Livestock and grain production functions from a new data source
by William Jess Haley

A thesis submitted to the Graduate Faculty in partial fulfillment of the requirements for the degree of
MASTER OF SCIENCE in Agricultural Economics
Montana State University
© Copyright by William Jess Haley (1967)

Abstract:
This study is concerned with the anlaysis and evaluation of a new source of data, namely income tax data, for analyzing the problems of Montana agriculture. The study is intended to determine whether statistically significant coefficients of production can be derived from the tax data and whether the results make reasonable economic sense. The motivation for doing the study was an interest in exploring the possibility of income tax data serving as a less expensive and more convenient source of data than has been previously available. Heretofore, most farm production economic studies have relied on data from survey schedules. Obtaining data through specifically designed surveys is expensive and time consuming, and all too often the information is of questionable accuracy.

In approaching the income tax data, two production functions were statistically derived; one for grain farms and another for livestock farms. These functions were then compared to those derived from previous studies that used other types of Montana data and similar variables. The similarity of the compared results was used to determine whether reasonable estimates could be derived from,the tax data. Statistical criteria were used in evaluating the reliability of the data.

The thesis yields the conclusion that income tax returns represent an acceptable data source, and that the production functions derived were reasonably descriptive of Montana grain and livestock farming.

227

LIVESTOCK AND GRAIN PRODUCTION FUNCTIONS FROM A NEW DATA SOURCE

by

WILLIAM JESS HALEY

A thesis submitted to the Graduate Faculty in partial fulfillment of the requirements for the degree

of

MASTER OF SCIENCE

in

Agricultural Economics

Approved:

Clarence W. Jensen
Head, Major Department

Clarence W. Jensen
Chairman, Examining Committee

Louis D. Smith
Graduate Dean

MONTANA STATE UNIVERSITY
Bozeman, Montana

August, 1967

ACKNOWLEDGMENTS

The author is very much indebted to Professor Maurice C. Taylor for his interest and advice on this thesis. Special thanks are due to Dr. William D. Diehl of the Bureau of Business and Economic Research for laying the groundwork in obtaining the data and for offering helpful comments during the beginning of this study. The author wishes to thank Professor Richard McConnen and Dr. Clarence Jensen, the other members who served on the thesis committee.

Thanks are due to the members of the State Board of Equalization for the kind permission to take the sample of income tax returns.

Appreciation is expressed to my wife for typing the rough draft of this paper and for being a constant source of encouragement.

Any errors or omissions in this study are the responsibility of the author.

TABLE OF CONTENTS

		Page
PART I:	INTRODUCTION	1
	A. The Problem	1
	B. Framework of Analysis	4
	C. Hypothesis	5
	D. Sequence of Analysis	5
PART II:	PROCEDURAL REVIEW	8
	A. Specification Bias	10
	B. Multicollinearity	15
	C. The "Regression Fallacy"	18
	D. Methods Available	21
	E. Alternative Types of Functions	25
PART III:	THE ANALYSIS	30
	A. The Data	30
	B. The Variables	35
	C. The Functions	40
PART IV:	SUMMARY AND CONCLUSIONS	55
	A. Summary	55
	B. Conclusions	58
	C. Recommendations for Future Research	59
APPENDICES		60
	A.	61
	B.	63
	C.	67
	D.	70
	E.	72
LITERATURE CITED		74

LIST OF TABLES

Number		Page
I	Labor for Grain Farms	38
II	Labor for Livestock Farms	38
III	Elasticities of Production and Marginal Productivities for Grain Farms	44
IV	Elasticities of Production and Marginal Productivities for Livestock Farms	47
V	Some Comparable Values from the Fienup and Heady-Shaw Studies	52

LIST OF FIGURES

Number		Page
1	. . . . .	24
2	Family Labor for Grain Farms.	41
3	Family Labor for Livestock Farms.	42

ABSTRACT

This study is concerned with the anlaysis and evaluation of a new source of data, namely income tax data, for analyzing the problems of Montana agriculture. The study is intended to determine whether statistically significant coefficients of production can be derived from the tax data and whether the results make reasonable economic sense. The motivation for doing the study was an interest in exploring the possibility of income tax data serving as a less expensive and more convenient source of data than has been previously available. Heretofore, most farm production economic studies have relied on data from survey schedules. Obtaining data through specifically designed surveys is expensive and time consuming, and all too often the information is of questionable accuracy.

In approaching the income tax data, two production functions were statistically derived; one for grain farms and another for livestock farms. These functions were then compared to those derived from previous studies that used other types of Montana data and similar variables. The similarity of the compared results was used to determine whether reasonable estimates could be derived from the tax data. Statistical criteria were used in evaluating the reliability of the data.

The thesis yields the conclusion that income tax returns represent an acceptable data source, and that the production functions derived were reasonably descriptive of Montana grain and livestock farming.

PART I

INTRODUCTION

Input-output analyses have been favorite research areas for economists for many years. Such research has been among the more fruitful of economic research efforts in terms of tangible results.

There are, however, many significant obstacles in production function research that tend to subject the apparent good results to question. Some of these problems or obstacles are discussed in Part II of this thesis. One of the primary sources of inconsistency in input-output research is the raw data from which production functions are derived. The over-all goal of this thesis will be to make an evaluation of a new source of data, namely the data from Montana personal income tax returns.

A. The Problem

Data collection is an extremely important phase of production function research. If one has bad data, no amount of statistical cultivation will yield meaningful results. The most sophisticated techniques known may be used in the analysis and the results will have no substance if the data are bad to start with. In short, shabby data can only lead to shabby results.

In deriving farm production functions, researchers have used the survey schedule method of data collection almost exclusively. This method is both costly and time consuming and in addition, one does not know how reliable his data are after the schedules are filled out. If the schedule inquires into areas which are generally considered to be private, the

possibility of the answers being either overstated or understated, or not given at all certainly exists. The problem of this thesis is to introduce a new and much less expensive source of data.

The Montana personal income tax returns provide this new source of data. These returns provide a wealth of information (both farm and non-farm) concerning income and expense. The information is punched on computer cards and is readily accessible. A description of the data available from the Montana income tax returns is presented in Part III.

