigned.

d. Obtain the two mean squares by dividing the sum of squares obtained in steps 1 and 3 by the appropriate number of degrees of freedom.

e. F equals the ratio of the two mean squares. If it differs significantly from zero, the sample indicates that the assigned values should not be used. If it does not differ significantly from zero, we have no reason to doubt that the sample came from a population for which the true values were equal to the assigned values and the assigned values can be used.

5. A test for overidentifying restrictions in overidentified equations within a system of equations, developed by Anderson and Rubin (4, p. 56), is described in Friedman and Foote (40, pp. 79-81) and elsewhere.

Application of the F-test to Some Problems in Regression Analysis

A test to determine whether one or more regression coefficients differ from an assumed value.--Suppose we have an equation of the following type:

$$C = a + b_1P + b_2Y, \tag{247}$$

in which C equals consumption, P equals price, Y equals income, and all variables are expressed in logarithms. We wish to test whether b_1 can be assumed to equal -1, that is, that the price elasticity is equal to unity.

The usual least-squares regression analysis is run and the following value computed: $(\Sigma C^2 - \bar{C}\Sigma C)(1-R^2)$. The degrees of freedom attached to this sum of squares is N - 3, where N is the number of years included in the analysis. The following series next is obtained: $C' = C + P$. This is the value that C would have in each year if b_1 were equal to -1 and Y and the unexplained residual were at their given levels. A simple correlation between C' and Y then is computed and the value $(\Sigma C'^2 - \bar{C}'\Sigma C')(1-r^2)$ obtained. Obtain the difference between these two sums of squares. The degrees of freedom attached to this sum of squares equals 1, as a value was assigned to 1 parameter. F is computed as described in steps 4 and 5 above and looked up in an F-table, using N - 3 and 1 degrees of freedom.

In this case, Student's t could be used instead. The t-value would be obtained as follows:

$$t = \frac{b_1 - (-1)}{s_{b_1}} \tag{248}$$

and looked up in the usual table, using N - 3 degrees of freedom. But if values were to be assumed for more than one of the parameters and a single test for all of the assigned values simultaneously were desired, the F-test would be needed.

A test to ascertain whether the corresponding regression coefficients in three or more multiple regression equations all are equal.--This example and the following one are taken from Meinken (70, pp. 100-102). The test was applied to analyses of factors that affect the price of corn, oats, and barley, respectively, in summer. In each analysis, a composite supply factor based on the iterative approach discussed beginning on page 122 was developed and used as an independent variable. Two additional independent variables were used in each analysis--livestock production and prices of livestock and livestock products. The hypothesis to be tested was that each of the three independent variables affects prices of corn, oats, and barley similarly. The following equations more definitely indicate the nature of the hypothesis:

$$\left.\begin{aligned} X_0' &= a' + b_{01.23}'X_1' + b_{02.13}'X_2 + b_{03.12}'X_3 \\ X_0'' &= a'' + b_{01.23}''X_1'' + b_{02.13}''X_2 + b_{03.12}''X_3' \\ X_0''' &= a''' + b_{01.23}'''X_1''' + b_{02.13}'''X_2 + b_{03.12}'''X_3' \end{aligned}\right\} \tag{249}$$

where

X_0' is the price of corn during June to September,

X_0'' is the price of oats during July to October,

X_0''' is the price of barley during July to October,

X_1', X_1'', and X_1''' are the respective weighted supply variables,

X_2 is the production of livestock during July to December, and

X_3 and X_3' are the average prices of livestock and products for the months to which the respective X_0's apply.

The hypothesis to be tested is that the following relationships hold simultaneously:

$$\left.\begin{aligned} b_{01.23}' &= b_{01.23}'' = b_{01.23}''' = \ddot{b}_{01.23} \\ b_{02.13}' &= b_{02.13}'' = b_{02.13}''' = \ddot{b}_{02.13} \\ b_{03.12}' &= b_{03.12}'' = b_{03.12}''' = \ddot{b}_{03.12} \end{aligned}\right\} \tag{250}$$

The following steps are involved in making the test:

1. For each of the three analyses, compute the unexplained sum of squares. In each case, this equals $\Sigma\ (X_0 - \bar{X}_0)^2\ (1 - R^2_{0.123})$. Add the results. The degrees of freedom attached to this sum of squares is N-P, where N is the total number of observations in the several analyses and P is the number of restrictions, that is the number of coefficients involved in the 3 equations in (249). In this case, each analysis was based on 20 years. Hence, N = 60, and P = 12 (9 regression coefficients plus 3 constant terms).

2. Combine the respective sums of squares and cross-products, after correcting for the respective means, for the three analyses. In the notation used above

$$\Sigma\ \ddot{x}_0^2 = \Sigma\ x_0'^2 + \Sigma\ x_0''^2 + \Sigma\ x_0'''^2, \text{ etc.} \tag{251}$$

Rerun the analysis, using these totals in obtaining the regression coefficients. Compute the unexplained sum of squares for the combined analysis, which equals $\Sigma\ \ddot{x}_0^2\ (1 - \ddot{R}^2_{0.123})$. The degrees of freedom attached to this sum of squares is N-6, the 6 representing the three regression coefficients from the combined analysis plus the 3 constant terms for the separate equations.

3. Subtract the result in step 1 from that in step 2 and divide by the difference between the respective degrees of freedom. In this case, the difference in degrees of freedom is 6 [N-6 - (N-12)]. This represents the mean square resulting from differences among the regression coefficients in the three analyses.

4. Divide the final result in step 1 by the degrees of freedom attached to it, namely N-12, in this case. This represents the error or remainder mean square.

5. Compute the ratio between the mean square in step 3 and the mean square in step 4, and compare this with the tabular values in an F table, using 6 and N-P degrees of freedom. If the ratio obtained is larger than the tabular value at the 5-percent point, the differences between the regression coefficients from the respective equations are statistically significant and the hypothesis is rejected. In this case, the ratio was smaller than the tabular value, as would have been expected, and the hypothesis was accepted.

Had we wished to include in (250) the hypothesis that a' = a'' = a''' = ä, the same test could be used, except that deviations from a common mean would have been used in step 2 and the degrees of freedom attached to it would be N-4, the 4 representing the three regression coefficients from the combined analysis plus the one constant term. The divisor in step 3 then would be 8 [N-4 - (N-12)], and appropriate adjustments would be required in step 5.

<u>A test to ascertain whether a particular set of regression coefficients in three or more regression equations are equal</u>.--This test was applied to analyses of factors that affect the ratio of prices of oats, barley, and sorghum grains, respectively, to prices of corn. In each analysis, the respective supply ratio was used as an independent variable. The hypothesis to be tested was that the regression coefficients for the supply ratio were identical for the three analyses. The test was complicated somewhat by the fact that two of the regression equations involved only a single independent variable, whereas the third involved three. The three equations can be written as follows:

$$\left.\begin{aligned} X_0' &= a' + b_{01}' X_1' \\ X_0'' &= a'' + b_{01}'' X_1'' \\ X_0''' &= a''' + b_{01.23}''' X_1''' + b_{02.13}X_2 + b_{03.12}X_3, \end{aligned}\right\} \qquad (252)$$

where

X_0' is the ratio between the price of oats and the price of corn,

X_0'' is the ratio between the prices of barley and that of corn,

X_0''' is the ratio between the price of sorghum grains and that of corn,

X_1', X_1'', and X_1''' are the ratios between the supply of the respective items and of corn, and X_2 and X_3 are related variables.

The hypothesis to be tested is that the following relationship holds:

$$b_{01}' = b_{01}'' = b_{01.23}''' = \ddot{b}_{01.23}. \qquad (253)$$

The steps involved are the same as those given above, except that the sums of squares and cross-products involving X_2 and X_3 in step 2 are based only on the variables from the third equation in (252), and P = 8 (5 regression coefficients plus 3 constant terms). The divisor in step 3 equals 2 [N-6 - (N-8)], and in step 4 equals N-8. Appropriate adjustments are made in step 5. In this case, the ratio between the mean square in step 3 and that in step 4 was larger than the tabular value at the 5-percent point, as would have been expected, and the hypothesis was rejected.

<u>A test to ascertain whether the composite effect of several variables in an analysis is statistically significant</u>.--This example is based on an unpublished study made by the Stanford Research Institute of the effects of an advertising and promotion program for a certain commodity. A regression analysis was run using eight separate variables that related to promotion, in addition to price, consumer income, and temperature. Wrong signs and coefficients that did not differ from zero by a statistically significant amount were obtained for a number of the variables relating to promotion. Hence, a test was desired of whether the eight variables considered as a group had a

significant effect. This test was run in the following way. The initial analysis contained 12 variables in all. The study was rerun, omitting the 8 promotional variables. The square of the multiple correlation coefficient from each of these studies was used as an indicator of the proportion of the initial variation in the dependent variable explained. These values are shown in the first column of table 14; the appropriate degrees of freedom are placed in the second column. The increment explained by the additional eight variables is obtained by subtraction and, naturally, eight degrees of freedom are used to obtain this increment. The unexplained variation for the 12-variable analysis is obtained by taking 1 minus the explained variation. An estimate of the mean squares is obtained in each case by dividing the figure in the first column by the degrees of freedom in the second column. The F-ratio is obtained by taking the ratio of the two numbers shown in the last column. An F-ratio of 2.69 is indicated, compared with a 1-percent critical value of 2.72. Thus, we would expect to get an F-ratio of this magnitude only slightly more than 1 percent of the time if, in fact, the 8 additional variables had no more effect on the multiple correlation coefficient than would occur if they were completely unrelated to the dependent variable. Thus, the 8 variables collectively have a statistically-significant effect on the dependent variable, even though it was impossible to obtain statistically significant results for each of them separately.

