ANALYTICAL TOOLS for Measuring Demand

by Richard J. Foote and Karl A. Fox

Agriculture Handbook No. 64

UNITED STATES DEPARTMENT OF AGRICULTURE
AGRICULTURAL MARKETING SERVICE
Washington, D. C. January 1954

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. 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 was 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 and some of the notes in the Appendix, of 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.

Most of this material was presented at a series of seminars held during 1951-53 for staff members of the former Bureau of Agricultural Economics working in this field. Suggestions made at the seminars were incorporated and the material was brought up to date to include later developments. Certain sections were prepared by other staff members; these are indicated by footnotes. Helpful suggestions were received from various members of the staff.

For sale by the Superintendent of Documents, U. S. Government Printing Office
Washington 25, D. C. - Price 50 cents

CONTENTS

Page

ANALYTICAL TOOLS FOR MEASURING DEMAND

by

Richard J. Foote, Agricultural Economic Statistician, and Karl A. Fox, Chief,
Statistical and Historical Research Branch
Agricultural Marketing Service

INTRODUCTION

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 section discusses a way to make a preliminary survey of the cause-and-effect relationships to be expected. Here is where the researcher brings to bear his knowledge of the commodity and his understanding of theory.

The report then discusses some of the techniques that can be used to test the assumed model and to give quantitative estimates, or forecasts. Specific examples are cited when needed to indicate how a particular technique should be used.

DIAGRAMMATIC MODELS OF THE SUPPLY-DEMAND STRUCTURE

Diagrams that show the flow of commodities from producer to consumer, in terms either of physical products or of marketing channels, or both, have been in use for many years. Similar diagrams that show the economic forces or relationships which affect a given commodity or group of commodities have been developed in the last several years by staff members of the former Bureau of Agricultural Economics. Four such charts are shown in figures 1 to 4. These were taken from Fox (20) 1/, Hermie (31), and Armore (2). Similar charts have been prepared for a number of other commodities. Charts of this kind are closely related to the "path coefficient" diagrams which were used by Sewall Wright in the early twenties. When such diagrams are formalized into a set of equations which indicate the variables involved in each of several supply and demand relationships, they become the "models" discussed in the monographs of the Cowles Commission for Research in Economics.

Such diagrams are useful in several ways: (1) They help the analyst in thinking through the basic factors and relationships involved, (2) they aid in preparing a logical writeup of the economic structure of the industry, and (3) they help the reader to follow fairly complex relationships and

---

1/ Numbers in parentheses refer to Literature Cited, p. 72.

discussions. Any statistical analyses that are run should be based directly on the relationships indicated in the diagrams except for such modification as may be indicated in the diagrams after the research work is under way.

Figures 1 and 2 indicate the kind of statistical questions that can be discussed in terms of such diagrams. Each shows the demand-supply structure for a certain type of perishable crop. It is assumed that each crop marketed under the plan diagramed in figure 1 will be sold in a single outlet - watermelons make a good example - and that each crop marketed under the plan diagramed in figure 2 will be sold in part in the fresh market and in part in processed form, that the farm or local market price is identical in the two outlets, and that the retail price in either form is not significantly affected by the retail price or consumption of the other form. This situation may apply approximately to consumption of table grapes, which are used in fresh form and also for making wine and other alcoholic beverages. Each diagram is based on the assumption that the total supply will not be affected by the price during the harvesting and marketing seasons.

If the purpose of an analysis is to estimate the expected price associated with given values for such variables as size of crop and consumer income, a satisfactory answer in the case illustrated by either diagram may be obtained by a least-squares regression analysis with price dependent and other variables independent. Such an equation will give unbiased estimates of the elasticity of demand and other structural coefficients if, and only if, the supply and consumption for a given period are not affected significantly by the price during the corresponding marketing season. If figure 1 and the assumptions noted above apply, these conditions will be met approximately. In this instance, retail price ordinarily would be considered to be determined by consumption (or production) and consumer income. In the case illustrated in figure 2, single equations could be used to measure the interrelationships among consumption, price, and income in either the fresh or processed market, given the amount that moved through each of these outlets. However, a simultaneous system of equations would be needed to estimate the relative proportion of the crop that could be expected to move through each outlet in any given year. Such a system would at the same time yield a measure of the different price and income elasticities of demand prevailing in each outlet.

Conditions under which single equations or simultaneous equations should be used are discussed in greater detail in a later section, and this subject is discussed fully in Fox (20).

STATISTICAL CONSIDERATIONS IN SETTING UP THE ANALYSIS

A number of statistical decisions must be made before an analysis dealing with the factors that affect price or consumption of a given commodity can be run. In this field, many decisions must be based mainly on judgment. Alternative methods exist, but in many cases methodological experts have not agreed as to the best procedure to be followed. But in certain cases

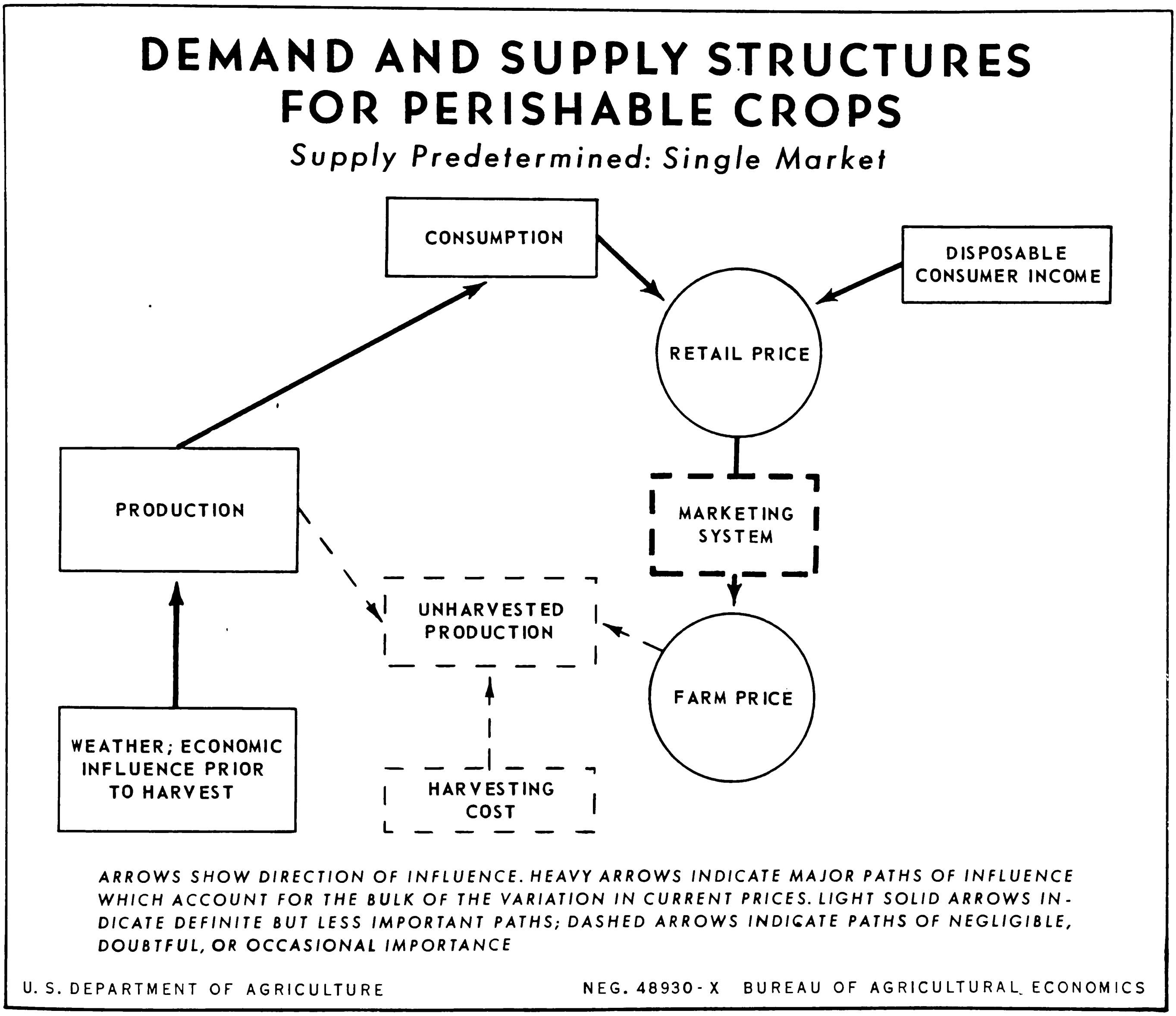

Figure 1.- If production is not affected by price during the marketing season and if all of a crop is sold in a single outlet, in general a single-equation analysis will give an unbiased estimate of the elasticity of demand.

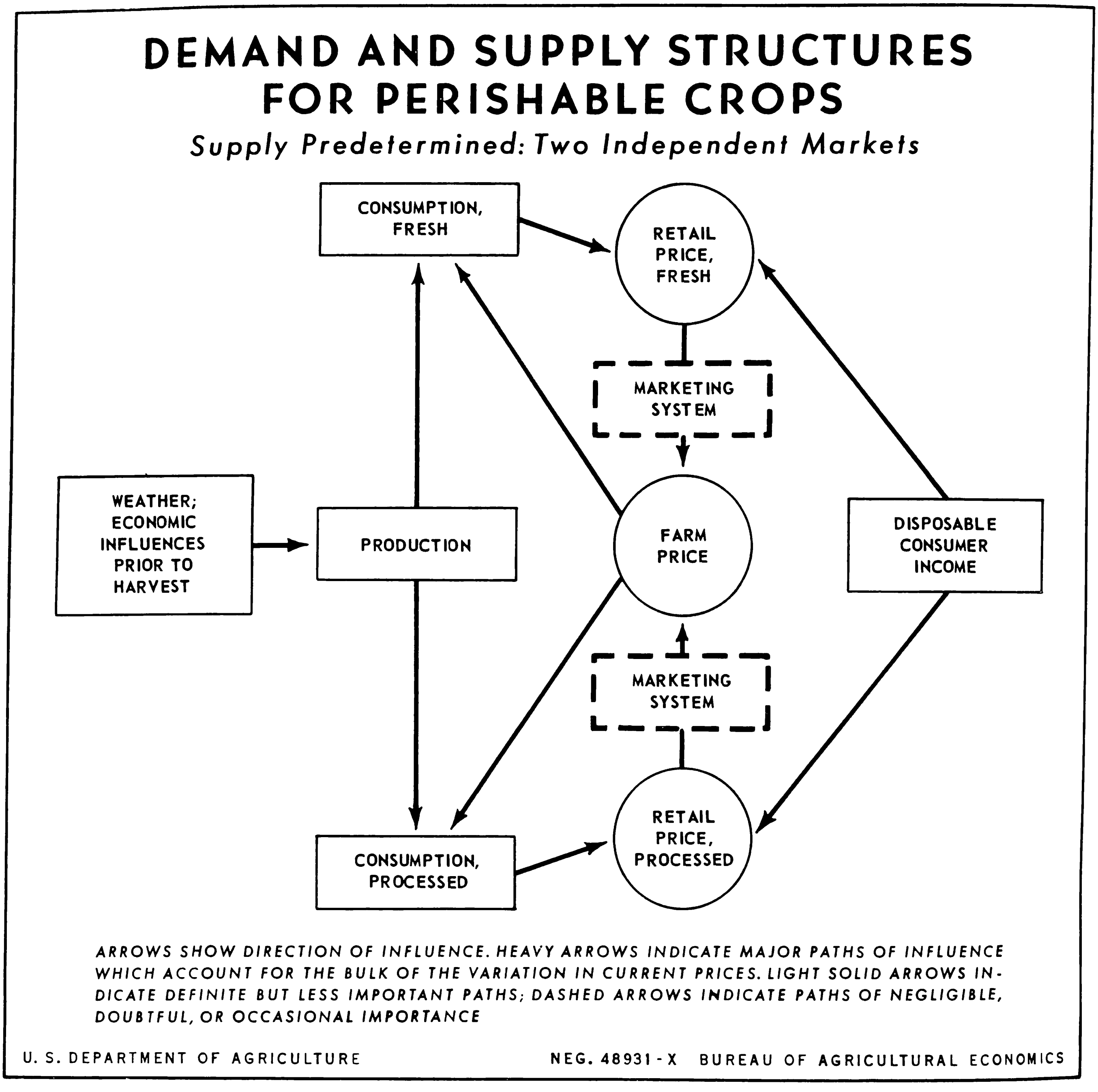

Figure 2.- The expected farm price associated with given values of size of crop and consumer income frequently can be estimated by a single-equation analysis even for crops used in multiple outlets. But a system of simultaneous equations would be needed to estimate the relative proportion of the crop that will move through each outlet and to yield unbiased estimates of the different price and income elasticities of demand prevailing in each.

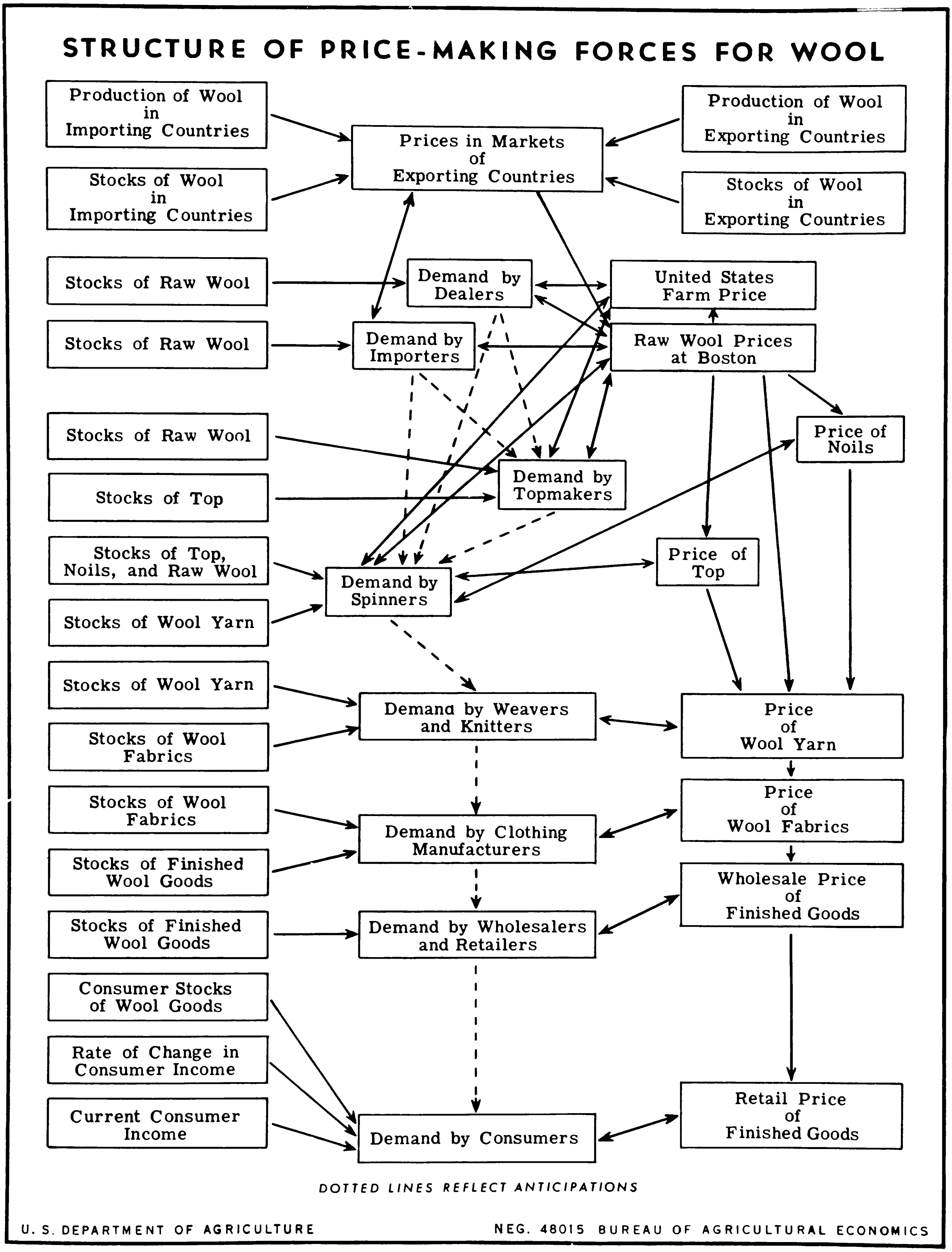

Figure 3.- The demand for wool at the farm level is derived from the combined demands of the many processors and users of wool. Domestic prices tend to exceed world prices by approximately the amount of our tariff, but prices rise and fall with changes in world supply and demand.

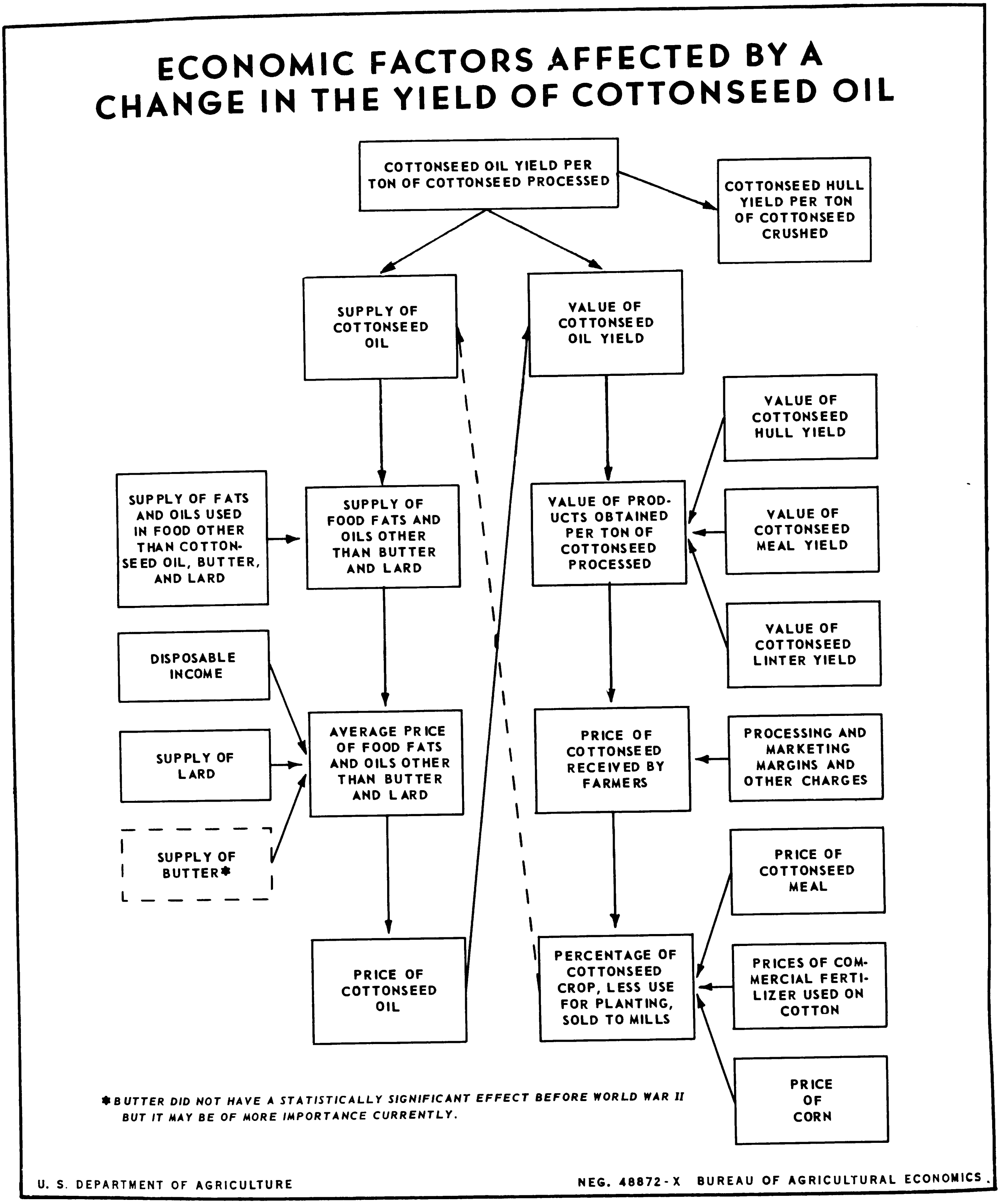

Figure 4.- The effects of an increase in the yield of cottonseed oil per ton crushed on total returns received by farmers takes place through two channels. One affects the quantity of oil obtained per ton of seed crushed and therefore directly affects its value; the other affects the price of the oil by increasing its total supply. The net effect on the price received by farmers for cottonseed must allow for both aspects.

particular conditions that are known to prevail with respect to the commodity studied will 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 Period of Time

Prior to about 1945, most analyses of the factors affecting the price or consumption of a given commodity were based on annual data for either calendar or crop years. In certain instances some other time period may be preferable. For example, analyses of the factors that affect prices of corn were run separately for the periods November-May and June-September, with allowance in the latter period for the effects of new-crop supplies of oats and barley and for the expected size of the new corn crop. (See Foote (15).) The period chosen should be long enough to average out the effects of irregular or nonmeasurable factors and short enough to insure that a homogeneous set of factors will be operating. Pearson and Vial (39) give monthly analyses for many livestock products. They show that the effects of production and stocks on price, for example, differ greatly from month to month and give reasons for expecting this to be so.

Years to be Included

In general, as many years as possible should be included in an analysis unless important changes in institutional or nonmeasurable factors are believed to have affected the relationships. "Abnormal" years, such as those in which price ceilings or rationing were in effect, should be omitted from the analysis but the calculated values should be checked to see that they are in line with expectations. If a sufficiently long series is available, the postwar years could be omitted from an analysis to see whether the prewar relationships hold in the postwar period. If they do, the analysis can be used with greater confidence. If this is done, the calculated values for the postwar years should be obtained in all cases. Except in extreme cases, decisions regarding the years to be included should be made before running the analysis.

Use of Prices at the Local Market, Wholesale or Retail Level

Retail prices should be used in an analysis designed to measure consumer demand at the retail level. In measuring domestic demand for items used to a large extent by industry, such as fats and oils or cotton, a wholesale price might be preferred. Farm or local market prices then can be determined by a simple correlation between farm and retail or wholesale prices. Such a procedure can be expected to give improved results whenever marketing margins change abnormally. If this method is used, the error of estimate for farm prices will be a function of the error of estimate in the equation that explains retail or wholesale price and the error in the equation that connects farm price with the retail or wholesale price.

Choice of the Quantity Variable

The type of quantity variable chosen will depend mainly on the purpose of the analysis. If the analysis is designed to measure factors that affect domestic demand, domestic consumption should be used. If price is used as the dependent variable, total supply of the commodity plus supplies of closely competing commodities (either separately or combined) probably should be used as the supply variable, except when consumption is highly correlated with production. If exports are subsidized, supplies available for domestic consumption and carryover may be the applicable supply variable.

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. 2/ 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 changes in exports and stocks are negligible, the problem of isolating the domestic demand curve is simpler than when major changes occur in these outlets. For most commodities, domestic demand is less elastic than total demand, the difference depending to a considerable extent on the average proportion of changes in production that is represented by changes in domestic consumption.