There are two apparent reasons why income tax returns may be an extremely valuable source of data. The first reason is the economy of data collection. Cost is an important consideration in any study and in most studies a large percentage of the cost is attributable to data collection. If this collection cost can be reduced more time and effort can be devoted to the part of research where fruit is borne, the analysis. The second reason is somewhat more subjective. An argument can be made that income tax data are more reliable than data collected from survey schedules. It is not the purpose of this thesis to test this proposition directly but some empirical evidence will be forthcoming to support the proposition that the data from income tax returns are at least as good as survey schedule data.

Income tax returns are made out according to tax laws. These laws provide the taxpayer with incentives to answer the questions on the tax return correctly. It seems reasonable to assume that most taxpayers will state their income as low as the law will allow and their expenses as high as the law allows. An individual firm that does not state its income

and expenses this way will not be maximizing profits. There are no doubt counter examples to this assumption just as there are to most assumptions made in economic theory, but if reasonableness prevails, these counter examples are minimal. This assumption leads to the conclusion that the bias in the sample is consistent throughout the sample. It can be expected that if income is biased it will be biased downward, and that if expenses are biased, they will be biased upward. But if most taxpayers are maximizers the bias will be uniform and therefore predictable.

An empirical attempt to show that the income tax data may be used as a reasonable substitute for data gathered from survey schedules is made by comparing the results of this study to the results of a study of Montana grain farms and cattle ranches published by Earl O. Heady and Russell Shaw. 1/ The Heady-Shaw study was made in 1950 and the results were published in 1955. The sample used by Heady and Shaw was taken from a more restricted area 2/ than was the sample for the present study, but comparisons of a general nature can be made. Still another study from the same data used by Heady and Shaw but with different functions was done by Darrell F. Fienup. 3/ The Fienup study provides another comparison for evaluating the income tax return approach.

1/ Earl O. Heady and Russell Shaw, Resource Returns and Productivity Coefficients in Selected Farming Areas of Iowa, Montana, and Alabama, (Ames, Iowa, Agricultural Experiment Station, Research Bulletin, 425, April 1955).

2/ Ibid., p. 337.

3/ Darrell F. Fienup, Resource Productivity on Montana Dry-land Crop Farms, (Montana State College, Agricultural Experiment Station, Mimeograph Circular 66, June 1952).

B. Framework of Analysis

Motivation for this study must be identified since research without specific motivation serves a useful purpose only by accident.

The purpose of this study is to attempt to determine the usefulness of income tax data in establishing economic relationships through analysis. The motivation for this is twofold. If a pool of reliable and easily accessible data is established, this should provide incentives for comprehensive usage in nearly all phases of economic research. The number of possible research projects in which the data could be used either in a primary way or as a supplement to data from other sources is very large.

Using the income tax data as a supplement to data gathered by other means has a particular appeal. It seems plausible that financial information about an enterprise is at best, difficult to obtain through an interview or schedule. If this information is obtained, it is usually subject to question. If, however, the researcher has the financial information and needs only to supplement it with other less private information through the interview or schedule technique, the probability of getting that information is good. If questions are asked that do not offend the person or invade his privacy the quality of the answers will be higher than if the questioned person were asked about intimate financial detail. In addition, having the financial information before making out a questionnaire will in most cases shorten its length and thereby relieve another chronic problem associated with the schedule or interview technique of collecting data.

The second part of the twofold objective for this study is to derive a production function for both livestock and grain farms in Montana. Very little has been done in this area using Montana data and it is hoped this study will in some part remedy this neglect. Also, by comparisons to the Fienup 4/ and Heady-Shaw 5/ studies, some indication of technological growth in Montana may be apparent.

C. Hypothesis

Income tax data along with supplementary land data will yield significant regression coefficients when the data are used to explain the relationship between inputs and outputs.

D. Sequence of Analysis

Part II is concerned with: (1) problems that occur when using stochastic functions; (2) a brief review of methods available for deriving input-output relationships; and (3) alternative equation forms for use in production function analysis.

One problem dealt with in Part II concerns the effect of excluding the management variable from the regression analysis. It will be shown that the exclusion of this variable tends to bias returns to scale downward and, in the case of the Cobb-Douglas function will bias the coefficient of production for capital upward.

4/ Fienup, op. cit.

5/ Heady and Shaw, op. cit.

The problem of multicollinearity has subjected many studies to serious question. A discussion of the effects of multicollinearity and the circumstances under which it becomes a problem is also included in Part II. Multicollinearity, when high enough, makes interpretation of the results difficult because the influence of each variable taken separately is difficult if not impossible to measure. A reasonable criterion for determining whether multicollinearity is a serious problem will be suggested.

Statistically derived cost functions have been criticized by Professor Friedman because of what he calls the "regression fallacy". Such criticism is difficult to counter. The "regression fallacy" involves the downward bias created by such random occurrences as hail and drouth. If, for example, a substantial portion of a random sample of farms lost much of their crop because of hail, a cost curve derived from this sample would indicate "economies of scale" when in reality the apparent "economies" were a result of hail damage. It will be shown that when a Cobb-Douglas production function is used, the bias will occur in the constant term.

The remainder of Part II will be directed toward the available methods of deriving input-output relations and to alternative forms of production functions. Both simultaneous equations and regression analysis for a single equation model will be discussed as possible methods of deriving the relationships needed in this study. The Spillman function, the Cobb-Douglas function, the Quadratic function, and the Square Root function will be considered as possible forms of stochastic production function for application to the income tax data.

Part III contains the analysis of the income tax data. Since the data are vital to the thesis, they will be described and defined at length. The various shortcomings of the data will be detailed and the methods used for minimizing the effect of these shortcomings will be explained. Types of functions used and the results derived will be explained and compared to previous results derived from Montana data. Some indication of how reasonable the results of this study are when compared to previous results should be apparent.

Part IV will contain the summary and conclusions and recommendations for further research.

PART II

PROCEDURAL REVIEW

In recent years there have been significant developments in the estimation of stochastic production functions. These development have given the economist powerful tools for the interpretation and statistical justification of derived input-output relationships. However, these developments are not without their attendant problems. In the following paragraphs one of the important developments, the effect of omitting the management variable, will be discussed. The problem of multicollinearity and the "regression fallacy" will also be discussed. 1/

It is common in empirical work to make do with something less than economic theory would indicate. The problem is in general one of measurement, or more specifically the lack of measurement. Implicit in the problem of measurement is the task of identification. For example, in production theory, output is generally considered to be functionally related to four kinds of variables--land, labor, capital, and management. Measurement of the management variable has thus far presented the researcher with insurmountable difficulties primarily because of lack of a unit of measure in order to quantify management. The absence of the management variable in a production relationship tends to bias the coefficients of the included variables if the management variable does

---

1/ For a discussion of other significant statistical problems see: J. Johnston, Econometric Methods, (McGraw-Hill Book Co., New York, 1960), Chapters 6, 7, and 8.

not vary in the same proportion as the included inputs, thus presenting a problem of identifying both the bias and its direction. This problem has a direct bearing on this thesis since the management variable was not included. Fortunately, a method of determining the effect of this omission has been developed and will be discussed later.

