AGRICULTURAL ECONOMICS RESEARCH

A Journal of Economic and Statistical Research in the
Bureau of Agricultural Economics and Cooperating Agencies

Volume II | JANUARY 1950 | Number 1

Factors Affecting Consumption of Fats and Oils, Other Than Butter, in the United States[1]

By Sidney J. Armore and Edgar L. Burtis

Consumption forecasts for fats and oils are essential in appraising the outlook for oilseeds and guiding farmers' acreage plans. They are also valuable to businessmen who produce, handle, or use fats and oils or who deal in oilseeds. Recognizing the extent of interchangeability among fats and oils, this paper presents an analysis of the factors affecting total consumption of fats and oils, except butter. An equation for forecasting total fat and oil consumption, per person, and techniques which should prove useful as a guide in the analysis of demand for particular groups of fats and oils are likewise presented.

CONSUMPTION of fats and oils other than butter in the United States has been closely associated with industrial production and trend. More than 97 percent of the variance in annual consumption, per person, from 1922 to 1940 has been accounted for in a statistical analysis of annual data that represent these factors and fats and oils prices. Prices of fats and oils during that period, however, apparently had little effect on total consumption.

Factors Used in the Analysis

The specific variables used in the statistical analysis were as follows:

Description	*Designation*
Domestic disappearance of all fats and oils, except butter, divided by the population of the United States on July 1 (pounds)	X_1
Bureau of Agricultural Economics index of wholesale prices of 26 major fats and oils (excluding butter) deflated by the Bureau of Labor Statistics wholesale price index for all commodities (1935–39=100)	X_2
Federal Reserve Board index of industrial production divided by the population of the United States on July 1 (1935–39=100)	X_3
Year (1922=1)	X_4

Actual consumption of fats and oils cannot be computed accurately because data on dealers' and consumers' inventories are lacking. In most years, however, a good working approximation to consumption is afforded by the disappearance of fats and oils from factories and warehouses. This disappearance is computed from data on production, factory and warehouse stocks, and foreign trade, including shipments to United States Territories.

[1] This article reports results of the first study undertaken on fats and oils on the project, "Factors Affecting Production, Prices, and Uses of Fats, Oils, and Oilseeds, Including Peanuts," financed by Research and Marketing Act funds. The authors are indebted to Richard O. Been for aid in the development of the mathematical aspects of this paper. Submitted for publication September 1949.

865648—50——1

Wholesale rather than retail prices of fats and oils were necessarily used for the price variable because hardly any of the fats going into nonfood products are priced at retail. Moreover, total demand for fats and oils can best be measured at the wholesale level, where the demands from all fat-using industries converge to "make" the price through competition for available supplies.

The FRB index of industrial production was chosen as a measure of the strength of demand for fats and oils (that is, as the "demand shifter"). Changes in industrial production are closely paralleled in nearly all years by changes in the nonfood uses of fats and oils, which account for far more of the variation in total disappearance than do changes in food uses. An analysis of factors affecting disappearance of food fats might well show that an index of consumer income is the proper demand shifter for food fats. Even if this is true, using industrial production as the demand shifter in the present analysis probably does not introduce serious error, because industrial production is highly correlated with "real" consumer income.

Time was included as an independent variable to measure the effects of factors other than price and industrial activity, to the extent that the combined influence of such other factors may have progressively increased or decreased during the period studied. Factors associated with time might include shifts in consumers' tastes and developments in industrial technology which resulted in new uses for or new products competitive with fats and oils.

The period studied was 1922–40. The beginning year was determined by availability of the fat-and-oil wholesale price index, which begins in 1922. Also, data on disappearance before 1920 are less reliable than those for the years afterward. The year 1941 was not used because of evidences of an abnormal accumulation of inventories by dealers and consumers during that year. This accumulation apparently was at least partly balanced by reductions in inventories in 1942, but none of the years 1942–46 was usable in the analysis owing to price ceilings and rationing. The years 1947 and 1948 were reserved to test the applicability of the analysis to the postwar period.