Choice of the Dependent Variable

If an analysis is designed primarily for use in forecasting and it is assumed that no change in structure has taken place over the period included in the analysis or for which the forecast is made, a single equation, using as the dependent variable the item for which a forecast is desired, will be appropriate. Marschak (35, pp. 1-8) discusses a simple example which illustrates the effect of changes in structure on the type of statistical analyses needed under specified conditions. If the analysis is designed mainly to estimate the elasticity of demand or other structural coefficients and a single-equation approach is applicable (see p. 39), the independent variables should be those that are "predetermined," as by weather, by economic forces in an earlier period, or by broad economic forces, such as those determining total consumer income.

2/ A proof of this is given in Appendix note 1.

Annual production, or production plus carry-in, of many crop and livestock products could be used as an independent variable, according to this criterion. If consumption and production are almost identical, as with some perishable commodities, or are very highly correlated, consumption could be used as an independent variable. If prices are set by Government action, as by a support program, price would be used as the independent variable and consumption as the dependent variable in an analysis designed to measure the elasticity of demand in the domestic market.

Choice of the Demand Shifter

Disposable income is the commonly accepted demand shifter for most consumer goods. The Federal Reserve Board index of industrial production is used for items such as certain inedible fats and oils which are to a considerable extent industrial raw materials. Particularly for durable or semi-durable goods, some analyses indicate that the change in income from the preceding year as well as the current level of income is important. (See Atkinson (4).) Changes in distribution of income also may be important for certain items. Different kinds of demand shifters are needed for producers' goods. For example, in an analysis of the factors that affect prices of corn, prices of livestock, and the number of animal units fed were used as separate demand shifters. (See Foote (15).) Animal units are significant partly because of the importance of diminishing returns in livestock feeding. In general, the use of a retail or wholesale price series as a demand shifter is frowned upon, but this use appeared to be justified on theoretical grounds in analyzing the price of corn. Attempts to develop satisfactory shifters for storage and export demand have been largely unsuccessful. The 4-curve equilibrium approach can be used to indicate why a given volume of exports or imports will occur. However, this implies a system of simultaneous equations. 3/

Use of Arithmetic or Logarithmic Variables

Linear (arithmetic) and logarithmic equations are the principal functional forms used in commodity analysis. For many commodities, linear equations yield elasticities which, when translated into total value-supply curves, are more consistent with theory at the extremes than those given by logarithmic equations. 4/ Logarithmic equations have the mechanical advantage of yielding constant elasticity curves and in many analyses they appear to fit the data better than arithmetic relationships. In general, logarithmic equations should be used when the relationships between the variables are believed to be multiplicative or the relations are believed to be more stable in percentage than in absolute terms or both. If the relationship

---

3/ This approach is discussed in Thomsen and Foote (45, pp. 232-240), based largely on material developed by Wells (51).

4/ In most cases no data are available for these extreme values and there is little interest in them.

between the variables is believed to be additive, arithmetic variables should be used. Differences between calculated values from arithmetic and logarithmic equations may be fairly small in the interwar period, but an attempt to project the analyses to high postwar price and income levels may reveal strong reasons for selecting the logarithmic form. The choice between arithmetic or logarithmic variables has nothing to do with the decision to run the analysis in terms of actual data or first differences. If logarithmic variables appear logical in a first-difference analysis, they would be equally logical in an analysis based on actual data (deviations from the mean). Use of certain other types of curves, some of which are hard to describe mathematically, has been proposed, particularly in the analysis of supply response. In general, the type of functional form to be used should be decided before the analysis is run. Except in a few special cases, the form of the equation cannot be determined from the data.

Use of Actual Data or of First Differences

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 prewar average to the postwar period.

First differences also are used in: (1) Cases where strong trend factors tend to overshadow the effects of economic variables (as in the study of factors that affect production of milk); (2) cases where the failure to allow for common trends would imply a relationship that does not actually exist (for example, the upward trend in consumption of evaporated milk and the downward trend in the price ratio of evaporated milk to fluid milk yields a pseudo "elasticity of substitution" of 1.76); (3) cases in which two variables are more closely correlated with an unrealized third factor than with each other (for example, the coefficient of determination between prices of tankage and of corn is 0.75 when based on actual data for 1926-42 and 1947-49, and 0.02 when based on first differences for the same years and for oats and corn, for which a high degree of correlation would be expected, the coefficient of determination based on actual data is 0.96 and on first differences 0.87); and (4) for certain technical reasons such as the reduction of correlation between the independent variables and of the serial correlation in residuals.

First differences may be hard to use in connection with certain long-range forecasts and in some types of analyses of relative benefits to be obtained from specified Government programs when no level for a preceding year can be assumed. The chief criterion for choosing between first differences and actual values (other than those given previously) for forecasting purposes is the standard error of the forecast from each equation form for the relevant values of the independent variables. If the residuals from an "actual value" equation are serially correlated to the extent of +0.5 or greater, the standard error of estimate will generally be smaller for the first-difference equation. The degree of intercorrelation between the independent variables also affects the relative forecasting value of first-difference and actual value equations.

Use of Total or of Per Capita Data

To avoid confusing the time trend for population with that due to other effects, per capita data probably should be used whenever they are applicable. For certain items, such as the consumption of cigarettes, care must be used in deciding what population groups to include in the population series.

Use of Deflated Data

At least two problems are involved in the use of deflated data: (1) This approach assumes a one-to-one relationship between the original series and the deflator, whereas certain series are known to swing more widely during a business cycle than other series; and (2) deflation of a value index by a price index with fixed weights cannot yield a quantity index expressed in "constant" dollars of the base period.

But Prest (41, pp. 33-38) and Stone (44, pp. 287-313) both used deflated income series in their demand analyses, which were based on careful theoretical formulations, although Stone concluded that it does not follow from a statistical point of view that the deflated series will lead to more reliable results. He says: "It is to be expected that both will lead to more or less the same, but not identical, results." Prest concurs in this view. In any analysis, it is important that the variables included be consistent. For example, if industrial production is used as a demand shifter in an analysis of consumption, prices should be deflated; if an income series is used as a demand shifter, both income and price could be deflated or both could be used in a nondeflated form.

Alternative Methods That Allow for Changes Over Time

Tinbergen (46, pp. 1-29, 193-209) indicates that time concepts enter into an analysis in four ways: (1) Certain variables may affect the dependent variable with a certain time lag; (2) if the effect is greater or less as the length of time varies, separate measurements might be made of the long- and short-time effects; (3) the relative importance of the long- and short-time effects may be affected by the "horizon," that is, the period over which producers, dealers, and consumers look ahead, which, for example, will affect the willingness of dealers and farmers to store a large crop; and (4) time may enter into an analysis as a measure of sources of continuous systematic variation that were not introduced explicitly into the equation. Time lags are handled in the equation by including lagged values of the variable; for example, using prices in each of 3 preceding years in a supply-response analysis.

The importance of the systematic variation may be measured in one of two ways: (1) Time may be included as a separate variable in the analysis and, if its partial regression coefficient fails to differ significantly from zero, it may be omitted; (2) time may be omitted from the initial analysis, in which case the final residuals should be checked to see whether they exhibit a trend over time. If they do, another variable which explains them can be added or, if such data are not available, time may be used as a catch-all variable. As with deflation, it is hard to indicate definitely which method is to be preferred.

If first differences are used, the constant value in the equation represents the linear effects of time as, if different from zero, it implies that there would be some change in the dependent variable from the preceding year even if there were no changes in the independent variables. The statistical significance of the constant value can be tested by use of the standard error of the function at the point at which each of the independent variables is equal to zero. Formulas for this coefficient for analyses based on two or three variables are given by Foote (14, pp. 4, 6). This test may be biased, as residuals from first-difference analyses tend to be negatively correlated so that the formulas given by Foote tend to underestimate their variance.

MEASURING DEMAND FOR SUBSTITUTE (COMPETING) OR COMPLETING COMMODITIES

Schultz (42) proposed comparing the coefficients of variation for the price and supply ratios for two competing commodities to measure the degree of substitution. If the two items were nearly perfect substitutes, the price ratios would be nearly constant, whereas the supply ratios would be expected to vary considerably. If the two items were complementary, the price ratios would vary more than the supply ratios. Thus, the ratio of the coefficients of variation for the two ratios would vary between zero and one for substitute commodities and for complementary goods it would be greater than one.

Peters and Van Voorhis (40, pp. 78-79) raised an objection to this test as follows:

> "In spite of the fact that this measure has received considerable attention from statistical workers, the authors have doubts of its value. For the mean may be distorted by a padding of all the scores. Consider the series of scores: 0, 3, 8, 12, 15, 20, 25, 29; and the series 20, 23, 28, 32, 35, 40, 45, 49. The mean of the first array is 14 and that of the second is 34. The coefficient of variation of the first is 68 while that of the latter is only 28. Nevertheless the variabilities of the two distributions are precisely the same, the distortion in coefficients of variation being due solely to the padding of the scores in one of the arrays."

In view of these objections, the Schultz test is not recommended. The methods subsequently discussed appear to be better.

Consumer Goods

Suppose we are dealing with a consumer-demand equation of the following form:

$$q_1 = a_1 + b_{11}p_1 + b_{12}p_2 + c_1y, \tag{1}$$

where q_1 is the quantity bought of commodity 1; p_1 and p_2 are prices paid for commodities 1 and 2 respectively; and y is consumer income.

For practical purposes, the two commodities may be considered as (1) independent, (2) competing, or (3) completing, depending upon whether b_{12} is respectively (1) zero, (2) positive, or (3) negative. 5/ When two commodities "compete," as in case (2), and the price of commodity 2 rises, people will consume more of commodity 1. In equation (1), b_{11} is always negative and c_1 is almost always positive (except in the case of so-called "inferior goods").

Suppose that the consumer-demand equation for commodity 2 is

$$q_2 = a_2 + b_{21}p_1 + b_{22}p_2 + c_2y. \tag{2}$$

One would logically expect b_{21} to have the same sign as b_{12}, 6/ or to equal zero when b_{12} equaled zero. This means only that if commodity 2 competes with commodity 1, commodity 1 competes with commodity 2.

Equations (1) and (2) are written appropriately for an individual consumer, who adjusts his purchases (q_1 and q_2) to the market prices (p_1 and p_2) which he must take as given. But these equations generally are not appropriate for deriving statistical demand curves for the market as a whole.

For many agricultural commodities, the market supply in a given year is determined by weather and by the economic influences that are operating when decisions as to planting or breeding are made. When this is the case, supplies of the two commodities are among the "independent" variables in the

---

5/ A definition based on utility theory, called to the author's attention by Frederick V. Waugh, is given in Appendix note 2.
6/ According to the utility theory, b_{21} should equal b_{12}. See Appendix note 2.

market-demand functions, and current market prices adjust to the given supplies. The market-demand equation may be written as

$$P_1 = A_1 + B_{11}Q_1 + B_{12}Q_2 + C_1Y \tag{3}$$

$$P_2 = A_2 + B_{21}Q_1 + B_{22}Q_2 + C_2Y, \tag{4}$$

in which the letters have the same meanings as applied to market averages as did those in equations (1) and (2) as applied to individual consumers. The signs that indicate demand relationships in equation (3) are as follows:

1. Independent in demand: $B_{12} = 0$

2. Competing in demand: $B_{12} < 0$; (that is, an increased supply of commodity 2 depresses the price of commodity 1);

3. Completing in demand: $B_{12} > 0$; (that is, an increased supply of commodity 2 increases the price of commodity 1).

In general, B_{21} should have the same sign as B_{12} (see footnote 6), or equal zero when B_{12} equals zero.

If the variables Q_1 and Q_2 are strictly "predetermined" (that is, are not affected significantly by price during the marketing season), equations (3) and (4) are "structural equations" in the interpretation of the Cowles Commission; that is, they are true demand equations. Production or supply (production plus carry-in) will meet this condition in many cases. If consumption equals production, as with some highly perishable commodities, it is also a predetermined variable. If consumption does not equal production (because of small or moderate variations in stocks) but is highly correlated with production, its use as a predetermined or independent variable probably is still satisfactory for most practical purposes.

In some problems, we want to ascertain not only whether two commodities are competing, but whether they are perfect substitutes. In equation (3), this would imply that $B_{11} = B_{12}$. Kuznets and Klein (34) tested the hypothesis that domestic lemons (Q_1) and imported lemons (Q_2) were not perfect substitutes during a certain period. In their analysis the coefficients equivalent to B_{11} and B_{12} were found to be:

$$B_{11} = \underset{(.0026)}{-0.0196}, \text{ and } B_{12} = \underset{(.0049)}{-0.0172},$$

in which the figures in parentheses are standard errors of the regression coefficients. The difference between B_{11} and B_{12} is nonsignificant; that is, it could have occurred by chance even though the commodities were, in fact, perfect substitutes on a pound for pound basis. Therefore, the authors concluded that domestic and imported lemons could be combined into a single total supply variable.

It appears that the ratio $\underline{B_{12}}/\underline{B_{11}}$ would be a reasonable measure of the closeness of competition or substitution between two commodities, if a pound of one was equivalent to a pound of the other in a given end use. The ratio would range from 1 for perfect substitutes down to zero for commodities that were independent in demand. For example, an analysis expressing the price of beef ($\underline{P_1}$) as a function of the consumption of beef ($\underline{Q_1}$) and the consumption of pork ($\underline{Q_2}$) yielded the following coefficients:

$$B_{11} = -0.4334, \text{ and } B_{12} = -0.1929, \text{ or } \frac{B_{12}}{B_{11}} = 0.445.$$

The results indicate that pork competes with beef but that it is far from a perfect substitute for it.

Examples of demand equations for red meats, chicken and turkey, and western and eastern apples showing the effects of supplies of competing commodities upon the price of a given commodity may be found in appendix notes 3 and 4. An exploratory analysis of competition among different cuts and grades of beef is discussed in appendix note 5.

Some Statistical Problems

Problems in Interpreting Logarithmic Coefficients

If two commodities are about equal in terms of total consumption, their logarithmic regression coefficients should stand in about the same ratio as their arithmetic coefficients. Otherwise, to test the closeness of substitution by means of the ratio $\underline{B_{12}}/\underline{B_{11}}$, it is necessary to multiply the ratio of the logarithmic coefficients, $\frac{B'_{12}}{B'_{11}}$ by the ratio $\frac{Q_1}{Q_2}$. Based on the definition of elasticity of demand, it can be shown that

$$\frac{B'_{12}}{B'_{11}} \cdot \frac{Q_1}{Q_2} = \frac{B_{12}}{B_{11}} .$$

Thus, in equation (70) of appendix note 3, Q_1, which applies to lamb, is about 6 pounds and Q_2, which applies to other meats, is about

120 pounds. For this analysis,

$$\frac{B_{12}}{B_{11}} = \frac{-.65}{-.50} \cdot \frac{6}{120} = 0.065,$$

which suggests that lamb is highly differentiated from other meats in consumer demand.

Apparently Nonreversible or Asymmetrical Competition

In parallel analyses of retail prices of beef and pork, the following equations and standard errors of the regression coefficients were obtained:

$$P_b = \underset{(.11)}{-1.08}\, q_b \; \underset{(.07)}{-0.42}\, q_p^{7/} + \underset{(.06)}{0.88}y \qquad (5)$$

$$P_p = \underset{(.13)}{-0.02}\, q_b^{8/} \; \underset{(.08)}{-1.16}\, q_p + \underset{(.07)}{0.90}y. \qquad (6)$$

Equation (5) implies that the supply of pork has a highly significant competitive impact upon the price of beef. But equation (6) shows no significant impact of the supply of beef upon the price of pork!

Possible <u>partial</u> explanations are: (1) As the supply of beef varies less from year to year than the supply of pork, it is harder to establish significant regression coefficients upon q_b than upon q_p. (2) The true regression of P_p upon q_b <u>might</u> be as high as -0.30, and a value as far off as -0.02 could still occur by chance about 2 or 3 times in 100. (But a value of -0.02 would have been about as likely if the true regression had been +0.30 instead of -0.30.)

Both of these "possible explanations" are negative. Based on the statistical results alone, the most plausible conclusion would be that the competition between beef and pork is asymmetrical--a large supply of pork depresses the price of beef, but a large supply of beef affects the price of pork very little. Yet a research worker could hardly recommend that an administrator charged with supporting the price of hogs pay no attention to the possible effects of a record large run of cattle.

Equations (3) and (4) could be fitted simultaneously, imposing the condition that $B_{21} = B_{12}$ and thereby <u>forcing</u> a symmetrical result, consistent with utility theory. When this was done for the analyses of beef and pork, the following results were obtained:

7/ In equation (5), q_p includes veal and lamb as well as pork.

8/ In equation (6), q_b includes veal and lamb as well as beef.

$$P_b = -1.09\ q_b - 0.32\ q_p + 0.89\ y \tag{5.1}$$

$$P_p = -0.32\ q_b - 1.16\ q_p + 0.95\ y. \tag{6.1}$$

But this approach has its dangers and should not be attempted unless these equations also are fitted separately, for comparative purposes, without this special constraint.

Reconciliation of Demand Equations for Individual Competing Commodities with an Aggregative Demand Equation for the Group

Retail demand equations for each of two competing commodities are given by equations (3) and (4). Suppose an aggregative analysis in terms of an index number of the two prices and the sum of the two quantities also is computed. Here

$$P_t' = (w_1P_1 + w_2P_2)' = A_t + B_t\ (Q_1 + Q_2) + C_tY_1, \tag{7}$$

in which $\underline{w_1} + \underline{w_2} = 1$ are the weights used in computing the index numbers of prices. For given values of $\underline{Q_1}$, $\underline{Q_2}$, and $\underline{Y}$, the weighted sum of equations (3) and (4) ought to equal, at least approximately, equation (7):

$$w_1P_1' = w_1A_1 + w_1B_{11}Q_1 + w_1B_{12}Q_2 + w_1C_1Y \tag{3.1}$$

$$w_2P_2' = w_2A_2 + w_2B_{21}Q_1 + w_2B_{22}Q_2 + w_2C_2Y. \tag{4.1}$$

In particular, the weighted sum of the four $\underline{B}$'s should (approximately) equal $\underline{B_t}$:

$$B_t \cong (w_1B_{11} + w_2B_{22}) + (w_1B_{12} + w_2B_{21}). \tag{8}$$

The terms "$\underline{w_1B_{11}}$" and "$\underline{w_2B_{22}}$" represent the <u>direct</u> influences of each supply upon the price index via its <u>own</u> price. The terms "$\underline{w_1B_{12}}$" and "$\underline{w_2B_{21}}$" represent the <u>indirect</u> influences of each supply upon the price index via the price of the <u>competing</u> commodity. Unless the commodities are perfect substitutes, the sum of the direct influences should exceed the sum of the indirect influences.

Furthermore, the sum of the direct influences is the weighted average own-price flexibility of the two commodities. In general, this sum is smaller than $\underline{B_t}$, the price flexibility of the group. That is, on the average, the demands for individual members of a competing group of commodities are more elastic than the demand for the group as a whole.

Two applications of this approach were made, one to meat animals and the other to meats at retail. In the analysis based on prices received by farmers, $\underline{B_t} = -1.266$; sum of direct influences $= -1.089$; sum of indirect influences

= -0.160; total effect, supplies of four meats, = -1.249, which is almost identical with $\underline{B}_t$. In the analysis based on prices at retail, $\underline{B}_t$ = -1.50; sum of direct influences = -1.068; sum of indirect influences = -0.332; total effect, supplies of four meats, = -1.400. Results of the retail analysis in particular indicate that, on the average, demands for individual meats are significantly more elastic than the demand for all meat as an aggregate.

Some Observations on Completing Goods

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:

$$q_s = a_1 + b_1 p_s + c_1 y, \text{ and} \tag{9}$$

$$q_s = a_2 + b_2 p_s + c_2 y + d \sum_{i=1}^{m} w_i q_i \ , \tag{10}$$

in which the last term of equation (10) is an index of consumption of sugar-using foods weighted by the quantity of sugar ($\underline{w}_i$) customarily used with a unit of each food. Both price and income elasticities of demand for sugar (given a certain consumption level for sugar-using foods) would be expected to be very small so that, in general, $\underline{b}_2$ would be considerably smaller than $\underline{b}_1$ and $\underline{c}_2$ would be considerably smaller than $\underline{c}_1$.

Other examples of completing goods may be found in rayon and cotton <u>in the very short run</u> (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.

Measurement of Competition when Supplies are not Predetermined

If marketings of two commodities during a given season are influenced by their current prices, an "identification" problem of the type analyzed in Cowles Commission literature exists. In some cases, four equations would need to be fitted simultaneously--two demand curves and two supply curves. If the two commodities compete in demand, the two demand curves would be identical with equations (3) and (4). The two commodities might or might not compete for productive resources on the supply side.

In 1941, Wells (51) worked out an eight-equation model to measure the incidence of tariffs on two competing products. The eight equations included domestic and foreign demand and supply functions for each commodity. The coefficients of these equations were estimated on a judgment basis, and an equilibrium solution was found for all prices and quantities under specified initial conditions. A change in tariff rates on either commodity would move the entire system to a different equilibrium position, which could be derived readily from the model. The same model could be applied to changes in transportation rates within a given country, with two commodities produced and consumed in each of two regions.

Pure Theory of Related Demands

Schultz (42, pp. 569-654) discusses some theoretical criteria for related demands on the assumption that consumers are "rational." Statistical results seldom are as neatly symmetrical as is implied in these theories of rational behavior. Also, the numerical difference between the "Hotelling conditions" and the "Slutsky conditions," which receives much attention in Schultz's two chapters, is small for commodities that account for less than 10 percent or so of disposable income. (Pork and beef, two of the major foods, account for only 2 or 3 percent each of disposable income.) Slutsky's condition allows specifically for the fact that a drop in the price of a commodity increases the real income of its consumers. Hotelling's condition lumps together the direct effect of a change in price upon consumption (income remaining constant) and the indirect effect resulting from the change in real income which the change in price entails. (See appendix note 2.)

Producer Goods 9/

Certain theoretical concepts involved in the analysis of the relations between competing goods are the same whether the commodities are used by producers or by consumers. Other concepts differ somewhat. The following methods were developed chiefly to study competition among the minor feed grains--oats, barley, and sorghum grains--with each other and with corn. Similar methods could be applied with only minor modifications to certain problems that involve consumer goods.

The following methods were used:

1. The ratio between the price of the minor feed grain and the price of corn was expressed as a function of the supply ratio of the two feeds and certain related variables. If the two items were perfect substitutes, their price ratio would equal a constant and the regression coefficient on the

---

9/ Most of the material in this section is taken from Meinken (36).

supply ratio would equal zero. Thus, the size of the regression coefficient is a measure of the degree to which the minor feed grain cannot be substituted for corn. Certain limitations in the use of this approach to study the degree of substitutability are discussed in Morrissett (37).

2. The price of the minor feed grain was expressed as a function of the price of corn, the supply ratio, and the same related variables as used in method 1 above. Here corn is thought of as the dominant price-making factor in the feed-grains market. Coefficients for the other variables would be interpreted in the same way as in method 1. In general, the coefficient of multiple determination will be considerably higher for this method, but the implied standard error of estimate for the price of the minor feed grain should be identical for the two analyses. Neither method yields good forecasting formulas, as the price of corn must be known (or estimated) in each case. The methods are primarily of value in studying the degree of substitutability among related items and the factors that affect price ratios.

