
FARM COST STUDIES
IN THE UNITED STATES
MERRILL K. BENNETT

22650

FARM COST STUDIES IN THE UNITED STATES

MISCELLANEOUS PUBLICATIONS
OF THE
FOOD RESEARCH INSTITUTE

No. 1. *Stale Bread Loss as a Problem of the Baking Industry*
[February 1923]

No. 2. *The American Baking Industry, 1849–1923, as Shown in the Census Reports*
[September 1925]

No. 3. *Combination in the American Bread-Baking Industry, with Some Observations on the Mergers of 1924–25*
[January 1926]

No. 4. *Farm Cost Studies in the United States: Their Development, Applications, and Limitations*
[June 1928]

EUROPEAN SALES AGENTS
GREAT BRITAIN: P. S. King & Son, Ltd., Orchard House, 14, Great Smith Street, Westminster, S.W. 1, London.

MISCELLANEOUS PUBLICATIONS
No. 4 , , , , June 1928

FARM COST STUDIES *in* THE UNITED STATES

THEIR DEVELOPMENT, APPLICATIONS, *and* LIMITATIONS

By

MERRILL K. BENNETT
Research Associate, Food Research Institute

FOOD RESEARCH INSTITUTE
STANFORD UNIVERSITY, CALIFORNIA

The Food Research Institute was established at Stanford University in 1921, jointly by the Carnegie Corporation of New York and the Trustees of the Leland Stanford Junior University, for research in the production, distribution, and consumption of food.

DIRECTORS

CARL LUCAS ALSBERG JOSEPH STANCLIFFE DAVIS
ALONZO ENGLEBERT TAYLOR

PRINTED IN THE UNITED STATES OF AMERICA
BY STANFORD UNIVERSITY PRESS

FOREWORD

Broadly stated, the purpose of the present study is to describe and evaluate the work which has been and is being done in the United States in the collection and analysis of statistics of farm costs of production. How such data are collected, who collects them, what purposes they have been intended to serve, how far they are suitable to serve these purposes: all these are topics of some interest and importance. Not long ago agitation was common for the fixation of prices by government on the basis of "cost of production plus a reasonable profit"; today changes in tariff duties involve determination of the difference between costs of production at home and abroad; for many years organized effort has been made to increase farm efficiency through the study of farm costs of production. It is desirable to examine the propriety of employing farm cost data for important purposes like these.

A long-continued study of the literature of farm costs has convinced the present writer that, if the possibilities and limitations of farm cost investigation are to be clearly recognized, more consideration ought to be given to the general theory of cost-and-price relationships in agriculture. Those who have compiled and analyzed the statistics of farm costs have ordinarily approached their problems from the point of view of accountancy. Primarily they have sought accurate methods of expressing costs; and too few attempts have been made to synthesize the existing doctrines of price theory with the statistical material which they have gathered. It is perhaps more proper to say that in many instances the theoretical doctrines have been misinterpreted. A major purpose of the present paper is to point out these misinterpretations.

Consequently a large part of the subsequent discussion is of a theoretical nature.

The study is addressed broadly to farmers and legislators as well as to professional students of economics and farm management. The use of public funds to gather data serviceable for increasing farm efficiency, for price fixing, or for tariff making, is or ought to be a matter of general concern. It has been the intention of the writer to discuss some intricacies of the theory of prices in such a manner that the general reader can follow the reasoning; and the writer hopes that this aim has been achieved. But the study was not designed merely for popular consumption. It is hoped that teachers of economics will find material—historical, statistical, and theoretical—not elsewhere available. It is hoped, in short, that any reader can find herein a sounder and more thorough discussion of the significance of agricultural costs of production than he can discover in any other single source.

The investigation falls naturally into four parts. Chapters i–iii include a generalized description of the forms of statistical data to be found in farm cost studies and of the sources of such data; a historical review of farm cost investigation; and an analysis of the objectives to which investigators have addressed themselves. Chapters iv–vi provide a groundwork for the critical examination of the usefulness of cost study in achieving these objectives—the matter with which we shall be chiefly concerned. In these chapters are considered the methods of collecting cost data, the significance of accounting principles, the nature of variations in farm costs of production, and the causes of the observed variations, with special reference to the question as to how far these variations are subject to control by farmers. Chapters vii–ix deal with cost studies undertaken in furtherance of the general objective of farm management—to increase farm efficiency, or to provide farmers with a basis for making

the best possible choice of enterprises and methods. Chapters x–xii deal with cost studies undertaken in general to influence prices. Here the discussion is more largely theoretical. The relation of agricultural costs to agricultural prices is considered first, and thereafter the theories underlying the use of farm cost data in price fixing and in tariff making.

Conclusions are summarized briefly in chapter xiii. An appendix deals with a technical aspect of farm management cost analysis, the recent proposal to apply the multiple or partial correlation analysis to farm cost data.

Several topics of some importance are not considered in the present study. The first is the subject of account keeping for farmers. Agricultural educational agencies have long been concerned to teach farmers how to keep accounts—either simple financial accounts, by single or double entry, or more involved cost accounts. This subject is excluded not because it is an unimportant type of work involving cost data, but because it has had only an indirect connection with the process of collecting and analyzing statistics from groups of farms.