Results of the Analysis

Analysis of the annual data for 1922–40, by multiple correlation techniques, led to the following equation as the best expression for the relationship between consumption, industrial production, time, and price:

$$(1)\text{----------}\quad X'_1 = -17.036 - 0.014X_2 + 31.713 \log X_3 + 6.691 \log X_4.$$ [2]

---

[2] Correlation coefficients and other statistical measures for this equation are as follows:

$$R_{1.234}=0.986 \qquad r_{12.34}=-0.26$$
$$\overline{R}_{1.234}=0.983 \qquad r_{13.24}=0.96$$
$$\overline{S}_{1.234}=0.642 \qquad r_{14.23}=0.97$$
$$\beta_{12.34}=-0.06 \qquad t_{\beta_{12.34}}=t_{b_{12.34}}=-1.1$$
$$\beta_{13.24}=0.72 \qquad t_{\beta_{13.24}}=t_{b_{13.24}}=13.9$$
$$\beta_{14.23}=0.68 \qquad t_{\beta_{14.23}}=t_{b_{14.23}}=14.7$$

The price of butter was tested as a fourth independent variable, on the hypothesis that butter is the principal competitor of other fats and oils. However, acceptable results were not obtained from correlations including the price of butter. Either these correlations indicated that the price of butter had no significant effect on total disappearance of other fats and oils or they were rejected because (1) the intercorrelation among independent variables was extremely high or (2) disappearance was explained largely by trend.

The highest multiple correlation in the subset of independent variables was $R_{2.34}$ (=0.586). A relatively low degree of intercorrelation among the independents, as indicated by the value of $R_{2.34}$, gives stability to the statistical constants of the full set, particularly to the standard errors of the regression coefficients, and so enhances confidence in the reliability of the results. Problems introduced by a high intercorrelation among the independent variables have been discussed by Frisch and by Waugh and Been. FRISCH, R. STATISTICAL CONFLUENCE ANALYSIS BY MEANS OF COMPLETE REGRESSION SYSTEMS. Oslo, 1934; WAUGH, F. V., and BEEN, R. O. ON THE VALIDITY OF AN ESTIMATE FROM A MULTIPLE REGRESSION EQUATION, an unpublished paper. A digest of part of this paper, by the same authors, SOME OBSERVATIONS ABOUT THE VALIDITY OF MULTIPLE REGRESSIONS, appears in the Statistical Journal of the College of the City of New York, Vol. 1, No. 1, pp. 6–14 (January 1939).

The semi-logarithmic relationships between X_1 and X_3 and between X_1 and X_4 were used in the equation because residuals from linear regressions suggested a curvilinearity best expressed mathematically by using the logarithms of X_3 and X_4 (fig. 1). When this was done, there was significant improvement in the multiple and partial correlation coefficients and an increase in the t-values of the regression coefficients.

The partial correlation coefficients, the β-values, and the standard errors of the β's indicate that industrial production and trend largely determined disappearance. The effect of price on disappearance is not statistically significant, but the regression coefficient for price is shown as the "most probable" value.

Elasticity of Consumption With Respect to Price ("Elasticity of Demand")

Generally speaking, as the price of a commodity increases the amount consumed tends to decrease. Conversely, as the price decreases, consumption tends to increase. However, the effect on consumption of a given change in price varies widely for different commodities. A convenient measure of the effect of price changes on consumption is the relationship of the percent change in consumption to the percent change in price. This is known as the coefficient of elasticity. In general, a coefficient of elasticity of 1 indicates that consumption and price tend to change in the same proportion (unit elasticity of demand); a coefficient greater than 1 indicates that consumption changes proportionately more than price (elastic demand); and a coefficient less than 1 indicates that consumption changes proportionately less than price (inelastic demand). A coefficient of zero indicates that consumption is not responsive to changes in price (perfectly inelastic demand).

Equation (1) indicates that elasticity of total demand for fats and oils, other than butter, is close to zero. The coefficient of elasticity cannot be determined exactly since statistical tests show that the relation between price and consumption indicated by the equation is subject to a fairly wide range of error. According to the equation, a decrease of 10 points in the deflated price index was accompanied, during the period 1922–40, by an average increase of only 0.14 pound per person in disappearance, after allowance is made for the effects of industrial production and time. With all factors at the 1922–40 average, the elasticity of demand indicated by the equation would be only 0.026.[3]

A consideration of the nature of the demand for fats and oils, other than butter, in each of the major uses tends to support the statistical evidence of a very inelastic total demand. Total demand for food fats other than butter probably is very inelastic, despite the probability of a relatively elastic demand for margarine. Margarine is mainly used as a spread, in competition with butter, and prices of butter and margarine probably have an appreciable effect on the consumption of margarine. Before the war, however, margarine accounted for only about 6 percent of the total consumption of food fats, other than butter.