Table 14.--Analysis of variance of the added effect of 8 variables that relate to promotional activities

Type of variance	Proportion of initial variation (1)	Degrees of freedom (2)	Estimate of mean square (3)
Explained by--			
12 variables	0.498	92	---
4 variables	.380	100	---
Increment by the additional 8 variables	.118	8	0.0148
Unexplained by 12 variables	.502	92	.0055

VALIDITY OF AN ESTIMATE FROM A MULTIPLE REGRESSION EQUATION

Conditions Required for Valid Forecasts 103/

Under what conditions does the following equation give a valid estimate of X_1?

103/ This section is based on material contributed by Frederick V. Waugh, Director, Agricultural Economics Division, Agricultural Marketing Service, and published by Foote and Fox (29, pp. 35-36).

$$X_1 = a + b_{12.3} X_2 + b_{13.2} X_3. \tag{254}$$

1. There must be a significant "scatter" between X_2 and X_3 - that is, r_{23} must differ significantly from 1. This is not a matter of sampling error. A correlation of 0.99 may differ significantly from 1 so far as sampling is concerned. It is solely a matter of errors of observation. A statistician should know enough about his data to judge whether the observed scatter is larger than could happen as a result of errors in X_2 and X_3. If not, the scatter is nonsignificant, and the equation is worthless.

2. Usually the equation is invalid in the case of extrapolation beyond observed values of X_2 and X_3. This can be tested by drawing a simple scatter diagram for X_2 and X_3. If they are highly correlated, the observations lie within a narrow ellipse. The values of X_1 associated with combinations of X_2 and X_3 which lie within this ellipse can be estimated from the equation. Unless we are willing to extrapolate, we cannot estimate the value of X_1 associated with any combination of X_2 and X_3 lying outside the ellipse. A chi-square test described in subsequent paragraphs can be used to measure the degree of extrapolation involved in equations having more than 2 independent variables.

3. The error of a forecast of X_1 is composed of two parts: First, the standard error of estimate; and second the error associated with the regression plane,

$$X_1 - a - b_{12.3} X_2 - b_{13.2} X_3 = 0. \tag{254.1}$$

The first of these errors is constant; the second varies with the particular values of X_2 and X_3. The error of the regression plane is least at the center of the ellipse mentioned above. We could draw a series of ellipses around this center; as we moved away from the center, each successive ellipse would connect combinations of X_2 and X_3 for which the error of the regression plane would be equal. And each successive ellipse would indicate a larger error. The standard error of a forecast as described by Ezekiel (20, pp. 342-345) allows for both types of errors. It applies exactly only for a set of independent variables identical to that included in the original analysis. However, it may be assumed to hold approximately for other values that lie within this range.

A Chi-square Test to Estimate the Degree of Extrapolation in a Multivariate Scatter [104]

Each value of the several independent variables used in a multiple regression analysis may lie within the range included in the analysis, but when considered together, groups of variables may lie outside the range of the original set.

104/ The test here described was developed by Waugh and Been (101); the description of its computation and use is based in part on Armore and Burtis (6, pp. 7, 9).

X

In this event, the new group is an extrapolation from the original scatter just as much as if one of the values lay outside the range of the original set for that variable. Hidden extrapolation of this kind becomes difficult to detect graphically when there are three independent variables and practically impossible to discover with more than three.

The N observations of the n independent variables used in a regression analysis may be represented by N points scattered in n dimensional space. The pattern and degree of concentration of this scatter depend on the structure of intercorrelation among the independent variables as well as the variances of the variables. Waugh and Been suggest that for any number of independent variables, a chi-square can be calculated for each combination of observations to indicate its position with respect to the grouping tendency of the whole set of observed combinations, as defined by the pattern and degree of concentration of the observed scatter. When the values of all independent variables are at their means, chi-square equals zero. As the values depart from their means, chi-square increases. However, chi-square also depends on the structure of intercorrelation among the independent variables in such a way that it indicates the position of any given combination of values of the independent variables with respect to the grouping tendency of the whole set of observed combinations.

A value for chi-square can be obtained as a preliminary step in the computations of the standard error of a function or, if the latter has been computed, may be obtained from it. The following formulas show the relations:

$$s^2_{F1.2345} = s^2_{1.2345} \frac{(1 + \chi^2)}{N} \qquad (255)$$

$$\chi^2 = \frac{N s^2_{F1.2345}}{s^2_{1.2345}} - 1 \; . \qquad (255.1)$$

Methods for computing χ^2 in connection with equation (255) are obvious, given the formula for the standard error of a function [see, for example, Friedman and Foote (40, pp. 17-19)].

The theoretical probability of each chi-square can be found in a chi-square table, which is given in many statistical textbooks. This indicates the probability of occurrence of the given combination, or one farther from the grouping tendency, in sampling from a universe implied by the scatter of the base-period data. A separate chi-square is computed for each observation. If the computed value for observations outside the initial sample is larger than any of those for the period on which the analysis is based, then the observation must be considered an extrapolation. The higher the chi-square the greater the extrapolation.

USE OF SYSTEMS OF EQUATIONS FOR ANALYTICAL PURPOSES

In using systems of equations for forecasting or analytical purposes, the standard procedure is to insert values for the predetermined variables within the system and to simultaneously estimate the several endogenous variables. Methods for doing this are discussed by Friedman and Foote (40, pp. 81-85) and elsewhere. Other types of equations that may be useful in forecasting are discussed in the section beginning on page 128. In this section, we describe instead (1) methods by which coefficients from a variety of sources can be combined within a structural system to make specified analyses and (2) procedures by which a system can be modified to allow for structural changes that are believed to have taken place.

Assigning Values to Specified Parameters 105/

The material that follows presents a method by which a rough quantitative measurement can be made of how a price support program for farm products can contribute, directly or indirectly, to stability in the remainder of the economy. Figure 10 shows some of the major lines of influence along which the effects of a price support program may be transmitted to other parts of the economy. The coefficient beside each arrow represents the estimated percentage change in the variable to which the arrow points which is associated with a 1-percent change in the variable from which the arrow leads. Most of these "path coefficients" are based upon known factors, such as the weights of particular components of official index numbers, or the coefficients of statistical demand functions. Others seem reasonable, but could be checked by empirical analysis. One coefficients assumes a "multiplier" of 2, and one operates with a time lag. Values of the three coefficients marked with asterisks are pure assumptions.

As indicated in figure 10, the farm price support program during a period of recession would have three immediate effects: It (1) raises the average level of prices received by farmers; (2) reduces farm output, at least after the first 12 months of recession; and (3) reduces the commercial utilization of farm products.

Suppose, for example, that the direct effect of a price support program is to increase farm prices by 10 percent. If marketing margins remain constant, this will increase retail food prices about 5 percent. As retail food prices carry a weight of 30 percent in the consumer price index, that index will rise 1.5 percent.

The "influence" of the consumer price index upon wage rates is based on pure assumption. This index figures in some important wage contracts, and is widely used as a talking point in wage disputes. The coefficient in figure 10 implies that a 1.5 percent increase in the consumer price index would have the

---

105/ Material in this section is adapted from Fox (36, pp. 334-337).