3. The price of the feed grain was expressed as a function of the individual supplies of each of the competing grains and certain related factors. These equations are similar to equations (3) and (4) on page 14, except that different "demand shifters" were used.

The third approach was complicated by the fact that the basic demand equation for feed grains is believed to be of a logarithmic type, whereas the effects of the separate supply factors is believed to be additive. An iterative method was used which permits the combination of these two types of relations. Supply factors considered for the summer period during which new-crop oats and barley are marketed in volume but before the new crop of corn is available were: New-crop supply of oats, new-crop supply of barley, July 1 stocks of corn, and the forecast of the new corn crop. As an initial step, these variables were added together on a tonnage basis and used as a composite supply variable in the logarithmic analysis, together with two demand factors --animal units fed and prices of livestock and livestock products. The price of the feed grain was then adjusted for the effect of the demand factors and a linear analysis was run, using this adjusted price against the separate components of supply. These components were then weighted by the respective regression coefficients obtained from this analysis to obtain a second approximation to the composite supply variable. This variable then was used, along with the factors measuring demand, in a second approximation for the logarithmic analysis. The process was continued for two or three iterations until the results became stable. This method is described in detail in Foote (15).

If two competing commodities were perfect substitutes, their prices would move perfectly together. Changes in factors of demand would result in identical price responses for each and a changed supply of one would affect the price of each by the same amount. In multiple regression analyses of the type described in method 3 above, with price as the dependent variable, the regression coefficients for the demand variables in both analyses would be identical and the regression coefficients (or weights) for the supply variables in each analysis also would be identical. If the price of one was expressed as a ratio to the price of the other, the ratio would always be 1, regardless of the ratio between the supplies of the two commodities. If the price ratio were related to the supply ratio in a linear regression analysis, the constant would be 1 and the regression coefficient for the supply ratio would be 0, thus giving rise to an infinitely large elasticity of substitution.

But suppose a given quantity of one resulted in a 10 percent larger output than the same quantity of the other in any given end use, regardless of the level of substitution. Suppose also that the demand functions had the following form:

$$P_1 = -b_{11}Q_1 - b_{12}Q_2 \tag{11}$$

$$P_2 = -b_{21}Q_1 - b_{22}Q_2. \tag{12}$$

Then $b_{11} > b_{12}$, $b_{12} = b_{21}$, $b_{21} > b_{22}$, $\frac{b_{11}}{b_{12}} = \frac{b_{21}}{b_{22}} = 1.1$, $b_{11}b_{22} = b_{12}^2$, and P_1 equals 110 percent of P_2, regardless of the values of Q_1 and Q_2.

This is clear from the example that follows. Suppose the following coefficients are substituted in equations (11) and (12):

$$P_1 = 1100 - 110Q_1 - 100Q_2 \tag{11.1}$$

$$P_2 = 1000 - 100Q_1 - 90.91Q_2. \tag{12.1}$$

Note that for these values, $b_{11}b_{22} = b_{12}^2$. If in these equations the given values of Q_1 and Q_2 are substituted, price ratios are obtained as follows:

Q_1	Q_2	$\frac{Q_1}{Q_2}$	P_1	P_2	$\frac{P_1}{P_2}$
1	1	1.0	890	809.09	1.10
1	2	.5	790	718.18	1.10.

Similar results are obtained for any value of Q_1 and Q_2.

If this price ratio were related to the supply ratio in a linear regression analysis, the constant would always be 1.1 and the regression coefficient again would be 0, thus giving rise to an infinitely large elasticity of substitution. This is the common interpretation of the theory of substitute goods.

But in certain cases the relative contribution of a substitute commodity to the total output depends on the relative size of its consumption or use. For example, Henry and Morrison (30, p. 494) say with regard to oats as a substitute for corn in feeding hogs: "Ground oats are worth about as much as corn per 100 pounds when forming a rather small part of the ration, but when fed in large amounts, they are worth much less than corn,..." and "Numerous experiments have shown that oats have the highest value for pigs when ground oats form not over one-fourth of the ration. When thus fed to replace part of the corn in 20 trials with pigs in dry lot, the addition of ground oats increased the rate of gain a trifle."

In such cases, the relationship between the price ratio $(\frac{P_1}{P_2})$ and the supply ratio $(\frac{Q_1}{Q_2})$ results in a regression coefficient for the supply ratio less than 0, and an elasticity of substitution less than infinity. In the notation used above, $\underline{b_{11}}$ still would be greater than $\underline{b_{12}}$, $\underline{b_{12}}$ still could equal $\underline{b_{21}}$, and $\underline{b_{11}b_{22}}$ must be greater than $\underline{b_{12}^2}$ to satisfy the requirement that the price ratio varies inversely with the supply ratio. Suppose that the following coefficients are substituted in equations (11) and (12):

$$P_1 = 1100 - 110Q_1 - 100Q_2 \tag{11.2}$$

$$P_2 = 1000 - 100Q_1 - 110Q_2 \ . \tag{12.2}$$

Note that $\underline{b_{11}b_{22}}$ now is greater than $\underline{b_{12}^2}$.
Substituting in the equations under these conditions, the following price ratios are obtained:

Q_1	Q_2	$\frac{Q_1}{Q_2}$	P_1	P_2	$\frac{P_1}{P_2}$
1	1	1.0	890	790	1.127
1	2	.5	790	680	1.162
1	3	.3	690	570	1.211 .

Under the circumstances, $\frac{P_1}{P_2}$ varies inversely with $\frac{Q_1}{Q_2}$. 10/

10/ Ezekiel (10, p. 179) found the ratio of the retail price of pork to the retail price of beef to be significantly correlated (inversely) with the ratio of the supply of pork to the supply of beef. Schultz (42, p. 584), using the same data, determined the following demand equations for beef and pork:

$$P_b = 77.4 - 13.3Q_b - 4.3Q_p + .49I$$

$$P_p = 68.8 - 5.4Q_b - 7.5Q_p + .48I,$$

These three possibilities are illustrated through the use of input-output relationships in figure 5. In this figure, the vertical and horizontal axes simultaneously measure input of feed and output of livestock product. For simplicity of presentation, livestock products are assumed to be measured in units such that 1 pound of corn yields 1 unit of product.

If the ration is composed of 100 pounds of corn, the output is 100 units. If the ration is composed of, say, 80 pounds of corn and 20 pounds of a perfect substitute, the output is still 100 units, as any combination of corn and a perfect substitute will yield 100 units of product, so long as the total input is 100. In this case, the output function is the straight line A illustrated, intersecting the Y and X axes respectively at 100. Under this condition, the demand functions for corn and the perfect substitute would be equivalent to those discussed in the first case referred to on page 20, and the elasticity of substitution would be infinite.

If corn were worth 110 percent of some other substitute, at all levels of relative supply, the output curve would intersect the Y axis at 100 and the X axis at 90.91, as shown by the straight line B. The demand functions for corn and the substitute would be similar to that illustrated by equations (11.1) and (12.1). Again the elasticity of substitution would be infinite.

If corn (or, conversely, the substitute) possessed varying values depending on the level of relative supply, the output function would be curvilinear as illustrated by curve C. The demand functions for corn and the substitute in question would take the form shown by equations (11.2) and (12.2), and the elasticity of substitution would be less than infinity. The rate that productivity (or utility) increases or decreases per unit of input for the substitute commodity from any given level of relative supply determines the elasticity of substitution at that level.

---

in which P_b and P_p are composite retail prices of beef and pork, respectively, in cents per pound, Q_b and Q_p are total consumption of federally inspected beef and pork, respectively, in billions of pounds, and I is an index of payrolls lagged by 3 months. If I is assumed to be constant, and given values of Q_b and Q_p are substituted in these equations, the following results are obtained:

Q_b	Q_p	$\frac{Q_b}{Q_p}$	P_b	P_p	$\frac{P_b}{P_p}$
1	1	1.0	59.8	55.9	1.07
1	2	.5	55.5	48.4	1.15
1	3	.3	51.2	40.9	1.25

It should be noted that in these equations $b_{11} > b_{12}$, $b_{12} \cong b_{21}$, and $b_{11}b_{22} > b_{12}^2$ in the notation used here.

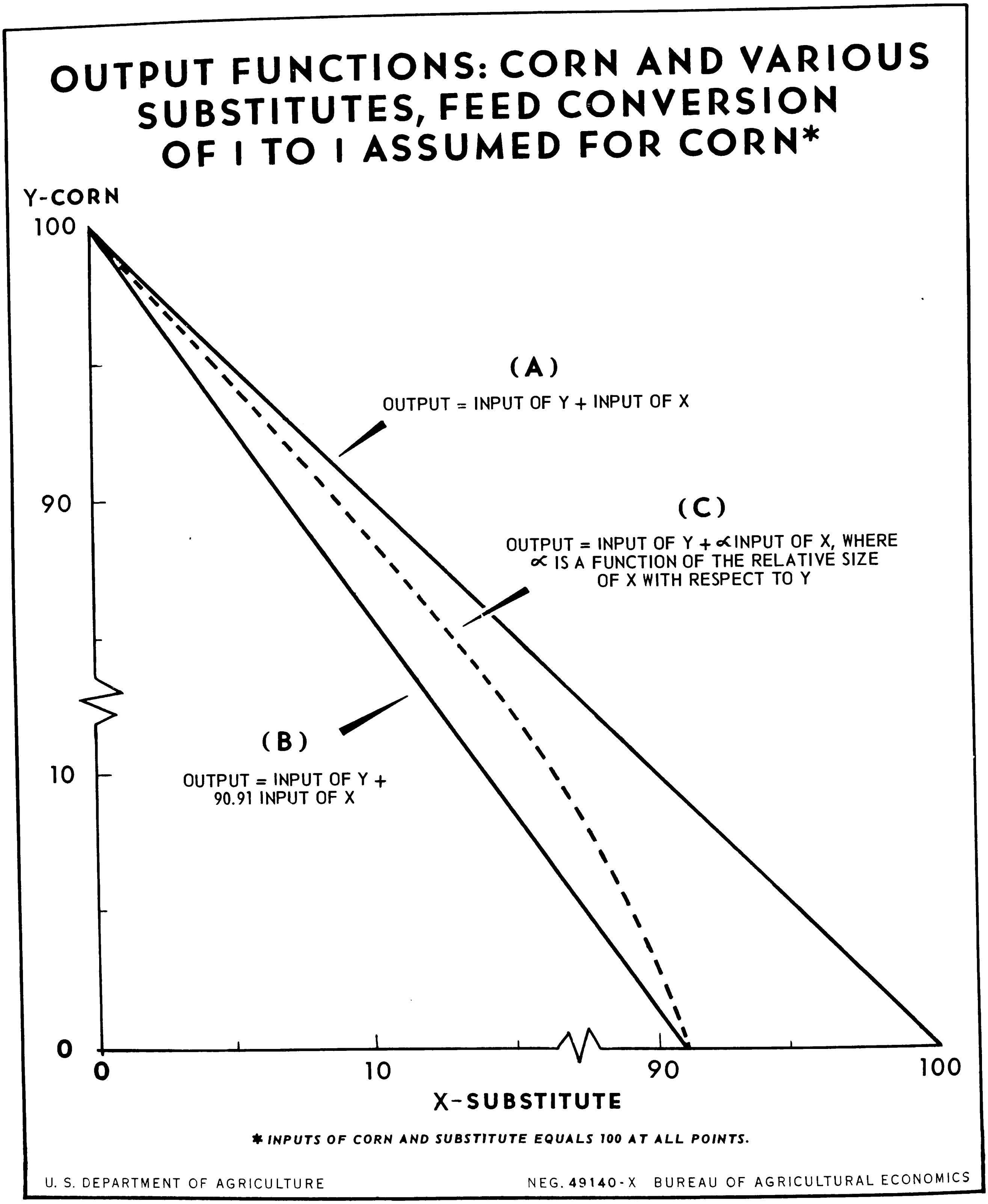

Figure 5.- Functions A and B illustrate perfect substitutes for corn, the feeding value of the substitute grain in function A being equal to 1 pound of corn at all levels of substitution and in function B to 0.909 pounds of corn at all levels of substitution. Function C shows the relation for a substitute grain for which feeding values vary relative to corn, depending on the level of substitution.

The statistical analyses run by Meinken (36) and Foote (15) indicate that the four feed grains compete with each other but that they are not perfect substitutes. This is consistent with available experimental evidence on feeding.

Knowledge of whether a substitute commodity has a constant or varying marginal rate of substitution is important from a practical viewpoint. Farmers and feed manufacturers alike are interested in obtaining least-cost rations while maintaining feeding efficiency. These usually can be approximated only if the manner in which one commodity substitutes for another is known. Prices of the substitute feeds also would be needed. Heady and associates (27, 28) have for some years been conducting feeding experiments designed to determine, among other things, substitution coefficients between various feedstuffs. These coefficients enable farmers and feed manufacturers to take advantage of changes in relative prices in determining least-cost rations.

STATISTICAL CONSIDERATIONS IN INTERPRETING THE COMPLETED ANALYSIS

Considering the Unexplained Residuals

In the 1920's and the early thirties, analysts expected to explain practically all of the variation in the dependent variable and frequently obtained multiple correlations of 0.99 or above. 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. The usual least-squares analysis assumes that the independent variables are known without error. In forecasting, better results are obtained by the least-squares method than by any other single-equation approach, regardless of the degree of error in the independent variables. But if interest is centered on measuring coefficients of elasticity and particularly if a "reversible" equation is desired, use of a "weighted" regression equation which allows for the relative errors in the several variables may be preferred. This problem is considered in more detail below.

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 find the relative importance of the nature of the chosen curves in causing residuals. Methods for preparing such charts are given in Foote (14, pp. 11-15) and examples of their use are shown by Thomsen and Foote (45, pp. 290-297).

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 mentioned on page 35. (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 that indicate the degree of partial correlation mentioned previously 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 thirties.

Residuals for years omitted because of price ceilings or rationing programs should be checked for consistency with known conditions. Residuals for all years should be approximately random when plotted against time. If the residuals appear to be serially correlated, a test should be made for this. Sometimes the addition of another variable, or a shift to the use of first differences, will reduce the degree of serial correlation in the residuals. An example of the former is given in Armore (2, pp. 30 and 57-58). The latter is discussed in Cochrane and Orcutt (7).

Considering the Nature of the Statistical Coefficients

In general, the direction and general shape of the regression curves should be consistent with expectations. "Wrong" signs or shapes may occasionally lead to a revamping of the underlying theory but they are more likely to indicate the need for a different approach in the analysis. Technical considerations may indicate that regression coefficients should be of a certain size. More confidence can be placed in an analysis if the results are consistent with factors of this sort. The constant value in a first-difference analysis should be consistent in sign and approximate magnitude with expectations, based on known long-term trends that affect the dependent variable, provided it differs significantly from zero. The appropriate error term to use in connection with such a test is the standard error of forecast when the independent variables equal zero. (See Foote (14, pp. 4, 6)

The Use and Interpretation of Tests of Significance

Yule and Kendall (52, p. 437) say, "It cannot be over-emphasised 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 (52, pp. 437-438) say, "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. But in general, the requirement of normality is not very restrictive.

If tests of significance as applied to correlation measurements are to be valid, the residuals should be randomly distributed. Certain modified tests have been suggested for cases in which the residuals are serially correlated, but, in general, it is preferable to use a "transformation", such as a shift to first differences, to eliminate the serial correlation in the residuals.

Some Common Tests

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. However, *t* could be used equally well to test for significant deviations from any other assumed value, such as a slope of -1 for a supply-price 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 (13, pp. 381-386). Snedecor (43, 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

$$\chi^2 = \Sigma \frac{(x-m)^2}{m},$$

in which *x* is the observed and *m* is the expected frequency.

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 two 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 feed grains discussed on page 20. Steps involved in this test are given in Foote (15, pp. 37-39).

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:

1. 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.

2. Minimize the sum of squares after assigning values to such coefficients as are to be tested.

3. Find the difference between the first and the second sums of squares. The degrees of freedom attached to this difference equal the number of coefficients to which a value was assigned.

4. 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.

5. *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.

Meinken (36) used this approach to determine whether related coefficients in a series of regression analyses differed significantly from each other. It also could be used in checking whether price and income elasticities differ from month to month or in a prewar and postwar period.

Application of the F-test to a Problem in Regression Analysis

Suppose we have an equation of the following type:

$$C = a + b_1 P + b_2 Y, \tag{13}$$

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, when 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 would be 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 equal 1, as a value was assigned to 1 parameter. F would be 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:

$$t = \frac{b_1 + 1}{s_{b_1}}$$

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 was desired, the F-test would be needed.

Allowing for Disturbances or for Errors in the Data

Most research analysts are inclined to appraise the success of a regression analysis by the level of the simple or multiple coefficient of determination (r^2 or R^2) obtained. But if errors of measurement are known to be present in the data, or if some minor variables are known to be omitted from the regression equation, a less-than-perfect correlation must be expected.

Consideration of the following cases illustrates the effects of errors in the variables and disturbances (omitted variables) on the regression and correlation coefficients. The model in each case illustrates the mechanism by which the observed data were generated; the true regression coefficient is b_o; and $\hat{x}_1$ and $\hat{x}_2$ are true values of the variables expressed as deviations from their respective means. Disturbances are indicated by the symbol d, and errors in the variables by $\underline{e_1}$ and $\underline{e_2}$. Results shown hold precisely only for large samples.

Alternative Cases

1. Perfect correlation: Neither disturbances nor errors:

Model: $$\hat{x}_1 = b_o \hat{x}_2$$

Proportion of variance explained: $$r^2 = \frac{b_o^2 \Sigma \hat{x}_2^2}{\Sigma \hat{x}_1^2} = \frac{b_o^2 \Sigma \hat{x}_2^2}{b_o^2 \Sigma \hat{x}_2^2} = 1$$

Least-squares regression coefficient: $$b = \frac{\Sigma \hat{x}_1 \hat{x}_2}{\Sigma \hat{x}_2^2} = b_o .$$

2. Disturbances (due to omitted variables) but no errors in the data:

Model: $$\hat{x}_1 = b_o \hat{x}_2 + d;$$

The disturbances, $\underline{d}$, are assumed to be uncorrelated with $\hat{\underline{x}}_2$.

Proportion of variance explained: $$r^2 = \frac{b_o^2 \Sigma \hat{x}_2^2}{b_o^2 \Sigma \hat{x}_2^2 + \Sigma d^2} < 1$$

Least-squares regression coefficient: $$b = \frac{\Sigma \hat{x}_1 \hat{x}_2}{\Sigma \hat{x}_2^2} = b_o.$$

3. Disturbances, and also random errors in the dependent variable:

Model: $$\hat{x}_1 = (\hat{x}_1 + e_1) = b_o \hat{x}_2 + d, \text{ or}$$

$$\hat{x}_1 = b_o \hat{x}_2 + (d - e_1)$$

Proportion of variance explained: $$r^2 = \frac{b_o^2 \Sigma \hat{x}_2^2}{b_o^2 \Sigma \hat{x}_2^2 + \Sigma d^2 + \Sigma e_1^2} < 1$$

Least-squares regression coefficient: $$b = \frac{\Sigma(\hat{x}_1 + e_1)\,\hat{x}_2}{\Sigma\,\hat{x}_2^2} = \frac{\Sigma\,\hat{x}_1\,\hat{x}_2}{\Sigma\,\hat{x}_2^2} = b_o$$

As errors of measurement are assumed random, $\Sigma\,\hat{x}_1\,e_2 = 0$.

4. Random errors in independent variables; no disturbances and no errors in the dependent variable:

Model: $$\hat{x}_1 = b_o\,(\hat{x}_2 + e_2) = b_o\,\hat{x}_2 + b_o\,e_2$$

Proportion of variance explained: $$r = \frac{b_o^2\,\Sigma\hat{x}_2^2}{b_o^2\,\Sigma\hat{x}_2^2 + b_o^2\,\Sigma\,e_2^2} < 1$$

Least-squares regression coefficient: $$b = \frac{\Sigma\hat{x}_1\,(\hat{x}_2 + e_2)}{\Sigma(\hat{x}_2 + e_2)^2} = \frac{\Sigma\hat{x}_1\,\hat{x}_2}{\Sigma\hat{x}_2^2 + \Sigma e_2^2} \neq b.$$

In case 4, b is biased toward zero: its absolute value in a large sample will be smaller than that of the true regression coefficient, b_o. In contrast, neither errors in the dependent variable nor disturbances which are uncorrelated with x_2 produce any bias in the least-squares regression coefficient.

5. Disturbances and random errors in both variables:

Model: $$\hat{x}_1 + e_1 = b_o\,(\hat{x}_2 + e_2) + d, \text{ or}$$

$$\hat{x}_1 = b_o\hat{x}_2 + (d - e_1 + b_o\,e_2)$$

Proportion of variance explained: $$r^2 = \frac{b_o^2\,\Sigma\hat{x}_2^2}{b_o^2\,\Sigma\hat{x}_2^2 + \Sigma d^2 + \Sigma\,e_1^2 + b_o\,\Sigma e_2^2} < 1$$

Least-squares regression coefficient: $$b = \frac{\Sigma(\hat{x}_1 + e_1)(\hat{x}_2 + e_2)}{\Sigma(\hat{x}_2 + e_2)^2} = \frac{\Sigma\hat{x}_1\,\hat{x}_2}{\Sigma\,\hat{x}_2^2 + \Sigma e_2^2} \neq b_o,$$

provided the errors and disturbances are uncorrelated with each other or with the true values of x_1 and x_2. As in case 4, the least-squares regression coefficient is biased toward zero.

286234 O - 54 - 5

The implications of the five cases may be summarized as follows:

1. In general, the unexplained variation in a regression analysis comes from a combination of disturbances (omitted variables or factors) and errors in the data. Each additional source of random error or disturbance tends to reduce the correlation coefficient.

2. Random errors in the dependent variable do not bias least-squares regression coefficients; random errors in the independent variables bias regression coefficients toward zero.

3. If we have a basis for estimating the degree of error in each independent variable, we can correct the least-squares regression coefficients for bias simply: Instead of using the observed sum of squares (x_2^2), we use this sum minus the estimated error component (e_2^2). The original bias depends on the proportion of the observed variation in x_2 which is due to errors of measurement.

4. The above correction will also increase the coefficient of determination, r^2. A correction for errors in the dependent variable would also increase r^2, but would not affect the regression coefficient.

5. Disturbances have real causes. They can be reduced only by identifying some of these causes and including them in the regression analysis. Thus, if any disturbances are believed to exist, the adjustment for errors should not be large enough to produce perfect correlation between the corrected variables.

Examples of Regression Equations Adjusted for Estimated Errors of Measurement

The notation used in these examples is obvious. The retail price of eggs is indicated by $P_e^{(r)}$; the farm price of turkeys, by $P_t^{(f)}$, etc. Consumer income is represented by y. Numbers in parentheses under the regression coefficients are their standard errors.