Another question that inhibits stochastic production functions has to do with the functional relationship between inputs. This problem becomes particularly acute with the labor and capital inputs. When labor is functionally related to capital, what statisticians call "multicollinearity" results. Labor and capital may be so highly correlated that disentangling their separate influences is extremely difficult if not impossible. This makes statements about the relative effects of the two variables on output statistically and economically meaningless unless from a theoretical standpoint it is expected. Some means of making a judgment as to whether to multicollinearity presents an obstacle to meaningful economic statements must be used. A means of handling this problem will be taken up in a later section.

A point raised by M. Friedman 2/ and G. Stigler 3/ on the use of cross-sectional accounting data for estimating cost functions has been very difficult to counter. Although the question has been raised in

---

2/ Business Concentration and Price Policy, (Princeton, 1952), pp. 230-38.

3/ G. J. Stigler, The Theory of Price, (New York, 1952), pp. 143-44.

connection with cost functions, some analysis on the effect of the regression fallacy on a production function derived in a cross-sectional study from accounting data is in order and will be taken up in Section C.

A. Specification Bias 4/

It was pointed out at the beginning of this chapter that in empirical work we must make do with something less than the ideal suggested by theory. In input-output research we generally make do without the benefit of a measured management input variable. An attempt to shed some light on the question of how much we can say about a variable for which we cannot measure follows.

In a regression analysis one attempts to estimate the regression coefficients. When fitting a production function of the Cobb-Douglas type, these regression coefficients sum to the elasticity of production. It will be shown that when we know something about the management variable a priori, general statements as to the bias generated by excluding that variable can be made.

Consider two multiple regression equations:

$$y = Xa + u \qquad (2\text{–}1)$$

$$y = X^*a^* + u^* \qquad (2\text{–}2)$$

where: y is $(n \times 1)$ $\quad a$ is $(p \times 1)$ $\quad a^*$ is $(k \times 1)$
X is $(n \times p)$ $\quad X^*$ is $(n \times k)$ $\quad k < p$

4/ This section based on: Zvi Griliches, "Specification Bias in Estimates of Production Functions"; Journal of Farm Economics, February 1957, pp. 8-20.

Assume that (2–1) is the true equation and because of lack of data or units of measurement, (2–2) is the equation to be estimated. To illustrate the difference between X and X^* let X be n x 5 and X^* be n x 4.

$$X = \begin{bmatrix} 1 & x_{12} & x_{13} & x_{14} & x_{15} \\ \cdot & & & & \cdot \\ \cdot & & & & \cdot \\ \cdot & & & & \cdot \\ \cdot & & & & \cdot \\ 1 & x_{n2} & x_{n3} & x_{n4} & x_{n5} \end{bmatrix}$$

$$X^* = \begin{bmatrix} 1 & x^*_{12} & x^*_{13} & x^*_{14} \\ \cdot & & & \cdot \\ \cdot & & & \cdot \\ \cdot & & & \cdot \\ \cdot & & & \cdot \\ 1 & x^*_{n2} & x^*_{n3} & x^*_{n4} \end{bmatrix}$$

The columns represent variables and the rows represent observations. In this case, the columns of X would represent land, labor, capital, and management and the columns of X^* would represent land, labor, and capital. So the difference in X and X^* is the management variable.

It is necessary to know the relation between a^* and the true parameters, a. The estimate for a^* is

$$a^* = (X^{*\prime}X^*)^{-1}X^{*\prime}y \qquad (2\text{–}3)$$

Substituting for y,

$$E(a^*) = E[(X^{*\prime}X^*)^{-1}X^{*\prime}(Xa+u)] = (X^{*\prime}X^*)^{-1}X^{*\prime}Xa \qquad (2\text{–}4)$$

The expectation of the estimates have now been expressed in terms of the true parameters and each estimate is a weighted sum of all the true parameters. If we had exogenous information we may be able to say something about these weights.

Letting

$$P = (X^{*'}X^{*})^{-1}X^{*'}X \tag{2-5}$$

then we have $E(a^{*}) = Pa$ from (2-4). In order to get at the desired results, think of PX^{*} as the X matrix in the general linear model. Each row of the P matrix can be thought of as the regression coefficients associated with the columns of X^{*}. Therefore,

$$x_{ij} = p_{i1}x^{*}_{1j} + p_{i2}x^{*}_{2j} + \ldots + p_{ip}x^{*}_{pj} + v^{*}_{j} \tag{2-6}$$

Where p_{ij}'s are functions of the estimated variables, x_{ij}'s are the true variables, and v^{*}_{j} is the least squares residual. By knowing the size and sign of the p_{ij}'s the size and sign of the bias in the estimates can be determined.

Using the preceding argument as a reference point, it is now necessary to look at the specific problem of bias in returns to scale and the coefficients of production when the management variable is omitted from the analysis.

It first must be established whether bias exists and if so, where. It seems plausible to assume that there is some positive correlation between the excluded variable and some of the included variables. This will bias at least one of the coefficients upward and depending on the

size of this bias will leave returns to scale biased downward or not biased at all. The expected bias in returns to scale is readily seen by the following. Let

$$e = \sum_{i=1}^{p} a_i \qquad (2\text{–}7)$$

be the true returns to scale and

$$\hat{e} = \sum_{i=1}^{p-1} a_i^* \qquad (2\text{–}8)$$

the estimated returns. 5/ The expected differences between (2–8) and (2–7) will yield the bias:

$$\text{Bias} = E(\hat{e} - e) = E\left(\sum_{i=1}^{p-1} a_i^* - \sum_{i=1}^{p} a_i\right) = \sum_{i=1}^{p-1} a_i - \sum_{i=1}^{p} a_i = -a_p$$

If there is no correlation between the included variables and the excluded variable, $-a_p$ will represent the bias in returns to scale but all the estimated coefficients of production will be unbiased. The following discussion takes up the more likely possibility of some correlation between at least one of the included variables and the excluded variable.