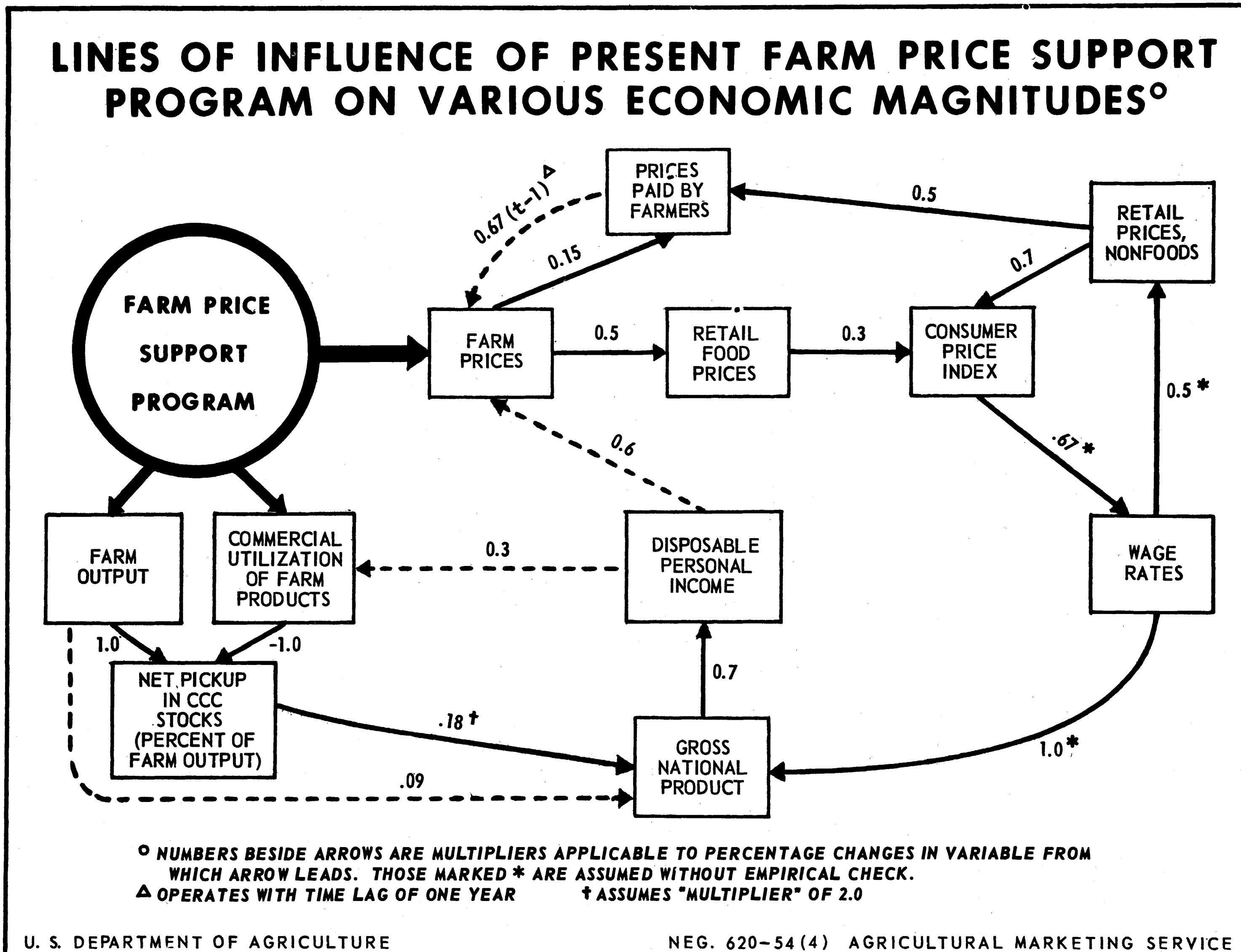


Figure 10.--Coefficients from a variety of sources can be inserted into a structural system and the results used for analytical purposes. Coefficients shown in this chart are used to measure effects of a price support program for farm products on the general economy during an economic recession.

effect of maintaining wage rates 1 percent higher than they would otherwise have been. The influences of wage rates upon gross national product and upon the retail prices of nonfood products also rest on assumption. The latter coefficient assumes that wages constitute about 50 percent of value added in manufacturing and distributing processes, and that most nonfarm prices are administered in such a way that these direct wage costs are covered, even during a recession.

Figure 10 implies that the initial or direct increase in the consumer price index generates a further increase in the same index. An increase in the consumer price index increases wage rates, which increase nonfood prices, which enter the consumer price index with a weight of 70 percent. Hence, the total effect upon the consumer price index of an increase in farm prices consists of the direct influence plus this "feed-back" effect.

An initial increase of 10 percent in prices received by farmers leads directly, through prices of purchased livestock, feed, and food products, to something like a 1.5 percent increase in the index of prices paid by farmers. There is also an indirect effect operating through the consumer price index, wage rates, and retail prices of nonfood products. This effect is only about a fourth as large as the direct one.

The dotted arrow running from "prices paid" to "farm prices" reflects the use of the prices paid index as a basis for setting price supports. As such supports are announced in advance of the planting season, this coefficient operates with a time lag of one year. Under the present price support program, the direct influence of a 1 percent increase in the prices paid index would apply to products accounting for only 45 percent of cash farm income; hence the direct effect on the average level of all farm prices would be only 0.45 percent. The coefficient of 0.67 shown in figure 10 allows for the influence of price support levels for feed grains upon the unsupported prices of meat animals, poultry and eggs.

Figure 10 also shows three chains of influence of price supports upon disposable personal income, particularly that of nonfarm people. An increase in disposable income raises prices of those farm products which are not supported and whose market supplies at any given time are fixed; it also increases commercial utilization (but not prices) of farm products which are in surplus at their applicable support prices. The direct effect of the farm price support program in raising the disposable income of farm operators is not adequately allowed for in figure 10.

The net increase in Commodity Credit Corporation stocks as a result of the price support program represents an injection of money from outside the private economy. It is equivalent to a purchase of goods by the Federal Government, with, in the practical situation, no simultaneous increase in government revenue. In figure 10 we apply a multiplier of 2 to the annual rate of increase in CCC price support stocks.

The model in figure 10 contains a number of implicit "path-multipliers." If we follow the arrow from farm prices through the consumer price index and back through disposable personal income to farm prices, we find that the initial increase in farm prices generates a secondary increase 4.2 percent as large as the first one. The second increase would generate a third order increase, and so on. Using a well-known formula for the sum of a power series, the final effect of a 1-percent increase in farm prices along this path would be equal to 1/(1-0.042), or 1.044 percent.

Some minor additional effects of the same sort are found if we consider the secondary "loops" centering around the consumer price index and disposable personal income. The chain of influences from farm prices through prices paid by farmers and back again involves a similar power series, but with a time lag of 1 year between the change in prices paid and the next-order change in farm prices. The effects of the price support program upon farm output (by means of acreage restrictions) and upon the net increase in CCC price support stocks may be additive to the initial effect of the program upon farm prices and may not involve power series multipliers.

In a stationary equilibrium, eliminating the effect of the one time lag in the system, it appears that if the existing price support program initially or directly increased farm prices by 10 percent it would result in a final level of farm prices about 12.3 percent higher than they would have been in the absence of a price support program. This figure depends, of course, on the coefficients in figure 10, and would be altered if some of these coefficients were revised.

Modifying Structural Systems 106/

Two sets of statistical analyses are basic for the studies reported in this section, and these are supplemented by certain other analyses. The first set is an equation that shows the effect of certain factors on the price of corn from November through May, when marketings are heaviest. The other set is a system of 6 equations that shows the simultaneous effect of 14 given variables on domestic and world prices for wheat and on domestic utilization for food, feed, export, and storage of wheat for the July to June marketing year. The supplemental analyses include studies of (1) normal seasonal variation in prices and (2) relationships among prices at local and specified terminal markets. These are mentioned in later sections.

The analysis for corn is described by Foote (26, pp. 5-12). The following variables were used:

X_0 - price per bushel received by farmers for corn, average for November to May, cents.

106/ Material in this section is adapted from Foote and Weingarten (31).

X_1 - total supply of feed concentrates for the year beginning in October, million tons.

X_2 - grain-consuming animal units fed on farms during the year beginning in October, millions.

X_3 - price received by farmers for livestock and livestock products, index numbers (1910-14=100), average for November to May.

The following regression equation applies:

$$\text{Log } X_0' = -0.95 - 1.82 \log X_1 + 1.71 \log X_2 + 1.36 \log X_3. \qquad (256)$$

For any given year, if expected values for X_2 and X_3 are inserted, this equation can be written in the following way:

$$\text{Log } X_0' = \log A_1 - 1.82 \log X_1, \qquad (256.1)$$

where $\log A_1 = -0.95 + 1.71 \log X_2 + 1.36 \log X_3$ for that year. In the rest of this paper, the form shown by equation (256.1) is used. The reader should remember, however, that the applicable value for $\log A_1$ must be obtained from equation (256).

The system of equations for wheat is described in detail by Meinken (71, pp. 36-50). Because of space limitations, a list of all variables taken as given for this system of equations cannot be included here. Included among these given variables is the price of corn, but, as is shown later, the system can be modified to include corn prices among the variables that are simultaneously determined within the system.

Variables that are assumed to be determined simultaneously within the original system of equations for wheat include the following:

P_w - wholesale price per bushel of wheat at Liverpool, England, converted to United States currency, cents.

P_d - wholesale price per bushel of No. 2 Hard Winter wheat at Kansas City, cents.

C_f - domestic use of wheat for feed, million bushels.

C_e - domestic net exports of wheat and flour on a wheat equivalent basis, million bushels.

C_s - domestic end-of-year stocks of wheat, million bushels.

C_h - domestic use of wheat and wheat products for food by civilians on a wheat equivalent basis, million bushels.

All variables relate to a marketing year beginning in July. C_s is assumed to apply to stocks held in commercial hands. When a price-support program is in effect, end-of-year stocks under loan or held by the Commodity Credit Corporation are computed as a residual.

P_w is assumed to depend directly on certain given variables; hence its value in any year can be obtained by a direct solution of a single equation similar to equation (256) for corn. It then can be treated as though it were given. The values of the given variables and the calculated value of P_w for any year can be substituted in each equation. By making computations similar to those used in obtaining log A_1, new constant terms can be obtained for each equation. The equations then can be written conveniently in the following form. These equations bear the same relation to the original equations as equation (256.1) does to (256).

$$C_h + C_f + C_e + C_s = A_2 \qquad (257)$$

$$C_h + 0.0015LP_d = LA_3 \qquad (258)$$

$$C_f + 2.5P_d = A_4 \qquad (259)$$

$$C_e + 7.8P_d = A_5 \qquad (260)$$

$$C_s + 411(P_d/I_d) = A_6 \qquad (261)$$

Two given variables are involved in these equations. They are (1) L, the total population eating out of civilian supplies, in millions, and (2) I_d, wholesale prices of all commodities in this country as computed by the Bureau of Labor Statistics (1926=100). They cannot be included in the modified constants because they appear as a multiplier or divisor, respectively, of P_d.

By subtracting the last 4 equations from equation (257) and solving the resulting equation for P_d, the following formula is given:

$$P_d = \frac{A_2 - LA_3 - A_4 - A_5 - A_6}{-0.0015L - (411/I_d) - 10.3} \cdot \qquad (262)$$

Once a value for P_d is obtained, equations (258) to (261) can be solved directly, after inserting values for L and I_d, to obtain the 4 price-determined utilizations.