1. Retail price of eggs:

a. Unadjusted:

$$P_e^{(r)} = -\ 0.010 - \underset{(0.48)}{1.83}\, q_e + \underset{(0.15)}{1.24}\, y \qquad (14)$$

$$R^2 = 0.80 \ : \ r^2_{pq.y} = 0.48 \ : \ r^2_{py.q} = 0.80$$

$$\bar{S} = 0.029$$

b. Adjusted (Assuming random measurement errors of 1 percent in production of eggs and disposable income, and 0.5 percent in prices of eggs. As year-to-year changes in production of eggs are small, the assumed measurement errors account for 15 percent of the observed variance in production, but only

1.5 percent for income and 0.2 percent for price.)

$$P_e^{(r)} = -0.010 - \underset{(0.44)}{2.34}\, q_e + \underset{(0.13)}{1.34}\, y \tag{14}$$

$$R_a^2 = 0.87 : r_{pq.y_a}^2 = 0.64 : r_{py.q_a}^2 = 0.87$$

$$\bar{S}_a = 0.023$$

1. Farm price of turkeys:

a. Unadjusted:

$$P_t^{(f)} = 0.024 - \underset{(0.26)}{1.44}\, q_t + \underset{(0.21)}{0.94}\, y \tag{15}$$

$$R^2 = 0.86 : r_{pq.y}^2 = 0.78 : r_{py.q}^2 = 0.68$$

$$\bar{S} = 0.046$$

b. Adjusted (Assuming random measurement errors of 2 percent in the price of turkeys, 2.5 percent in production of turkeys, and 1 percent in disposable income. The assumed variance resulting from errors in the data for production amounted to nearly 10 percent of the observed variance, but only about 1 percent for price and income.)

$$P_t^{(f)} = 0.028 - \underset{(0.20)}{1.58}\, q_t + \underset{(0.16)}{0.95}\, y \tag{15}$$

$$R_a^2 = 0.92 : r_{pq.y}^2 = 0.87 : r_{py.q}^2 = 0.80$$

$$\bar{S}_a = 0.034$$

The errors of measurement in the above examples were based on judgments of commodity and price specialists. If they are approximately correct, they are helpful in these ways: (1) They indicate that a third to a half of the variance unexplained by the unadjusted equation may come from errors in the data. There is no point in trying to explain this part of the variation with additional real variables or unusual manipulations of data. But this variability need not be ascribed to irrationality of consumers or other unknown causes; (2) By getting a better approximation to the true regression, better estimates of such things as costs of alternative price-support programs can be made. These are ultimately determined as "universe" values--the totality of price-support purchases and costs--and not as sample values subject to error.

Forcing Perfect Correlation to Obtain a Reversible Demand Curve

This can be thought of as a special case of adjustment for errors in the data. It involves the assumption that there are no disturbances (or omitted variables) and that the correct functional form (such as arithmetic or logarithmic) has been chosen. Under these assumptions the only reason for less-than-perfect correlation is the presence of random errors in each variable. The relative level of error in each variable (that is, the percentage of the observed variance that is due to random errors) is estimated, and the equation is solved mathematically for the absolute level of error that must be assumed in order to produce perfect correlation.

The following example relates to the estimation of domestic demand for all food at retail:

1. Unadjusted:

$$p_f^{(r)} = -0.004 - \underset{(0.49)}{2.00}\, q_f + \underset{(0.08)}{0.91}\, y \tag{16}$$

$$R^2 = 0.86 : r^2_{pq.y} = 0.51 : r^2_{py.q} = 0.86$$

$$q_f = -0.000 - \underset{(0.06)}{0.25}\, p_f^{(r)} + \underset{(0.05)}{0.25}\, y \tag{17}$$

$$R^2 = 0.66 : r^2_{qp.y} = 0.51 : r^2_{qy.p} = 0.65$$

If there is a single demand curve for all food at retail, the price flexibility at any point of the curve is exactly equal to the reciprocal of the elasticity of demand at that point. However, equations (16) and (17) give (1) a price flexibility of -2.0 and (2) a demand elasticity of -0.25. The reciprocal of the latter figure is -4.0, which is twice as large as the first estimate of price flexibility. Is there a value somewhere between these two figures which represents a best estimate of the price flexibility? This is given (on certain assumptions) by the adjusted equations.

2. Adjusted:

a. First assumption (Errors as percentage of observed variance: Prices of food 3 percent; income 12 percent; consumption of food 50 percent. Perfect correlation requires errors only 39 percent as large as these):

$$p_f^{(r)} = a_1 - 2.97\, q_f + 0.95\, y \tag{16.1}$$

$$q_f = a_2 - 0.34\, p_f^{(r)} + 0.32\, y \tag{17.1}$$

$\underline{R}^2 = 1$ in each case; demand elasticity is exact reciprocal of price flexibility, except for rounding.

b. Second assumption (Assumed relative errors in price and income same as before; error in consumption of food estimated at 25 percent of observed variance. Perfect correlation requires measurement errors only 51 percent as large as these):

$$p_f^{(r)} = a_{12} - 2.70\ q_f + 0.92\ y \tag{16.2}$$

$$q_f = a_{22} - 0.37\ p_f^{(r)} + 0.34\ y \tag{17.2}$$

$R^2 = 1$ in each case; demand elasticity is exact reciprocal of price flexibility.

This method is not recommended for general use; but when such rigid assumptions seem justified, it should give the best (statistical and judgmental) estimate of the true reversible demand curve, if one exists.

Validity of an Estimate from a Multiple Regression Equation 11/

Only the simplest case is considered here - a linear equation in three variables:

$$X_1 = a + b_{12.3}\ X_2 + b_{13.2}\ X_3\ . \tag{18}$$

Under what conditions does this equation give a valid estimate of X_1?

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 non-significant, 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 Waugh and Been (50) or Armore and Burtis (3, pp. 7-9) 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,

---

11/ This section was prepared by Frederick V. Waugh, Director, Agricultural Economics Division, Agricultural Marketing Service.

$$X_1 - a + b_{12.3} X_2 + b_{13.2} X_3 .$$

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 (12, 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 SURVEY OF DEMAND ELASTICITIES BASED ON SINGLE-EQUATION ANALYSES

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 variously as

$$\eta = \frac{\Delta q}{\Delta p} \cdot \frac{p}{q} \left[\text{which is equivalent to } \left(\frac{\Delta q}{q}\right) \div \left(\frac{\Delta p}{p}\right) \right]$$

$$= \frac{d\ (\log q)}{d\ (\log p)} .$$

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 changes 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.

Terminology regarding regressions of price upon income is not standardized. Some analysts refer to the ratio of a percentage change in price to the corresponding percentage change in income as "price flexibility with respect to income." The ratio of a percentage change in consumption to the corresponding percentage change in income, whether in a time-series analysis or a family budget study, is usually referred to as "income elasticity of demand" or elasticity of demand with respect to income.

Some writers apply the term, "income elasticity of demand", to the ratio of a percentage change in retail expenditures to the corresponding percentage change in consumer income. This use is confusing, as traditionally elasticity of demand has been associated with price rather than value relationships. The authors prefer to refer to a percentage relationship between expenditures and income explicitly as "the elasticity of expenditures with respect to income."

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 which is a trivial object of expenditure (that is, one that takes up an almost infinitesimal fraction of total income) is likely to be highly inelastic. To support this, another factor that tends to make for inelasticity would have to be added, that is, that the commodity has no close substitutes. A commodity such as potatoes or onions, on this score, would have a less elastic demand than a commodity such as beef, pork, or chicken which has several fairly close substitutes. Numerical results from a group of analyses described in Fox (18, 20)are generally in accord with this assumption (tables 1 and 2). Whether two commodities are effective (economic) substitutes depends on consumer attitudes toward them and not directly upon their nutritional and physical attributes.

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 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.

Problems Involved in Interpretation

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) there is no necessary reason for the elasticity of demand in any one of these outlets to equal the reciprocal of "price flexibility with respect to production." An example of this is mentioned in Fox (18, pp. 70-71). Changes in commercial stocks and in net exports of meat tended to cushion the effects of a change in production of meat upon its retail price. The reciprocal of the price-production relationship suggest 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 were highly correlated with changes in production (r^2 = 0.94), but 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 probably not greater than -0.1. The demand for wheat as a livestock feed is inelastic as long as the price of wheat is considerably above the prices 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 twenties, 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 current conditions of extensive import and exchange regulations probably has 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.

The value of statistical regression coefficients depends upon the extent to which they enable us to act more 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 demand is unit elastic, this tends to make the costs to Government of the two methods of price support identical. If demand is highly inelastic, purchase and diversion will be less expensive to the Government; if demand is more than unit elastic, compensatory payments will presumably be less expensive to Government as well as more satisfactory to consumers. The reasons are discussed in detail in Fox (19). 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.

Coefficients Obtained

Numerous analyses based on data for 1922-41 are presented in Fox (18, 20). Results from these are summarized in tables 1 and 2.

The demand curves on which these tables are based were fitted by single-equation methods after considering the conditions under which each commodity was produced and marketed. Commodities with complicated patterns of utilization were treated partially or not at all.

The functions selected were straight lines fitted to first differences of logarithms of annual data. In most cases, retail price was taken as the dependent variable and per capita production and per capita disposable income undeflated as the major independent variables. Per capita consumption was substituted for production in some analyses.

The logarithmic form was chosen on the ground that price-quantity relationships in consumer-demand functions were more likely to remain stable in percentage than in absolute terms when there were major changes in the general price level. First differences (year-to-year changes) were used to avoid spurious relationships resulting from trends and major cycles in the original variables, and for their relevance to the outlook work of the Bureau of Agricultural Economics, which focuses on changes from one year to the next.

Results from these analyses, including projections into the post-World War II period, are discussed in detail in the references noted above. Tables 1 and 2 are believed to be largely self-explanatory, at least when combined with the discussions given in earlier parts of this handbook.

Most of the demand elasticities for livestock products at the retail price level range between -0.5 and -1.0 with respect to domestic consumption. If demand elasticities at the farm price level are derived from the elasticities of domestic food consumption with respect to retail price by dividing the elasticity at the retail price level by the flexibility of farm price with respect to retail price, they center around -0.5. But demand elasticities at the farm level with respect to total supply or production are greater than elasticities derived from domestic consumption alone, as the impacts of changes in production upon farm price are softened by adjustments in commercial stocks, exports, and imports.

THE SIMULTANEOUS EQUATIONS APPROACH

A modern econometric investigation consists of three major steps: (1) Specifying the model or system of relationships believed responsible for generating the observed data (for example, prices and consumption of a given commodity); (2) establishing the identifiability (that is, measureability) of individual equations or coefficients in the model; and (3) statistical estimation of the identifiable equations or coefficients.

A single-equation least-squares analysis of demand involves the same steps, implicitly if not explicitly. It assumes (1) that the demand function is such that one variable can be selected as dependent upon the others, and that all residual errors or disturbances are concentrated in the dependent variable; (2) that none of the independent variables in the demand function are, in fact influenced by or determined simultaneously with the dependent variable; (3) that the disturbances in the dependent variable tend to be normally distributed and not serially correlated.

In the 1920's and 1930's some price analysts were aware that these assumptions were not always satisfied. Particular departures from this model were considered

Table 1.- Food livestock products: Alternative estimates of elasticities of demand at retail and farm price levels, United States, 1922-41

Commodity	Elasticity of demand at retail price level: Least-squares coefficients: Consumption dependent 1/ (1)	Least-squares coefficients: Price dependent 2/ (2)	Adjusted coefficients 3/ (3)	Farmer's share of retail dollar, 1950 4/ (4)	Flexibility of farm price with respect to retail price 5/ (5)	Approximate elasticity of demand at farm price level: "Derived" coefficients 6/: Least-squares: Col.(1) ÷ Col.(5) (6)	"Derived" coefficients 6/: Least-squares: Col.(2) ÷ Col.(5) (7)	"Derived" coefficients 6/: Adjusted: Col.(3) ÷ Col.(5) (8)	Direct calculations 7/: Least-squares: Production dependent (9)	Direct calculations 7/: Least-squares: Price dependent (10)	Direct calculations 7/: Adjusted (11)
	Percent	Percent	Percent	Percent	Percent	Percent	Percent	Percent	Percent	Percent	Percent
Food livestock products:											
Meat:											
Pork	-0.81	-0.86	8/-0.87	9/64	1.75	-0.46	-0.49	8/-0.50	-0.44	-0.65	---
Beef	- .79	- .94	10/-1.04	11/70	1.74	- .45	- .54	10/- .60	- .54	- .84	---
Lamb	12/- .91	13/-2.00	---	64	1.06	- .86	13/-1.89	---	- .40	- .67	---
All meat:											
Deflated income	- .62	- .65	---	62	1.57	- .39	- .41	---	---	---	---
Actual income	- .64	- .67	---	62	1.57	- .41	- .43	---	- .44	- .62	---
Chicken	12/- .72	13/-1.33	14/-1.15	51	1.35	12/- .53	- .99	14/- .85	12/- .40	13/-1.61	---
Turkeys	---	---	---	---	---	---			- .61	- .83	15/-0.75
Eggs	- .26	- .55	15/- .43	70	1.08	- .24	- .51	- .40	- .21	- .44	15/- .34
All food livestock products:											
Deflated income	- .52	- .56	---	60	1.47	- .35	- .38	---	---	---	---
Actual income	- .56	- .61	---	60	1.47	- .38	- .41	---	16/ .33	16/ .41	---

1/ This coefficient has traditionally been called the "elasticity of demand" in least-squares demand analyses.
2/ This is the reciprocal of "price flexibility" in the traditional terminology. According to simultaneous equations theory, this coefficient is an unbiased estimate of the true demand elasticity in an uniequational complete model.
3/ For nature of adjustments, see footnotes on individual coefficients.
4/ If unit marketing margins are constant in dollars and cents, the "derived" demand elasticity at farm prices will be equal to the retail elasticity of demand multiplied by the percentages listed below.
5/ Average percentage change in farm price associated with a year-to-year change of one percent in retail price during 1922-41. May not apply precisely to 1950-1952 conditions.
6/ Represents the elasticity of final domestic consumer demand as reflected at the farm price level. This is usually less elastic than the total market demand.
7/ In most cases, these coefficients are estimates of the elasticity of total market demand.
8/ Based on simultaneous equations model.
9/ Pork including lard.
10/ Based on the author's interpretation of a simultaneous equations model.
11/ Choice grade only.
12/ Probably understates true elasticity.
13/ Probably overstates true elasticity.
14/ Based on a weighted regression allowing for errors in all variables and forcing a reversible (unique) demand function.
15/ Allowing for direct estimates of the level of measurement error in all variables, but without forcing a reversible (unique) demand function.
16/ Based on per capita consumption rather than production.

Table 2. - Selected crops: Alternative estimates of elasticity of demand at farm price level, United States, 1922-41 1/

Commodity	Least-squares coefficients using as the dependent variable: Production 2/	Least-squares coefficients using as the dependent variable: Price 3/
	Percent	Percent
Fruits:		
Deciduous:		
Apples	-1.21	-1.27
Peaches 4/	-1.18	-1.49
Cranberries 5/	-.57	-.67
All deciduous	-1.11	-1.47
Citrus:		
Oranges	-0.58	-0.62
Grapefruit	-.40	-.56
Lemons:		
Shipped fresh:		
Summer	-.29	-.40
Winter	-.61	-.72
All lemons	-.35	-.59
All citrus	-0.69	-0.76
All fruits	-0.82	-1.06
Potatoes, using as quantity variable:		
Production	-0.26	-0.28
Consumption 6/	-.22	-.27
Sweetpotatoes	-.74	-1.30
Onions:		
Late summer	-.28	-.34
All onions	-.39	-.44
Truck crops for fresh market:		
Winter	-0.45	-0.88
Spring	7/ -.30	7/ -1.05
Summer	-.42	-.58
Fall	-.41	-.60
Calendar year	-.59	-.97
Hay	-0.62	-0.72
Corn:		
1st analysis	-.43	8/ -.52
2d analysis	-.44	9/ -.79

1/ Represents elasticity of total market demand in most cases; derived demand for final domestic consumption would typically be less elastic.

2/ This coefficient has traditionally been called elasticity of demand in single-equation analysis.

3/ In traditional terminology, this is the reciprocal of "price flexibility." According to simultaneous equations theory, this coefficient is an unbiased estimate of the true demand elasticity in a uniequational complete model.

4/ Excluding California.

5/ Based on data for 1922-36.

6/ Elasticity measured at retail prices.

7/ Nonsignificant at 5-percent level.

8/ May reflect demand for all feed concentrates as an aggregate.

9/ More nearly reflects demand for corn given constant supplies of all other grains and feed concentrates.

in some detail. The more important of these theoretical discussions are mentioned in Fox (20). As late as 1942 opinion differed as to which variable should be placed in the dependent position in a least-squares analysis of demand, although Ezekiel followed a causal principle which in many practical cases led to the same choice as that indicated by modern econometric theory. But the probing and questioning of the single-equation, least-squares model continued on a piecemeal basis until 1943.

In that year, Haavelmo (25) put forward a general theory of econometric analysis which included the traditional single-equation model as a special case. This general theory successfully integrated the three problems of specification, identification, and estimation for systems of simultaneously determined variables. However, the new theory was stated in a relatively new terminology; it presumed a knowledge of matrix algebra; and it involved "maximum-likelihood estimation" rather than least squares. All of these elements were barriers to understanding, and hence to acceptance and application, of Haavelmo's new synthesis.

Haavelmo recognized these difficulties in a second publication issued in 1944:

> "We believe that, if economics is to establish itself as a reputable quantitative science, many economists will have to revise their ideas as to the level of statistical theory and technique and to the amount of tedious work that will be required, even for modest projects of research." (26, p. 114)

This statement has an unduly ominous ring. But it is true that the basic logic of the simultaneous-equations approach must be understood in order to use and interpret single-equation analyses with proper discrimination.

The three major problems of (1) specification, (2) identification, and (3) estimation are considered in turn.

Specifying the Economic Model

An economic model consists of a set of relationships between observed variables and a set of assumptions concerning the nature of variations in the data that are not explained by the systematic relationships but are caused by residual errors or disturbances, or both. The relationships between observed variables are called "structural equations" in the Cowles Commission literature. Koopmans (33) mentions four types of structural equations: (1) "Behavior equations", which describe "a certain type of economic decision taken by a certain category of economic agents." These include the demand and supply curves of economic theory; (2) "institutional equations," which describe behavior patterns set by law or rule; (3) "technical equations," which express the physical relation between input and output in production; and (4) "identities," which "should be classified as deriving directly from the definitions of the variables through the principles of economic accounting."

Many arbitary elements are involved in the specification of a model. Consumers, marketing agencies, and producing areas must be aggregated even when there may be sizable variations within these categories. The specification of a model also requires the classification of variables into "exogenous" and "endogenous" groups. Exogenous variables are those which represent forces outside the confines of the economic system (for example, weather) whereas endoggenous variables are those determined within the system of economic forces in a narrow sense (as prices or consumption). The <u>number of structural equations in a complete model must be equal to the number of endogenous variables whose values are to be explained by the equation system</u>. No attempt is made to develop equations explaining the exogenous variables; they influence the endogenous variables but are assumed to be not influenced by them. Structural equations also may include values of endogenous variables for preceding time periods. Lagged endogenous variables are similar to exogenous variables in that they influence current values of the endogenous variables but are not affected by them. For this reason the lagged endogenous and the exogenous variables are often referred to collectively as predetermined variables.

The enumeration of structural equations, the selection and classification of variables, and the choice of variables and functional forms in each equation collectively define the <u>economic</u> properties of a model. If each structural equation expressed a true and exact relationship, knowledge of the coefficients in each equation would enable us to find the precise values of all the endogenous variables that would result from a given set of values of the predetermined variables. As a matter of practice, we do not know the values of the coefficients and our ability to ascertain them depends upon the mathematical properties of the equations as well as the statistical properties ascribed to the unexplained residuals. Only the mathematical properties of the model are involved in the problem of identification.

Problems of Identification

If a structural equation contains only one nonlagged endogenous variable, its parameters or coefficients are identifiable. In this case, a least-squares equation expressing the endogenous variable as a function of the predetermined variables will yield appropriate estimates of the parameters. Such an equation is called a "uniequational complete model."

Suppose, however, that two endogenous variables appear in each of two equations. For example, suppose a set of price-quantity observations were generated by the following model:

$$\text{Demand:} \quad Q = a + bP + u \tag{19}$$

$$\text{Supply:} \quad q = \alpha + \beta P + v, \tag{20}$$

in which $\underline{u}$ and $\underline{v}$ are regarded as random disturbances. From a statistical viewpoint we have <u>no</u> basis for distinguishing between these two equations. Both contain the same variables and both have random disturbances. It can be shown that the least-squares regression of $\underline{Q}$ upon $\underline{P}$ is, in an indefinitely large sample,

$$B = \frac{b\,\sigma_v^2 - (b+\beta)\, r_{uv}\,\sigma_u\sigma_v + \beta\sigma_u^2}{\sigma_v^2 - 2r_{uv}\,\sigma_u\,\sigma_v + \sigma_u^2}\;. \tag{21}$$

If the demand curve has not shifted (that is, if $\sigma_u = 0$), $\underline{B} = \underline{b}$; if the supply curve has not shifted (that is, if $\sigma_v = 0$), $\underline{B} = \beta$. But in general $\underline{B}$ will not equal either of the structural coefficients and neither of them are identifiable. If equations (19) and (20) constitute the true and complete model there is, in general, no way by which $\underline{b}$ and β can be estimated from the data.

The situation is different if each equation includes a different predetermined variable. Suppose that the structural equations have the following form:

$$\text{Demand:} \quad Q = a + bP + cY + u \tag{22}$$

$$\text{Supply:} \quad Q = \alpha + \beta P + \gamma Z + v, \tag{23}$$

where $\underline{u}$ and $\underline{v}$ again are random disturbances. $\underline{Y}$ might represent consumer income and $\underline{Z}$, livestock numbers on January 1. If $\underline{Q}$ and $\underline{P}$ are expressed as functions of the predetermined variables and the disturbances and each of the variables is expressed in deviation form, we obtain the "reduced form" of the model as follows:

$$p = -\left(\frac{c}{b-\beta}\right) y + \left(\frac{\gamma}{b-\beta}\right) z + \left(\frac{v-u}{b-\beta}\right) \tag{24}$$

$$q = -\left(\frac{c\ \beta}{b-\beta}\right) y + \left(\frac{\gamma b}{b-\beta}\right) z + \left(\frac{bv-\beta u}{b-\beta}\right) . \tag{25}$$

The coefficients of the reduced-form equations may be estimated by least squares. The structural coefficient β can be estimated as the ratio of the coefficients of $\underline{y}$ in equation (24) and (25); $\underline{b}$ can be estimated as the ratio of the coefficients of $\underline{z}$. Knowing β and $\underline{b}$, we can derive $\underline{c}$ and γ from the coefficients of equation (24); $\underline{a}$ and α can be estimated from the other coefficients, as means of $\underline{Q}$, $\underline{P}$, $\underline{Y}$, and $\underline{Z}$ for the sample are known. The model consisting of equations (22) and (23) is just identified: That is, our information is sufficient to make a unique estimate of each structural parameter or coefficient. In other models it may be possible to identify some of the structural equations but not others. Still other models may be "overidentified." In this case there are two or more ways of deriving a given equation, and the various possible individual estimates must be reconciled or averaged by a maximum likelihood method.