To determine the bias on each of the coefficients of production, consider the expectation of these coefficients as given by

$$E(a_i^*) = a_i + p_{ip} a_p \qquad (2\text{–}9)$$

Where p_{ip} is the i-p^{th} element of the P matrix. 6/ Assuming $a_p > 0$, the

---

5/ A discussion of this problem in a covariance analysis context is given by: E. Malinvaud, Statistical Methods of Econometrics, (Rand-McNally, Chicago, 1966), pp. 263-66.

6/ By partitioning P into an identity matrix and a column vector, P = [Ip] since X* is equal to X except for the last column.

sign and magnitude of p_{ip} will determine the bias of any particular a_i^*. It has therefore been determined that on the average, some of the coefficients will be overestimated. The unlikely possibility of a negative association between x_p and some of the x_i's is admitted but not considered. The effect of a negative association would be to underestimate some of the coefficients.

It has been established now that we can have bias in both the returns to scale and in the individual coefficients of production caused by specification error. The specification error encountered most often results from the omission of the management variable. It is hypothesized by Professor Griliches that

> . . .while it is reasonable to assume that certain supervisory functions will increase proportionately with total inputs, it does not seem reasonable that decision making and planning, which account for a substantial share of these inputs, will also increase proportionately. 7/

This would mean that returns to scale would be underestimated.

The question of how this estimation bias is distributed among the various input coefficients is a more difficult one. It seems plausible that firms with a relatively high level of management input would be less subject to capital rationing. This would mean there would be a positive correlation between the capital and management inputs. This line of reasoning leads one to the conclusion that in equation (2–9), p_{ip} is positive and hence returns to capital would be overestimated.

7/ Zvi Griliches, op. cit., p. 13.

In one of the few attempts to measure the management input, F. J. Reiss 8/ derived a function that confirms the above argument. The coefficient most influenced by the inclusion of the management variable was the capital input coefficient. It decreased from .174 with management excluded to .143 with management included. This difference, however, was not significant nor was the management coefficient. Griliches suggests that this may be due to the inherent difficulty in constructing an index to measure the management variable. 9/ Also, the sum of the elasticities went from .974 to 1.173 which was also not a significant difference. Although these results are by no means conclusive, they do give additional information about the characteristics of production functions and thereby allows hypotheses to be made from a sound logical foundation.

B. Multicollinearity

Multicollinearity becomes a problem in regression analysis when the independent variables are related to one another. The degree of relationship can vary from zero to an exact functional relation. In regression analysis, this relationship is measured by the correlation between the various independent variables. When this correlation becomes too high it is difficult, if not impossible, to disentangle the separate influences

---

8/ See: Earl B. Swanson, "Determining Optimum Size of Business From Production Functions", in Resource Productivity, Returns to Scale, and Farm Size, ed. by Earl O. Heady, Glenn L. Johnson, Lowell S. Hardin, (Iowa State College Press, 1956), p. 137, see footnote 10.

9/ Zvi Griliches, op. cit., p. 14.

of the correlated variables. In production theory, such interrelationships between inputs are natural and expected. One would certainly be very surprised for example, to discover a zero correlation between land and labor. On the other hand, it would be very surprising indeed, to find an extremely high correlation between these same variables, because of the substitutability of the inputs. The point to be made is that perfect multicollinearity is not a problem in derived production functions but that something less than perfect multicollinearity is a problem. The real question to be asked is: When does the problem become so severe that the results are unreliable?

In order to answer this question in a reasonably rigorous fashion, consider the following linear model.

$$y = \beta_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \varepsilon \qquad (2\text{–}10)$$

In order to estimate the betas, the product (X'X) must be inverted. This estimate is

$$\hat{\beta} = (X'X)^{-1}X'Y \qquad (2\text{–}11)$$

and the variance of $\hat{\beta}$ is

$$\text{var}(\hat{\beta}) = \sigma^2(X'X)^{-1} \qquad (2\text{–}12)$$

It can be readily seen that the existance of the inverse of X transpose X plays an important role in estimating beta and the variance of beta. If there exists a linear relationship between two or more of the variables

in our model then (X'X) is called a "singular matrix" and has no inverse. 10/ This can be determined by evaluating the determinant of (X'X). 11/ If $|X'X| = 0$ then (X'X) has no inverse. In this case, problems of serious nature are apparent. Obviously $\hat{\beta}$ or the variance of $\hat{\beta}$ cannot be estimated. But the situation is not hopeless if independent estimates of the dependent parameters can be obtained. 12/ This condition would be unusual in input-output situations and it will be dwelled on no further.

A much more common problem is the less severe case where the independent variables are correlated but not perfectly correlated. To demonstrate how ample warning of multicollinearity will be evidenced, consider the following. Assume a three-variable system in which one of the variables is dependent. Then,

$$\text{var}(\hat{\beta}_2) = \frac{\sigma^2}{\sum x_{2\cdot 3}^2} \qquad (2\text{–}13)$$

where $x_{2\cdot 3} = x_2 - b_{23}x_3$. The x's refer to deviations, not raw data. The b_{23} is the least squares estimate of the slope in two-space between x_2 and x_3.

As the correlation between x_2 and x_3 increases, the magnitude of $x_{2\cdot 3}$ decreases and hence, so does $\sum x_{2\cdot 3}^2$. Therefore, (2–13) will increase

---

10/ Sam Perlis, Theory of Matrices, (Reading, Mass., Addison-Wesley Publishing Co., 1952), p. 60.

11/ Ibid., p. 79.

12/ J. Johnston, op. cit., p. 202.

in magnitude and this provides a means of identifying the imprecision of the estimates. If the standard error of the betas is too large, multicollinearity presents a problem. Statistically, the researcher tends to make Type II errors. That is, he tends to accept the null hypothesis when it is not true.

One means of determining the relative size of the standard error is to calculate the coefficient of variation. If it exceeds some value based on the judgment of the researcher, then it may be concluded that the function has some multicollinearity. Should this strong relation between inputs be discovered, the reliability of the results will of course be diminished.

C. The "Regression Fallacy"

Although the "regression fallacy" has been brought up in connection with stochastic cost functions, the argument is germane to statistically derived production functions. The criticism has been made by both Stigler 13/ and Friedman. 14/ Friedman expresses his criticism as follows:

> Suppose a firm produces a product, the demand for which has a known two-year cycle, so that it plans to produce 100 units in year one, 200 units in year two, 100 units in year three, etc. Suppose also that the best way to do this is by an arrangement that involves identical outlays for hired factors in each year (no 'variable' cost). If outlays are regarded as total costs, average costs per unit will obviously be twice as large when output is 100 as when it is 200. If

---

13/ Stigler, op. cit.