The second excluded topic, closely allied to the first, is that of the development of sound and usable principles of accounting. Accounting principles are considered only from the point of view of their effect on cost data which are to be used in cost-and-price comparisons. Much discussion about appropriate accounting procedures has raged in the past among farm management investigators, and beyond a doubt changes in accounting procedures have resulted in rendering useless, because incomparable, data gathered chiefly by the route method from the same farms in successive years. But recently farm management cost investigators have practically abandoned attempts either to gather or to compare complete and accurate data on the money costs of producing particular products, so that accounting principles play a far smaller rôle in farm

management investigation than was the case even five years ago. It appears desirable in this study to examine the difficulties to be encountered in currently accepted analytical methods, and hence to ignore methods now abandoned and disputes chiefly of historical interest.

A third topic has not been given the consideration which it perhaps deserves. Farm cost data have more than once been employed in discussions of farm prosperity. But such utilization of data has usually been incidental; cost data have never been collected from specific groups of farms merely to afford comparisons between different farming areas, between different years, and between farming and other pursuits; and one must extend somewhat the meaning of the term "cost data" if one is to include the data on trends in agricultural wages, interest rates, and the like that have been employed in discussions of farm prosperity. The subject of farm prosperity, its measurement and causes, is far too broad a subject to be treated in a study dealing chiefly with data on the costs incurred by specific groups of farmers in producing specific products, especially since these data have not been employed extensively in broad discussions of farm prosperity.

Acknowledgments are due to Dr. Joseph S. Davis, of the Food Research Institute of Stanford University, for invaluable advice and suggestions. The study has been made possible only by the fact that it was conducted and financed as a part of the research program of the Food Research Institute. Through the generosity of the Institute, the author was enabled in the fall of 1924 to visit many organizations which were conducting farm cost studies; and the information secured on this trip, largely through the kindness of workers in the field too numerous to specify, has proved invaluable. For careful reading of the manuscript and many valuable suggestions, the writer is indebted to Professors H. E. Erdman of the University

of California, Bernard F. Haley of Stanford University, George A. Pond of the University of Minnesota, and to Dr. Carl L. Alsberg, Dr. Alonzo E. Taylor, and Dr. Holbrook Working, of the Food Research Institute. The charts were prepared by Douglas L. King.

M. K. B.

STANFORD UNIVERSITY, CALIFORNIA
June 1, 1928

CONTENTS

CHARTS AND TABLES

CHARTS

TABLES

APPENDIX TABLES

APPENDIX

CORRELATION ANALYSIS IN FARM MANAGEMENT RESEARCH

I

Within the past few years there has been an increasing disposition among experts in farm management and agricultural economics to employ elaborate forms of the correlation analysis as the best method of conducting research in farm management. A major aspect of the problem to which farm management investigators address themselves is that of evolving certain generalized statements respecting the factors affecting farm profits in well-defined areas. Such investigation has been conducted for many years, until recently by the device of "cross-tabulation" or "grouping and averaging"—that is, by carefully contrasting in tables the variations in "profits" successively with variations in such factors as the size of farms, yields, receipts from various crops, and the like.

It is now proposed largely to abandon the older cross-tabulation analysis, and to substitute multiple correlation analysis, using either the older method which assumes linear relationships between the variables, or a modified method permitting the use of curvilinear relationships. There can be no question that the methods of multiple correlation properly applied may yield much information which cannot be obtained by the cross-tabulation method. In particular, it permits a simultaneous determination of the separate effects of several related factors where the cross-tabulation method, like a simple correlation, attributes to a single factor the effects of that factor plus a part of the effects of all other factors that happen to be correlated with it.

But questions may well be raised regarding the usefulness of the partial or multiple correlation analysis to serve the general purpose of farm management research. This purpose is not merely to explain last year's profits on a specified group of farms, but also to lay down generalizations valid over periods of years concerning the relationship between profits and the factors affecting profits, and further to point out specifically to farmers what they ought to do to secure higher profits. The last is the true or ulti-

mate objective. Obviously it cannot be attained unless the first two steps in the analytical process are successfully undertaken.

Little has been written in criticism of the proposal to employ the partial correlation analysis in farm business analysis surveys. The literature on the subject has thus far dealt largely with exposition of the method and development and testing of the device, and it has emanated chiefly from advocates of the method. Yet there are certain significant limitations on the general procedure which might well be more clearly recognized, especially since farm management is a field of investigation wherein standardization of analytical methods proceeds rapidly through the influence of the United States Department of Agriculture. The adoption of the partial correlation analysis in farm management research has been urged by influential members of the Department. But in the judgment of the present writer it is a device at once too complicated, expensive, and uncertain to justify its adoption generally throughout the various research agencies, though it may well prove to be the most satisfactory of available methods in the hands of a few qualified investigators working on a mass of really adequate data. It seems desirable here to point out certain of the more obvious difficulties in applying the partial correlation analysis to farm management data.