Identification, then, is a matter of mathematics or logic which precedes statistical estimation. When statistical considerations are introduced additional difficulties arise. These include the situation in which two variables are logically distinct but happen to be highly correlated during the period of observation. The parameters or coefficients still could be accurately determined in an infinite sample, but the standard errors of parameters estimated from a small sample may be greatly increased by such intercorrelation.

Problems of Estimation

The preceding two sections have considered only the economic and mathematical properties of a simultaneous-equations model. The problems of statistical estimation are basically more difficult than those that arise in specifying its economic and mathematical properties.

As noted above, the number of structural equations in a complete model must be equal to the number of endogenous variables whose values are determined within the model. Each structural equation may be written as a function of the endogenous and predetermined variables which appear in it, and of a random disturbance.

The problem of statistical estimation is to identify values of the coefficients of the structural equations which maximize the probability of obtaining the observed values of the endogenous variables on the basis of an assumed (usually a normal) probability distribution of the disturbances. The theoretically ideal method for obtaining such values of the coefficients is that of maximum likelihood, using all of the information contained in the structural equations.

When the number of endogenous variables is large, the calculations involved in maximum likelihood estimation are laborious and expensive. This in itself would encourage the use of less expensive, even if less accurate, methods of computation. If the investigator does not require estimates of the structural coefficients but is content to obtain forecasts of the endogenous variables on the basis of given values of the predetermined variables and no changes in structure have taken place, a least-squares estimating equation, using a specified endogenous variable in the dependent position, is indicated. (See Marschak, (35, pp. 1-8) Such an equation gives the best linear unbiased estimate of this variable, given the specified values of all predetermined variables. This will not in all cases be the most efficient possible method of estimation, as information afforded by the structural equations is disregarded.

But it can be shown that when the structural equations in the system are just identified, the information contained in a set of least-squares equations of the type just described, one for each endogenous variable, is equivalent to that contained in the structural equations themselves. In this case it is possible to make two simple, straightforward, and unique transformations. One transforms structural equations into least-squares equations, each containing one endogenous variable; the other transforms the least-square estimates of coefficients back to estimates of structural coefficients. This approach is called the "method of reduced forms." Most applications of the simultaneous-equations approach, including that made by Girshick and Haavelmo (23), have used this method.

Applicability of Single-Equation Methods to Analyses Designed to Measure the Elasticity of Demand for Farm Products

On page 43 it was noted that if a structural equation contains only one nonlagged endogenous variable, its parameters or coefficients are identifiable and that a least-squares equation expressing the endogenous variable as a function of the predetermined variables will yield appropriate estimates of the parameters.

Consider the following rather typical consumer-demand function for an agricultural food product:

$$X_1(t) = b_0X_2(t) + c_0X_3(t) + d_1(t) \tag{26}$$

where $\underline{X}_1$ = retail price;

$\underline{X}_2$ = per capita consumption;

$\underline{X}_3$ = per capita disposable income;

$\underline{d}_1$ = random disturbance;

$\underline{b}_o$, c_o = true values of structural coefficients.

In order to show that a single least-squares equation involving these variables will give unbiased estimates of the structural coefficients in this function, we must show that the explanatory variables--consumption and disposable income--are in some sense predetermined, that is, that their values result from economic decisions made before harvest and from the action of exogenous or noneconomic variables such as weather. In addition, we must assume that the variance in retail price that is not explained by variations in consumption and disposable income is due to a random disturbance, representing the effect of minor variables omitted from the analysis, which is uncorrelated with the independent variables. If this is true, consumption and disposable income may be regarded as predetermined variables. As the disturbances are assumed to be reflected only in retail price, the demand function must be fitted with price in the dependent position.

The discussion that follows is intended to show that demand functions for many farm products probably meet the specification of the uniequational complete model. It cannot be shown affirmatively that the disturbances in a given case are uncorrelated with the explanatory variables, because the disturbances by definition are not directly observable. Thus it is possible that a noneconomic variable, such as summer temperature, which affects consumer demand for lemons, will be correlated with disturbances that arise from minor economic factors. Similarly, it is possible that production of apples, although determined by weather and by economic influences before harvest, will somehow be correlated with nonmeasurable disturbances in the demand function for apples during the subsequent marketing season. But there is no reason to expect that the disturbances will be dependent (in a probability sense) upon the variables in question.

From a commonsense viewpoint it appears that the nonanswerable question as to whether the disturbances are distributed independently of the explanatory variables can be disregarded unless there are special reasons for assuming the contrary. The answerable question as to whether certain variables entering into demand functions for farm products are predetermined in a logical sense, or nearly enough so to be used as explanatory variables without leading to seriously biased estimates of elasticities of demand is considered only briefly here. It is discussed in detail in Fox (20,22).

In ascertaining whether the single-equation approach can be used in deriving estimates of the elasticity of demand and similar coefficients, the following questions must be answered.

1. Is consumption a predetermined variable? One possibility is that consumption of a perishable commodity is precisely equal to its production and that production itself is a predetermined variable. It is evident that a variable which is identically equal to a predetermined variable may be regarded as predetermined.

A second case is that in which consumption of an item may differ from its predetermined production because of variations in exports, imports, or stocks, but in which consumption is highly correlated with production. If the disturbance or unexplained residual in the relationship between consumption and production is random, it appears that the degree of bias in the least-squares estimate of b_o in equation (26) will be fairly small.

But this situation could also be handled by a structural approach, as follows:

Assume a true demand function,

$$p = \beta q + u \tag{27}$$

and a consumption-production relationship,

$$q = \gamma z + v, \tag{28}$$

in which p, q, and z are, respectively, price, consumption, and production, all in deviation form; u and v are random disturbances uncorrelated with z which is a predetermined variable.

If we try to estimate β by fitting the least-squares regression of p on q, we obtain

$$b = \frac{\Sigma pq}{\Sigma q^2} = \frac{\beta \Sigma q^2 + \Sigma qu}{\Sigma q^2} = \beta + \frac{\Sigma \gamma zu + \Sigma vu}{\Sigma q^2} \; . \tag{29}$$

As $\Sigma \gamma zu = 0$ under our assumption, this is equivalent to

$$b = \beta + \frac{r_{uv}\sigma_u \sigma_v}{\sigma_q^2} = \beta + b_{uv}\frac{\sigma_v^2}{\sigma_q^2} \; . \tag{30}$$

If the disturbances u and v are not correlated, b is an unbiased estimate of β. If σ_v^2 is small relative to σ_q^2, the bias in b will be small, as it seems unlikely that the regression of disturbances in equation (27) upon disturbances in equation (28), b_{uv}, will be as large in absolute value as the structural coefficient, β, relating p and q.

In the present model, β might be estimated as follows:

By least squares, fit the reduced-form equations,

$$p = dz \text{ and} \tag{31}$$

$$q = cz. \tag{32}$$

As d is an estimate of $\beta\gamma$ and c is an estimate of γ, the estimate of the structural parameter β is given by d/c. The precision of this estimate will tend to increase as the size of the sample increases. This estimate should be a statistically consistent one even if q contains random measurement errors as well as the effects of disturbances.

2. Is production predetermined? For many products this question can be answered on logical grounds. The production of a crop is the product of planted acreage, which was influenced by economic and other considerations at or before planting time, and yield, which in most cases is largely a function of weather. For livestock products, the question might be approached on a partly statistical basis: What part of the observed variance in production can be explained by (1) variables whose values were actually known before the beginning of the current period, (2) variables whose values, though not known in advance, must clearly have been determined before the current period, (3) exogenous variables, such as weather and disease, and (4) errors of measurement. If by such a procedure we can explain 95 percent or so of the observed variation in production, we may conclude that, for practical purposes, production is a predetermined variable. The residual variation sets an upper limit to the possible endogenous or jointly determined element in production.

3. Is disposable income significantly affected by price or consumption of the given commodity? Disposable income is determined by a vast complex of economic decisions, some of which may be influenced by current prices of an individual commodity or group of commodities. But gross investment and Government expenditures in the current year are mainly predetermined variables, and these are considered to be largely responsible for major changes in disposable income. Hogs - a major agricultural commodity-account (as pork) for less than 3 percent of all consumer expenditures. It is improbable that the back-effects from consumption and prices of pork upon disposable income are great enough to rise above the level of measurement error in the income series.

Consumer-demand functions for many farm and food products approximately meet these requirements. For these items single least-squares equations can be used in deriving coefficients of elasticity. Other potentially simultaneous commodity structures may be broken down into a set of individual least-squares equations if these relationships operate in sequence (perhaps in terms of time units shorter than a year). The 4-equation model of the feed-livestock economy, discussed on p. 53, is an example of this.

However, the demand-supply structures for export crops, including fats and oils, for milk and dairy products, and for some fruits and vegetables with two or more major outlets involve two or more simultaneous equations. Structural coefficients estimated from single-equation demand functions fitted for such commodities are likely to be unreliable. Serious experimentation is needed to see whether reliable measures of elasticity can be derived, despite basic data and small sample limitations, by means of the simultaneous-equations approach. Anthony S. Rojko has developed some 3-equation models to explain the demand and price structure for individual dairy products. Price elasticities estimated from single-equation models were found to be substantially lower than corresponding estimates from the 3-equation model. Results of this research are to be published in the 1953 Proceeding Issue of the Journal of Farm Economics.

EXAMPLES INVOLVING THE USE OF MORE THAN A SINGLE EQUATION

The examples that follow were selected because they represent unusual approaches to the problem under consideration. Some of these analyses are concerned as much with supply as with demand, but the approach that was used appears to have wide application in related fields.

Simultaneous-Equation System for Pork, Beef, and Export Crops

Pork

Suppose we are interested in estimating a consumer-demand equation for pork, using calendar-year data. Farmers can alter average weights and numbers of hogs sold for slaughter somewhat during a given calendar year in response to prices of hogs during that year even though slaughter of hogs is largely determined by number of pigs saved and sows bred during the previous year. Similarly, meat-packers can alter their year-end inventories of pork in response to current-year influences. If we assume, therefore, that consumption of pork is determined simultaneously with prices of pork, we are led to a 2-equation model, as follows:

The structural equations are:

$$\text{Demand:}\quad p = \frac{1}{b}\,q + cy + u \tag{33}$$

$$\text{Supply:}\quad q = \beta p + \gamma z + v, \tag{34}$$

where p is price, y is disposable income, q is consumption, and z is an estimate of production based wholly on predetermined variables; u and v are random disturbances. Each equation is just identified. If the variables are in logarithmic form, b is the elasticity of demand and β the elasticity of supply.

The reduced-form equations are:

$$p = \left(\frac{cb}{b-\beta}\right) y + \left(\frac{\gamma}{b-\beta}\right) z + \left(\frac{bu+v}{b-\beta}\right) \tag{35}$$

$$q = \left(\frac{cb\beta}{b-\beta}\right) y + \left(\frac{b\gamma}{b-\beta}\right) z + \frac{b(\beta u+v)}{b-\beta}, \tag{36}$$

each of which can be fitted by least squares. As β will be calculated as the ratio of the coefficients of y in the two equations, our ability to establish a value of β that differs significantly from zero depends among other things upon whether the net regression of q upon y in equation (36) significantly differs from zero.

If we argue that equation (33) may be fitted by least squares without significant bias, we are arguing that β is equal to, or close to, zero. If $\beta = 0$, the reduced-form equations become:

$$p = cy + \left(\frac{\gamma}{b}\right) z + \left(u + \frac{v}{b}\right) \tag{35.1}$$

$$q = 0.y + \gamma z + (v), \tag{36.1}$$

from which the structural equations are

$$\text{Demand:} \quad p = \frac{1}{b} q + cy + u \tag{33.1}$$

$$\text{Supply:} \quad q = \gamma z + v. \tag{34.1}$$

For pork, with about 95 percent of the variation in hog production predetermined, the following results were obtained.

The reduced-form equations are:

$$p = \underset{(0.1091)}{-0.9581} z + \underset{(0.1026)}{0.9707} y + u; \; R^2 = 0.92 \tag{35.2}$$

$$q = \underset{(0.0670)}{0.8370} z \underset{(0.0629)}{-0.0641} y + v; \; R^2 = 0.91. \tag{36.2}$$

The two structural equations, assuming $\beta \neq 0$, are:

$$\text{Demand:} \quad p = -1.1447 q + 0.8974 y + u \tag{33.2}$$

$$\text{Supply:} \quad q = -0.0660 p + 0.7738 z + v. \tag{34.2}$$

The least-squares demand function is

$$p = \underset{(0.07)}{-1.16} q + \underset{(0.06)}{0.90} y + u; \; R^2 = 0.97 \tag{33.3}$$

and the "supply" function, assuming $\beta = 0$, is

$$q = \underset{(0.0670)}{0.8403} z + v; \; r^2 = 0.90. \tag{34.3}$$

Differences between the structural demand coefficients and those of the least-squares demand function are small in relation to the standard errors of the latter. The supply elasticity, β, is negative and does not differ significantly from zero.

Beef

The potential response of production of beef to current-year influences, and particularly to prices of beef and cattle, is greater than that of production of pork. Using a 2-equation model similar to that for pork, we find that the supply elasticity, β, does not differ significantly from zero. Assuming $\beta = 0$, we obtain the following structural demand function:

$$p = -0.96 q + 0.82 y - 0.43 z + u. \tag{33.4}$$

The least-squares demand function, using q directly and ignoring a possible simultaneous-supply function, is

$$p = -1.06\,q + 0.88\,y - 0.52\,z + u;\; R^2 = 0.95. \qquad (33.5)$$
$$\quad\;(0.12)\quad\;\; (0.06)\quad (0.09)$$

Although the differences between equations (33.4) and (33.5) are larger than in the case of pork, they do not exceed one standard error of the least-squares regression coefficients.

All Farm Products: Separating the Effects of Domestic and Export Demand

Assume that the domestic demand function for all farm products is given by

$$p_d = a + b\,q_d + c\,y\,, \qquad (37)$$

and the export demand by

$$p_e = \alpha + \beta\,q_e + \gamma\,e, \qquad (38)$$

in which q_d and q_e are domestic consumption and exports respectively; y is domestic income, and e is a measure of foreign demand which, under free-trade conditions, might be simply the total income of foreign consumers. Let us assume also that total disappearance ($q_d + q_e = q_t$) is a predetermined variable. Then the equilibrium price ($p_t = p_d = p_e$) for any given combination of q_t, y and e is

$$p = \left(\frac{\alpha b + a\beta}{b+\beta}\right) + \left(\frac{b\,\beta}{b+\beta}\right) q_t + \left(\frac{c\,\beta}{b+\beta}\right) y + \left(\frac{b\gamma}{b+\beta}\right) e. \qquad (39)$$

The following equation, which applies to the index of prices received by farmers for all commodities, involves approximately the same variables as does equation (39).

(40)

log (prices received = 2.812 -

1.658
(0.273) log (physical volume of farm marketings) +

1.241
(0.102) log (disposable income) +

0.142
(0.035) log (value of agricultural exports). $R^2_{1.234} = 0.97$.

This equation should be subject to interpretation in terms of separate domestic and foreign demand curves.

Let us assign the following values: $q_t = 100$; $y = 100$; $e = 100$; $p_t = 100$.

During the 1930's (and again in 1949-50) q_e averaged about 10 percent of q_t, and domestic sales (q_d) averaged 90 percent of q_t. Finally, we can assign some reasonable values to b, β, c, and γ, based on collateral evidence. We shall assume a domestic price flexibility of -2 with respect to q_d and 1.5 with respect, to y; also, a price flexibility of -1 with respect to q_e and 1 with respect to e. The resulting arithmetic slopes are b = 2.22; β = 10; c = 1.5 and γ = 1.0. Substituting in equation (39) we have

$$p_t = A - \left(\frac{22.20}{12.22}\right) q_t + \left(\frac{15.0}{12.22}\right) y + \left(\frac{2.22}{12.22}\right) \tag{39.1}$$

or

$$p_t = A - 1.817\ q_t + 1.230\ y + 0.182\ e.$$

The assumed set of structural coefficients is reasonably consistent with equation (40). If the assumed value of b is dropped to -2.0 (which means a domestic price flexibility of -1.8), the correspondence is still closer:

$$p_t = A - \left(\frac{20}{12}\right) q_t + \left(\frac{15.0}{12}\right) y + \left(\frac{2}{12}\right) e \tag{39.2}$$

or

$$p_t = A - 1.667\ q_t + 1.250\ y + 0.167\ e .$$

Two coefficients of this equation are almost identical with those of equation (40) and the third is within one standard error. It is evident that equation (40) is consistent with the assumed structural coefficients of the domestic and export demand functions when these are substituted in equation (39).

A System of Simultaneous Difference Equations Relating to the Feed-Livestock Economy

Demand and supply functions for some agricultural commodities imply an inner mechanism which, in the absence of other factors such as weather, would generate an endless series of price and production observations for successive years on the basis of specified initial conditions. A simple system of this sort that is familiar to agricultural economists is the "cobweb" model described by Ezekiel (11). For any given commodity, there is some interest in knowing whether this inner, or endogenous, mechanism is essentially a stabilizing factor, or whether it tends to perpetuate cycles of considerable, and possibly everincreasing, magnitude.

The Cyclic Structure of a Two-Equation Model

The simplest simultaneous equation system of this type would involve the following equations, in which each variable is expressed in terms of deviations from some mean or normal value:

$$\text{Demand equation:} \quad p_t = b\ q_t \tag{41}$$

$$\text{Supply equation:} \quad q_{t+1} = B\ p_t. \tag{42}$$

A mathematical method exists by which the nature of the inner mechanism can be determined from the coefficients of the equations. Essentially this involves finding the nature of the roots of an equation formed from these coefficients. In this instance, the root of the equation is bB. 12/ If this is negative, as will always be the case for this set of equations, one-period oscillations will occur. When bB is less than 1, the oscillations gradually drop down to zero; when bB is greater than 1, the oscillations become larger and larger; when bB equals 1, the oscillations are of constant magnitude.

Research by commodity analysts in the Agricultural Marketing Service indicates that for most agricultural commodities analyses run in terms of logarithms with prices at the farm level will yield coefficients for b between -0.7 and 3.5 (see tables 1 and 2) and for B between 0.1 and 0.5. Table 3 shows data for the first 5 years for price and quantity for systems of equations similar to (41) and (42) and having coefficients of specified size within these ranges. As, for these equations, b is a price flexibility and B is the elasticity of supply (when all variables are expressed in logarithms), the case in which $b = -2$ and $B = 0.5$ is equivalent to one in which the elasticity of demand is equal to the elasticity of supply. In such cases, as pointed out by Ezekiel (11), the oscillations are of constant magnitude.

A Four-Equation Model of the Feed-Livestock Economy

A system of simultaneous linear-difference equations was developed to study the effect of support programs for corn on prices of corn, production of livestock products, and returns to farmers from corn and livestock over a series of years. These equations can be used in sequence to follow through the effects of changes in feed supplies or consumer income on livestock production and on feed and livestock prices or to study the effects over time of specified support programs for corn.

12/ The way in which bB fits into the system is clear from the following:

$$p_1 = b\, q_1$$

$$q_2 = B\, p_1$$

$$p_2 = b\, q_2 = bB\, p_1 .$$

Extending this, we see that

$$p_3 = (bB\ )^2\ p_1 \text{ and, in general,}$$

$$p_t = (bB)^{t-1}\ p_1 .$$

The same "multiplier" (bB)would apply to year-to-year changes in both prices and quantity.

The following variables were used in the system of equations:

C - Price received by farmers for corn, which is assumed to be representative of the general price level for all feeds, in cents per bushel.

S - Supply of all feed concentrates, in million tons.

A - Number of grain-consuming animal units fed annually, in millions.

L - Price received by farmers for livestock and livestock products, index numbers (1910-14=100).

Q - Production of livestock and livestock products for sale and home consumption, index numbers (1935-39=100).

I - Personal disposable income, in billion dollars.

Table 3.- Successive values of price and quantity for a 2-equation simultaneous linear difference system with specified coefficients 1/

Time period	Coefficients									
	b = -0.7 B = .1 bB = -.07		b = -0.7 B = .5 bB = -.35		b = -3.5 B = .1 bB = -.35		b = -2.0 B = .5 bB = -1.0		b = -2.5 B = .5 bB = -1.25	
	q_t	p_t	q_t	p_t	q_t	p_t	q_t	p_t	q_t	p_t
1	1.00	-0.70	1.00	-0.70	1.00	-3.50	1.00	-2.00	1.00	-2.50
2	-.07	.05	-.35	.24	-.35	1.22	-1.00	2.00	-1.25	3.12
3	.00	-.00	.12	-.09	.12	-.43	1.00	-2.00	1.56	-3.91
4	--	--	-.04	03	-.04	.15	-1.00	2.00	-1.95	4.88
5	--	--	.02	-.01	.02	-.05	1.00	-2.00	2.44	-6.10

1/ Computed from unrounded data. p_t and q_t are expressed in terms of deviations from some mean or normal value.

All of the equations were fitted by single-equation least-squares techniques, using logarithmic data for approximately the crop years beginning 1922 through 1942. The use of single equations in fitting is permissible because the variables to the right of the equality sign in each equation are assumed to be either exogenous or lagged values of endogenous variables. In the following equations, Δ equals the change in a particular item from the preceding year. The numbers in parentheses underneath the regression coefficients are their respective standard errors.

$$\Delta C = 0.00373 - \underset{(0.24)}{2.36}\Delta S + \underset{(0.57)}{1.94}\Delta A + \underset{(0.18)}{1.13}\Delta L \qquad R^2 = 0.91 \tag{43}$$

$$\Delta A = -0.092 + \underset{(0.040)}{0.214}\Delta S - \underset{(0.032)}{0.185}\, C\ \underline{13/} + \underset{(0.036)}{0.207}\, L\ \underline{13/} \qquad R^2 = 0.86 \tag{44}$$

13/ Prices apply to the calendar year in which the crop year begins.

$$\Delta Q = 0.00369 + \underset{(0.090)}{0.562}\Delta A \qquad r^2 = 0.71 \qquad (45)$$

$$\Delta L = 0.00578 - \underset{(0.25)}{2.08}\Delta Q + \underset{(0.08)}{1.45}\Delta I \qquad R^2 = 0.96 \qquad (46)$$

ΔQ is assumed to be a direct function of ΔA, as feed fed per animal unit had the wrong sign and had no statistically significant effect on sales of livestock products. Therefore, equation (46) can be rewritten as:

$$\Delta L = -0.00190 - 1.17\Delta A + 1.45\Delta I. \qquad (46.1)$$

If this is done, equation (45) can be omitted.

The economic and statistical aspects of these equations are discussed in Foote (15).

Mathematical analysis similar to that described for the 2-equation model indicates that this system probably will involve one-period oscillations and that the fluctuations will tend to increase in amplitude. The algebraic computations required are discussed in detail by Foote (16,17).