14/ Friedman, op. cit.

> instead of years one and two, we substitute firms one and two, a cross-section study would show sharply declining average costs. When firms are classified by actual output, essentially this kind of bias arises. The firms with the largest outputs are unlikely to be producing at an unusually low level; on the average they are clearly likely to be producing at an unusually high level, and conversely for those that have the lowest output. 15/

The fallacy is that a statistically estimated average cost function would show economies of scale that, in reality, are not present.

It seems clear that Professor Friedman is thinking of random fluctuations rather than a seasonal or trade cycle pattern. 16/ If he were thinking of seasonal or cyclical variation, there would be no "regression fallacy" since all firms would be subject to the same pattern. Given that this is the case, Friedman's example would be quite applicable to a sample of farm firms with weather being the prime source of random variation. If the sample covered a large geographic area, it is quite plausible that some firms in the sample have bad weather conditions, hail, for example, and other firms have relatively good weather. Another source of variation would be government programs but this variation could be thought of as the type brought about by trade cycles. The "regression fallacy" would not exist because the pattern of government programs applies to all firms.

In order to identify where the bias would occur in a cost function derived from a Cobb-Douglas type production function, consider a production function of the following type:

15/ Business Concentration and Price Policy, op. cit., p. 236.

16/ A. A. Walters, "Expectations and the Regression Fallacy in Estimating Cost Functions", Review of Economics and Statistics, May 1960, p. 210.

$$P = AX_1^{b_1} X_2^{b_2} \qquad (2\text{–}14)$$

where: P = output

X_i = inputs, $i = 1, 2$

b_i = coefficients of production, $i = 1, 2$

A = constant

This production function will yield the following total cost function. 17/

$$C = \frac{r_1^{b_1} r_2^{b_2}}{A b_1^{b_1} b_2^{b_2}} P \qquad (2\text{–}15)$$

where: r = factor input prices, $i = 1, 2$. Friedman's criticism will cause a downward bias in C in the relevant range. Since inputs are assumed to remain fixed, the source of the downward bias must be A. As output decreases, A increases, and C decreases. Since the constant term rarely has an interesting economic meaning 18/ we may conclude that the "regression fallacy" has little effect on the input-output relationships of interest to this study.

One additional thought on this matter. It has been proven that no bias exists if one can assume a normal distribution of the random

---

17/ This derivation can be found in: J. M. Henderson and R. E. Quandt, Microeconomic Theory, (McGraw-Hill, New York, 1958), pp. 66-7.

18/ Zvi Griliches, op. cit., p. 19.

variations of the "regression fallacy". 19/ Without this assumption, however, special cases do exist that are biased. An example is the rectangular distribution. 20/

D. Methods Available

In addition to deriving a production function statistically, another method of arriving at information about production is available. Although no claim to having done justice to this method is made, at least a cursory look at simultaneous equations as it applies to estimation of input-output relationships is in order.

Simultaneous Equations Model

The simultaneous equations method for estimating coefficients of production can be said to be relatively unsuccessful thus far in economic research. 21/ The difficulty with estimation through simultaneous equations is inherent in the method. To develop a system of simultaneous equations one must specify exactly, using prior information, all the simultaneous relationships associated with the variable of interest. Professor Fisher, in his new book, had this to say:

> . . . it does not suffice to know that the equation to be estimated contains precisely a specified list of variables.

19/ J. J. Johnston, "Statistical Cost Functions: A Re-Appraisal", Review of Economics and Statistics, November 1958, pp. 349-50.

20/ Ibid., p. 350.

21/ Franklin M. Fisher, The Identification Problem in Econometrics, (McGraw-Hill Book Co., 1966), p. 2.

> It is also necessary to know what variables are contained in other simultaneously holding equations or to have other information about the equation in question. 22/

Professor Fisher goes on to point out why this method is necessary:

> One cannot hope to know the parameters of the equation in question on the basis of empirical observations alone, no matter how extensive and complete those observations may be. Observational information cannot 'identify' the equation to be estimated--cannot distinguish it from a host of other possibilities each as capable of generating the observed data as the true one. 23/

Through simultaneous equations the problem of identification comes up. Identification is defined as:

> . . . the problem of computing the parameters of the structure, which is presumed to have generated the observations of the endogeneous variables, from the parameters of the likelihood function. 24/

If the system of equations is exactly identified, then ordinary least squares will suffice to estimate the parameters. If the system is under-identified there exists no statistical means of estimating the parameters. If the system is over-identified, methods have been devised to handle the problem. Some of these are limited-information single equation, two-stage least squares, three-stage least squares, and full-information maximum likelihood.

This problem of identifying the system is by no means trivial and has been the "Achilles Heel" in all economic research using this model as a

---

22/ Franklin M. Fisher, op. cit., p. 2.

23/ Ibid, p. 2.

24/ Johnston, op. cit., p. 243.

means of estimating parameters. Although this method offers great hope of being an exceptionally accurate predictive device in future research, the mathematical and statistical problems at the present time are substantial. 25/

Stochastic Production Function

Stochastically determining a production surface seems, at this point in time, the most commonly used means available. This method is certainly not ideal, but few if any methods employed in economic research are ideal. The statistical production function attempts to imitate some relevant part of a theoretical function with the use of real world data. The theory of production is essentially the analysis of relative extremums of mathematical relationships. These extremums are of special interest to economists when they can be quantified. One method of quantification is statistical estimation. Since statistical estimation employs the sum of squares in minimizing the deviations about a mean, some observations will necessarily lie above the estimated production function and some will lie below. The difference between a theoretical function and a derived function is that the derived function does not represent a maximum for a given level of input (see Figure 1, B). Whereas the theoretical function does represent an upper bound on production. In Figure 1, A, production may occur anywhere in the shaded area.

25/ Johnston, op. cit., pp. 294-5.