II

Although the systematic collection and analysis of farm cost data began in 1902, correlation analysis was not proposed until 1917. In that year Tolley[1] urged its adoption in a publication explaining, with illustrations, the methods of computing both gross and net coefficients (linear). For several years the proposal seems to have received little attention, but, beginning with 1923, discussion and use of the correlation method became more common. Taylor,[2] in a journal article, discussed a partial correlation analysis of data from 965 farms in Iowa, and stressed the importance of employing partial rather than gross correlation because of the inter-relationship usually existing between the factors affecting farm profits. A few months later Ezekiel[3] discussed rather

[1] H. R. Tolley, *The Theory of Correlation as Applied to Farm-Survey Data on Fattening Baby Beef* (U.S. Department of Agriculture Bulletin 504), May 1917.

[2] C. C. Taylor, "A Statistical Analysis of Farm Management Data," *Journal of Farm Economics*, July 1923, V, 153–62.

[3] M. Ezekiel, "The Use of Partial Correlation in the Analysis of Farm Management Data," *Journal of Farm Economics*, October 1923, V, 198–213.

thoroughly, on the basis of the same data employed by Taylor, the general problems of farm profits analysis by the correlation method. He showed its superiority to the customary "grouping and averaging" or "cross-tabulation" method, endorsed Taylor's contention that partial correlation was necessary, demonstrated the need for careful definition of the variables, laid down general rules for the use of the method, and indicated the need for some means of handling partial correlations when relationships were known to be not rectilinear, but curvilinear—a situation to be expected in agriculture, where the law of diminishing returns operates. Ezekiel and Tolley[1] toward the end of 1923 explained in a technical paper a short method for computing the coefficients of net correlation and the multiple correlation coefficient, thus lessening the labor of computation to such a degree that the handling of many variables became feasible. Linear relationships were assumed.

Early in 1924 Ezekiel published a preliminary report on an analysis of farm management data from a dairying region in Pennsylvania; but although the author employed correlation analysis, this preliminary statement made no reference to the method. A few months later Tolley and Mendum again explained the use of simple rectilinear correlation, laying particular stress upon scatter-diagrams, lines of regression, and regression coefficients.[2] In September Tolley, Black, and Ezekiel outlined, in an important bulletin, the uses to which partial correlation analysis, both rectilinear and curvilinear, might be put and illustrated the methods to be employed.[3] Particular stress was laid upon the use of regression coefficients as forecasting devices, useful in the problem of aiding the farmer to achieve the highest possible future profits. In December Ezekiel[4] developed (independently of Mills) the index of correlation, and, employing it, demonstrated a method for handling partial curvilinear correlation.

[1] H. R. Tolley and M. J. B. Ezekiel, "A Method of Handling Multiple Correlation Problems," *Journal of the American Statistical Association*, December 1923, XVIII, 993–1003.

[2] H. R. Tolley and S. W. Mendum, *A Method of Testing Farm-Management and Cost-of-Production Data for Validity of Conclusions* (U.S. Department of Agriculture Circular 307), March 1924.

[3] H. R. Tolley, J. D. Black, and M. J. B. Ezekiel, *Input as Related to Output in Farm Organization and Cost-of-Production Studies* (U.S. Department of Agriculture Bulletin 1277), September 1924.

[4] M. Ezekiel, "A Method of Handling Curvilinear Correlation for Any Number of Variables," *Journal of the American Statistical Association*, December 1924, XIX, 431–53.

Up to this point the development had been chiefly in the perfection of a technique which permitted the application of correlation analysis to a multiple-variable problem wherein the relationships were curvilinear as well as rectilinear, and in the explanation of statistical methodology, illustrated from farm management data. In 1925 direct applications of the method began to appear. In June, Crickman[1] published a study of farm organization and management in Warren County, Iowa, basing it upon surveys taken in 1916 (832 farms), 1919 (477 farms), and 1922 (231 farms). Partial rectilinear correlation was employed, but the details of its use were not made clear until November, in a separate bulletin.[2] In October Hitchcock[3] applied simple linear correlation to survey data obtained chiefly from 189 farms in Vermont for the year 1922–23. In November Taylor and Hurd[4] employed rectilinear partial correlation in analyzing survey data from farms in Tama County, Iowa (965 farms 1913, 237 farms 1921; correlation was not applied to 210 records obtained for 1918). In the same month Vernon and Ezekiel,[5] using both rectilinear and curvilinear partial correlation, presented an analysis of records from 258 tobacco farms in Virginia for the year 1922–23, but omitted any detailed explanation of their analytical methods. In October Ezekiel, touching upon a different problem, had employed multiple correlation to demonstrate that operator's earnings constituted a figure more suitable as an index of financial success than labor income or per cent of return on investment.[6] In April 1926, Ezekiel published in final form the results of an analysis of 422 records obtained from

[1] C. W. Crickman, *Farm Organization and Management Studies in Warren County, Iowa* (Iowa Agricultural Experiment Station Bulletin 229), June 1925.

[2] C. W. Crickman, *A Partial Correlation Analysis of Farm Organization and Management Data from Warren County, Iowa* (Iowa Agricultural Experiment Station Research Bulletin 89), November 1925.

[3] J. A. Hitchcock, *A Study in Vermont Dairy Farming* (Vermont Agricultural Experiment Station Research Bulletin 88), November 1925.