Lard, shortening, and cooking and salad oils—the remaining major food fats—probably are only little affected by competition of butter, which is mainly used as a spread and usually is priced two to three times as high. Most of the use of these fats, moreover, is either in products having a very inelastic demand or in products to which the fat contributes only a small part of the total cost. Bread is the leading outlet for lard and shortening. As the demand for wheat in food uses is almost completely inelastic,[4] a very inelastic demand for bread may be inferred, and hence a very inelastic demand for the fat used in bread.

Frying is another major food use of fats and oils. Lard, shortening, margarine, and cooking oils are used for this purpose. In most commercial fried products, such as potato chips and salted nuts, the total cost of the product to the consumer probably is several times larger than the cost of the fat used. Therefore, variations in the price of fat probably have relatively little effect on total cost of these products and little effect on the quantity of fat used. A similar relationship exists between the cost of green salads and the salad dressing or salad oil used. Olive oil is an exception but it constitutes only a small fraction of the total quantity of salad oils consumed.

Soap manufacture accounts for about half of the total nonfood use of fats and oils. As substitutes for soap were unimportant during the period studied, demand for soap probably was also very inelastic with respect to price.

Other nonfood utilization of fats and oils comprises use in the manufacture of paints, varnishes, linoleum, lubricants and greases, core oils, synthetic resins, rubber, and many other industrial products. In most of these uses, cost of the fats and oils is only a small fraction of the total cost of the final product. For example, cost of the oils in paints, varnishes, and linoleum, used in building a new house is very small compared with the total cost of the house. Cost of fats used in lubricants for factory machinery is unimportant in relation

[3] Coefficient of point elasticity of consumption with respect to price $= -b_{12.34}\frac{X_2}{X'_1} = .014\frac{X_2}{X'_1}$; where X'_1 is per person consumption estimated from equation (1).

[4] AN ANALYSIS OF THE EFFECTS OF THE PROCESSING TAXES LEVIED UNDER THE AGRICULTURAL ADJUSTMENT ACT, prepared by the Bureau of Agricultural Economics for the Bureau of Internal Revenue, 1937, p. 25.