Allowing for an assumed maximum on exports.--Because of the effect of institutional forces in the world today, it is believed that exports from this country that exceed specified levels will result in retaliatory action on the part of other governments. So long as our exports remain below these levels, it is likely that the same kind of economic forces will apply as those in the pre-World War II years on which the analysis was based. This adjustment in the system of equations can be made easily. The equations are first solved with no restriction on exports. If the indicated figure for C_e is higher than the specified maximum, the following formula is used to estimate P_d. In this formula, the symbol E is used to indicate the assumed maximum for exports.

$$P_d = \frac{A_2 - LA_3 - A_4 - E - A_6}{-0.0015L - (411/I_d) - 2.5} \qquad (262.1)$$

The reader can easily verify that this formula is obtained by substituting $C_e = E$ for equation (260), then deriving the formula for P_d by the same algebraic process as that used in the previous case. The other utilizations are obtained in the same way as previously.

Including the price of corn as an endogenous variable in the system for wheat.--In most years, the quantity of wheat fed is so small in relation to the total supply of feed concentrates that the price of corn can be assumed to be determined independently of the price of wheat. If price supports and acreage control programs for wheat and corn were dropped, however, the quantity of wheat fed might increase significantly. Under these circumstances, the quantity of wheat fed depends partly on the price of corn, and the price of corn depends to some extent on the quantity of wheat fed. Hence, it seemed desirable to modify the system of equations for wheat so that the price of corn could be included among the simultaneously determined variables.

If the analysis for corn had been based on a linear, rather than a logarithmic, relationship, this could have been done easily. In the next few paragraphs we discuss how a linear relationship was derived from the logarithmic one for corn. The linear relation can be used as an approximation for the logarithmic if changes in X_1 from the initial value are small.

To simplify the discussion, we first rewrite equation (256.1) by substituting the letter b for the numerical value of the regression coefficient. Thus b = -1.82. The equation then reads:

$$\log X_0' = \log A_1 + b \log X_1. \qquad (256.2)$$

If we translate this equation into actual numbers (rather than logarithms) we obtain:

$$X_0' = A_1 X_1{}^b \ . \qquad (256.3)$$

We now borrow a notion from differential calculus. To get the slope of a curve at any given point, we need to evaluate the first derivative at that point. The first derivative of the function (256.3) with respect to X_1 is:

$$\frac{dX_0'}{dX_1} = bA_1X_1{}^{b-1} \ . \qquad (263)$$

Inserting the value for b, we get:

$$\frac{dX_0'}{dX_1} = -1.82A_1X_1^{-2.82} \qquad (263.1)$$

We wish to evaluate the slope of the line when X_1--total supply of feed concentrates--is at its expected level, for the particular analysis, making use of the appropriate value of A_1. As of the start of the analysis, we know all values that enter into X_1 except the quantity of wheat to be fed during the crop year, and that we can estimate approximately. In most instances, an error of as much as 100 percent in our advance estimate of the quantity of wheat fed will affect X_1 by only a few percentage points (as the quantity of wheat fed normally constitutes only about 2 percent of the total supply of feed concentrates) and will affect the estimate of the slope of the line even less. If the initial estimate of the quantity of wheat fed is found to be badly off, after making the computations for the system of equations, so that the computed linear relationship is a poor approximation to the true curve, we can always make a better approximation by using a revised value for X_1 and then making a new set of computations for the system. Let us designate the answer obtained from equation (263.1) as B. The reader should note that logarithms are needed to evaluate the expression $X_1^{-2.82}$.

We now wish to obtain a linear equation that has the slope B and that passes through the point on the original logarithmic curve at the chosen value for X_1. By substituting the estimated value of X_1 in equation (256.1), we can obtain an estimated value for X_0 at that point. Let us designate these numbers by the symbols $\hat{X}_0$, $\hat{X}_1$. We now write the equation of the desired linear relation as:

$$X_0' = (\hat{X}_0 - B\hat{X}_1) + BX_1. \qquad (264)$$

The reader who remembers his elementary analytical geometry will see that this is the equation of a line for which we know the slope and 1 point.

We must now effect some further transformations to make equation (264) apply to the variables included in the system of equations for wheat. For the combined analysis, all of X_1 is assumed to be given except the quantity of wheat fed. This can be allowed for in the equation by letting

$$X_1 = X_1'' + C_f''. \qquad (265)$$

BX_1'' then can be combined with the other constant terms in the equation. The symbol C_f'' is used because this is in terms of million tons, while C_f, as used in the system of equations for wheat, is in million bushels. The relationship between C_f'' and C_f is given by:

$$C_f'' = \frac{60}{2{,}000} C_f. \qquad (266)$$

In the system of equations for wheat, the price of corn, P_c, relates to 60 pounds of No. 3 Yellow at Chicago, average for July-December, in cents, whereas X_0 is the average price received by farmers per standard or 56-pound bushel, average for November-May, in cents. A relationship between P_c and X_0 can be developed in several ways, one of which follows: (1) Based on the computation discussed by Foote (26, p. 12), the season-average price received by farmers for corn equals approximately $X_0/0.95$. (2) Based on an analysis referred to by Foote, Klein, and Clough (30, p. 65), the annual average price of No. 3 Yellow corn at Chicago equals the annual price received by farmers for all corn times 1.05 plus 1.11 cents. (3) Based on index numbers of normal seasonal variation for No. 3 Yellow corn at Chicago as shown on page 50 of that bulletin, the July-December price at Chicago equals 1.017 times the annual price. (4) The price of 60 pounds of corn naturally equals 60/56 times the price of a standard bushel. By combining these relationships, we find that

$$X_0 = 0.83P_c - 1.004. \qquad (267)$$

If we make the three substitutions implied by equations (265), (266), and (267), we can rewrite equation (264) as

$$P_c = 1.2\ (\hat{X}_0 - B\hat{X}_1 + BX_1' + 1.004) + 0.036\ BC_f. \qquad (264.1)$$

By letting $A_7 = 1.2(\hat{X}_0 - B\hat{X}_1 + BX_1' + 1.004)$ and $b_{71} = 0.036B$, we can rewrite this as

$$P_c = A_7 + b_{71}C_f. \qquad (264.2)$$

The equation in this form is used in the rest of the discussion. In following it, one should keep in mind the substantial number of computations involved in obtaining A_7 and b_{71}.

We are now ready to consider the system of equations that includes (264.2). Referring to page 191, if equations (258), (260), and (261) are subtracted from equation (257), equation (268) shown below is given. Equation (259) now must be modified to show P_c as a separate variable. This is done by removing $2.5P_c$ from A_4 and transposing this term to the opposite side of the equality sign. The modified equation is designated as equation (259.1) in the system shown below, and the modified A_4, as A_4'. Equations (268), (259.1), and (264.2) can be written conveniently as follows:

$$C_f - (0.0015L + 7.8 + 411/I_d)P_d = A_2 - LA_3 - A_5 - A_6 \qquad (268)$$

$$C_f + 2.5P_d - 2.5P_c = A_4' \qquad (259.1)$$

$$-b_{71}C_f + P_c = A_7. \qquad (264.2)$$

If equation (264.2) is multiplied by -2.5 and subtracted from equation (259.1), the following equation results:

$$(1-2.5b_{71})C_f + 2.5P_d = A_4' + 2.5A_7. \qquad (269)$$

If equation (268) is multiplied by $(1-2.5b_{71})$ and subtracted from equation (269), a formula for P_d can be derived directly. To write this in algebraic symbols is somewhat complicated but, when working with numbers in an actual problem, it would be very simple. A value for C_f then can be obtained from equation (268), P_c can be obtained from equation (264.2), and the other price-determined utilizations for wheat can be obtained easily from the initial equations.

Allowing for a fixed negative differential between wheat and corn prices.--A linear relationship is believed to apply between the quantity of wheat fed and the price of wheat when the spread between the price of wheat and the price of corn used in the analysis is between zero and 40 cents per 60 pounds (the weight of a bushel of wheat).

For larger price spreads, requirements for wheat in poultry and other rations is more than the quantity indicated by the linear analysis. Thus, when use for feed is plotted on the vertical scale, a slope that becomes less steep is required. When the price of wheat approaches or falls below the comparable price of corn, use of wheat for feed increases rapidly and by more than that suggested by the linear relationship. For this part of the curve, a slope that becomes increasingly steep is required. When the price spread is outside the specified range, the quantity of wheat fed frequently can be estimated approximately by making use of a logarithmic relation between prices of wheat and quantity fed. The logarithmic relation can be modified for inclusion in the system of linear relations for wheat by the same general approach as described for the price equation for corn (see page 192).

When the price of wheat falls near or below that for corn, the demand for wheat for feeding is much more elastic than when the price is considerably above that for corn. The logarithmic analysis for wheat fed cannot be used with negative price differentials because the logarithm of a negative number is undefined. The following method was used instead: A 20-cent negative differential between prices received by farmers for wheat and corn seemed like a maximum, and the regression coefficient for $(P_d - P_c)$ in equation (259.1) was adjusted in such a way as to reduce the negative price differential to this level.