[4] C. C. Taylor and E. B. Hurd, *Farm Organization and Farm Profits in Tama County, Iowa* (Iowa Agricultural Experiment Station Research Bulletin 88), November 1925.

[5] J. J. Vernon and M. J. B. Ezekiel, *Causes of Profit or Loss on Virginia Tobacco Farms* (Virginia Agricultural Experiment Station Bulletin 241), November 1925.

[6] M. Ezekiel, "A Statistical Test of Measures of Farmers' Financial Success," *Journal of Farm Economics*, October 1925, VII, 399–413.

farms in southeastern Pennsylvania.[1] In all these publications there appears only one thoroughgoing attempt (Ezekiel, 1926) to employ the partial curvilinear correlation analysis. The literature consists chiefly of either (1) explanations and illustrations of some form of correlation analysis, intended to show what can be done with the various methods, or (2) thorough analyses, like Taylor's and Hurd's or Crickman's, which employ only rectilinear partial correlation. But curvilinear correlation differs from rectilinear only in so far as it accounts for relationships not rectilinear in nature between variables (indisputably often the case in agriculture); the logical processes underlying the analysis, and the interpretation of correlation coefficients, are essentially subject to the same principles. Since we are here concerned with some of these fundamental principles, we may proceed without special reference to the type of correlation employed. It is to be noted that the correlation analysis, as employed in farm management inquiry, is applied to data for one year only. No attempt has thus far been made to correlate time series.

III

The difficulties in the use of the partial correlation analysis in farm management research may be considered under four heads: (1) The securing of adequate data; (2) The choice of variables; (3) The imputation of causal relationships between variables; and (4) The interpretation of low-value coefficients.

Adequacy of Data.—The partial correlation analysis is one of the most intricate, detailed, and refined methods of statistical procedure, a device which may well be reserved for application only to basic data which may be accepted as reasonably accurate. Questions may be raised regarding the accuracy of the data which are available to farm management investigators.

The first requirement is an adequate number of records, since, broadly speaking, the validity of conclusions drawn from correlation coefficients depends upon the size of the probable errors of the coefficients and probable errors are larger as the number of records is smaller. Even within a relatively restricted and fairly homogeneous area farms vary widely in size, volume and type of production, and natural endowments. On general principles it may therefore be assumed that data from several hundred farms, probably from 400 to 1,000, are required if one is to be assured

[1] Mordecai Ezekiel, *Factors Affecting Farmers' Earnings in Southeastern Pennsylvania* (U.S. Department of Agriculture Bulletin 1400), April 1926.

that a representative sampling has been secured. Investigators who have employed correlation analysis have for the most part secured an adequate number of records. Nevertheless the requirement imposes a significant limitation on the use of the device.

This limitation arises from the fact that so large a number of records must be obtained either by a method yielding unduly inaccurate data, or by a method involving prohibitive expense—at least under present conditions.[1]

In order to obtain really accurate data on quantitative expenditures the route method must be used, but it is likely that one year's data from as many as fifty farms would cost $2,000 or $3,000 at a minimum. The expense of obtaining route data from 500 farms would be prohibitive. Consequently the survey method is the only practicable one. This method is accurate in so far as the farmer's memory is dependable and the field agent's skill is great in checking the farmer's answers, one against the other. Merely because of the number and kind of items which must be remembered and checked, there is reason to believe that a good deal of error must occur in the estimates given by individual farmers. These errors are not so significant as entirely to discredit the use of the survey method, but the liklihood of their presence is sufficiently great to give rise to the contention that correlation coefficients calculated from survey data probably need to be accepted with considerable reserve. The data are by no means strictly comparable with "errors of observation" in physical measurements, or even with enumerations. Under the survey method, of course, certain items are more trustworthy than others: areas and numbers of live stock, for example, are likely to be reported more accurately than values or quantitative expenditures. Even under the route method significant errors may occur. One would suppose that it ought to be possible to account for the disposition of wheat on farms with real accuracy. But Table 1A (p. 272), compiled from route cost records obtained from five farms in Jackson County, Kansas, in 1920, gives reason to question the accuracy even of the route method. Discrepancies in the recorded disposition of quantities of wheat produced are large, and one cannot ascertain precisely where the errors of recording occurred. This table, of course, is not presented as typical of the data obtained from route cost investigation; it is intended merely to show that the route method does not preclude errors, even

[1] Expense of gathering data would not be so serious a problem if more farmers kept accurate records under a uniform system, or if the co-operative service now operating in Illinois should come into common use.

when conducted with care, as the investigation yielding these data appears to have been.

If records err, memory may err also. The inexact quality of survey data ought to be remembered by investigators employing the correlation analysis. The danger is that refinement of the analytical procedures will create a disposition to ignore the matter; statements regarding inaccuracies of data have always been too uncommon in published studies.

TABLE 1A.—DISPOSITION OF WHEAT ON FIVE FARMS IN JACKSON COUNTY, KANSAS, 1920*

(Bushels)

Farm No.	Opening inventory	Produced	Purchased	Total supply	Closing inventory	Sales	Consumed on farm	Total disposition	Discrepancies
2.....	45	909		954		1,188		1,188	+234
10.....	8	385	29	422	10	176	28	214	—208
11.....		875		875		874	120	944	+119
12.....	70	680		750	455	146	46	647	—103
16.....		604		604	245	250	21	516	— 88

* Compiled from original route cost data furnished the Food Research Institute by the Kansas State Agricultural College.