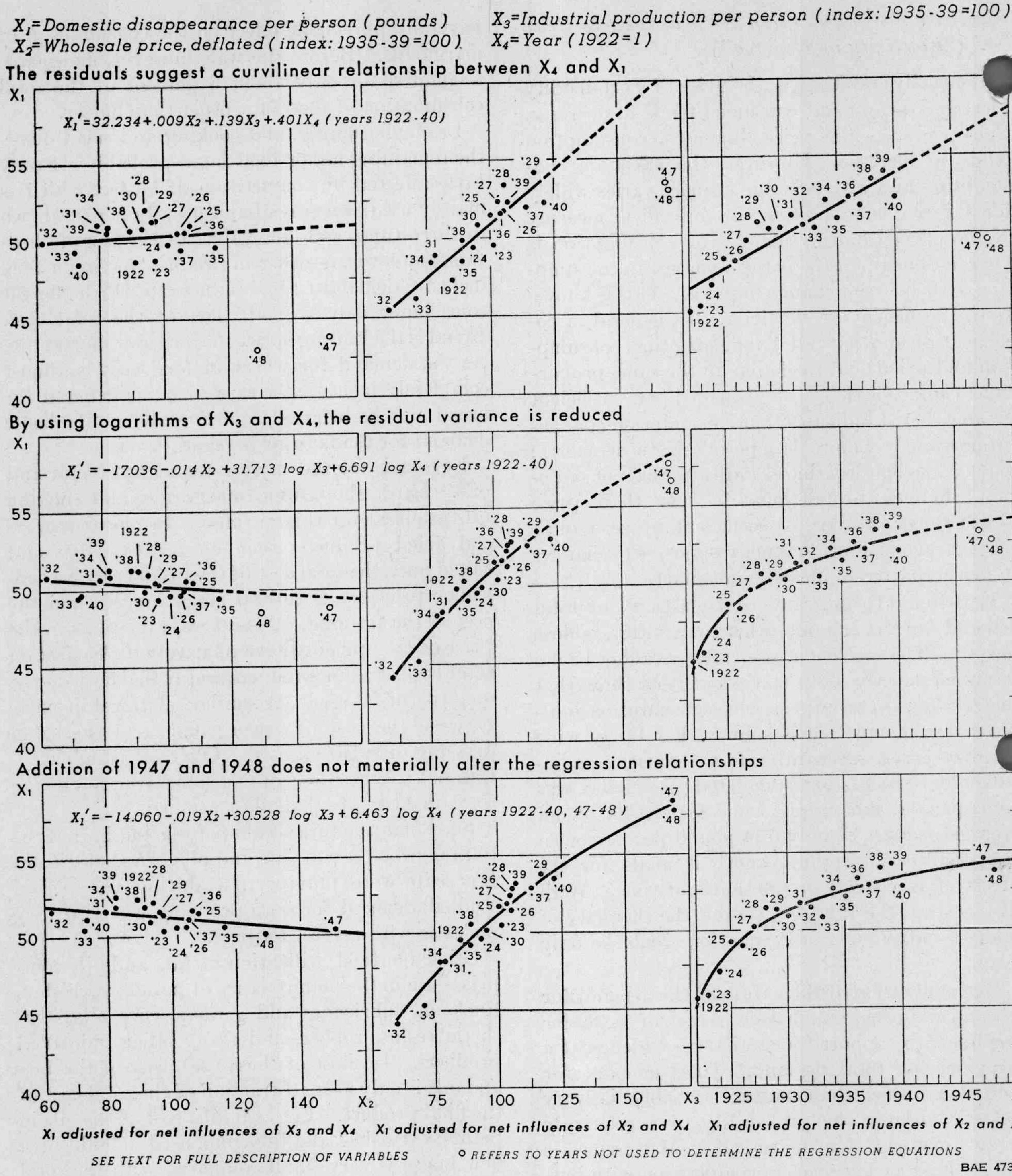

FIGURE 1.—NET INFLUENCE OF PRICE, INDUSTRIAL PRODUCTION, AND TIME ON THE CONSUMPTION OF FATS AND OILS OTHER THAN BUTTER—NET REGRESSIONS AND NET RESIDUALS.

to total cost of the products of the factory. A large percentage change in the price of fats and [oil]s would raise the total cost of the final product by only a very small percentage and probably would have little, if any, effect on the use of fats and oils in manufacturing the product. Hence, elasticity of demand for fats and oils in a wide variety of industrial uses may be expected to be close to zero.

The conclusion that elasticity of total demand for fats other than butter is close to zero does not preclude a higher coefficient of elasticity for particular fats or fat-and-oil products. Also, elasticity for many products, notably foods, is smaller at wholesale than at retail. Therefore, elasticity of demand for some fat-and-oil products, measured at the retail level, may well be materially higher than the elasticity suggested by the present analysis for total demand at the wholesale level.

Elasticity of Consumption With Respect to Industrial Production

The regression equation indicates that consumption increases with industrial production, after allowance is made for the net influence of time and price. However, the higher industrial production, the smaller is the effect on consumption of a given increase in industrial production. With the independent variables at their 1932 level, an increase [of] 10 percent in industrial production per person would have increased disappearance 3.0 percent; but when the independent variables are held at their 1940 level, the resulting increase in disappearance would have been only 2.4 percent. These percentages indicate income elasticities of 0.30 and 0.24, respectively.[5]

The regression of consumption on industrial production apparently combines a relatively elastic response of utilization in drying-oil and miscellaneous industrial products with a small response in food uses. Decrease in elasticity at higher levels of industrial production probably is due to behavior of the food component in the total.

An unpublished analysis indicates that use of fats and oils in drying-oil products varies closely with the index of industrial production and that elasticity in this use with respect to industrial production is about 0.9. Use of oils and fats in nonfood products other than soap and drying-oil products probably also is roughly proportional to the index of industrial production. This miscellaneous group of products includes lubricants and greases, fatty acids, and many others which are affected by the general level of industrial activity.

A rough approximation to prewar consumption of food fats and oils other than butter, per person, by income groups, has been worked out from consumer purchase studies. This shows that prewar consumption of food fats other than butter rose as income increased, up to the middle-income group, and then declined slowly with rising income. This relation between income and food-fat consumption may be responsible for the curvilinearity in the regression of consumption of all fats and oils, except butter, on industrial production (which is highly correlated with consumer income).

The income elasticity of food fats other than butter at the lower income levels, based on prewar consumer purchase studies, was roughly 0.2. The average income elasticity of consumption for all income groups was approximately zero.

Net Trend in Consumption

The regression equation indicates a net upward trend in consumption during 1922–40, after allowance is made for the effects of price and industrial production. The net trend was rapid in the 1920's but tapered off in the 1930's, and by 1940 it had nearly disappeared. The indicated net increase in disappearance from 1924 to 1925 due to trend was 0.84 pound per person; from 1934 to 1935, 0.22 pound per person; from 1939 to 1940, 0.16 pound per person.