Application to Analyses of Specified Corn-Loan Programs

In the analysis of the effects of specified loan programs for corn on the feed-livestock economy, the following general approach was used: (1) Certain assumptions were made regarding year-to-year changes in the exogenous variables, feed production and disposable income; (2) the effects of these on prices of corn and livestock, on production of livestock, and on total returns to farmers, within the assumed loan-program framework, were measured on a year-to-year basis; (3) results for the several loan programs and those arising without a loan program were compared, both by years and in total for a number of years. This method appears preferable to one in which results for specified loan programs are compared with actual prices for a given period. Many factors other than those allowed for in the model affect actual prices, so that the differences between results under a given loan program and actual prices reflect not only those due to the loan program as such, but also the effects of all factors not allowed for in the model. Under the system described above, the differences that are found are mainly due to the type of loan program assumed. But the results will depend to some extent on the particular pattern chosen for the exogenous factors, the starting levels for the various items, the particular equations used, and the length of time for which results are computed. All of these must be chosen as logically and as realistically as possible.

The following sequence was assumed in analyzing year-to-year changes in the several variables involved: (1) Prices of livestock for November to May are determined by the number of animal units fed on farms in the preceding year and current disposable income from November to May; (2) prices of corn for November to May are determined by the number of animal units in the preceding year, new-crop supplies (production plus stocks) of feed, and prices of livestock from step 1; (3) animal units fed are determined by new-crop supplies of feed and prices of corn and livestock for the calendar year in which the October to September corn-marketing year begins. Prices for the calendar year were obtained by weighting prices for November to May and June to October by the proportion of the sales for the calendar year obtained from each period; (4) prices of livestock from

June to October are determined by the number of animal units from step 3 and current disposable income for June to October; (5) prices of corn from June to October are determined by the number of animal units from step 3, the level of prices of livestock from step 4, and the supplies of feed used in step 2.

As all of the analyses were based on first differences of logarithms, the results are given in the form of percentage changes from the preceding year. Actual levels for any given year were obtained by applying these changes to the computed level in the preceding year. By this method, the study was continued year by year for as long as desired.

For each of the two periods within the corn-marketing year, four alternative price levels must be computed: A price equivalent to the loan rate, the release price, a computed price based on nonloan supply only, and a computed price based on total supplies. With a loan program in effect, prices are assumed to average at or above the loan equivalent. If the loan-equivalent price is above the computed price based on nonloan supply only, the loan-equivalent price prevails. Under such circumstances loan stocks will, in general, increase. If the computed price, based on nonloan supply only, is between the loan-equivalent price and the release price, the computed price will prevail and loan stocks will not change. If loan stocks exist and the release price is between the computed price based on total supply and the computed price based on nonloan supply only, the release price will prevail. In general, under such circumstances, loan stocks will be reduced during the year. If the computed price based on total supply is between the release price and the computed price based on nonloan supply only, computed prices based on total supply will prevail. Under such circumstances, loan stocks are reduced to zero by the end of that year. On the worksheets, the four prices were listed side by side and the prevailing price was circled. These circled prices were used in computing calendar-year average prices for use in the animal-units equation and in computing marketing-year average prices to ascertain the value of the crop.

To find the magnitude of the changes in loan stocks, the supply consistent with the loan rate at the end of the season was computed. If $\underline{K}$ is used to represent the parity index, $\underline{S}''$ is the supply that is consistent with the loan rate, and all variables are expressed in logarithms, the equation is as follows:

$$\Delta S'' = 0.00158 - 0.424\Delta K + 0.822\Delta A + 0.479\Delta L. \tag{43.1}$$

This is obtained by transposing equation (43) and substituting $\underline{\Delta K}$ for $\underline{\Delta C}$ and $\underline{\Delta S}''$ for $\underline{\Delta S}$. The supply consistent with the release price at the end of the season also is needed in those years in which the release price prevails from June through October. This can be obtained in a similar way.

If $\underline{S}''$, in million tons, is smaller than the nonloan supply available, the differences represent the quantity by which loan stocks will be increased. If the release price prevails from June through October, the supply consistent with the release price, in million tons, will be larger than the nonloan supply available, and the difference represents the quantity by which loan stocks will be reduced. As noted above, loan stocks will be reduced to zero in those years in which the computed price based on total supply is between the release price and the computed price based on nonloan supply only.

A similar approach could be used in analyzing the effects of specified loan programs for other crops.

A 10-region Model of the Feed-Livestock Economy 14/

Agricultural economists long have used statistical demand functions and estimating equations based on production and average prices for the country as a whole. Such aggregation disregards information concerning regional differences in prices, production, and consumption of the commodities involved. Many data exist on prices and production by regions, States, and in some cases even smaller geographic units. Marketing agencies and commodity specialists use this information in various, though generally informal, ways.

Agricultural outlook work has been a major undertaking of the Department of Agriculture for almost 30 years. The fruitfulness of this line of work probably can be increased if more accurate account is taken of regional differences in the situations that confront the different commodities. In addition, special problems center around changes in costs of transportation -- the major factor in geographical price differentials-- and their incidence upon consumers, producers, and transportation agencies. Incidence problems often are discussed piecemeal in terms of two or three shipping points competing for a particular market. But it is hardly possible by informal or qualitative methods to predict the effects of changes in a single freight rate, or in a limited number of freight rates, upon all producing areas and markets in the national economy. The simplified model of the livestock feed economy in this country which is presented here emphasizes the factors that determine prices and consumption of feed and the pattern of interregional trade in feed grains and other feed concentrates.

Data and Assumptions

Ten regions were defined in rough accordance with differences in the types of livestock and production of feed that are emphasized. The model includes a demand function for feed in each of the 10 regions and a structure of freight rates or transportation costs between all possible pairs of regions.

Demand functions.- Regional demand functions for feed were based upon a statistical analysis of demand for the country as a whole. This analysis indicated that a 1-percent increase in the supply of feed concentrates per grain-consuming animal unit was associated on the average with a 2-percent decrease in the average farm price of corn; also, that a 1-percent increase in prices of grain-consuming livestock and livestock products was associated on the average with about a 1-percent increase in the farm price of corn. This demand function was converted to arithmetic form in such a way that these price flexibilities were realized for approximately the actual values of United States farm prices of livestock and feed and production of feed concentrates in the 1949-50 feeding year. Specifically, this arithmetic demand function is as follows:

14/ A detailed discussion of this example is given in Fox (21).

$$P_c = 2.6873 - 3.50 \left(\frac{Q_f}{N_1}\right) + 0.0135\, P_1, \tag{47}$$

where P_c = United States average price of corn, in dollars per bushel;

Q_f = Total quantity of feed concentrates available for feeding, in million tons;

N = Number of grain-consuming animal units fed during the preceding season, in millions and

P_1 = Index of prices of grain-consuming livestock and livestock products, United States (average in 1950 = 100).

As prices of livestock show considerable geographical differences, an index of prices of grain-consuming livestock and livestock products was calculated for each of the 10 regions, using 1950 data. For convenience, the constant term in each regional demand curve is adjusted to include the specified value of livestock prices. This leaves a demand function in each region which includes as variables only prices of feed (corn) and consumption of feed per animal unit. Basic data on these variables are presented in table 4.

As an equilibrium condition, total consumption of feed in the country as a whole must equal the supply available for feeding. Hence, the demand functions must be further transformed in such a way that total consumption of feed in each region is made dependent upon the price of feed. These demand functions are given in table 5. A necessary condition for solution of this "geographical equilibrium" system is that the set of regional prices must cause the sum of the regional estimates of feed consumption to equal the total supply available for feeding in the country as a whole.

Freight rates.- Equilibrium prices also must be consistent with the structure of freight rates among regions. Freight rates are a special problem because the model mathematically implies that production and consumption of feed in each region are concentrated at a single point. In this analysis, freight rates were estimated on the basis of data on freight charges by mileage blocks, from the 1950 Interstate Commerce Commission waybill sample. The relationships used for estimating freight rates between regions are as follows:

$$X_{ij} = 5.6 + 0.0168\, M_{ij} \tag{48.1}$$

$$X_{ij} = 5.6 + 0.0224\, M_{ij} \tag{48.2}$$

$$X_{ij} = 5.6 + 0.0280\, M_{ij} \tag{48.3}$$

where X_{ij} = freight rate on corn from region i to j, in cents per bushel; and M_{ij}= distance in miles between centers of production of grain-consuming livestock in regions i and j.

Equation (48.1) was used when the ICC data indicated relatively low freight rates for corn in the railroad territories involved; equation (48.2) was used

Table 4.- Basic data on price and number of livestock and on supply of feed available, by regions, United States, specified years

Region	Price of livestock, 1950 1/	Supply of feed concentrates available, year beginning October 1949 2/	Grain-consuming animal units fed, year beginning October 1949	Feed supply per animal unit, year beginning October 1949
		1,000 tons	Thousands	Tons
Northeast	116.5	7,445	17,602	0.4230
Corn Belt	97.0	55,136	62,005	.8892
Lake	90.2	18,395	20,216	.9099
Northern Plains	91.8	18,430	16,893	1.0910
Appalachian	102.2	8,636	14,184	.6089
Southeast	111.1	4,571	8,981	.5090
Delta	101.0	3,176	6,364	.4991
Southern Plains	102.8	5,735	9,472	.6055
Mountain	99.6	3,296	3,638	.9060
Pacific	110.3	2,735	6,766	.4042
United States	100.0	127,555	166,121	.7678

1/ Index numbers (United States average in 1950 = 100). Regional average price of each major livestock product weighted by its appropriate contribution to total grain-consuming animal units in the region.
2/ Available for livestock feeding after eliminating non-feed uses and changes in end-of-year stocks.

Table 5.- Data relating to regional demand functions for feed concentrates under conditions applying approximately to the year beginning October 1949 1/

Region	Price per bushel of feed dependent:		Consumption of feed dependent:	
	Intercept (price if feed supply is zero)	Price in absence of interregional trade	Intercept (consumption if price is zero)	Change in consumption per dollar change in price
	Dollars	Dollars	Million tons	Million tons
Northeast	4.26	2.78	21.4229	-5.0289
Corn Belt	4.00	.88	70.8022	-17.7148
Lake	3.90	.72	22.5517	-5.7757
Northern Plains	3.93	.11	18.9544	-4.8263
Appalachian	4.07	1.94	16.4834	-4.0524
Southeast	4.19	2.41	10.7454	-2.5659
Delta	4.05	2.30	7.3665	-1.8182
Southern Plains	4.08	1.96	11.0292	-2.7062
Mountain	4.03	.86	4.1903	-1.0394
Pacific	4.18	2.76	8.0746	-1.9330
United States	4.04	1.35	191.6206	-47.4608

1/ All prices are per bushel of corn or an equal weight of other feeds. For simplicity, prices and freight rates for corn are taken as representative of average prices and freight rates for all feed grains and byproduct feeds.

when an intermediate level of freight rates was indicated; and equation (48.3) was used when the data suggested relatively high freight rates on corn, as well as for all paths over which actual shipments of corn were improbable or rare.

<u>Feed production and livestock numbers</u>.- Production of feed and numbers of livestock are used as predetermined variables in the explanation of feed prices, although the latter is strictly justified only during the early part of the feeding year. In this case the number of grain-consuming livestock in each region during the feeding year ended September 1950 was used. Actual data on 1949 production and carry-in stocks of feed were adjusted as follows: (1) Nonfeed utilization of feed grains in each region, as in food products or for seed, was substracted from the total supply of feed grains; (2) Exports of feed grains were allocated as nearly as possible to the regions of origin; (3) Carryover stocks of feed grains in each region at the end of the year were substracted from the beginning supply. The remaining (and by far the largest) quantity of feed grains, plus the total production of byproduct feeds, was assumed to be available for feeding domestic livestock during the 1949-50 feeding year. The resulting figures on both a total and a per animal-unit basis are shown in table 4.

It is evident that the model presented here is schematic and involves many simplifying assumptions. The effects of relaxing some of these are discussed later.

Equilibrium Solutions

The problem attacked in this section may be stated as follows: Given (1) the demand function for feed in each region and (2) the structure of transportation costs between regions, to find the equilibrium values of prices and consumption of feed in each region and the net quantities of feed shipped over each inter-regional path as a result of any specified set of regional values of feed production, livestock numbers, and livestock prices. This problem was solved for several different sets of initial conditions, illustrating the effects of changes in freight rates, changes in regional production of feed, and changes in the regional distribution of grain-consuming livestock.

A necessary requirement for equilibrium is that no region can increase its revenue by changing its pattern of consumption or shipment. If one region ships to another, the prices must differ by the cost of transportation between the two. If two surplus regions ship to the same deficit region, the difference between equilibrium prices in the surplus regions will equal the differences between their freight rates to the deficit region. Thus, an equilibrium solution for the whole system involves a precise structure of regional prices bound together by specific freight rates (except for regions which prove to be self-sufficient under the given conditions).

Theoretically, working out an equilibrium solution in a 10-region system could be laborious. Fortunately, intuition or judgment enables us to move rather directly toward the equilibrium arrangement.

<u>An Approximation to Actual Conditions in 1949-50</u>.- If the data on supplies of feed per animal unit in table 4 are examined, it appears that four regions probably will be surplus regions, or "sources", and that the other six probably will be deficit

regions, or "destinations." In general it appears likely that a given surplus region will ship to deficit regions in which it has a freight advantage over other, usually more distant, surplus regions.

The arrangement that emerges is simple and logical in most respects. The Corn Belt acts as a basing point for four deficit regions--the Northeast, Southeast, Appalachian, and Delta. Prices in these regions are equal to the Corn Belt price, plus freight. The Lake region supplies part of the requirements of the Northeast, and its price is equal to the price in the Northeast, minus freight. The northern Plains region serves as a basing point for the Pacific and southern Plains regions, and the Mountain region serves as an auxiliary source of supply for the Pacific region. The only additional item needed is a link connecting these two systems. In this case, it turns out that the Corn Belt, in equilibrium, would import some feed from the northern Plains and would reexport an equivalent quantity to other regions. Hence, the link is provided by the fact that the price in the Corn Belt is equal to that in the northern Plains, plus freight.

The technique used to obtain the equilibrium solution for this arrangement was as follows: (1) It was assumed that the arrangement just described did apply; (2) using the demand functions in table 5, the consumption of feed in each of the 10 regions was calculated on the basis of the assumed price in the Corn Belt (in this case $1.40 per bushel). But total consumption of feed at this price would not exhaust the available supply. As the assumed rigid structure of price differentials makes total consumption of feed in this country as a whole a linear function of the price in the Corn Belt, the adjustment in price that would be necessary to equate total consumption of feed with the available supply can be calculated immediately. This adjustment gives the equilibrium prices in each region, assuming the specified arrangement, and the equilibrium rates of feed consumption.

A comparison of the regional consumption estimates with the production estimates in table 4 gives at once the net imports or exports of each region. In conjunction with the assumed arrangement, the precise quantity of feed shipped over each interregional path can be determined. These figures are shown in table 6.

<u>A Forecasting Model</u>.- The usefulness of the 1949-50 equilibrium model for the purpose of forecasting changes in regional prices of feed and in interregional trade are now explored. Suppose that the 1949-50 model represents the actual situation in the year just past, and that numbers and prices of livestock are substantially unchanged during the current year. In July of the current year, well before harvesttime, published forecasts of production of feed grains are available. They indicate changes, region by region, from the previous year. If the new production forecasts are substituted into our model, we should obtain forecasts of the corresponding regional prices and consumption of, and net trade in, feeds. If sufficiently accurate, these forecasts should be of considerable value to farmers, marketing, transportation, and processing agencies, Government price-support and procurement agencies, and others.

But the July production forecasts are subject to error. They are estimates of the production of feed grain that would be expected if growing conditions during the remainder of the season (after July 1) were average. In some years,

Table 6.- Feed: Equilibrium solution for price, consumption, and net trade under conditions applying approximately to the year beginning October 1949 1/

Region	Price per bushel		Con-sumption	Pro-duction	Net trade	Net imports, by region of origin				
	Differ-ential from Corn Belt	Actual in equili-brium				Corn Belt	Lake	Northern Plains	Mountain	Total
	Dollars	Dollars	Million tons	Million tons	Million tons	Million tons	Million tons	Million tons	Million tons	Million tons
Northeast	0.216	1.521	13.78	7.44	-6.33	3.51	2.82	---	---	6.33
Corn Belt	.000	1.305	47.68	55.14	7.45	---	---	2/(1.80)	---	---
Lake	-.097	1.208	15.57	18.40	2.82	---	---	---	---	---
Northern Plains .	-.140	1.165	13.33	18.43	5.10	---	---	---	---	---
Appalachian	.158	1.463	10.56	8.64	-1.92	1.92	---	---	---	1.92
Southeast	.218	1.523	6.84	4.57	-2.27	2.27	---	---	---	2.27
Delta	.145	1.450	4.73	3.18	-1.55	1.55	---	---	---	1.55
Southern Plains .	.123	1.428	7.16	5.74	-1.43	---	---	1.43	---	1.43
Mountain	-.032	1.273	2.87	3.30	.43	---	---	---	---	---
Pacific	.266	1.571	5.04	2.74	-2.30	---	---	1.87	0.43	2.30
Total	---	---	127.56	127.56	0	3/9.25	2.82	5.10	.43	4/15.80

1/ Computations based on unrounded data.
2/ Under the assumed structure of freight rates, this quantity of feed is shipped to the Corn Belt and reshipped to deficit regions in addition to the 7.45 million tons classed as net exports from the Corn Belt.
3/ Includes 1.80 million tons received from northern Plains and reshipped to other regions.
4/ Excludes 1.80 million tons of imports into the Corn Belt offset by reexport.

weather may cause marked improvement or deterioration in crop prospects after July 1. Hence we are interested in the accuracy with which the pattern of equilibrium forecast in July anticipates the pattern that emerges after harvest.

For illustration, the pattern of equilibrium for "1949-50" was compared with patterns reflecting differences from 1949 production of feed grains by regions, based on (1) July 1947 production forecasts and (2) December 1947 production estimates. The July 1947 forecasts indicated a reduction of 18 million tons below reported production in 1949, and the December 1947 estimates indicated a further reduction of 6.5 million tons.

The regional price differentials were identical in all three cases. The July forecasts implied a price increase of 38 cents a bushel, and the December estimates indicated a further rise of 14 cents a bushel, corn equivalent.

But the net trade positions of the different regions showed striking changes. The July forecasts suggested that the Corn Belt would have a small surplus and that the northern Plains would ship much larger quantities than in 1949-50. The December estimates indicated that the Corn Belt would shift to a net deficit basis and that shipments from the northern Plains would be only moderately above those for 1949-50. The Appalachian region would shift to an almost self-sufficient basis. Under the assumed structure of freight rates, even after it had shifted to a deficit position, the Corn Belt would reship large quantities of corn received from the northern Plains. A reduction of freight rates out of the northern Plains by 1 to 7 cents a bushel on various freight paths would be required to eliminate this forwarding role of the Corn Belt.

The indicated volume of net trade declined much more sharply than did production of feed. Compared with 1949-50, the July forecasts implied a 14-percent drop in production of feed and a 29-percent drop in interregional shipments of feed. The December estimates showed a further drop of 6 percent in production and 21 percent in shipments. Total net trade evidently depends to a considerable extent upon the regional distribution of feed production in a given year as well as upon its general level.

Problems Involved in the Application of This Approach

In the livestock-feed economy, adjustments should be made for the following:

1. Allowance for short-run responses of livestock production to changes in supplies and prices of feed;
2. Allowance for longer run effects of increases in freight rates on production of livestock and feed;
3. Recognition that each livestock product has a geographical price-production-consumption equilibrium of its own;
4. Allowance for less-than perfect substitutability between different feed concentrates;
5. Allowance for possible differences in elasticities of demand for feed (a) by regions and (b) by classes of livestock;
6. Allowance for the existence of price-dependent demands for storage and export of feed;
7. Removal of the assumption that production of feed in a large region is concentrated at a single point; and
8. Allowance for seasonal elements in the equilibrium price structure.

For example, the South uses its own grain production early in the crop year, because of storage problems, and gets most of its "imports" of feed late in the crop year.

It remains to be seen whether a useful compromise can be effected between the complexities of a full description of reality and the need for a model sufficiently aggregated so that it can be manipulated and interpreted without undue expense. A partial answer might be found in terms of the nature of policy decisions, either private or Governmental, which might be made on the basis of an equilibrium analysis or on the basis of forecasts from such a model.

It should be emphasized that the most important and most complex of the commodity structures in our agriculture was purposely chosen. Livestock and feed products account for approximately 60 percent of total cash receipts from farm marketings in this country, and livestock products account for about an equal percentage of total retail expenditures for food. There may be many simpler and more appropriate applications of spatial equilibrium analysis to other farm products. Some unpublished analyses along these lines have been made by the Bureau of Agricultural Economics for oranges, potatoes, and celery. When the number of shipping points to be considered is relatively small, it may be feasible to use specific freight rates between given sources and destinations instead of relying upon some partly arbitrary procedure of estimation.

The 10-region model of the livestock-feed economy presented here is an exploratory venture. The potentialities of spatial equilibrium models for economic forecasting and outlook work are far from clear, both as to the degree of precision that may be obtained and the commodities for which such an approach may be helpful. Such models appear to have considerable value (1) for teaching principles to students of transportation and marketing and (2) for showing to policy officials the central features and tendencies of a geographically extended demand and supply structure which might be affected by their decisions. These expository uses probably are enough to justify considerably more experimentation with spatial equilibrium models based on actual data for specified commodity groups.

The Technique of Linear Programming 15/

Linear programming may be the key to many economic problems, including problems in farm management and in agricultural marketing. But most of the literature on the subject is abstract and mathematical. Few practical applications have yet been made and most of these concern military programs, so the results are classified. A partial exception is the recently published monograph by Charnes, Cooper, and Henderson. They point out that "linear programming is concerned with optimal planning of interdependent activities subject to a complex of restrictions" (5, p.8). An example which involves the highest profit combination for certain mixes of nuts is described. The first 10 pages, which give

15/ This section was prepared by Frederick V. Waugh, Director, Agricultural Economics Division, Agricultural Marketing Service. Further details regarding this example are given by Waugh (49).

an insight into the basic approach of the method, can be understood by research workers who are not versed in mathematics.

The general idea can be illustrated by the example of dairy feeds. Relevant information concerning the following is summarized in a convenient form in table 7.

1. Required amounts of total digestible nutrients (TDN), digestible protein (DP), calcium (Ca) and phosphorus (P) in a 24-percent dairy feed, according to Henry and Morrison (30).

2. Amounts of each of these nutrients in 100 pounds of several feeds, and

3. Average prices of these feeds for 1949-50, wholesale at Kansas City.

Table 7.- Proportion of requirements in a 24-percent dairy feed supplied by $1 worth of each feed

Feed	Nutritive factor				Weight 1/
	Total digestible nutrients	Digestible protein	Calcium	Phosphorous	
Corn	0.441	0.136	0.040	0.168	0.417
Oats	.375	.187	.170	.201	.397
Milo	.495	.203	.066	.206	.459
Bran	.423	.321	.312	.900	.467
Middlings	.436	.332	.176	.434	.410
Linseed meal	.272	.400	.511	.336	.262
Cottonseed meal	.268	.464	.268	.513	.281
Soybean meal	.286	.504	.335	.238	.270
Gluten	.395	.412	.879	.471	.385
Hominy	.448	.158	.412	.417	.394

1/ Proportion of 100 pounds that can be bought for $1.00, or the reciprocal of the price.

First, we shall concentrate upon the four nutritive factors.