Even if accurate data could be secured by the route method, the propriety of basing conclusions upon one year's data may be questioned. Farmers within a given area do not tend to maintain their rank in labor income from year to year, but fluctuate widely from high to low positions on the curve of incomes. The same can probably be said with respect to other income figures—the dependent variables in the correlation analysis—such as "operator's earnings." Now if investigators are seeking to develop permanently valid generalizations, it is clear that this inconsistency of farmers in maintaining their rank in income gives rise to the necessity of employing data for more than one year. Size of farms, for example, would not show considerable change from one year to the next, while labor income would change. Thus in one year a high correlation might be found between labor income and size of farms, while in the next year it might prove low. Similar difficulties might appear in correlating acres in corn with labor income, for, though acreage might remain constant, yield on some farms might be high in one year, low in the next, on account of soil conditions; and changes in yield might affect labor incomes

so that the correlation of acreage in corn with labor income proved quite different in the two years.

It would appear that a thoroughgoing correlation analysis ought to be based on data for several years, at least if investigators seek to develop general principles of farm management. Thus far no such analysis has been attempted, for investigators have apparently been inclined to believe that farmers consistently maintain their positions on the income curve. Until correlation analyses have been run on three or four years' data from the same area, we shall have only an inadequate notion of the factors usually significant in their bearing on profits in that area. In the judgment of the present writer, two or three such analyses (using one year's data at a time, with the same variables each year but not necessarily data from the same farms; and comparing the correlation coefficients so as to determine whether or not the same factors are significant each year) ought to be undertaken in preference to many analyses of one year's data. Such analyses of data obtained in carefully selected areas would probably contribute more definite and dependable knowledge of the factors affecting farm profits than we now possess despite 25 years of farm management research. But only a few such inquiries ought to be conducted in view of their expense and of the paucity of capable statisticians.

Choice of Variables.—The correlation analysis involves at the outset a set of definitions of variables. Hypotheses must be formed respecting the form of farm income which is to be analyzed, and the factors which may reasonably be supposed to affect it. Summary figures representing each factor must be compiled from the raw data. Table 2A (p. 274) displays the summary figures which have actually been employed in the more extensive correlation analyses. The table emphasizes the fact that the problem of explaining farm profits is a multiple-variable problem; the most elaborate analysis includes 20 variables.

An analysis including variables defined as those of Table 2A, however, constitutes scarcely more than an approach to the problem in hand. Even after correlation coefficients or coefficients of determination have been calculated, all that one knows is that, roughly speaking, certain of the defined factors apparently bore a closer relationship to profits than certain others. Most of the variables are not simple, but complex. Thus "receipts" variables are composites of volumes of produce sold and prices received. Volume of sales may depend on acreage and yield; yield on soil,

TABLE 2A.—VARIABLES EMPLOYED IN SELECTED CORRELATION ANALYSES OF FACTORS AFFECTING FARM PROFITS*

Ezekiel, Taylor, Hurd (1913 data)	Crickman (1921 data)	Taylor, Hurd (1921 data)	Ezekiel (1922–23 data)
Labor income	Profits	Profits	Operator's earnings
Acres operated	Acres operated	Acres operated	
Crop acres per horse	Acres pasture per animal unit		
Acres in crops			Acres in crops
Percentage of area in corn	Acres of corn	Percentage of area in corn	
Percentage of area in hay and pasture	Acres of hay and pasture	Percentage of area in hay Percentage of area in pasture	Acres in pasture
Percentage of area in small grains	Acres of small grains	Percentage of area in oats	
Productive animal units	Productive animal units		Number of cows
Crop index	Crop index	Corn yield per A. Oat yield per A.	Crop index
Live stock index	Live stock index	Lbs. pork produced Lbs. beef produced	
Percentage of receipts from dairy	Percentage of receipts from dairy	Receipts from live stock per 100 acres	Receipts from dairy herd
Percentage of receipts from hogs	Percentage of receipts from hogs		Receipts from hogs
Percentage of receipts from cattle	Percentage of receipts from cattle		Receipts from beef cattle
			Receipts from sheep Receipts from poultry
Percentage of receipts from crops	Percentage of receipts from crops	Receipts from corn per 100 A. Receipts from oats per 100 A.	Receipts from crops
Months of man labor	Months of man labor		Labor index
Total capital	Value of real estate per A.	Rent charge per A.	
Percentage of working capital is percentage of total capital		Building charge per A. Machinery value per A.	
		Hog price per pound Cattle price per pound Pounds of dead hogs	Percentage of dairy feed purchased