The algebra involved in obtaining the adjusted coefficient is rather complicated and need not be shown in detail here. The general approach is as follows: (1) By making use of relationships between prices received by farmers and prices at specified terminal markets, the algebraic value for $P_d - P_c$ that is equivalent to a negative spread of 20 cents at the farm level can be obtained. Let this algebraic value equal M. (2) Equation (259.1) is modified to substitute a regression coefficient for $P_d - P_c$ that is unknown for the value of 2.5 used under normal circumstances. Call this coefficient K. (3) M is substituted for $P_d - P_c$ in equation (259.1) and $P_d - M$ is substituted for P_c in equation (264.2). This eliminates P_c from the equations. (4) Equations (268), (259.1), and (264.2) now contain 3 unknowns--C_f, P_d, and K. As some of the equations may be nonlinear, part of the solution of them may need to be made graphically. Once values for C_f, P_d, and K have been obtained, the other desired unknowns can be obtained easily. A regression coefficient of 11 instead of 2.5 was obtained in one set of analyses.

Effects of a multiple-price plan for wheat.--Meinken (71, pp. 49-50) describes how his system of equations can be used to study the effect of multiple-price plans. Suppose a 2-price plan is in effect under which wheat used for domestic food consumption is sold at a price equivalent to 100 percent of parity, while the remaining wheat sells at a free-market price. The amount of wheat used for food could be estimated from equation (258) (see page 191) based on a value for P_d equivalent to the parity price. Suppose this amount is C_h. The equation $C_h = C_h$ is substituted for equation (258), and the system is solved for the other variables in the same way as described on page 191.

If the Government were to place a floor under the "free" price at say 50 percent of parity, an estimate of the quantity of wheat going under the support program at this price, if any, could be obtained as a residual after computing the expected utilizations and commercial carryover. If the Government established a price for wheat used for food and a lower price for wheat used for feed, an approach similar to that described above could be used to solve for the expected utilizations for food and feed and the free-market price at which the remaining wheat would sell.

LITERATURE CITED

(1) Aitkin, A. C.
1934-35. On Least Squares and Linear Combinations of Observations. Royal Soc. of Edinburgh, Proc. 55:42-48.
(2) Allen, R. G. D.
1956. Mathematical Economics. 768 pp., illus. London.
(3) Anderson, R. L.
1954. The Problem of Autocorrelation in Regression Analysis. Jour. Amer. Statis. Asso. 49:113-129.
(4) Anderson, T. W., and Rubin, Herman
1949. Estimation of the Parameters of a Single Equation in a Complete System of Stochastic Equations. Annals Math. Statis. 20:46-63.
(5) Armore, Sidney J.
1953. The Demand and Price Structure for Food Fats and Oils. U. S. Dept. Agr. Tech. Bull. 1068, 69 pp., illus.
(6) ______, and Burtis, Edgar L.
1950. Factors Affecting Consumption of Fats and Oils, Other than Butter, in the United States. Agr. Econ. Research. 2:1-9, illus.
(7) Basmann, R. L.
1957. A Generalized Classical Method of Linear Estimation of Coefficients in a Structural Equation. Econometrica. 25:77-83.
(8) Bennion, E. G.
1952. The Cowles Commission's "Simultaneous-Equation Approach": A Simplified Explanation. Rev. Econ. and Statis. 34:49-56, illus.

(9) Berry, Charles H., Anderson, E.O., Baker, Donald J., Brinegar, George K., Dowd, Leonard R., and Johnson, Stewart
1956. The Demand for Fluid Skim Milk. Conn. Agr. Exp. Sta. Bull. 325, 31 pp., illus.

(10) Brandow, G. E.
1956. A Statistical Analysis of Apple Supply and Demand. Penn State Univ. Agr. Econ. and Rural Sociology Rpt. 2, 23 pp. [Processed.]

(11) Breimyer, Harold F.
1957. On Price Determination and Aggregative Price Theory. Jour. Farm Econ. 39:676-694, illus.

(12) Bronfenbrenner, Jean
1953. Sources and Size of Least-Squares Bias in a Two-Equation Model. In Hood, Wm. C., and Koopmans, Tjalling C., ed., Studies in Econometric Method, Cowles Commission for Research in Economics Mongr. 14, 324 pp., illus.

(13) Chernoff, Herman, and Divinsky, Nathan
1953. The Computation of Maximum-Likelihood Estimates of Linear Structural Equations. In Hood, Wm. C., and Koopmans, Tjalling C., ed., Studies in Econometric Method, Cowles Commission for Research in Economics Mongr. 14, 324 pp., illus.

(14) Christ, Carl F.
1956. Aggregate Economic Models: A Review Article. Amer. Econ. Rev. 46:385-408.

(15) Cochrane, D., and Orcutt, G. H.
1949. Application of Least Squares Regression to Relationships Containing Autocorrelated Error Terms. Jour. Amer. Statis. Asso. 44:32-61, illus.

(16) Cramér, Harald
1955. The Elements of Probability Theory and Some of Its Applications. 281 pp. New York.

(17) Durbin, J.
1951. Discussion on Paper by Morgan and Corlett. Jour. Royal Statis. Soc., Ser. A. 64:354-355.

(18) ______, and Watson, G. S.
1951. Testing for Serial Correlation in Least Squares Regression. Biometrika. 38(1-2):159-177, illus.

(19) Ezekiel, Mordecai
1928. Statistical Analyses and the "Laws" of Price. Quart. Jour. Econ. 42:199-225, illus.

(20) ______
1941. Methods of Correlation Analysis. Ed. 2, 531 pp., illus. New York.

(21) Farrell, M. J.
1954. Some Aggregation Problems in Demand Analysis. Rev. Econ. Studies. 21:193-203.

(22) ______
1954. The Demand for Motor-Cars in the United States. Jour. Royal Statis. Soc., Ser. A. 117:171-201.

(23) Ferber, Robert
1949. Statistical Techniques in Market Research. 542 pp., illus. New York.

(24) Fisher, Irving
1925. Our Unstable Dollar and the So-called Business Cycle. Jour. Amer. Statis. Asso. 20:179-202.

(25) Foote, Richard J.
1953. The Mathematical Basis for the Bean Method of Graphic Multiple Correlation. Jour. Amer. Statis. Asso. 48:778-788.

(26) ______
1953. Statistical Analyses Relating to the Feed-Livestock Economy. U. S. Dept. Agr. Tech. Bull. 1070, 41 pp., illus.

(27) ______
1955. A Comparison of Single and Simultaneous Equation Techniques. Jour. Farm Econ. 37:975-990, illus.

(28) ______
1957. Avoiding Nonsense Correlations in Time Series Analysis. U. S. Agr. Mkt. Service. 9 pp., illus. [Processed.]

(29) ______, and Fox, Karl A.
1954. Analytical Tools for Measuring Demand. U. S. Dept. Agr. Agr. Handb. 64, 86 pp., illus. [Processed.]

(30) ______, Klein, John W., and Clough, Malcolm
1953. The Demand and Price Structure for Corn and Total Feed Concentrates. U. S. Dept. Agr. Tech. Bull. 1061, 79 pp., illus.

(31) ______, and Weingarten, Hyman
1956. How Research Results Can Be Used To Analyze Alternative Governmental Policies. Agr. Econ. Research. 8:33-43.

(32) Fox, Karl A.
1951. The Measurement of Price Support Cost. Jour. Farm Econ. 33:470-484.

(33) ______
1953. The Analysis of Demand for Farm Products. U. S. Dept. Agr. Tech. Bull. 1081, 90 pp., illus.

(34) ______
1953. A Spatial Equilibrium Model of the Livestock-Feed Economy in the United States. Econometrica. 21:547-566.

(35) ______
1954. Use of Economic Models in Appraising Foreign Trade Policies. Jour. Farm Econ. 36:944-958.

(36) ______
1956. The Contribution of Farm Price Support Programs to General Economic Stability. In Policies to Combat Depression, A Report of the National Bureau of Economic Research, 417 pp., illus.

(37) ______
1956. Econometric Models of the United States. Jour. Political Economy. 64:128-142.

(38) ______, and Taeuber, Richard C.
1955. Spatial Equilibrium Models of the Livestock-Feed Economy. Amer. Econ. Rev. 45:584-608.

(39) Foytik, Jerry
1951. Characteristics of Demand for California Plums. Hilgardia. 20:407-527, illus.

(40) Friedman, Joan, and Foote, Richard J.
1955. Computational Methods for Handling Systems of Simultaneous Equations. U. S. Dept. Agr. Agr. Handb. 94, 109 pp., illus. [Processed.]

(41) Friedman, Milton
1957. A Theory of the Consumption Function. National Bureau of Economic Research General Series 63. 243 pp., illus.

(42) Gerra, Martin J.
1955. Supplementary Information on Procedure Used for Major Commodity Groups. In Measuring the Supply and Utilization of Farm Commodities, U. S. Dept. Agr. Agr. Handb. 91, 114 pp., illus. [Processed.]

(43) Godwin, Marshall R.
1952. Customer Response to Varying Prices for Florida Oranges. Fla. Agr. Exp. Sta. Bull. 508, 24 pp., illus.

(44) Gurland, John
1954. An Example of Autocorrelated Disturbances in Linear Regression. Econometrica. 22:218-227.

(45) Haavelmo, Trygve
1943. The Statistical Implications of a System of Simultaneous Equations. Econometrica. 11:1-12.

(46) Harrington, A. H., and Gislason, Conrad
1956. Demand and Quality Preferences for Deciduous Fruits. Jour. Farm Econ. 38:1405-1414, illus.

(47) Hermie, Albert M.
1951. Prices of Apparel Wool. U. S. Dept. Agr. Tech. Bull. 1041, 48 pp., illus.