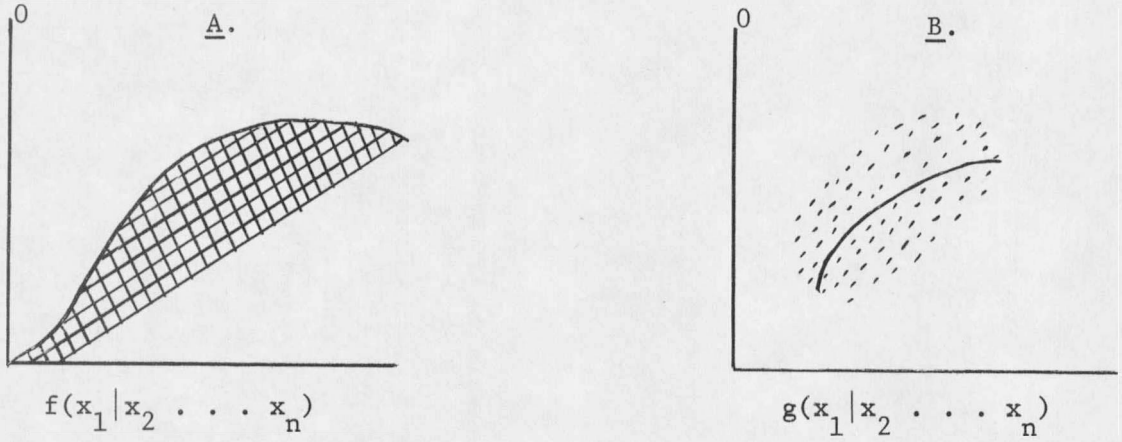


Figure 1.

Two methods are available to directly get at the notion of a theoretical production function. One of these methods is to run successive least squares regressions throwing out observations which lie below the fitted surface. This has the effect of lifting the function shown in Figure 1, B upward. The other method developed by M. J. Farrell employs the concept of an efficient production function. 26/ By finding the observed upper bound of a production surface, one finds the real world counterpart of the theoretical production function.

It will also be noticed that the derived function does not include extremely low levels of input or extremely high levels of input. Therefore, statements about its characteristics in areas of output for which no data exist are extremely hazardous at best. The only portion of the function that is statistically relevant is that portion where production

---

26/ M. J. Farrell, "The Measurement of Productive Efficiency", Journal of the Royal Statistical Society, Part III, March 1957, p. 253.

takes place. So it seems apparent that statistically derived production functions are certainly something less than ideal. The question is: Do they provide us with reasonable and reliable estimates of the production coefficients theory suggests? On this point it can only be said that statistically derived functions are better than nothing. Beyond that the argument becomes subjective and a matter of individual preference. And as it happens, in this thesis the preference is for a statistically derived function.

E. Alternative Types of Functions

There are several different types of regression equations that might be used to explain input-output relationships. We will consider four types.

The Spillman Function

Equation (2-16) is a form of production function commonly called the Spillman Function.

$$Y = M - AR^X \qquad (2\text{-}16)$$

M is the maximum yield which can be attained by use of the particular input; A is the yield with zero input, R is the ratio by which successive increments are added to total output, and X is the quantity of input. 27/ The same function with two variables inputs is given by

---

27/ Earl O. Heady, "Technical Considerations in Estimating Production Functions", in Resource Productivity, Returns to Scale, and Farm Size, ed. by Earl O. Heady, Glenn L. Johnson, and Lowell S. Hardin, (Iowa State College Press, Ames, 1956), pp, 8 and 9.

$$Y = M(1 - R_1^{X_1})(1 - R_2^{X_2}) \qquad (2\text{–}17)$$

Generalization to n variable is an obvious possiblity.

There are certain implicit assumptions contained in this type of function. First, since the function asymptotically approaches the limit of M (i.e., AR^X tends toward zero as X becomes large), it never reaches a maximum in a finite sense. The second assumption is that the elasticity of production changes as output changes, but that the ratio of marginal products is constant over all ranges of output. 28/ Symbolically this assumption is stated as follows:

$$Ep = \frac{dY}{dX}\left(\frac{X}{Y}\right) = (-AR^X \text{Log } R)\frac{X}{Y} = \frac{-XAR^X \text{Log } R}{M - AR^X} \qquad (2\text{–}18)$$

This is the elasticity of production for equation (2–16). Equation (2–18) is obviously not a constant. If we now look at the derivative of (2–16) it will be noticed that it varies as X varies and in multiples of R. Therefore,

$$\frac{dY_i}{dX_i} = R\left(\frac{dY_{i-1}}{dX_{i-1}}\right) \qquad (2\text{–}19)$$

and by rearrangement we get,

28/ Earl O. Heady, and John L. Dillon, Agricultural Production Functions, (Iowa State Press, 1961), p. 77.

$$R = \frac{\frac{dY_i}{dX_i}}{\frac{dY_{i-1}}{dX_{i-1}}} \qquad (2\text{—}20)$$

Equation (2—20) is a constant, with the i subscripts denoting levels of input of X.

One of the primary difficulties with this type of function is inherent in the equation itself. One can readily imagine the obstacles in getting numbers for R_i in (2—17) and still maintain a reasonable degree of reliability.

The Cobb-Douglas Function

The Cobb-Douglas function for the one variable case is as follows:

$$Y = AX^b \qquad (2\text{—}21)$$

Y is output, A is a constant term, X is the input, and b is the production coefficient, and in the case of one input the elasticity of production. For the case of n inputs,

$$Y = AX_1^{b_1} X_2^{b_2} \ldots X_n^{b_n} \qquad (2\text{—}22)$$

and the elasticity of production is

$$e = \sum_{i=1}^{n} b_i \qquad (2\text{—}23)$$

When e is less than one, we have decreasing returns to scale, when e is equal to one, we have constant returns, and when e is greater than one, we have increasing returns. This phenomenon can be illustrated as follows:

$$\frac{dY}{dX} = bAX^{b-1} = \frac{bAX^b}{X} \qquad (2\text{–}24)$$

$$e = \frac{bAX^b}{X}\left(\frac{X}{Y}\right) = b\left(\frac{Y}{X}\right)\left(\frac{X}{Y}\right) = b \qquad (2\text{–}25)$$

It will also be noticed that the marginal product is constant if and only if b = 1. If b is greater than one, MP increases as X increases and if b is less than one, MP decreases as X increases.