* Sources: For Column 1, M. Ezekiel, "The Use of Partial Correlation in the Analysis of Farm Cost Data," *Journal of Farm Economics*, V, 202; and C. C. Taylor and E. B. Hurd, *Farm Organization and Farm Profits in Tama County, Iowa* (Iowa Agricultural Experiment Station Research Bulletin 88), November 1925, p. 321. For Column 2, C. W. Crickman, *A Partial Correlation Analysis of Farm Organization and Management Data from Warren County, Iowa* (Iowa Agricultural Experiment Station Research Bulletin 89), November 1925, p. 12. For Column 3, Taylor and Hurd, *loc. cit.* For Column 4, Mordecai Ezekiel, *Factors Affecting Earnings in Southeastern Pennsylvania* (U.S. Department of Agriculture Bulletin 1400), April 1926, p. 55.

fertilization, precipitation, or cultivation, and others; while price (aside from general influences) must be affected more or less by time of marketing and quality, perhaps by bargaining ability. That a high correlation is found between profits and receipts from one source or another is therefore not a finding of great significance in itself; the analysis must be pushed farther if one is to point out specifically what factors affect receipts, and in addition indicate to farmers what they both ought and can do to increase receipts.

When this stage of the analysis is reached (in the published studies only those by Ezekiel have pressed so far) difficulties are certain to arise in two directions: securing sufficiently accurate data on details, and distinguishing controllable from uncontrollable factors. As a practical matter it is not feasible by any method to obtain data on soil quality, precipitation, and temperature on each of a large group of farms; and accurate data on feeding practices and labor utilization are probably not to be had by the survey method. Consequently the correlation analysis must remain to some degree an inadequate analysis. It must remain so unless very heavy expenditures are to be made in securing data adequate to explain all that farm management investigators seek to explain. Such expenditures ought not to be incurred in more than a few instances, especially in view of cheaper alternative methods of attacking the same problem. Hence the multiple correlation analysis ought not to be regarded as an analytical device suitable for general adoption in the agricultural experiment stations.

If the correlation analysis is to render to farmers the sort of services which investigators apparently contemplate, the definition of variables ought to be such that measurement is possible of the extent to which factors that farmers can control influence profits. To establish the proposition that yield as summarized in "crop indexes" or "live-stock indexes" influences income to such and such a degree is doubtless a useful contribution to knowledge; but it is not a proposition of major interest to individual farmers. The analysis must be carried farther in order merely to explain yield, and farther still if farmers are to be informed precisely how to increase yields and increase profits at the same time. Broadly speaking, one might suppose that yield is dependent not only upon largely controllable factors like seed selection, date and depth of planting, fertilizing, closeness of rows, and methods of cultivation, but also upon uncontrollable factors like precipitation,

temperature, and chemical composition and texture of soils. Suppose hours of labor per acre appears not to be correlated with yield. Does this demonstrate that intensity of cultivation has no noticeable effect; or does it mean that an hour of one man's labor is not the same as another's for reasons of differences in physical constitution and that this is the only cause of variation in hours of labor per acre; or does it mean that farmers with low-lying fields were forced to spend more time in soil preparation than others of the group because of heavy rainfall, and yet secured the same yields on the average? It is clear that what investigators seek is to indicate procedures whereby farmers can increase their profits if they will. But it is equally clear that only the most detailed records will provide data adequate to permit the requisite separation of controllable from uncontrollable influences; and one may doubt if thoroughly adequate data could be obtained by any means whatsoever.

Imputation of Causal Relationship.—Most writers on the theory of correlation recognize that the process does nothing more than to measure, more precisely than is otherwise possible, the degree to which two variables vary with each other. The correlation coefficient enables us to say with some precision how far high, medium, and low orders of one variable are associated with high, medium, and low orders of the other. Nothing but the presence or absence of association is measured. No light whatever is thrown upon the nature of the association; it may be directly causal and compulsory, or wholly incidental, or a mixture of the two. Suppose farm profits were to be correlated with length of farmers' fingers, and a coefficient of +.4 emerged. Nothing in the theory of correlation compels us to infer that the one is the cause of the other; we are assured merely that, for the data in hand, long fingers were associated somehow or other with high profits. And if a high correlation is found between yield and profits, nothing is proved but that a high degree of association exists; it need not be a causal relation. That the relation is in fact causal must be established on quite different grounds.

Yet to assume that high correlation proves high causal relationship is a tempting procedure in economic analysis, simply because economic inquiry is largely directed toward explanation of effects by causes. This is, as we have seen, the chief object in farm cost inquiry. We wish to determine what factors influenced profits. We meditate upon the possible causal influences and fix upon a list, say of twenty. Some of these we regard as more im-

portant than others, but on account of the complexities of their relationship, we are uncertain which to choose. On carrying through a partial correlation analysis, we reach coefficients measuring the net association of each independent variable with the dependent variable. We scan the list; and at this point the chief danger of the process arises. For, observing that certain coefficients are of higher value than the others, we argue that causal importance is shown by the rank of the coefficients; yet no such thing is proved. We merely know that various degrees of association exist. It might still be true that the factor with, say, the third highest coefficient was in fact the one most important in its causal influence on profits. The process of correlation provides no dependable means for separating association of a causal nature from association of an accidental nature.