(48) Hicks, J. R.
1946. Value and Capital. 2d ed., 340 pp., illus. Oxford, England.

(49) Hildreth, Clifford, and Jarrett, F. G.
1955. A Statistical Study of Livestock Production and Marketing. Cowles Commission for Research in Economics Mongr. 15, 156 pp., illus.

(50) Hopp, Henry, and Foote, Richard J.
1955. A Statistical Analysis of Factors That Affect Prices of Coffee. Jour. Farm Econ. 37:429-438, illus.

(51) Houthakker, H. S.
1956. The Pareto Distribution and the Cobb-Douglas Production Function in Activity Analysis. Rev. Econ. Studies. 23:27-31.

(52) Jasper, A. William
1954. Increasing Sales of Fresh Fryers Through Improved Merchandizing Practices. N. Y. (Cornell) Agr. Expt. Sta. A. E. 954, 37 pp., illus. [Processed.]

(53) Judge, George C.
1956. Competitive Position of the Connecticut Poultry Industry. No. 7. A Spatial Equilibrium Model for Eggs. Conn. (Storrs) Agr. Exp. Sta. Bull. 318, 36 pp., illus.

(54) Kempthorne, Oscar
1952. The Design and Analysis of Experiments. 631 pp. New York.
(55) Kendall, M. G.
1951. Regression, Structure, and Functional Relationship. Biometrika. 38:11-25.
(56) Klein, Lawrence R.
1946. Remarks on the Theory of Aggregation. Econometrica. 14:303-312.
(57) ______
1952-53. On the Interpretation of Professor Leontief's System. Rev. Econ. Studies. 20:131-136.
(58) ______
1953. A Textbook of Econometrics. 355 pp., illus. Evanston, Ill.
(59) ______
1955. On the Interpretation of Theil's Method of Estimating Economic Relationships. Metroeconomica. 7:147-153.
(60) ______, and Goldberger, A. S.
1955. An Econometric Model of the United States, 1929-1952. 165 pp., illus. Amsterdam.
(61) Koopmans, Tjalling C.
1953. Identification Problems in Economic Model Construction. In Hood, Wm. C., and Koopmans, Tjalling C., ed., Studies in Econometric Method, Cowles Commission for Research in Economics Mongr. 14, 324 pp., illus.
(62) Kuh, Edwin, and Meyer, John R.
1955. Correlation and Regression Estimates When the Data are Ratios. Econometrica. 23:400-416, illus.
(63) Ladd, George W.
1956. Effects of Shocks and Errors in Estimation: An Empirical Comparison. Jour. Farm Econ. 38:485-495.
(64) Lindley, D. V.
1947. Regression Lines and the Linear Functional Relationship. Jour. Royal Statis. Soc. Suppl. 9:218-244.
(65) Lowenstein, Frank, and Simon, Martin S.
1954. Analyses of Factors That Affect Mill Consumption of Cotton in the United States. Agr. Econ. Research. 6:101-110, illus.
(66) Marschak, Jacob
1953. Economic Measurements for Policy and Prediction. In Hood, Wm. C., and Koopmans, Tjalling C., ed., Studies in Econometric Method, Cowles Commission for Research in Economics Mongr. 14, 324 pp., illus.
(67) Marshall, Alfred
1948. Principles of Economics, 8th ed. 871 pp., illus. London. First published 1920.
(68) May, K.
1946. The Aggregation Problem for a One Industry Model. Econometrica. 14:285-298.
(69) ______
1947. Technological Change and Aggregation. Econometrica. 15:51-63.

(70) Meinken, Kenneth W.
1953. The Demand and Price Structure for Oats, Barley, and Sorghum Grains. U. S. Dept. Agr. Tech. Bull. 1080, 102 pp., illus.

(71) ______
1955. The Demand and Price Structure for Wheat. U. S. Dept. Agr. Tech. Bull. 1136, 93 pp., illus.

(72) ______, Rojko, Anthony S., and King, Gordon A.
1956. Measurement of Substitution in Demand from Time Series Data--A Synthesis of Three Approaches. Jour. Farm Econ. 38:711-735, illus.

(73) Meyer, John R., and Miller, Henry Lawrence, Jr.
1954. Some Comments on the "Simultaneous-Equations" Approach. Rev. Econ. and Statis. 36:88-92, illus.

(74) Mood, Alexander McFarlane
1950. Introduction to the Theory of Statistics. 433 pp., illus. New York.

(75) Nerlove, Marc
1957. A Note on Long-Run Automobile Demand. Jour. of Mkt. 22:57-64, illus.

(76) ______
1958. Distributed Lags and Demand Analysis for Agricultural and Other Commodities. U. S. Dept. Agr. Agr. Handb. 141, 121 pp., illus. [Processed.]

(77) Orcutt, G. H.
1948. A Study of the Autoregressive Nature of the Time Series Used for Tinbergen's Model of the Economic System of the United States, 1919-1932. Jour. Royal Statis. Soc. Supp. 10:1-53, illus.

(78) ______, and Cochrane, Donald
1949. A Sampling Study of the Merits of Autoregressive and Reduced Form Transformations in Regression Analysis. Jour. Amer. Statis. Asso. 44:356-372.

(79) Ostle, Bernard
1954. Statistics in Research. 487 pp., illus. Ames, Iowa.

(80) Pubols, Ben H.
1954. Factors Affecting Prices of Apples. Agr. Econ. Research. 6:77-84, illus.

(81) Rojko, Anthony S.
1957. The Demand and Price Structure for Dairy Products. U. S. Dept. Agr. Tech. Bull. 1168, 252 pp., illus.

(82) Sackrin, S. M.
1957. Income Elasticity of Demand for Cigarettes: A Cross-Section Analysis. Agr. Econ. Research. 11:1-9, illus.

(83) ______, and Conover, Arthur G.
1957. Tobacco Smoking in the United States in Relation to Income. U. S. Dept. Agr. Mkt. Research Rpt. 189, 216 pp., illus. [Processed.]

(84) Shuffett, D. Milton
1954. The Demand and Price Structure for Selected Vegetables. U. S. Dept. Agr. Tech. Bull. 1105, 133 pp., illus.

(85) Siegel, Sidney
1957. Nonparametric Statistics. Amer. Statis. 11:13-19.
(86) Snedecor, George W.
1940. Statistical Methods. Ed. 3, 422 pp., illus. Ames, Iowa.
(87) Stone, Richard
1954. The Measurement of Consumers' Expenditures and Behavior in the United Kingdom, 1920-1938. 448 pp., illus. Cambridge, England.
(88) Suits, Daniel B.
1955. An Econometric Model of the Watermelon Market. Jour. Farm Econ. 37:237-251, illus.
(89) ______, and Goldberger, Arthur
1955. A Statistical Model for 1955. Mich. Business Rev. 7:25-28.
(90) Theil, H.
1953. Estimation and Simultaneous Correlation in Complete Equation Systems. Netherlands Centraal Plan - Bureau. The Hague, Netherlands. [Processed.]
(91) ______
1954. Estimation of Parameters of Econometric Models. Internatl. Statis. Inst. Bull. 34:122-128.
(92) ______
1954. Linear Aggregation of Economic Relations. 205 pp., illus. Amsterdam.
(93) Tobin, James
1952. A Survey of the Theory of Rationing. Econometrica. 20:521-553.
(94) United States Bureau of Agricultural Economics
1953. Consumption of Food in the United States, 1909-52. U. S. Dept. Agr. Agr. Handb. 62, 249 pp., illus.
(95) ______, and Production and Marketing Administration
1949. The Agricultural Estimating and Reporting Service of the United States Department of Agriculture. U. S. Dept. Agr. Misc. Pub. 703, 266 pp., illus.
(96) Wagner, Harvey M.
1954. A Monte Carlo Study of Estimates of Simultaneous Linear Structural Equations. Tech. Rpt. 12, Dept. of Econ., Stanford Univ., 56 pp. [Processed.]
(97) Watson, G. S.
1955. Serial Correlation in Regression Analysis, I. Biometrika. 42:327-341.
(98) ______, and Hannan, E. J.
1956. Serial Correlation in Regression Analysis, II. Biometrika. 43:436-448.
(99) Waugh, Frederick V.
1956. A Partial Indifference Surface for Beef and Pork. Jour. Farm Econ. 38:102-112, illus.
(100) ______
1957. Graphic Analysis in Agricultural Economics. U. S. Dept. Agr. Agr. Handb. 128, 69 pp., illus. [Processed.]

(101) ______, and Been, R. O.
1939. Some Observations About the Validity of Multiple Regressions. Statis. Jour., N. Y. City College. 2:6-14.

(102) Wold, Herman
1938. A Study in the Analysis of Stationary Time Series. 214 pp. Uppsala.

(103) ______
1949. On Least Squares Regression With Autocorrelated Variables and Residuals. Trans. Internatl. Inst. Statis. 32(2):277-289.

(104) ______
1956. Causal Inference from Observational Data: A Review of Ends and Means. Jour. Royal Statis. Soc. 119:28-61, illus.

(105) ______, and Faxer, P.
1957. On the Specification Error in Regression Analysis. Annals Math. Statis. 28:265-267.

(106) ______, and Jureen, Lars
1953. Demand Analysis. 358 pp., illus. Stockholm.

(107) Working, Elmer
1927. What Do Statistical "Demand Curves" Show? Quart. Jour. Econ. 4:212-235, illus.

(108) Yule, G. Udney, and Kendall, M. G.
1937. An Introduction to the Theory of Statistics. Ed. 11, 570 pp., illus. London.