Several developments in the fats and oils field during 1922–40 are probably reflected in the net trend. The ratio of fat to flour in commercial bread baking was increased from about 2 percent in the early twenties to about 3 percent in 1940.[6] As commercial bread baking constitutes a major outlet for lard and shortening, this probably exerted a strong influence toward increasing the level of total demand for these fats. A second

[5] Coefficient of point elasticity of consumption with respect to industrial production$=\frac{0.434}{X'_1} b_{13.24}=\frac{13.763}{X'_1}$; where X'_1=per person consumption estimated from equation (1) and $0.434=\log_{10} e$.

[6] These estimates were derived from Census of Manufactures data by N. Jasny in an unpublished study PRICE AND COST OF BREAD IN THE UNITED STATES AND OTHER COUNTRIES.

development was the expanding use of washing machines during 1922–40, which increased sales of flaked and granulated laundry soaps at the expense of laundry bars. The fat content of the former type of soap is considerably larger and resulted in an increased demand for fats and oils in soap. Another development tending to increase demand for fats and oils, especially during the 1920's, was a rapid increase in consumption of green salads. Consumption of lettuce in the United States rose from less than 11 pounds per person in 1922 to nearly 17 pounds in 1929 and, except for the depression period, continued at about 16 to 17 pounds per person until 1942. This undoubtedly was reflected in a rising level of demand for salad oil.

Projection to the Postwar Period

Projection of a statistical analysis of economic data is particularly hazardous when the change in economic conditions and disruption of consumers' habits are as great as that which took place during and immediately after World War II. Substantial extrapolation of the base-period variables is likely to be necessary. In addition, factors which were not significant in the base period may have become significant in the later period. The probable importance of new factors can be appraised only by judgment based on familiarity with the industry studied. However, the degree of extrapolation of the independent variables beyond the range of the base-period scatter can be determined by a chi-square test proposed by Waugh and Been.[7]

Application of this test shows that a χ^2, as large as the one computed from values of independent variables observed in 1947, could be expected to occur only seven times in a million through random sampling from the universe implied by the 1922–40 data. A χ^2, as large as the one computed for the 1948 combination of observations, could be expected only once in a thousand times.

On the assumption that the relationships determined for the period 1922–40 will hold for 1947 and 1948, despite the extreme extrapolation involved, the standard error of forecast can be computed for each year.[8] Aside from the effects of extrapolation and the influence of new variables, the odds are 2 out of 3 that a forecast of consumption will fall 1 standard error or less from the "true" value; and 99 out of 100 that it will fall 2 standard errors or less from the "true" value. The standard error of forecast for 1947 is 1.0 pound per person. The forecasted disappearance for 1947 from equation (1) was 60.9 pounds per person. The actual disappearance in 1947 was 60.3 pounds per person. In 1948, the forecast was 61.4 pounds per person, the standard error was 0.9 pound per person, and the actual disappearance was 60.3 pounds per person.

The relatively close agreement of the forecasts with actual disappearance in 1947 and 1948 may have occurred despite the operation of factors which are not included in the forecasting equation and which have become important since 1940. These possible factors are the price of butter and the use of synthetic detergents, synthetic resins, and chemical bread softeners.

The price of butter may have had considerable influence on disappearance of other fats and oils in 1947 and 1948. No significant influence could be discovered when the data for 1922–40 were analyzed; but in 1947 and 1948 the price of butter was much higher than in 1922–40. Consumption of butter in 1948, at 10 pounds per person, was more than 6 pounds below the 1935–39 average. Consumption of margarine in 1948, at slightly more than 6 pounds per person, was about 2.5 pounds above the prewar average, in terms of fat; and this increase undoubtedly was related to the high price of butter.

On the other hand, synthetic detergents have become serious competitors of soap since the prewar period. Household use of synthetic detergents became rather general during the war and increased rapidly in 1947 and 1948. By the latter year, the use of synthetic detergents for all purposes amounted to roughly 15 percent of total use of "synthetics" and soap. Some synthetic detergents are partly derived from fats, but others,

[7] WAUGH, F. V., and BEEN, R. O. ON THE VALIDITY OF AN ESTIMATE FROM A MULTIPLE REGRESSION EQUATION, an unpublished paper. A digest of part of this paper, by the same authors, SOME OBSERVATIONS ABOUT THE VALIDITY OF MULTIPLE REGRESSIONS, appears in the Statistical Journal of the College of the City of New York, V. 1, No. 1, pp. 6–14, January 1939. The formula for calculating chi-square in this usage is given in note 2B, at the end of this paper.