The literature of correlation analysis of farm management data is not free from this fallacy; and no writer seems to have given specific warning against it. Vernon and Ezekiel list receipts per acre of tobacco, area in tobacco, and receipts from other crops as important in the order named for their causal influence on earnings, and the conclusion is apparently based on the relative size of the coefficients.[1] Crickman, clearly arguing from the size of coefficients, says, "it is found that the most important factors, in order of importance, are the production per animal unit, efficiency in the use of man labor, value of the real estate per acre, ... crop yields, and the amount of pasture used to carry one animal unit."[2] Hitchcock, presenting a tabular statement of correlation coefficients, says: ". . . . the relative size of the several coefficients indicates the relative importance of the several factors."[3] Almost every writer speaks of the independent variables as causal, the dependent as resultant.

Correlation is in this respect a dangerous tool. It should not be employed as an infallible means of discovering and measuring the relative importance of causal relationship, for it leads easily to what may be called the fallacious argument from association. It is a particularly insidious method in a problem where the object is largely to disentangle many obscure causal relationships; and farm cost investigation is a problem of this sort. To attack farm

[1] Vernon and Ezekiel, *op. cit.*, p. 45.

[2] Crickman, *A Partial Correlation Analysis of . . . . Data from Warren County, Iowa* (Iowa Agricultural Experiment Station Research Bulletin 89), November 1925, p. 2.

[3] Hitchcock, *op. cit.*, p. 46.

cost analysis by guessing at factors likely to influence profits, correlating these by any method whatever, and gauging the causal importance of the factors by reference to the relative sizes of the correlation coefficients is theoretically unsound, but tempting because it seems decisive. There is real danger that such a rule-of-thumb procedure should gain adherents in a field where standardization of analytical methods is customary. Correlation is not a mechanical substitute for logic. In the end it provides only a statistical description. As Keynes says, ". . . . not unnaturally, the more complicated and technical investigations become, the more prone enquirers are to mistake the statistical description for an inductive generalization."[1] The evidence indicates that advocates of correlation analysis of farm cost data have fallen into this error.

From another point of view, however, correlation bids fair to perform a service not otherwise available. Tolley and Mendum regard it as essentially a check "on the validity of data under study . . . ."[2] In the light of our earlier remarks, it is clear that hypotheses regarding causal relationships are not infallibly verified by the device. But it may nevertheless be argued that although the presence of high correlation by no means constitutes certain proof of causal relationships, its absence may be employed to disprove hypotheses. Thus if very great causal relationship is expected, and practically no correlation is found, there is good reason to question the hypothesis. Use has been made of this method of arguing to show that size of farm has less influence on profits than has commonly been supposed. In this respect correlation seems serviceable in curtailing the number of erroneous conclusions which without it might gain acceptance. But it is unwise, and indeed impossible, to press this use of the device very far. If one expected a coefficient of +.8 and obtained no correlation, then it could properly be said that a hypothesis was disproved. But the same could not be said where coefficients in the middle and lower ranges were expected, and only slightly lower coefficients were obtained. The disproof of the hypothesis would here be no more certain than the proof of a hypothesis would be when "some" causal relationship was predicted, and "some" correlation was obtained.

Interpretation of Low-Value Coefficients.—We have already seen that the problem of explaining farm profits by correlation analysis is a multi-variable problem. In such a problem we have

[1] J. M. Keynes, *A Treatise on Probability*, p. 329.
[2] Tolley and Mendum, *op. cit.*, p. 13.

every reason to expect that most of the coefficients obtained in the analysis will be of low value. It is impossible that each of, say, 20 variables should show a large effect on farm profits. From what evidence we have, it is clear that farm management investigators must grapple with the problem of interpreting coefficients of low value, or at least the problem of determining which coefficients to dismiss from discussion as unimportant. Table 3A (p. 280) displays the most elaborate published list of net coefficients of correlation,[1] together with their probable errors.

Accepted statistical practice calls for the exercise of judgment in determining when a coefficient of correlation is sufficiently high to warrant the inference that it indicates a relationship of real importance. This is true of any form of coefficient, whether simple, multiple, partial, rectilinear, or curvilinear; and it is equally true of the "coefficient of determination."[2] Customarily judgment is guided by calculating the probable error and by ascertaining whether or not the coefficient is appreciably larger than its probable error. There appears to be some divergence of opinion among authorities regarding the proper method of calculating the probable error of a coefficient of partial correlation, much less of a coefficient of determination, though the formula $PE = \frac{.6745\ (1-r^2)}{N}$ is accepted for application to coefficients of gross correlation. Apparently farm management investigators, when they have calculated probable errors at all, have employed this formula. But it is a further query, with respect to the relation of the coefficient to its probable error however calculated, which concerns us here.

Writers on statistics, when they discuss what size of coefficient may properly be regarded as significant, are inclined to say vaguely that relationship is established when the coefficient is "appreciably greater" or "three or four times larger" or "at least four times larger" than its probable error. It is true that they are speaking of gross, not net, coefficients. Discussion of the significance of partial coefficients as tested by their probable errors is not to be encountered in any but the more technical statistical textbooks. But presumably authorities would follow much the same line of reasoning in adjudging the significance of a net

[1] In *U.S. Department of Agriculture Bulletin 1400*, Ezekiel presents list of coefficients of determination; the coefficients of correlation are not shown, and hence cannot be presented in Table 3A.