APPENDIX

Proof That the Elasticity of Price Transmission Is Less Than One in Specified Cases 107/

Case 1: Constant dollar margin.--In this case

$$P_r - P_w = c, \tag{270}$$

where c is a constant. It follows that

$$\frac{\partial P_r}{\partial Q_c} = \frac{\partial P_w}{\partial Q_c} = 0 \tag{271}$$

and, therefore, A = 0; and

$$\frac{\partial P_w}{\partial P_r} = 1. \tag{272}$$

Consequently

107/ This section was written by Marc Nerlove, agricultural economic statistician, Agricultural Marketing Service.

$$E_{P_rP_w} = 1 - \frac{c}{P_r} < 1 , \tag{273}$$

since both c and P_r are positive. It follows from equations (271) and (273) that

$$\frac{E_{P_rP_w}}{1 - A \cdot B} < 1 , \tag{274}$$

so that

$$E_{Q_cP_w} < E_{Q_cP_r} . \tag{275}$$

Case 2: Constant percentage margin.--In this case

$$P_r - P_w = a P_w , \tag{276}$$

where a is a constant. It follows that

$$\frac{\partial P_r}{\partial Q_c} = \frac{\partial P_w}{\partial Q_c} = 0 \tag{277}$$

and

$$\frac{\partial P_r}{\partial P_w} = 1 + a , \tag{278}$$

so that

$$\frac{E_{P_rP_w}}{1 - A \cdot B} = 1 . \tag{279}$$

In this case, therefore,

$$E_{Q_cP_w} = E_{Q_cP_r} . \tag{280}$$

Case 3: The margin is a positive monotonic increasing function of Q_c.--In this case

$$P_r - P_w = f(Q_c) , \tag{281}$$

where $f' \geq 0$. It follows that

$$\frac{\partial P_r}{\partial Q_c} = f' \geq 0 , \tag{282}$$

so that $A \geq 0$, and further that

$$\frac{\partial P_w}{\partial P_r} = 1 , \tag{283}$$

so that

$$E_{P_rP_c} = 1 - \frac{f(Q_c)}{P_w} < 1 \, , \tag{284}$$

since $f(Q_c)$ and P_w are positive. From equations (282) and (284) we have, therefore, that

$$\frac{E_{P_rP_w}}{1 - A \cdot B} < 1 \, . \tag{285}$$

Consequently,

$$E_{Q_cP_w} < E_{Q_cP_r} \, . \tag{286}$$

Charts that Indicate Approximately the Degree of Partial Correlation

The formulas in this section directly apply to a 5-variable problem. However, they may be applied to a 3- or 4-variable problem if the following adjustments are made:

(1) Subscripts which apply to variables not included in the analysis and which are to the right of the point may be omitted, that is, $b_{12.345}$ for a 5-variable problem is written as $b_{12.34}$ in a 4-variable problem and as $b_{12.3}$ in a 3-variable problem.

(2) Terms which include coefficients having subscripts to the left of the point and which apply to variables not included in the analysis should be omitted, that is, the coefficient corresponding to $b_{15.234}$ in a 5-variable problem does not appear in a 3- or 4-variable problem.

(3) Charts which involve variables having subscripts to the left of the point and which apply to variables not included in the analysis should be omitted.

In obtaining the charts, computational time is saved by first computing the unexplained residual, d_i, for each observation. As this residual is the same in each chart, the adjusted values of the dependent variable can be obtained from the following formulas:

$$X_{1.345_i} = d_i + b_{12.345} X_{2_i} + (\bar{X}_1 - b_{12.345} \bar{X}_2) \tag{287}$$

$$X_{1.245_i} = d_i + b_{13.245} X_{3_i} + (\bar{X}_1 - b_{13.245} \bar{X}_3) \tag{288}$$

$$X_{1.235_i} = d_i + b_{14.235} X_{4_i} + (\bar{X}_1 - b_{14.235} \bar{X}_4) \tag{289}$$

$$X_{1.234_i} = d_i + b_{15.234} X_{5_i} + (\bar{X}_1 - b_{15.234} \bar{X}_5) . \tag{290}$$

The following scatter diagrams are made:

1. $X_{1.345}$ is plotted on the vertical scale and X_2 on the horizontal scale and a line based on the following equation is drawn through the point $(\bar{X}_2, \bar{X}_1)$:

$$X_{1.345} = \bar{X}_1 + b_{12.345} (X_2 - \bar{X}_2). \tag{291}$$

2. $X_{1.245}$ is plotted on the vertical scale and X_3 on the horizontal scale and a line based on the following equation is drawn through the point $(\bar{X}_3, \bar{X}_1)$:

$$X_{1.245} = \bar{X}_1 + b_{13.245} (X_3 - \bar{X}_3). \tag{292}$$

3. $X_{1.235}$ is plotted on the vertical scale and X_4 on the horizontal scale and a line based on the following equation is drawn through the point $(\bar{X}_4, \bar{X}_1)$:

$$X_{1.235} = \bar{X}_1 + b_{14.235} (X_4 - \bar{X}_4). \tag{293}$$

4. $X_{1.234}$ is plotted on the vertical scale and X_5 on the horizontal scale and a line based on the following equation is drawn through the point $(\bar{X}_5, \bar{X}_1)$:

$$X_{1.234} = \bar{X}_1 + b_{15.234} (X_5 - \bar{X}_5). \tag{294}$$

The charts obtained in this manner are the required ones, the first indicating approximately $r_{12.345}$, the second indicating $r_{13.245}$, the third indicating $r_{14.235}$ and the fourth indicating $r_{15.234}$.

When working with analyses based on logarithms or first differences of logarithms, if the variables in the charts are retained in their logarithmic form no modifications are required. Frequently, it is preferred to show the variables in the charts in their original form. In such cases, all computations up to the point of plotting should be made with the variables given as logarithms. The anti-logarithms of the values to be plotted are then obtained. In working with analyses based on first differences of logarithms, the number 2 should be added to the calculated values before finding the anti-logarithms. This eliminates any negative numbers. The anti-logarithms are then given in terms of a percentage of the preceding year. The regression lines in the charts become curves when plotted in terms of the original values.

In connection with such charts, the final residuals frequently are plotted against time in a separate chart. When working with logarithmic analyses the d_i are computed with the variables given as logarithms. The number 2 is added to the calculated value and the anti-logarithm obtained. This then shows the percentage which the actual value is of the calculated values. The same procedure is used for analyses based on first differences of logarithms.

GLOSSARY 108/

Aggregates.--Totals or index numbers of microvariables.

Aggregative variables.--See aggregates.

Analogy approach.--A method of studying relations between microtheory, aggregates, and macrotheory by use of analogies. (See page 86.)

Argument of a function.--One of the independent variables upon whose value that of a function depends. For example, in the function $y = f(x_1, x_2, \ldots x_n)$ the arguments are $x_1, x_2, \ldots x_n$.

Autocorrelation.--See serial correlation.

Autoregressive process.--A sort of inverse of a process of moving summation. (See page 153.)

Best estimates.--See contrasting definition given under efficient estimates.

Chi-square.--A coefficient that can be used to determine whether a series of frequencies differ significantly from a theoretical or expected set of frequencies. (See page 178.)

Complete model.--A model that contains one equation for each endogenous variable. In general, complete models are required if we wish to derive from them equations to be used for analytical purposes or prediction.

Consistency approach.--A method of studying relations between microtheory, aggregates, and macrotheory. Given any two of the foregoing, the third is determined in such a way that it is consistent with the other two. (See page 85.)

Consistent estimates.--Estimates of statistical coefficients obtained in such a way that the average value for many large samples equals the value that would be obtained from a similar calculation based on the combined evidence of all possible samples. For unbiased estimates, the same property holds when estimates are made from samples of any size.

Correlogram.--The set of autocorrelation coefficients generated by letting the time lag range over the integers from $-\infty$ to ∞ .

Degree of identification.--See identification.

108/ Definitions given here refer only to terms included in this handbook. In cases where an exact definition would require a large amount of space, a condensed explanation is given instead, possibly with a cross-reference to the text. The reader is presumed to be acquainted with terms covered in a first course in statistics that includes multiple and partial correlation and regression and with elementary terms that relate to matrices.

Derived demand equation.--A demand equation that applies at a different level within the market from that which is believed to hold in the structural model. Derived demand equations normally are estimated statistically by making use of partially-reduced form equations. (See page 23.)

Derived elasticity coefficient.--A coefficient that relates to a derived demand or supply equation.

Distributed lag.--If an economic cause produces its effect over a number of time periods, we say that the effect occurs with a distributed lag. If the entire effect takes place within a single time period, this can be considered a limiting case.

Economic model.--A set of structures or equations consistent with the assumptions developed by a research analyst from economic theory and existing factors that relate to a particular commodity area.

Efficient estimates.--Estimates of statistical coefficients obtained in such a way that their average standard error for many large samples is as small as possible. For "best" estimates, as defined in this handbook, the same property holds when estimates are made from samples of any size.

Elasticity of expectations.--The proportional change in expected future prices of a commodity in relation to the proportional change in its current price. (See page 112.)

Elasticity of price transmission.--The elasticity of the retail price with respect to the price paid producers.

Elasticity of substitution.--A combination of direct and cross price elasticities and of income elasticities which appears to be of dubious value as a measure of the ease of substitution between two commodities. (See page 89.)