[8] Waugh and Been, ibid. The formula for the standard error of forecast is given in note 2C at the end of this paper.

which now probably account for a large part of the total, do not contain any fat derivative.

The increasing use of chemical bread softeners was brought to public attention by hearings on standards for bread, held by the Food and Drug Administration in late 1948 and early 1949. A leading type of bread softener is polyoxyethylene stearate, derived partly from petroleum and partly from fats. It is claimed that this type of softener produces the same shortening effect in baked goods as is produced by several times its weight in fat and also retards drying out thus prolonging the shelf life of bread. It was alleged at the hearing that the softener was being used as a replacement for fat in baked goods rather than merely as a supplement. Testimony indicated that sales of the polyoxyethylene type of softener in early 1949 may have been at a rate as high as 10 million pounds annually. If used to replace fat, this quantity would have a relatively minor effect on the market for lard and shortening; but with a continuing increase in sales there would be a major depressing effect on the demand for lard and shortening.

Synthetic resins are being used in increasing quantities in paints, varnishes, and linoleum. Alkyd resins, the leading type of synthetic used in these products, are composed on an average of about 50 percent oils or fatty acids. Other resins contain no fat or fat derivatives.[9]

Technical Notes

Note 1: Use of Chi-square to Estimate Degree of Extrapolation in a Multivariate Scatter[10]

Regression analysis is based on observed data. The regression equation describes the relationship between the dependent and independent variables within the limits of the observations. Therefore, in using the regression equation to make a forecast, we need to know whether the new combination of values for the independent variables is inside or outside the range of observations from which the regression was calculated. A forecast beyond the observed range is to be used only with caution, this caution increasing with the degree of extrapolation from the limits of the observations.

When there is only one independent variable, the position of a new value with respect to the original set of values is easily observed. With a pair of independent variables, the position of a pair of new values can be noted when plotted as a point on a scatter diagram and compared with the scatter of the original pairs of values. In this case, the new value for each variable may be within the range of the original set of that variable, but when considered together, this new pair may lie outside of the scatter of the original pairs. In this event, the new pair is an extrapolation from the original scatter just as much as if one of the values lay outside the range of the original set for that variable. Hidden extrapolation of this kind becomes difficult to detect graphically when there are three independent variables and practically impossible to discover with more than three.

The N observations of the n independent variables used in a regression analysis may be represented by N points scattered in n dimensional space. The pattern and degree of concentration of this scatter depend on the structure of intercorrelation among the independent variables as well as the variances of the variables. Waugh and Been have suggested that for any number of independent variables, a chi-square can be calculated for each combination of observations to indicate its position with respect to the grouping tendency

---

[9] Regression equation (1) has been re-computed with 1947 and 1948 included in the base period. The results, which are close to those based on 1922–40 only, are as follows:

$$(2)\text{----}X'_1=-14.060-0.019X_2+30.528\ \log X_3+6.463\ \log X_4$$

$$R_{1.234}=0.990 \quad r_{12.34}=-0.37$$
$$\bar{R}_{1.234}=0.989 \quad r_{13.24}=0.96$$
$$\bar{S}_{1.234}=0.662 \quad r_{14.23}=0.97$$
$$\beta_{12.34}=-0.99 \quad t_{\beta_{12.34}}=-1.7$$
$$\beta_{13.24}=0.74 \quad t_{\beta_{13.24}}=13.6$$
$$\beta_{14.23}=0.54 \quad t_{\beta_{14.23}}=15.1$$
$$R_{2.34}=0.766 \quad R_{3.24}=0.786 \quad R_{4.23}=0.337$$

The forecast of total domestic disappearance in 1949 for all fats and oils except butter, based on equation (2), assuming a 10-percent decline in industrial production from 1948, and a 25-percent decline in the deflated index of prices of fats and oils, would be 59.9 pounds per person, with a standard error of forecast of 0.8 pound. The chi-square value for the assumed 1949 set of values of the independent variables indicates that this set lies inside, but near the outer boundary of the scatter of the base-period data and, consequently, that the 1949 forecast should be used cautiously. As a result of the upward trend in sales of synthetic detergents and a larger prospective production and consumption of butter in 1949 than in 1948, disappearance of fats and oils other than butter may fall near the lower end of the range indicated by two standard errors of forecast (that is, 59.9±1.6).

[10] Based on WAUGH, F. V., and BEEN, R. O., ON THE VALIDITY OF * * *, op. cit.