[2] The coefficient of simple correlation multiplied by the partial regression coefficient when expressed in standard deviations.

coefficient or correlation or a coefficient of determination as they follow when speaking of gross coefficients of correlation; and we may properly assume that if there is no generally accepted test of

TABLE 3A.—NET COEFFICIENTS OF CORRELATION (RECTILINEAR) AND THEIR PROBABLE ERRORS, OBTAINED IN SELECTED FARM COST CORRELATION ANALYSES*

Variables (965 farms in Iowa, 1913)	r	PE	Variables (231 farms in Iowa, 1921)	r	PE
LABOR INCOME[a]			PROFITS[a]		
Acres in crops	+.319		Total acres	+.1194	±.0437
Acres not in crops	+.044				
Productive animal units	—.039		Productive animal units	+.1479	±.0434
Months of man labor	—.275		Months of man labor	—.4031	±.0368
Total capital	—.188		Value of land per acre	—.2655	±.0412
Percentage working capital is of total	+.141				
Percentage of area in corn	+.185		Acres in corn	—.0375	±.0443
Percentage of area in small grains	+.154		Acres in small grains	—.0655	±.0442
Percentage of area in hay and pasture	+.152		Acres in hay and pasture	—.1569	±.0433
Percentage of receipts from dairy	—.091		Percentage of receipts from dairy	—.1919	±.0427
Percentage of receipts from cattle	—.167		Percentage of receipts from cattle	—.0886	±.0440
Percentage of receipts from swine	—.188		Percentage of receipts from swine	—.0074	±.0444
Percentage of receipts from crops	—.047		Percentage of receipts from crops	—.0620	±.0442
Crop acres per horse	—.085		Acres of pasture per animal unit	—.1973	±.0426
Live stock index	+.528		Live stock index	+.5257	±.0321
Crop index	+.300		Crop index	+.2433	±.0417

Where r=0, the probable error is 1.022

* Sources: For Column 1, C. C. Taylor and E. B. Hurd, *Farm Organization and Farm Profits in Tama County, Iowa* (Iowa Agricultural Experiment Station Research Bulletin 88), November 1925, p. 321; for Column 2, C. W. Crickman, *A Partial Correlation Analysis of Farm Organization and Management Data from Warren County, Iowa* (Iowa Agricultural Experiment Station Research Bulletin 89), November 1925, p. 12.

[a] Dependent variables.

the significance of a gross coefficient, such a test for a net coefficient is also lacking.

Some authorities (notably King, Jerome, and Day) following Bowley, say that a coefficient of correlation ought to be "six times

larger" than its probable error before significance can be attached to it—not "appreciably larger," or "three or four times larger," or "at least four times as large." There is in fact no precise rule-of-thumb method of testing the significance of a coefficient, either simple or partial, by comparing it with its probable error. There is rather a considerable difference of opinion. What one authority might regard as a significant coefficient would presumably be discarded as unworthy of attention by others. Day voices the opinion that "coefficients under 30 give very little indication of any definite connection between the variables."[1]

Now it is notable that out of 30 coefficients, only two run above .5, only three above .4, only four above .3; only four out of these thirty coefficients can be said to indicate an association really worth considering. If four times the probable error is to be the test, thirteen (or perhaps twelve) out of the sixteen in the first column may be regarded as significant, and six out of the fourteen in the second column—a total of nineteen out of thirty. If six times the probable error be the proper test, the number of trustworthy coefficients is reduced to fourteen.

Evidently, then, the lack of a dependable rule-of-thumb test is of some importance to farm cost investigators. There is always danger of perceiving relationships, and discussing them, where their very existence is doubtful and can be disputed. The danger is particularly prevalent in farm cost analysis, not only because the probable error test is often applied too blindly, but also because the emergence of so few high-value coefficients tempts one to seek for significance in those of lower and more doubtful values. It is easy, in the search for significant conclusions, to lean too heavily upon the tangible coefficients so laboriously computed, meanwhile forgetting the inherent inaccuracies of the original data, and the uncertain nature of the probable error tests.

IV

It is clear, then, that limitations partly of a practical and partly of a theoretical nature must be encountered in any attempt to establish the partial correlation analysis as the standard ana-

[1] E. E. Day, *Statistical Analysis*, p. 209. Dr. Day is speaking here of gross coefficients of correlation, and it may be unfair to assume that he would say the same thing of net coefficients. He also suggests that the probable error test is not properly applicable where the data do not constitute a random sample. In the judgment of the present writer, farm survey data cannot strictly be regarded as random samples; but it is unnecessary to consider the matter here.

lytical device in farm management research among the agricultural experiment stations. In view of these limitations of the partial correlation analysis, it seems proper to urge that it should be employed only by skilled investigators working with adequate accumulations of data. The device is not a magical machine, which, once set up, will turn out significant conclusions regardless of its operation, and at moderate expense. It is a better device than the cross-tabulation analysis; but the co-operative farm management service is also a better device. This service widely used, and supplemented by partial correlation analyses sparingly used, would in the judgment of the present writer constitute a more fruitful approach to the farm management problem than any general procedures which have hitherto been employed. But no method of analysis ought to be regarded in and of itself as the road to solution of farm management problems.

INDEX

CHECKED
2003-04