The General Quadratic Function

The General Quadratic function

$$Y = a + b_1X_1 + b_2X_2 - b_3X_1^2 - b_4X_2^2 + b_5X_1X_2 \qquad (2\text{–}26)$$

shown here for two variables, provides much greater flexibility than either the Cobb-Douglas or Spillman functions. The General Quadratic form assumes nothing about the marginal products or returns to scale. Both can change with changes in input, and the function allows a positive but diminishing product, following a negative marginal product or declining total output. A cubed term can be added to give increasing marginal product as well as decreasing marginal products. 29/

29/ Earl O. Heady, "Technical Considerations in Estimating Production Functions", op. cit., p. 10.

The Square Root Function

This function is very much like the quadratic function in appearance but considerably different in results.

$$Y = a - b_1X_1 - b_2X_2 + b_3\sqrt{X_1} + b_4\sqrt{X_2} + b_5\sqrt{X_1X_2} \qquad (2\text{–}27)$$

In order to use this function the data must be such that the researcher suspects large marginal products over small inputs and small marginal products over large quantitites of inputs. 30/ In other words, the production surface will increase relatively fast for small inputs and then the increase will slow down radically for large inputs.

30/ Earl O. Heady, op. cit.

PART III

THE ANALYSIS

A. The Data

Since this is a study based on a random sample of farms, it is necessary to define the population from which the sample was drawn. The population consisted of all tax returns filed in the state of Montana in 1964. From this population a sample of 5,045 returns was taken without regard to type. That is, the sample included personal, partnership, farm, and in some cases, corporate returns.

The farm returns in the sample were identified by the type of tax schedule filed. All persons who receive farm income or incurred farm expenses are eligible to file their tax return using schedule F (form 1040). All tax returns using schedule F were selected as the initial farm sample. (See Appendix A.)

Initial Deletions

Before the initial deletions, 669 schedule F returns were contained in the original sample of 5,045 tax returns. Certain returns were eliminated from the sample for various reasons.

First, all partnerships were deleted from the 669 schedule F returns because of inherent duplications. The reason for the duplication is that the return number was used as the control number, and each partner in a partnership has a different return number. This resulted in the possibility of the sample having as many duplicate schedule F returns as there were partners in any given partnership operation. It might be argued that

since the sample was random, the a priori probability that two or more partners of the same firm would be in the sample would be low, and that the gain in the number of observations through the inclusion of partnerships would be worth the risk of duplication. In the process of inspecting the data, several instances of duplication were discovered and traced to partnerships. On the basis of this inspection, the partnership returns were deleted from the sample. In addition, in many partnerships the land is owned by individuals and if partnerships were included in the sample, only that portion of land owned by the partner in the sample would be included in the sample thereby creating an error in the land input data.

Second, returns which indicated that the individual had non-farm income and no farm income but had incurred farm expenses were also deleted from the sample. An example of such a case is the farmer just beginning in business. He may have purchased the land too late to raise spring wheat, but expenses for summer fallowing and seeding of winter wheat would have been incurred. Such a taxpayer probably held some type of non-farm job before purchasing the farm and may have kept that employment as a source of income until the farm began producing income the following year.

Returns which indicated both non-farm and farm income but no farm expenses were also deleted from the sample. Farmers who sold out but received income from last year's crop would fit in this category. Retired farmers or farmers with rental arrangements would also be part of this group. It was found that a substantial percentage of this

group were people who received income from retained oil or mineral rights on land previously sold.

There were also a few cases in which the individual farmer had filed an amended tax return. In these cases, the only figure shown on the amended schedule F was the debit or credit to the specific income or expense account amended. It was necessary to combine the original return and all the amended returns. In some cases there were more than one amended return. The return, aggregated with its amendments, was then included in the sample.

The Land Data

The land data needed was determined by the sample of farm returns. This information was obtained from the counties by sending a form to be filled out by the county agents. (See Appendix B.) The forms were returned with the land owned by each taxpayer listed and classified as indicated by county records. All farm land is classified into six categories. These are: tillable irrigated, non-irrigated farm, non-irrigated continuously cropped, wild hay, grazing, and timber land. For purposes of this study the land data were aggregated into three categories. Tillable irrigated was one category, non-irrigated farm and non-irrigated continuously cropped was another, and wild hay, grazing, and timber land was the third category.

The land data furnished by the counties was for the year 1965. There was no convenient way to get the farm acreages for 1964. But the size of

farms in Montana did not change substantially between 1964 and 1965. The number of farms by size for 1959 and 1964 is shown in Table I, Appendix E, along with the percentage change in the number of farms by size. Since the number of acres farmed increased only .05 percent per year and the average size farm increased about 1.8 percent between 1959 and 1964, 1/ using 1965 land data to represent 1964 doesn't appear to introduce much error. See Appendix E for a more comprehensive treatment of this problem.

In the case where whole farm operations changed hands between 1964 and 1965 and the farmer that sold out was in the sample, the record of land classification was returned with a note indicating no farm land owned. That taxpayer was then deleted from the sample.

The problem with gathering land data in this fashion is that nothing is known about the amount of land leased or rented. This problem was solved by selecting all those taxpayers indicating rent as an expense on their tax return. If the amount of rent paid exceeded $100, a questionnaire was sent to the taxpayer asking how much land in acres he leased in the year 1964. (See Appendix B.) The question was framed so as to classify the land as tillable-irrigated, non-irrigated cropped, or grazing. When the questionnaire was returned, the leased land was aggregated with the land owned. Those farms that did not return the questionnaire were deleted from the sample.

1/ U. S. Bureau of the Census, Census of Agriculture, Statistics for the State and Counties, Montana, (U. S. Government Printing Office, Washington, D. C., 1967), p. 8.

Merging Land and Tax Data

Various problems with the sample were noticed during the process of inspecting the data. Certain methods were used to select observations from the sample and aggregate the data in the desired fashion. Once this rather laborious but necessary task was completed it was relatively simple to employ the use of electronic accounting machines (EAM) and a computer to delete and aggregate the data in the desired fashion.

The initial deletions from the farm tax returns were made by the use of a collator. The collator board was wired so as to select out all unmatched return numbers, whether from the master file or detail file. When this process was complete each observation had all the information required for aggregation of the income tax data. In addition to providing a fast means of determining the observations to be included for further processing, several keypunching errors were discovered and corrected.