TABLE 1.—*Domestic disappearance of all fats and oils excluding butter, per person, and factors used in multiple correlation analysis to explain disappearance, 1922–48*

Year	Domestic disappearance of all fats and oils excluding butter		Wholesale prices of 26 major fats and oils excluding butter (1935–39=100)		Industrial production (1935–39=100)		Time (X_4)	Wholesale prices of all commodities (1935–39=100)	Estimated domestic disappearance of all fats and oils excluding butter, per person, based on—		
	Total	Per person (X_1)	Undeflated	Deflated [1] (X_2)	Total	Per person (X_3)			Linear regression fitted to 1922–40 data [2]	Curvilinear regression fitted to 1922–40 data [3]	Curvilinear regression fitted to 1922–40, 1947–48 data [4]
	Million pounds	*Pounds*							*Pounds*	*Pounds*	*Pounds*
1922	4, 841	43. 7	109	91	73	85	1	120. 0	45. 2	42. 9	43. 1
1923	5, 229	46. 4	121	97	88	101	2	124. 8	47. 9	47. 2	47. 2
1924	5, 359	46. 6	123	101	82	93	3	121. 7	47. 2	47. 2	47. 2
1925	5, 734	49. 2	138	108	90	100	4	128. 4	48. 6	48. 9	48. 8
1926	5, 876	49. 7	129	104	96	105	5	124. 1	49. 7	50. 3	50. 2
1927	6, 074	50. 7	115	97	95	103	6	118. 4	49. 7	50. 6	50. 6
1928	6, 322	52. 1	113	94	99	106	7	120. 0	50. 5	51. 5	51. 4
1929	6, 544	53. 4	113	96	110	117	8	118. 2	52. 5	53. 2	53. 1
1930	6, 234	50. 4	99	93	91	95	9	107. 2	49. 8	50. 8	50. 7
1931	6, 073	48. 7	72	79	75	78	10	90. 6	47. 7	48. 5	48. 6
1932	5, 719	45. 5	50	62	58	60	11	80. 4	45. 5	45. 4	45. 8
1933	5, 916	46. 8	59	72	69	71	12	81. 8	47. 5	47. 9	48. 0
1934	6, 291	49. 5	76	82	75	77	13	92. 9	48. 8	49. 1	49. 2
1935	6, 421	50. 2	115	116	87	88	14	99. 3	51. 0	50. 7	50. 5
1936	6, 924	53. 7	106	106	103	104	15	100. 2	53. 6	53. 3	53. 1
1937	7, 051	54. 4	112	105	113	113	16	107. 1	55. 2	54. 6	54. 4
1938	6, 847	52. 4	87	89	89	88	17	97. 5	52. 0	51. 6	51. 5
1939	7, 297	55. 4	79	82	109	107	18	95. 7	55. 0	54. 6	54. 4
1940	7, 428	55. 9	72	73	125	122	19	97. 5	57. 4	56. 7	56. 5
1941	8, 643	64. 5	116	107	162	157	20	108. 3	[5] 62. 9	[5] 59. 8	[5] 59. 3
1942	7, 995	59. 0	149	122	199	191	21	122. 6	[5] 68. 2	[5] 62. 4	[5] 61. 8
1943	8, 003	58. 2	154	120	239	226	22	127. 9	[5] 73. 4	[5] 64. 9	[5] 64. [illegible]
1944	8, 236	59. 3	154	119	235	220	23	129. 1	[5] 73. 0	[5] 64. 7	[5] 64. [illegible]
1945	8, 031	57. 2	155	118	203	188	24	131. 3	[5] 68. 9	[5] 62. 7	[5] 62. 0
1946	7, 777	54. 7	197	131	170	155	25	150. 2	[5] 64. 9	[5] 59. 9	[5] 59. 3
1947	8, 741	60. 3	285	151	187	168	26	188. 7	[5] 67. 1	[5] 60. 9	60. 1
1948	[6] 8, 870	[6] 60. 3	264	[6] 129	[6] 192	[6] 169	27	[6] 204. 7	[5] 67. 5	[5] 61. 4	60. 7

[1] Deflated by the BLS index of wholesale prices for all commodities (1935–39=100).
[2] $X'_1 = 32.234 + 0.009 X_2 + 0.139 X_3 + 0.401 X_4$.
[3] $X'_1 = -17.036 - 0.014 X_2 + 31.713 \log X_3 + 6.691 \log X_4$.
[4] $X'_1 = -14.060 - 0.019 X_2 + 30.528 \log X_3 + 6.463 \log X_4$.
[5] These years not used in the regression analysis.
[6] Preliminary.