Assumptions

The following assumptions are required:

1. A dairyman is satisfied with any combination of feeds that supplies all the nutritive requirements.

2. A combination is acceptable if it supplies more than the required amounts of some elements.

3. The listed feeds can be bought at exactly the quoted prices.

4. No other feeds are available.

5. A negative quantity of any feed cannot be bought.

The Problem

What feeds, and how much of each should he buy to satisfy all requirements at the minimum cost?

The Analysis

1. The least expensive source of a single nutrient.- Take total digestible nutrients, for example. The largest number in the first column in table 7 is 0.495, indicating that $1 worth of milo will buy 0.495 times the TDN requirement. The required amount of TDN could be obtained from 1/0.495 dollars = $2.02 worth of milo. This is the least expensive source of TDN.

Before going on, we note two points:

a. Some one feed will always supply a single nutrient at less expense than any possible combination of two or more feeds. In this case, straight milo will supply TDN at less expense than any combination, such as milo and bran, or corn and middlings.

b. If the least expensive source of one nutrient happens to supply or oversupply the requirements of another nutrient, then that single feed is the least expensive source of both nutrients together. In this example, $2.02 worth of milo does not completely supply any of the other requirements. But suppose that the price of middlings were reduced enough so that middlings became the least expensive source of TDN. This would increase all the figures in the middling row. Then if we bought enough middlings to supply the TDN requirements, we would also supply enough DP and P. Thus, if the price of middlings were reduced, it alone would become the least expensive source of TDN, DP, and P together. Any combination of two or more feeds that would supply these three nutrients would cost more.

2. The least expensive source of a pair of nutrients.- A graphic analysis of this problem can be found in Waugh (49, p. 304). It leads to three criteria:

a. A feasible combination is one that does not involve negative purchases of any feed. A nonfeasible way to supply TDN and DP would be to buy a large positive quantity of middlings and a large negative quantity of oats. But we rule out such solutions.

b. A necessary combination of two feeds is one that supplies two nutrients at less expense than either single feed. Thus, gluten alone supplies TDN and DP at less expense than any feasible combination of linseed meal and gluten, so that the latter would not be a necessary combination.

c. A graphic test is given to determine the least-cost, feasible, and necessary combination of two feeds meeting two requirements.

Suppose we have found the least-cost, feasible, and necessary combinations of two feeds supplying the requirements of TDN and DP. In this case it is about

35 cents worth of middlings, and about \$2.14 worth of gluten. Then:

d. No combination of three or more feeds will supply these two nutrients at less cost and

e. If the least-cost, feasible, and necessary combination supplying TDN and DP happens also to supply requirements of some other nutrient it is the least-cost, feasible, and necessary combination for all three nutrients.

In this case it happens that 35 cents worth of middlings and \$2.14 worth of bran supply enough Ca and P, so it is defined as a complete combination, as far as nutrients are concerned. But this combination weighs only 97 pounds. The feed manufacturer may have another requirement--a combination that would meet all four nutritive requirements and also weigh 100 pounds.

In this case he could find a combination of two feeds that would do this. He could just meet the requirements for DP and weight with 39 cents worth of bran and \$2.12 of gluten. (Use of this additional requirement would necessitate redrawing the diagram used in the graphic analysis.)

3. The general case.- The graphic criteria have been put into mathematical terms and extended to define the least-cost, feasible, necessary combination of $\underline{m}$ inputs to supply $\underline{n}$ requirements ($\underline{m} \leq \underline{n}$). One procedure involves the following steps. Find the single input that will supply any one requirement at least-cost; test to see if it happens to supply other requirements also. If not, find the combination of two inputs which is the least-cost, feasible, necessary way of meeting any pair of requirements. Again test whether it supplies other requirements in full. If not, proceed to 3, 4, or more inputs. At some point, it will be found that some combination of $\underline{p}$ inputs exactly meets $\underline{p}$ requirements and also happens to supply or oversupply all other requirements. The analysis is over. We have found the least-cost, feasible, necessary combination of inputs meeting all $\underline{n}$ requirements.

General Observations

The feed problem is a very simple case. But it illustrates a problem that is basic to most economic research. In general, our resources are limited. We search for ways of using existing resources to maximize real net income or something like it. Or we search for ways of meeting stated objectives at the least cost or something like it.

Research workers at the University of North Carolina are using linear programming to find the most profitable combination of enterprises on certain types of farms. Railway Age (1) discusses the minimization of cost in distributing empty freight cars. Presumably many problems in agricultural marketing essentially involve linear programming. There are two main difficulties: (1) Most economic problems are big--they involve many inputs and many outputs. Here the electronic computer could be the answer; (2) It is ordinarily difficult or even impossible to get an accurate statement of either input-output relations or requirements (objectives). In case of dairy feeds, we probably know fairly well the average nutritive content of most common feeds. But the requirements may not be fully stated.

DEMAND ANALYSES BASED ON PURCHASES BY INDIVIDUAL CONSUMERS

For many years, commodity analysts have used family-budget data applying to a single unit of time in measuring elasticity of income. Such data cannot be used to measure elasticity of price as, in general, only an average price for the period is published.

Since October 1949 the United States Department of Agriculture has published monthly and quarterly reports of household purchases of fresh citrus fruits, frozen and canned juices, and dried fruits (47,48). These reports are based on figures obtained under contract from the Market Research Corporation of America. The monthly reports relate to 4-week periods and show for each product the total quantity bought by householders, average price paid, percentage of all families that purchased, average number of purchases made during the period by buying families, and the average quantity bought per purchase, based on a sample of approximately 4,200 families blownup to a United States total basis. The quarterly reports relate to 13-week periods and show similar data by geographic regions and type of retail outlet. Several research projects that would make use of the data on which these reports are based are being conducted or considered by the Agricultural Marketing Service in cooperation with the University of California. Some of these studies are outlined below.

Regression Analyses Based on National Aggregate Data by Months

George M. Kuznets of the University of California has run a number of analyses, using both arithmetic and logarithmic data, from the published monthly reports on citrus. Considering the short period of time involved, some of these turn out surprisingly well. The "own-price" elasticities of demand for frozen orange concentrate and canned single-strength juice turn out to be highly significant. There appears to be significant competition between fresh oranges and frozen orange concentrate. The evidence for competition between fresh oranges and canned juice or between canned and frozen orange juice is not conclusive. The analyses for fresh oranges are weak. Panel data on prices of fresh oranges are given in cents per dozen, with no adjustment for size. Use of retail prices published by the Bureau of Labor Statistics gave better results, but the price elasticities were still barely significant. For the studies discussed below, data on the price of fresh oranges will be expressed in terms of cents per pound. Based on outside information the prices per pound of oranges appear to be relatively uniform regardless of size. Thus, statistical analyses based on them should be more reliable than those based on price per dozen.

Regression Analyses Based on Aggregates for Selected Groups of Families

A major purpose of these studies is to ascertain the relative elasticities of demand for fresh oranges and for frozen juice and the ultimate potential demand for frozen juice. Separate regression analyses will be made for panel families who regularly use frozen juice and for those who seldom or never use frozen juice.

Quasi-Descriptive Studies of Consumers with Specified Characteristics

Ogren (38) analyzed household purchases of citrus products by 500 urban families from November 1948 through October 1949. In addition to those aspects normally covered by a one-period budget study, he determined the frequency of purchase of the various items by income groups and size of family throughout the period of the study. For example, he found that less than half of the families who bought frozen orange juice concentrate on a trial basis became regular purchasers. But about half of those who became regular buyers of frozen juice discontinued purchases of fresh oranges or canned juice, or both. Total purchases of all citrus products increased when families became regular users of frozen juice. Frozen juice is now used much more widely that was the case during the time period covered by the Ogren study. A similar study covering a longer time period and possibly a larger sample is planned. Questions the study will attempt to answer include: (1) What is the typical pattern of consumption of frozen juice for a family from the time purchases first start until family consumption levels off on a flat plane, and what happens to consumption of other citrus products; and (2) What part of the rapid increase in consumption of frozen juice represents new consumers and what part represents increased consumption on the part of habitual users?

Regression Analyses Based on Purchases of Individual Families by Months

So far as we know, this type of study has not been attempted previously. If successful, it would open up a wealth of information to commodity analysts. The unit of observation would be purchases of a single family within a single month. The large number of observations would permit a high degree of stratification and the use of relatively complex equations. But individual consumers are known to behave erratically, so that low correlations would be expected from such a study. Whether the large number of observations would more than offset the low correlations could be determined only by experimentation. Several pilot studies are proposed which would serve to test the feasibility of the full study discussed here. Preparation of punch cards from data available in the files of the Market Research Corporation of America has been completed. These cards are to be used for the pilot studies and certain of the other studies discussed above, but they also contain all of the information needed for the full study as discussed here.

Strata

Separate analyses are proposed for rural and for urban consumers within each of 5 geographic regions. Small towns in which frozen juice is not readily available would be included in the rural classification. Separate analyses also are proposed for summer and winter, as California navel oranges differ essentially from California Valencias. Thus 20 separate strata would be considered (5 x 2 x 2).

Items to be Considered

(1) California-Arizona fresh oranges, (2) Florida, Texas, and unspecified oranges, (3) fresh grapefruit, (4) frozen orange juice, (5) canned orange juice, (6) canned grapefruit juice, (7) canned orange-grapefruit blend, (8) canned tomato juice, (9) canned pineapple juice, (10) juice equivalent of fresh lemons, canned lemon juice, shelf-pack lemonade base, and frozen lemonade base.

Analyses to be Run

Separate equations for each strata are proposed with consumption of each of the first 6 items in turn as dependent variables. Thus, 6 equations would be determined for each strata. The following would appear as independent variables in each equation: (1) Prices for each of the 10 commodity groups (intercorrelations do not appear to be unduly high), (2) income per family(based on annual data), (3) number in family, including boarders (based on annual data), (4) percentage of stores within the area (weighted by volume of business) that handle frozen orange juice, (5) month within the season (this would be a "0-1 variable" such that an additive factor for each month representing a composite of those factors normally associated with seasonal variation would be determined). Each equation would contain between 16 and 19 variables, the number depending on the number of months in the winter or summer season.

A problem in connection with such analyses would be finding the prices that confronted consumers for those items they did not buy in any given month. Prices that confronted nonbuying families could be estimated from the prices paid by buying families. As an alternative, data on retail prices from the Bureau of Labor Statistics might be used for all families.

The major purpose of these studies is to estimate elasticities and cross-elasticities of demand for a complex group of competing items. In this instance, the problem is complicated by the fact that certain of the competing items have been available for only a few yearsand that availability of and demand for these items has shifted upward and probably is still doing so. Thus conventional time-series analyses are inapplicable. Whether the proposed approach will provide a feasible solution is yet to be determined.

DEMAND ANALYSES BASED ON SALES FROM INDIVIDUAL STORES

Experiments in retail stores have been used for many years to learn the effect on sales of various innovations, such as prepackaging or packaging in various sizes or various types of containers, or to determine relative sales of varying qualities or degrees of ripeness of specified commodities. But only a few such studies have tried to ascertain the effect of changes in prices on sales. Studies along this line that have come to our attention are noted below.

Studies Relating to Specified Fruits

Apples

In a study by Cravens (8, 9), 19 retail food stores in Detroit were visited on each of 50 shopping days in the winter of 1950-51. Data were obtained on sales, prices, and display space for apples and other fruits. Grade inspections were made of the apples. Prices of apples and the size of the apple display were the chief factors affecting day-to-day changes in the volume of sales of apples relative to sales of all fruit. A 1-percent increase in the relative price of apples was associated with a 1-percent decrease in the relative quantity of apples sold. The appearance of the apples also affected sales.

The effect of price on sales was greater in medium-sized and large stores than in small stores.

Oranges

In a study by Godwin (24) sales of oranges associated with positive and negative deviations in price of 5, 10, and 15 cents a dozen from the established market price were analyzed. Thus 7 levels were tested. To permit the use of a latin square, the test was run in 7 large retail stores in a city in Kentucky. By analysis of variance, the effect of price on sales was measured after allowing for differences in the purchase patterns of customers among stores and seasonal changes in purchase rates. Weekly data were taken over a 7-week period in May-June 1952. Quality was held constant. Elasticity ranged between 0.7 and 1.5, with an average elasticity of 1.16, although a logarithmic (or constant elasticity) curve apparently would have fitted the data equally well. A coefficient of determination between sales and price of 0.94 was obtained.

Fresh Fruits

A pilot study covering 2 stores in Denver, Colo. was conducted in the summer of 1952 by the Technical Committee of the Western Regional Deciduous Fruit Project. Data were collected by days during a 10-week period on volume of sales, price, extent of display space, color, ripeness, condition, size, and method of display for each fresh fruit (excluding watermelons) sold at any time during the period. Daily sales of each item were divided by the value of total sales of all fruits to eliminate factors associated with overall store traffic. The latter 6 items were rated on a scale ranging from 1 to 9, with most ratings falling in the 3-7 range. Many color pictures were taken to ascertain whether a uniform rating system could be developed. A major aim of this study was to find the effect of individual factors of quality on sales. but it is realized that price effects should be allowed for before determining these relationships. A measure of cross-elasticities between the various fruits also is desired. It is doubtful whether all of these desires can be met even with a substantial increase in the number of stores included in the study. However, the analyses given below indicate that some useful information probably could be derived from studies of this type.

Comments Regarding Specific Analyses

Three items from the Denver study were analyzed on the assumption that they would illustrate certain basic principles. In view of the limited amount of data available from the pretest, the results should be considered only as illustrative of the kind of information that may be obtained from such studies and the types of analyses that might be desirable.

A check was first made to determine whether weekend sales for specific items differed significantly in their relation to sales of all fruit from such sales on weekdays. If, for example, consumers tended to buy melons or grapes, say, chiefly on weekends, significant differences would show up even after dividing daily sales of the item by sales of all fruit. Weekend averages of the "quantity divided by value" series tended to be higher than weekday averages for both cantaloups and Thompson seedless grapes, but the differences were not statistically significant when tested by analysis of variance. Graphic analyses for both fruits indicated that little was gained by analyzing these periods separately. In the regression analyses discussed below, no differentiation was made between weekdays and weekends.

In an attempt to determine which factors of quality appeared to be most important, scatter diagrams were made between the "quantity divided by value" series and price for cantaloups and Thompson seedless grapes. No single factor appeared to be consistently associated with positive or negative deviations from the apparent regression line on these charts. Hence, in the mathematical analyses discussed below, the choice of variables was made largely on a judgment basis. Certain variables were eliminated because they showed practically no variation over the period for which data were available.

In each of the analyses discussed below, the dependent variable (X_0) is the quantity of the item sold divided by the value of sales of all fruits, numbers in parentheses are the standard errors of the respective regression coefficients, and r_i refers to the highest order partial correlation between the dependent variable and the i th independent variable. A double asterisk indicates that these coefficients differ significantly from zero at the 5-percent probability level and have the correct sign; a single asterisk indicates that they do not differ significantly from zero at the 5-percent probability level but have the correct sign. In each case, reasons for the choice of the particular variables and the implications of the analysis are given immediately following the presentation of the results. No attempt was made to check these analyses to learn whether curvilinear relationships other than those used would give improved results. It is possible that improvements would have been indicated by such a check.

1. Cantaloups - data for 59 days for each of the 2 stores, analysis based on logarithms.

   X_1 - price of cantaloups, X_2 - ripeness of cantaloups, X_3 - maximum daily temperature in Denver.

   $$X_0 = 2.4 \underset{(0.12)}{-0.68} X_1 \underset{(0.17)}{-0.26} X_2 \underset{(0.43)}{-0.23} X_3 \qquad (49)$$

   $^{**} r_1 = -0.44$, $^{*}r_2 = -0.14$, $r_3 = -0.05$, $^{**}r = 0.49$.

For cantaloups it was believed that ripeness might be the main quality factor. Temperature was included on the assumption that melon and ice cream is a tempting dessert on extremely hot days.

The analysis yielded the wrong sign on temperature. Price appears to be the dominant factor of the variables included but ripeness also was of some importance.

2. Cantaloups - average data for 10 weeks for each of the 2 stores, analysis based on logarithms.

X_1 - price of cantaloups, X_2 - ripeness of cantaloups.

$$X_0 = 2.1 \underset{(0.14)}{-0.70} X_1 \underset{(0.44)}{+0.05} X_2 \tag{50}$$

$$^{**}r_1 = -0.76,\ r_2 = 0.03,\ ^{**}R = 0.77.$$

This was run primarily to compare results from an analysis based on weekly averages with one based on daily data. The percentage of variation in sales explained by price and by all factors was considerably higher for the analysis based on weekly data, reflecting the tendency of random influences to average out over the longer period. The regression coefficient on price was almost identical for the two analyses. Effects of quality factors (ripeness) were negligible in the weekly analysis, however, reflecting the small variability in this factor when weekly averages were used.

3. Cantaloups - average data for 10 weeks for each of the 2 stores, analysis based on logarithms.

   X_1 - price of cantaloups, X_2 - sales of cantaloups divided by value of all fruit sales in the preceding week.

$$X_0 = 2.1 \underset{(0.16)}{-0.70} X_1 \underset{(0.18)}{+0.08} X_2 \tag{51}$$

$$^{**}r_1 = -0.74,\ r_2 = 0.11,\ ^{**}R = 0.82.$$

This was run to learn whether the magnitude of sales in the preceding week had a significant effect on sales in the current week. Ripeness was omitted, as it had a nonsignificant effect in the preceding analysis. Sales in the preceding week had a negligible effect and the regression coefficient had the wrong sign. This factor could be expected to be more important for certain other fruits.

4. Thompson seedless grapes - data for 59 days for one store, analysis based on actual data.

   X_1 - price of Thompson seedless grapes, X_2 - quantity of other grapes sold divided by value of all fruit sales (zero values included), X_3 - average of ratings for ripeness, color, condition and size for Thompson seedless grapes.

$$X_0 = 0.99 - \underset{(0.006)}{0.02} X_1 - \underset{(0.20)}{0.24} X_2 - \underset{(0.01)}{0.001} X_3 \tag{52}$$

$$^{**}r_1 = -0.33,\ ^{*}r_2 = -0.15,\ ^{*}r_3 = -0.02,\ ^{**}R = 0.33$$

Elasticity of demand at the means equals - 0.55.

This is the first of two analyses designed to ascertain whether the effects on sales of individual competing commodities were measurable. Prices of

Thompson seedless and other grapes were almost identical over the period, so that the price for "other" grapes could not be used in the analysis. Quantity of other grapes (divided by value of sales of all fruit) was used instead. Zero values were included. As the logarithm of zero is not specified, the analysis was run in terms of arithmetic data. Use of such data in any case gives a more readily interpretable regression coefficient for X_2. The analysis indicates that sales of 1 pound of other grapes tended on the average to reduce sales of Thompson seedless grapes by one-fourth of a pound. Price of Thompson seedless grapes was the chief factor affecting sales, quantity of other grapes sold had some effect, and the effect of the indicated variations of quality of Thompson seedless grapes was negligible. As no one factor of quality was believed to be of major importance, an average of ripeness, color, condition, and size was used.

5. Apricots - data for 28 days for one store, analysis based on logarithms.

   X_1 - price of apricots, X_2 - price of Elberta peaches, X_3 - average of ratings for ripeness, color, condition, and size for apricots.

$$X_0 = 1.85 - \underset{(0.72)}{0.31}\, X_1 + \underset{(2.05)}{0.90}\, X_2 - \underset{(0.57)}{1.00}\, X_3 \qquad (53)$$

$$*r_1 = -0.09, \; *r_2 = 0.09, \; *r_3 = -\,0.33, \; R = 0.45.$$

This was designed to find whether Elberta peaches are a major competitor of apricots. Prices of the two fruits were not highly correlated over this period, so that each could be used in the analysis. As in the case of grapes, an average of ripeness, color, condition, and size was included for apricots. An average of these showed considerable variation for apricots (partly because of the high correlation between the several factors of quality), but it would have shown little variation for Elberta peaches. The analysis indicates that a 1-percent change in the price of peaches had about 3 times as much effect on sales of apricots as did a 1-percent change in the price of apricots although the standard errors of these coefficients are sufficiently large so that these differences would not be statistically significant. All of the signs were correct in this analysis but, partly because of the small number of observations available, none of the coefficients differed significantly from zero. Of the three factors, quality appeared to be most important.

Comments Regarding Experimental Design 16/

Two criticisms of the experimental approach to the measurement of consumer demand, similar to that used by Godwin, should receive attention inasmuch as they have been serious deterrents to its use. A properly designed experiment measures the importance of each.

16/ This section is based on a talk by Glenn L. Burrows, Mathematical Statistician, Agricultural Marketing Service.

First, with only a sample of stores cooperating in an area, there is no assurance that, with increased prices in test stores, consumers will not shift their purchases to other stores in the same area. The validity of the price-quantity relationship suffers in direct proportion to the extent to which a shift takes place, and the possibility of such shifts places practical limits upon the range of price differentials for which valid results can be obtained. But fortunately with increasing price differentials, any such shift would manifest itself in a decline in the number of buyers, a result that can be verified and tested. Data presented by Godwin (24) on the number of customers associated with each price differential indicated that this was not a serious problem in his study. An additional check on this point could be made by maintaining a simple count, probably during only a sample of business hours, of purchasers of the item per 100 customers in neighboring nontest stores. This problem does not appear to be as serious as many fear, although conclusive evidence will have to await further experimentation along these lines.

The second criticism is harder to deal with. It arises from the sluggish response of purchases of consumers to changes in prices. If the test period is too long, some of the advantages of this technique are lost because of adjustments in incomes, prices of competing commodities, and similar factors during the test period. This problem has been at least partly solved with the innovation of experimental designs incorporating provisions for the measurement of carryover effects. This approach was used by Henderson (29) in a retail store study but, unfortunately, the explanation of the carryover feature is inadequate. The method was described in detail by Cochran, Autrey, and Cannon (6), although their explanation is not always easy to follow.

Even in cases in which the carryover effects are not important, the changeover, or rotational-type, experiments, of which the Godwin analysis is an example, yield more precise measures of the effects being studied. Store-to-store variability does not enter into experimental error as it would with the usual comparative type of experiment.

The following quotation from Kempthorne (32), p. 7) indicates the area within which designed experiments may be most useful. "The real distinction between two of the applications of statistics, the design of experiments and sample surveys [of which analyses based on time series, in a sense, form a part], is that, in the design of experiments applied to a problem, the populations that are studied are formed by the experimenter in a specified way, whereas, in a sample survey dealing with the same problem, the population under study has arisen from a set of forces, the relation of which to the forces under consideration is unknown...[A survey] can demonstrate the existence of associations between characteristics in the population, but...the existence of an association between attributes *X* and *Y* in the population in no way suggest that attribute *X* can be altered to a specified value by altering attribute *Y* in a particular way. In an experiment we determine whether altering attribute *X* has an effect on attribute *Y*, and this is the knowledge that is necessary for any action program...Survey work can be very useful in cases in which a deductively formulated theory exists and it is desired to estimate some parameters in the theory. It is,however, difficult to visualize how a theory can be started without some experimentation on which to base the original abstractions."

Analysis of demand is concerned to a considerable degree with the estimation of parameters, such as coefficients of elasticity. But our knowledge of behavior patterns for individual consumers is sufficiently limited that much probably can be learned by the further use of designed experiments in retail stores.