Endogenous variables.--Variables whose values are assumed to be correlated with the unexplained residuals in the structural equation in which they occur. Endogenous variables frequently are referred to as those that are simultaneously determined by the system of equations.

Error models.--Statistical models in which the variables are assumed to be subject to error but in which the true variables are assumed to be related in a functional way.

Evolutionary process.--A stochastic process for which the sets of distribution functions depend on the particular point from which we measure time. All stochastic processes which are not stationary are evolutionary.

Exogenous variables.--Variables in a particular economic structure other than those that are assumed to be endogenous.

Exponential function.--A mathematical function or its statistical equivalent in which at least one variable appears as an exponent.

Family of curves.--A group of curves of a specified type for which the value of a specified constant differs from curve to curve.

First differences.--Transformation of variables to express them as a change from the preceding time period. The section beginning on page 29 discusses circumstances under which they should be used.

Fisher's z transformation.--See z-transformation.

Formal approach.--A method of studying relations between microtheory, aggregates, and macrotheory. Given a microtheory and some distributional assumptions concerning some or all of the relevant variables which serve to differentiate the individual units, both the appropriate aggregates and the macrotheory are derived. (See page 86.)

F-test.--A test used to determine whether the assembly of data in various classes defined by certain restrictions exhibit on the average greater differences between members of different classes than between members of the same class. (See pages 178-183.)

Full information approach.--A maximum likelihood method for obtaining estimates of the structural coefficients for each equation in a system of equations. Estimates of all coefficients in all equations in the system are obtained simultaneously. In general, the computations are formidable, so the method is seldom used.

Identification.--A mathematical property of an equation that indicates whether the structural coefficients can be estimated by statistical means. Degree of identification refers to whether the equation is underidentified, just identified, or overidentified. (See page 61.)

Identities.--Structural equations which are presumed to hold exactly.

Instrumental variables.--Predetermined variables from an equation system that are chosen for use as multipliers when fitting equations by the instrumental variable approach. (See page 65.)

Just identified equation.--An equation that contains just enough variables in relation to all of the variables in an equation system so that the structural coefficients can be determined uniquely from the regression coefficients in the reduced form equations. (See page 62.)

Laurent matrix.--A matrix which not only is symmetric but for which the elements are equal along any northwest-southeast diagonal.

Limited information approach.--A maximum likelihood method for obtaining estimates of the structural coefficients for equations that are overidentified.

The coefficients usually are estimated for one equation at a time, with the simultaneity implied by the system taken into account in the computations, but information on the particular variables that appear in each of the other equations in the system is ignored. The estimates are statistically consistent and as efficient as any other based on the same amount of information. (See page 63.)

Linear homogeneous function.--When we consider linear relationships between two variables, this is a mathematical function for which one variable is a constant proportion of the second variable. The term also is applied to statistical relationships that involve random errors provided the regression line passes through the zero intercept.

Linearized form.--As used in this handbook, this term refers to nonlinear combinations of endogenous variables that have been transformed into linear approximations. (See page 71.)

Macrotheory.--That body of economic theory that deals with relationships that should hold between aggregative variables.

Markov process.--A commonly discussed stochastic process. (See page 154.)

Maximum likelihood.--A commonly used mathematical procedure for obtaining formulas to estimate statistical coefficients. Coefficients are derived in such a way as to maximize a likelihood function. The results are known to be statistically consistent and efficient. Statistical coefficients obtained from such formulas are called maximum likelihood estimates.

Microtheory.--That body of economic theory that is concerned with the behavior of individuals, either consumers or production or marketing firms.

Microvariables.--Variables that relate to individuals, either consumers or production or marketing firms.

Model.--A set of structures or equations that are compatible with the investigator's advance assumptions about the statistical universe from which a set of economic data are drawn. (See "economic" and "statistical" models.)

Monotonic function.--A mathematical function of the relation of one variable to another such that the values of the first variable increases or decreases steadily as the second variable increases or decreases.

Moving average random process.--A moving summation process for which the summation is taken over a finite number of terms. The name is unfortunate, as this is not an average in the usual sense, but it has become established.

Moving summation random process.--A process obtained by adding together the elements of a purely random process. (See page 153.)

Nonparametric tests.--Tests of significance that are independent of the nature of the distribution from which the sample was drawn.

Null hypothesis.--A hypothesis, commonly used in connection with tests of significance, that a specified coefficient or value equals zero.

One-tailed test.--A test of a hypothesis for which the region of rejection is wholly located at one end of the distribution of the test statistic as, for example, when we test only for positive serial correlation.

Overidentified equation.--An equation having mathematical properties such that alternative estimates of its structural coefficients can be obtained from the regression coefficients in the reduced form equations. (See page 63.)

Partial indifference curve or surface.--An indifference curve that represents substitution possibilities of two commodities for which consumers are indifferent after adjusting for the effects of changes in income or other variables that might affect the relationship.

Partially-reduced form equation.--An equation obtained by the algebraic substitution of an equation for a variable in a second equation. Equations of this sort normally are used only when data are lacking on the variable for which the substitution is made.

Polynomial.--See power function.

Power function.--A mathematical function or its statistical equivalent in which at least one variable is raised to a power greater than 1.

Predetermined variables.--Variables that are believed to be uncorrelated with the unexplained residuals in the structural equations in which they appear. They include exogenous variables and lagged values of endogenous variables.

Purely random stochastic process.--A stochastic process for which the error terms are assumed to be completely independent of one another. (See page 152.)

Random stochastic process.--See purely random stochastic process.

Realization of a process.--Any particular sequence of a process. For example, the error terms of an economic relationship are a realization of a stochastic process.

Recursive approach.--A method of fitting recursive systems such that consistent estimates of the coefficients are obtained. (See page 64.)

Recursive system.--A system of equations of a form such that by successive steps each of the equations can be transformed into one that contains only a single endogenous variable other than those which have been treated as dependent in prior analyses. (See page 64.)

Reduced equations.--Equations that do not contain distributed lags that have been obtained by algebraic manipulation from equations that do contain such lags.

Reduced form equations.--Equations that result when each endogenous variable in a system of equations is written as a linear function of all of the predetermined variables in the system. Depending on the circumstances, they may be (1) algebraically derived from the structural coefficients or (2) fitted by least squares.

Reduced form method.--A method that yields estimates of structural coefficients that are statistically consistent and efficient for equations that are just identified. Reduced form equations are fitted by least squares and the structural coefficients obtained by an algebraic transformation. (See page 62.)

Sector model.--A model that relates to a particular sector of the economy. A sector analysis for hogs might be included in a model of the livestock economy and a sector analysis for all of agriculture might be included in a model of the general economy.

Serial correlation.--The correlation between a series of observations and the same series lagged by one or more units of time.

Shock models.--Statistical models in which the variables are assumed to be known without error but are related in a stochastic way.

Spatial equilibrium.--A technique for studying geographical price equilibriums and flows of commodities between regions or countries. (See page 169.)

Stationary process.--A stochastic process for which the probability laws depend on time in such a way that if we replace time as measured from a fixed point by a time variable measured from another time point, the probability laws remain the same. (See page 150.)

Stationary process to the second order.--A process for which the mean, variance, and correlogram are all independent of time. Stationarity implies stationarity to the second order, but a process stationary to the second order may be evolutionary. (See page 151.)

Statistical model.--A set of structures or equations consistent with both economic and statistical specifications of the research analyst. (See pages 7-8.)

Stochastic process.--A generalization of the notion of a random variable. It expresses the ordered development (usually although not necessarily through time) of a series of events subject to random influences. (See pages 149-154.)

Stochastic relations or equations.--Equations that include a set of unexplained residuals or error terms whose direction and magnitude are usually not known exactly for any particular set of calculations, but whose behavior on the average over repeated samples can be described or assumed.

Stochastic vector process.--A stochastic process for which each component of x is interpreted as a vector.

Structural change.--A change in the process by which a set of economic variables is believed to be generated.

Structural equations.--Individual equations which define the process by which a set of economic variables are believed to be generated.

Structure.--The process by which a set of economic variables is believed to be generated.

Student's t.--See t-test.

Substitution elasticity.--See elasticity of substitution.

Trace of a matrix.--The sum of its diagonal elements.

Transformed variables.--Variables that have been modified in some way. Common transformations include (1) deflation, (2) conversion to logarithms, and (3) conversion to first differences. A particular series may be transformed in more than one way.

t-test.--Student's t is defined as the difference between any particular value assumed to be the true value in the universe for the statistical measure examined and the value determined from the sample divided by the standard error of this coefficient. (See page 178.)

Two-rounds estimates.--A method of obtaining estimates of the coefficients in equations that contain two or more endogenous variables that represents a modification of the instrumental variables approach so that all the predetermined variables in the system are used in the instrumental set. (See page 65.)

Two-tailed test.--A test for which the region of rejection comprises areas at both extremes of the sampling distribution of the test function as, for example, when we test for both positive and negative serial correlation.

Unbiased estimates.--See contrasting definition given under consistent estimates.

Underidentified equation.--An equation having mathematical properties such that the structural coefficients cannot be estimated by statistical means.

Variance.--The square of a standard error.

Zero-one variables.--Variables that have a value of 0 in one period and a value of 1 in another period. For a description of cases in which they are used, see page 22.

z-transformation.--A transformation developed by Fisher for use in testing hypotheses that relate to correlation coefficients.

INDEX

INDEX--Continued

Page

INDEX--Continued

☆ U. S. GOVERNMENT PRINTING OFFICE : 1958 O—470624