Compiled as follows: Domestic disappearance of all fats and oils excluding butter from data on production, stocks, exports, imports, and shipments to United States territories and possessions; from reports of Bureau of the Census, Fish and Wildlife Service, and U. S. Department of Agriculture.

Index of wholesale prices of all fats and oils excluding butter (1935–39=100)—computed by the Bureau of Agricultural Economics from prices in the National Provisioner, The Journal of Commerce (New York), Oil, Paint, and Drug Reporter, Chicago Journal of Commerce, and reports of the U. S. Department of Agriculture; for method of computation see USDA Technical Bulletin No. 737 (September 1940).

Wholesale price of butter—from reports of the Bureau of Agricultural Economics.

Index of industrial production—U. S. Federal Reserve Board.

Index of wholesale prices of all commodities—Bureau of Labor Statistics.

of the whole set of observed combinations (as defined by the pattern and degree of concentration of the observed scatter). When the values of all independent variables are at their means, chi-square equals zero. As the values depart from their means, chi-square increases. However, chi-square also depends on the structure of intercorrelation among the independent variables in such a way that it indicates the position of any given combination of values of the independent variables with respect to the grouping tendency of the whole set of observed combinations.

Note 2: Computation of Certain Statistics

A. Intercorrelation among the independent variables

If the multiple regression constants have been computed by the method suggested by Waugh,[11] estimates of intercorrelation among the independent variables can be determined readily from the "*P*-table" (the reciprocal correlation matrix). After eliminating the dependent variable from the *P*-table,[12] select the highest new P_{ii} value. Translated into terms of the multiple correlation coefficient, this indicates the maximum degree of intercorrelation among the independent variables.

B. Chi-square

Chi-square for any combination of independent variables can likewise be calculated from the *P*-table after the dependent variable has been eliminated. Denote each element of the new *P*-table (for the independent variables only) as p_{ij}. Form a square matrix, Q_{ij}, with elements q_{ij}, by dividing each element p_{ij} by the standard deviations $\sigma_i\sigma_j$. Compute $x_i = X_i - \overline{X}_i$ (the deviation from the mean), for the observed value of each independent variable. Then, if n equals number of variables used in the regression analysis and X_1 is the dependent variable:

$$\chi^2 = \sum_{i=2}^{n} \sum_{j=2}^{n} q_{ij} x_i x_j.$$ [13]

The theoretical probability of each chi-square can be found in a chi-square table, such as Elderton's[14] or Fisher's.[15] This indicates the probability of occurrence of the given combination, or one farther from the grouping tendency, in sampling from a universe implied by the scatter of the base-period data.

C. Standard error of a forecast

The standard error of forecast, adjusted for degrees of freedom, may be computed from the following equation:

$$\bar{\sigma}_f^2 = \bar{S}_{1.2\ldots n}^2 \left(1 + \frac{1+\chi^2}{N}\right).$$

Where: $\bar{S}_{1.2\ldots n}$ = standard error of estimate of the regression equation, adjusted for degrees of freedom.

χ^2 = chi-square computed for the given combination of values of the independent variables.

n = number of variables used in the regression analysis.

N = number of observed combinations on which the regression equation is based (that is, the number of years used).

The standard error of forecast is a combination of the standard error of estimate and the standard error of the regression function. The $\bar{\sigma}_f$ concept and an alternative method of computation are presented by Ezekiel.[16]

[11] WAUGH, F. V. A SIMPLIFIED METHOD OF DETERMINING REGRESSION CONSTANTS. Amer. Statis. Assoc. Jour., December 1935.

[12] WAUGH, F. V. THE ANALYSIS OF REGRESSION IN SUBSETS OF VARIABLES. Amer. Statis. Assoc. Jour., December 1936. In this usage, Waugh's equation (2) should be adjusted to read: $P_{ij)k(} = \frac{P_{ij}P_{kk} - P_{ik}P_{jk}}{P_{kk}}$.

[13] An illustration of the above method of computing χ^2 is available on request.

[14] PEARSON, K. TABLES FOR STATISTICIANS AND BIOMETRICIANS, Part 1, Ed. 3, 1930, pp. 26–8. In this table, n' = the number of independent variables plus 1.

[15] FISHER, R. A. STATISTICAL METHODS FOR RESEARCH WORKERS, Ed. 8, 1941, pp. 110–11. In this table, n' = number of independent variables.

[16] Ezekiel, M., METHODS OF CORRELATION ANALYSIS, Ed. 2 (1941), pp. 341–7.

865648—50——2