LITERATURE CITED

(1) Anonymous
1953. Operations Research in Distributing Empty Cars. Railway Age. 134:73-74.

(2) Armore, Sidney J.
1953. The Demand and Price Structure for Food Fats and Oils. U. S. Dept. Agr. Tech. Bul. 1068, 69 pp., illus.

(3) __________, and Burtis, Edgar L.
1950. Factors Affecting Consumption of Fats and Oils Other than Butter, in the United States, U. S. Bur. Agr. Econ., Agric. Econ. Research. 2:1-9.

(4) Atkinson, L. Jay
1950. The Demand for Consumers' Durable Goods. U. S. Dept. Commerce, Survey of Current Business. 30:5-10, illus.

(5) Charnes, A., Cooper, W . W., and Henderson, A.
1953. An Introduction to Linear Programming. 74 pp., illus. New York.

(6) Cochran, W. E., Autrey, K. M., and Cannon, C. Y.
1941. A Double Change-over Design for Dairy Cattle Feeding Experiments. Jour. Dairy Sci. 24:937-951.

(7) Cochran, D., and Orcutt, G. H.
1949. Application of Least-square Regression to Relationships Containing Autocorrelated Error Terms. Amer. Statis. Assoc. Jour. 44:32-61 illus.

(8) Cravins, M. E., Jr.
1952. Selling Michigan Apples. Mich. Agr. Expt. Sta. Special Bul. 382, 30 pp., illus.

(9) __________
1952. Studies in Midwest Apple Marketing, I. Retail Merchandising of Apples. Mich. Agr. Expt. Sta. Special Bul. 378, 75 pp., illus.

(10) Ezekiel, Mordacai
1933. Some Considerations on the Analysis of Prices of Competing or Substitute Commodities. Econometrica. 1:172-180, illus.

(11) __________
1938. The Cobweb Theorem. Quart. Jour. Econ. 52:255-280, illus.

(12) __________
1941. Methods of Correlation Analysis. Ed. 2., 531 pp., illus. New York.

(13) Ferber, Robert
1949. Statistical Techniques in Market Research. 542 pp., illus. New York.

(14) Foote, Richard J.
1952. Calculation of Partial and Multiple Regression and Correlation Coefficients - 3 to 5 Variables. U. S. Bur. Agr. Econ. 15 pp. (Processed.)

(15) __________
1953. Statistical Analyses Relating to the Feed-livestock Economy. U. S. Dept. Agr. Tech. Bul. 1070, 41 pp., illus.

(16) ____________
1953. A Four-equation Model of the Feed-livestock Economy and Its Endogenous Mechanism. Jour. Farm Econ. 35:44-61, illus.

(17) ____________
1953. A Four-equation Model of the Feed-livestock Economy and Its Endogenous Mechanism - A Correction. Jour. Farm Econ., 35:613-615.

(18) Fox, Karl A.
1951. Factors Affecting Farm Income, Farm Prices, and Food Consumption, U. S. Bur. Agr. Econ., Agr. Econ. Research. 3:65-82.

(19) ____________
1951. The Measurement of Price Support Cost. Jour. Farm Econ. 33:470-484.

(20) ____________
1953. The Analysis of Demand for Farm Products. U. S. Dept. Agr. Tech. Bul. 1081, 90 pp., illus.

(21) ____________
1953. A Spatial-equilibrium Model of the Livestock-feed Economy in the United States, Econometrica. 21:547-566.

(22) ____________
1954. Structural Analysis and the Measurement of Demand for Farm Products. Review Econ. and Statis., v. 36 (In press.)

(23) Girshick, Meyer A., and Haavelmo, Trygve
1947. A Statistical Analysis of the Demand for Food: Examples of Simultaneous Estimation of Structural Equations. Econometrica. 15:79-111.

(24) Godwin, Marshall R.
1952. Customer Response to Varying Prices for Florida Oranges. Fla. Agr. Expt. Sta. Bul. 508, 24 pp., illus.

(25) Haavelmo, Trygve
1943. The Statistical Implications of a System of Simultaneous Equations. Econometrica. 11:1-12.

(26) ____________
1944. The Probability Approach in Econometrics. Econometrica, v. 12, Supplement, 118 pp., illus.

(27) Heady, Earl O., and Olson, Russell O.
1952. Substitution Relationships, Resource Requirements and Income Variability in the Utilization of Forage Crops. Iowa Agr. Expt. Sta. Research Bul. 390, pp. 864-938, illus.

(28) ____________, Woodworth, R., Catron, O., and Ashton, G.
1953. Productivity and Substitution Coefficients in Pork Output. Jour. Farm Econ. 35:341-354, illus.

(29) Henderson, P. L.
1952. Application of the Double Change-over Design to Measure Carryover Effects of Treatments in Controlled Experiments. Cornell University, Methods of Research in Marketing Paper 3, 86 pp. (Processed.)

(30) Henry, W. A., and Morrison, F. B.
1948. Feeds and Feeding, A Handbook for the Student and Stockman, rewritten by Frank B. Morrison, Ed. 21, unabridged, 1,307 pp., illus. Ithaca, N. Y.

(31) Hermie, Albert M.
1951. Prices of Apparel Wool. U. S. Dept. Agr. Tech. Bul. 1041, 48 pp., illus.

(32) Kempthorne, Oscar
1952. The Design and Analysis of Experiments, 631 pp., Illus. New York.

(33) Koopmans, Tjalling
1945. Statistical Estimation of Simultaneous Economic Relationships. Amer. Stat. Assoc. Jour. 40:448-466.

(34) Kuznets, George M., and Klein, Lawrence R.
1943. A Statistical Analysis of the Domestic Demand for Lemons, 1921-41. Giannini Found. Agr. Econ. Rept. 84, 112 pp., illus. (Processed.)

(35) 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 Monogr. 14, 324 pp., illus.

(36) Meinken, Kenneth W.
1953. The Demand and Price Structure for Oats, Barley and Sorghum Grains. U. S. Dept. Agr. Tech. Bul. 1080, 103 pp., illus.

(37) Morrissett, Irving
1953. Some Recent Uses of Elasticities of Substitution - a Survey. Econometrica. 21:41-62, illus.

(38) Ogren, Kenneth E.
1951. Analysis of Household Purchases of Citrus Products by 500 Urban Families, November 1948-October 1949, U. S. Bur. Agr. Econ. 51 pp., illus. (Processed.)

(39) Pearson, Frank A., and Vial, Edmond E.
1946. Prices of Dairy Products and Other Livestock Products. 154 pp., illus. Ithaca, N. Y.

(40) Peters, Charles C., and Van Voorhis, Walter R.
1940. Statistical Procedures and Their Mathematical Bases. 516 pp., illus. New York and London.

(41) Prest, A. R.
1949. Some Experiments in Demand Analysis. Review Econ. and Statis. 31:33-49.

(42) Schultz, Henry
1938. The Theory and Measurement of Demand. 817 pp., illus. Chicago.

(43) Snedecor, George W.
1940. Statistical Methods. Ed. 3. 422 pp., illus. Ames, Iowa.

(44) Stone, Richard
1945. The Analysis of Market Demand. Royal Statis. Soc. Jour. 108:286-382, illus.

(45) Thomsen, Frederick Lundy, and Foote, Richard Jay
1952. Agricultural Prices. Ed. 2. 509 pp., illus. New York.

(46) Tinbergen, J.
1951. Econometrics. Transl. from the Dutch by H. Ryken Van Olst. 258 pp., illus. Philadelphia.

(47) United States Agricultural Marketing Service.
1949-53. Consumer Purchases of Fruits and Juices, Washington, D. C. (Processed.)

(48) ______________
1949-53. Consumer Purchases of Fruits and Juices by Regions and Retail Outlets. Washington, D. C. (Processed.)

(49) Waugh, Frederick V.
1951. The Minimum-cost Dairy Feed. Jour. Farm Econ. 33:299-310, illus.

(50) ______________, and Been, Richard O.
1939. Some Observations About the Validity of Multiple Regressions. Statis. Jour. 2:6-14, illus.

(51) Wells, C. F.
1941. The Incidence of Tariffs on Competing Products. U. S. Bur. Agr. Econ., Stat. and Agr. Ser. 2, 28 pp., illus. (Processed.)

(52) Yule, G. Udney, and Kendall, M. G.
1937. An Introduction to the Theory of Statistics. Ed. 11. 570 pp., illus. London.

APPENDIX

Note 1. Proof That if X Equals Y_1 Plus Y_2 and Simple Regressions Are Run Between X and Y_1 and Y_2, Respectively, the Sum of the Slopes Will Equal 1 and the Sum of the Constant Values Will Equal 0

1. The following two simple regression equations are fitted:

$$Y_1 = a_1 + b_1 X \tag{54}$$

$$Y_2 = a_2 + b_2 X. \tag{55}$$

Let the small letters, $\underline{x}$, $\underline{y}_1$ and $\underline{y}_2$ represent deviations of $\underline{X}$, $\underline{Y}_1$, and $\underline{Y}_2$ from their respective means. From basic regression formulas,

$$b_{y_1x} = \frac{\Sigma y_1 x}{\Sigma x^2} \tag{56}$$

$$b_{y_2x} = \frac{\Sigma y_2 x}{\Sigma x^2}. \tag{57}$$

Also, the regression of $\underline{x}$ on itself is

$$b_{xx} = \frac{\Sigma xx}{\Sigma x^2} = 1 . \tag{58}$$

Since $x = y_1 + y_2$, equation (58) can be rewritten as

$$b_{xx} = \frac{\Sigma x(y_1 + y_2)}{\Sigma x^2} = \frac{\Sigma y_1 x}{\Sigma x^2} + \frac{\Sigma y_2 x}{\Sigma x^2} . \tag{58.1}$$

But by equations (56) and (57), the two right-hand terms are equal to b_{y_1x} and b_{y_2x} respectively. Hence.

$$b_{xx} = b_{y_1x} + b_{y_2x} = 1. \tag{58.2}$$

The same demonstration can be extended to cases in which x is the sum of any number of variables.

2. In equation (58.1) above, it was assumed without proof that $x = y_1 + y_2$. This follows from the fact the $X = Y_1 + Y_2$ for every observation. If the observations are summed and then divided by their number to obtain arithmetic means,

$$\frac{\Sigma X}{n} = \frac{\Sigma Y_1}{n} + \frac{\Sigma Y_2}{n}, \text{ or} \tag{59}$$

$$\bar{X} = \bar{Y}_1 + \bar{Y}_2, \tag{60}$$

in which the bars denote arithmetic means. For any given observation $X = Y_1 + Y_2$. X may be subtracted from the left-hand term and $(\bar{Y}_1 + \bar{Y}_2)$ from the right-hand term without destroying the equality. Hence,

$$X - \bar{X} = (Y_1 - \bar{Y}_1) + (Y_2 - \bar{Y}_2) \tag{61}$$

which, by definition, is (62)

$$x = y_1 + y_2.$$

3. From basic regression formulas,

$$a_{y_1x} = \bar{Y}_1 - b_{y_1x}\,\bar{X} \tag{63}$$

$$a_{y_2x} = \bar{Y}_2 - b_{y_2x}\,\bar{X} \tag{64}$$

$$a_{xx} = \bar{X} - b_{xx}\,\bar{X} = 0, \text{ as } b_{xx} = 1\ . \tag{65}$$

Adding equations (63) and (64) gives

$$a_{y_1x} + a_{y_2x} = (\bar{Y}_1 + \bar{Y}_2) - (b_{y_1x} + b_{y_2x})\,\bar{X}\ . \tag{66}$$

But this equation also may be written as

$$a_{y_1x} + a_{y_2x} = \bar{X} - b_{xx}\bar{X} = 0. \tag{66.1}$$

Hence, the sume of the a values is equal to zero. This demonstration also can be extended to any number of subdivisions of x.

Note 2. Demand Relationships in Utility Theory

A theoretical criterion for determining the nature of the demand relationship between two commodities is given by $\frac{\partial^2 U}{\partial x_1 \partial x_2} \lesseqgtr 0$, where U, total utility, is an increasing function of the quantities consumed (x_1 and x_2) of each of the two commodities. This statement implies that $\frac{\partial U}{\partial x_1} > 0$ (and also that $\frac{\partial U}{\partial x_2} > 0$). The expression $\frac{\partial^2 U}{\partial x_1 \partial x_2}$ shows the effect upon the rate of increase in utility per unit of x_1 of consuming additional units of x_2. If an additional unit of x_2 reduces the per unit contribution of x_1 to consumer satisfaction, the two commodities are competing, and $\frac{\partial^2 U}{\partial x_1 \partial x_2} < 0$. If an added unit of x_2 has no effect on the per unit contribution of x_1 to consumer satisfaction, the commodities are independent in demand, and $\frac{\partial^2 U}{\partial x_1 \partial x_2} = 0$. Finally, if an added unit of x_2 increases the per unit contribution of x_1, the two commodities are completing, and $\frac{\partial^2 U}{\partial x_1 \partial x_2} > 0$.

This criterion is in terms of the utility surface of an individual consumer and is not subject to statistical measurement without severe restrictions.

Footnote 6, page 13, said that, according to utility theory, the coefficients b_{12} and b_{21} of equations (3) and (4) respectively should be equal in all cases if consumers are rational. This result follows from the basic criterion stated above. If utility is assumed to be measurable, we have $\frac{\partial U}{\partial x_1} = p_1$ and $\frac{\partial U}{\partial x_2} = p_2$. As the order of differentiation is immaterial, we have also that $\frac{\partial^2 U}{\partial x_1 \partial x_2} = \frac{\partial^2 U}{\partial x_2 \partial x_1}$. But this is equivalent (on the measurability assumption) to $\frac{\partial p_1}{\partial x_2} = \frac{\partial p_2}{\partial x_1}$. The left-hand term of this equality is equivalent to b_{12} and the right-hand term to be b_{21}. This is "Hotelling's condition," presented by Schultz (42).

Note 3. Some Demand Interrelationships in the Meat and Poultry Group 17/

The notation is symbolic. For example: $p_m^{(r)}$ stands for the price of all meat at retail; q is quantity (per capita consumption except in equations (6)

17/ All variables expressed as first differences of logarithms of annual observations, 1922-41, except for turkeys for which the period is 1929-41.

and (6w)); and y is disposable personal income per capita. The subscripts are m = all meat, p = pork, b = beef, l = lamb, c = chicken, and t = turkey. The subscript pvl stands for pork, veal and lamb, and pbv stands for pork, beef, and veal. Numbers in parentheses are standard errors of the regression coefficients.

Meats

$$p_m^{(r)} = \underset{(0.08)}{-1.50}\, q_m + \underset{(0.03)}{0.87}\, y \tag{67}$$

$$R^2 = 0.98$$

$$p_p^{(r)} = \underset{(0.13)}{-0.02}\, q_{bvl} \underset{(0.07)}{-1.16}\, q_p + \underset{(0.06)}{0.90}\, y \tag{68}$$

$$R^2 = 0.97$$

$$p_b^{(r)} = \underset{(0.11)}{-1.08}\, q_b + \underset{(0.06)}{0.88}\, y \underset{(0.07)}{-0.42}\, q_{pvl} \tag{69}$$

$$R^2 = 0.95$$

$$p_l^{(r)} = -0.01 \underset{(0.14)}{-0.50}\, q_l + \underset{(0.06)}{0.78}\, y \underset{(0.14)}{-0.65}\, q_{pbv} \tag{70}$$

$$R^2 = 0.94$$

Chickens

$$p_c^{(r)} = \underset{(0.18)}{-0.75}\, q_c + \underset{(0.09)}{0.76}\, y \underset{(0.16)}{-0.42}\, q_m \tag{71}$$

$$R^2 = 0.86$$

$$p_c^{(r)} = -0.01 -0.87\, q_c + 0.82\, y -0.43\, q_m \tag{71.1}$$

$R^2 = 1.00$ (Forced result, assuming perfect correlation of random errors in all variables.)

Turkeys

$$p_t^{(f)} = 0.02 - \underset{(0.25)}{1.21}\, q_t + \underset{(0.20)}{1.06}\, y - \underset{(0.48)}{0.97}\, q_c \tag{72}$$

$$R^2 = 0.90$$

$$p_t^{(f)} = 0.02 - \underset{(0.19)}{1.09}\, q_t + \underset{(0.11)}{1.11}\, y - \underset{(0.28)}{0.96}\, q_c - \underset{(0.11)}{0.23}\, q_p \tag{72.1}$$

$R^2 = 0.97$ (Special analysis allowing for effects of estimated errors in all variables, but without forcing perfect correlation.)

Comments

According to these equations, retail prices of beef and lamb were substantially affected by supplies of other meats--particularly pork. But supplies of beef, veal, and lamb had no statistically significant influence on the retail price of pork.

The retail price of chickens was significantly affected by the supply of red meats as well as by the supply of chicken.

No retail price series is available in the case of turkeys. The unadjusted analysis (equation (72)) suggests that prices of turkeys are influenced by the supply of chicken. Supplies of pork in October-December may also influence turkey prices but this is not so clearly established. (See equation (72.1).)

Supplies of poultry probably have some reverse effect upon prices of meat but no attempt was made to measure it. This effect probably cannot be measured from historical data, mainly because the absolute importance of poultry meat, in terms of pounds per capita, was only a fifth or a sixth that of meat. The effect of supplies of poultry upon prices of meat is not likely to stand out over the effects of errors in the data and variables omitted from the analysis. Nevertheless, it is clear that poultry meat can and does serve to some extent as a substitute for red meat when supplies of the latter are short.

Note 4. An Approach To The Measurement of Competition Between Apples From Different Producing Areas 18/

The variables are season average farm prices (p), total production (q) and total disposable income (y).

p_w = price in Washington State

q_w = production in 11 western States

p_n = price in New York State

q_e = production in all other than the 11 western States.

Price and production variables do not refer to strictly comparable areas, so the equations should be taken as illustrative only.

$$p_t^{(f)} = -0.02 - \underset{(0.04)}{0.79}\, q_t + \underset{(0.12)}{1.04}\, y \qquad (73)$$

$$R^2 = 0.96$$

18/ All variables are in terms of first differences of logarithms of crop-year data, 1921-41.

$$p_w^{(f)} = -\,0.03 - \underset{(0.31)}{1.19}\, q_w + \underset{(0.48)}{1.16}y \tag{74}$$

$$R^2 = 0.55$$

$$p_w^{(f)} = -\,0.02 - \underset{(0.20)}{0.84}q_w - \underset{(0.07)}{0.38}q_e + \underset{(0.30)}{1.33}y \tag{74.1}$$

$$R^2 = 0.83$$

$$p_n^{(f)} = -0.01 - \underset{(0.08)}{0.68}q_e + \underset{(0.33)}{1.23}y \tag{75}$$

$$R^2 = 0.81$$

$$p_n^{(f)} = -\,0.02 - \underset{(0.24)}{0.21}q_w - \underset{(0.09)}{0.65}q_e + \underset{(0.36)}{1.23}y \tag{75.1}$$

$$R^2 = 0.82$$

Equation (73) implies a slightly elastic demand for apples during the 1921-41 period.

Equations (74) and (74.1) indicate that the price of apples in Washington State is strongly influenced by the production of apples east of the Rocky Mountains as well as in Washington and other western States. Production in the east averaged larger than production in the west. If apples from the two areas were perfect substitutes bushel for bushel, q_e (in the logarithmic form of equation (74.1)) should have a larger coefficient than q_w. The smaller coefficient obtained reflects significant differences in market demand.

Equations (75) and (75.1) imply that production of western apples had little effect on the price of apples in New York State. This may be due in considerable part to the fact that western production, much of which is on irrigated land, was more stable than eastern production. The variance of $\underline{q_e}$ during 1921-41 was almost 8 times as large as that of $\underline{q_w}$.

A more refined analysis would consider (1) a larger number of producing areas, with price and production variables properly associated; (2) the seasonal pattern of marketings from each area; (3) competition between varieties in specific markets; and (4) the fact that relative prices even of identical commodities in widely separated areas would be influenced by changes in relative production in the two areas even though prices in consuming centers were the same.

Note 5. Suggestions for Exploratory Analyses of Competition between Grades of Beef Cattle and between Grades and Cuts of Beef

Correlations between retail prices of different cuts of beef for 1922-41 appear to indicate significantly different elasticities of demand. Some simple approaches to determine the extent to which this is true follow:

1. Assume the same percentage variation from year to year in quantities sold of two retail cuts of beef, that is, that they represent constant percentages of the carcass. Calculate the standard deviation of logarithmic price changes in each cut.

2. As a first approximation, the elasticity of demand for each cut may be written as

$$\eta_1 \cong \frac{\sigma_{q_1}}{\sigma_{p_1}} \tag{76}$$

$$\eta_2 \cong \frac{\sigma_{q_2}}{\sigma_{p_2}}, \tag{77}$$

where the q's and p's are the quantity sold and price at retail, respectively.

Even if σ_q is unknown, the ratio between η_1 and η_2 could be estimated from the relationship:

$$\frac{\eta_1}{\eta_2} = \frac{\sigma_{p_2}}{\sigma_{p_1}}. \tag{78}$$

3. In the case of round steak and rib roast, the following regression equation (based on logarithmic first differences) is obtained:

$$p^{(r)}_{\text{rib roast}} = a \underset{(0.031)}{+1.187} p^{(r)}_{\text{round steak}}. \tag{79}$$

$$R^2 = 0.988.$$

As b differs significantly from 1, it appears likely that the elasticities of demand also differ significantly.

4. However, the different price behavior of the two cuts may be partly due to different income elasticities, or to different competitive relationships with pork, veal, and lamb. An equation explaining the average retail price of all cuts of beef is as follows:

$$p_b^{(r)} = a -1.06\, q_b + 0.88y -0.52\, q_{pvl} \tag{80}$$

$$R^2_{1.234} = 0.95.$$

The price of each individual cut of beef might be expressed as a function of the same three independent variables and a comparison made of the various coefficients as evidences of differential price flexibilities.

5. Competition between grades of beef (wholesale carcass prices) and grades of cattle could be approached statistically by various means. One

difficulty is lack of adequate data for the total supply of slaughter cattle and beef by grades.

One approach is simply to compare relative prices of different grades of (say) beef steers at Chicago with relative numbers slaughtered. A more formal approach would be to express the price of each grade as a function of slaughter of the same and other grades and of a general demand factor such as disposable income. A preliminary inspection of data at Chicago on prices and numbers of steers by grades suggests that the degree of competition between grades is limited on a year-to-year basis. This would be consistent with a fairly rigid stratification of the consumer market on a grade basis. Leading hotels and restaurants may use Prime grade as a matter of policy; some major retail chain stores are said to sell only Choice grade beef;and so on. Some price competition doubtless occurs between adjacent grades. But if sufficiently accurate data on quantities sold were available, even the adjacent grades might appear to be imperfect substitutes and competition between (say) Prime and Medium grades might be negligible.

(6) The available data may not be adequate to give solidity to a complete demand structure for beef differentiated by grades and cuts, although they might provide some useful insights. To the extent that price analyses based on time series were differentiated by cuts and grades, comparisons could be made with the family budget data for individual cuts. Also, inferences might be made as to the demand for individual grades on the basis of differences in average prices paid for given cuts by different family income groups.

U. S. GOVERNMENT PRINTING OFFICE: O—1954