The next step in the processing was to aggregate and transfer the tax information from four cards to one card. Once this was accomplished, the tax data was ready to be processed. The next process was preparing the land data to be merged with the tax data. A program was written to aggregate the land into the three categories previously mentioned and punch this information out along with the return number. By the use of the collator, land and tax data for each return number were matched and a subsequent program punched out the land and tax data on a single card. This one card deck containing all the appropriate land and tax information for each farm in the sample was then divided into two groups, one

containing farms which derived at least half of their income from grain farming and the other containing farms which derived at least half of their income from livestock operations. Some farms derived more than half their income from something other than livestock or grain, sugar beets for example. These farms were also deleted from the sample leaving those farms which derived 50 percent or more of their income from grain crops and those farms deriving 50 percent or more of their income from livestock. These two groups were used in deriving the "grain function" and the "livestock function".

B. The Variables

In both the livestock and grain functions, all variables used were significant at the 5 percent level. A specific description of the variables, with income as the dependent variable and land, labor, and capital the independent variables follows.

Income

In both functions income was considered the total revenue derived from livestock and grain. Because of the heterogeneous nature of observations in the "grain" and "livestock" categories of operation, this type of aggregation seems to make more economic sense than including only grain or livestock income. It would not be reasonable to attempt to explain grain income alone using inputs that were used to produce both grain and livestock income. This fact was clearly illustrated by the difference in statistical results, first using only income derived from

the primary part of the operation then using income derived from the total production process. The results were superior when income derived from total production was used.

Land

The land input for the livestock function includes both grazing land and cropland. An inspection of the data revealed that ranchers with considerable cropland tended to have small expenditures for feed and hay. Cropland was therefore considered to be a proxy for purchased feed and hay, and cropland was accordingly considered to be an appropriate livestock input. Grazing land was excluded from the land input in the grain function. In examining the grain farm data it was found that a few operators had a sizable amount of grazing land and little, if any, livestock income. This created an unwarranted bias that could be corrected by eliminating grazing land from the grain production function.

To normalize the various categories of land, this was done by adding the average assessed values for categories of land and calculating the percentage that the value of each land type was of the total value. 2/ The following equations were used to calculate total land.

2/ State of Montana Twenty-First Biennial Report of the Montana State Board of Equalization, (State of Montana, December 1964), pp. 38-41. Since the land information was in terms of acres and since grazing land, irrigated cropland and non-irrigated cropland should not have equal weight in the production function, it was necessary to weight the types of land according to their respective importance.

Grain:

$$TL = .71(X1) + .29(X2)$$

where: TL = Total Land

X1 = Irrigated Land

X2 = Non-irrigated Cropped land

Livestock:

$$TL = .66(X1) + .27(X2) + .07(X3)$$

where: X1 = Irrigated Land

X2 = Non-irrigated Cropped Land

X3 = Grazing land

Each of the categories of land (X1, X2, X3) were aggregated as previously noted.

Labor

Labor proved to be the most difficult variable to adjust in a reasonable way. The problem was lack of any information about the magnitude of family labor. Any payment for hired labor was, of course, indicated as hired labor expense on the tax return. The problem was to determine how family labor varied as income or size of operation varied. This problem was solved by making the ratio of family to hired labor a function of income.

Table I shows the process of arriving at the ratio of family to hired labor on grain farms in Montana and Table II shows the same thing for livestock farms.

TABLE I. LABOR FOR GRAIN FARMS a/

Class	II	III	IV	V
Operators Working	1291	2139	1348	546
Others Working 15 hrs. or more	654	860	614	184
Others Working less than 15 hrs.	816	1203	698	300
TOTAL	2761	4202	2660	1030
Regular Hired Workers	1260	888	279	30
Estimated Family Workers	1501	3314	2381	1000
Ratio of Family to Hired Workers	1.19	2.68	8.53	33.33
Average Income	27,001	14,342	7,484	3,743

TABLE II. LABOR FOR LIVESTOCK FARMS. a/

Class	II	III	IV	V
Operators Working	629	1034	881	521
Others Working 15 hrs. or more	329	488	417	201
Others Working less than 15 hrs.	558	816	558	247
TOTAL	1516	2338	1856	969
Regular Hired Workers	830	585	160	45
Estimated Family Workers	686	1753	1696	924
Ratio of Family to Hired Workers	1.21	3.00	10.60	20.53
Average Income	27,741	14,279	7,302	3,668

a/ U. S. Bureau of Census, op. cit., p. 94.

When these ratios are plotted with the corresponding average incomes for the income classes 3/ and the points are connected with a smooth curve, the line approximates a rectangular hyperbola. Both hyperbolas for the grain and livestock farms are represented by the dotted lines in Figures 2 and 3.

Since this method of estimating family labor 4/ is rough at best, it was decided that connected linear lines would be adequate for gaining the needed information. A series of three lines were used for the grain farms (Figure 2), and four lines were used for the livestock farms (Figure 3). The equations used are represented by the solid lines in Figures 2 and 3. They were derived by connecting the points listed in Tables I and II.

Family labor is actually determined by substituting total income for X in the appropriate equation and calculating Y. Y is then multiplied by the total hired labor yielding family labor. Total labor used as the input in production is then the sum of family labor and hired labor.

To select the appropriate equation, the level of income must be determined for the grain function; if income was less than $7,530 the equation Y_1 of Figure 2 was used; if income was between $7,530 and $13,440 the equation Y_2 was used, and if income was over $13,440

3/ The income classes represented in Tables I and II are: Class II $20,000 to $29,999; Class III $10,000 to $19,999; Class IV $5,000 to $9,999; and Class V $2,500 to $4,999. Classes I and IV become asympototic to their respective axis. That is, Class I has a very small ratio and Class IV has a very large ratio.

4/ What is referred to here as "family labor" is actually an estimate of labor other than what the Census calls "regular hired labor".

equation Y_3 was used. For the livestock function, equation Y_1, in Figure 3 was used, if income was less than $3,400. If income was between $3,400 and $7,200, equation Y_2 was selected, and if income was between $7,200 and $14,200, equation Y_3 was used. Equation Y_4 was used for incomes over $14,200.

Capital

The capital input consisted of the following expenses: maintenance repairs, interest, feed purchases, seeds, plants, fertilizers, lime, machine hired, supplies purchased, breeding fees, veterinary services, medicine, gas, fuel, oil, storage, taxes, insurance, utilities, rent, freight, conservation expense, retirement plans, vehicle license and tax, farm car expense, cleaning, treating, spraying, and depreciation. No adjustments were made to this input.

C. The Functions

As was pointed out before, two production functions were derived, one for operations deriving 50 percent or more of their income from livestock and one for operations deriving 50 percent or more of their income from grain sales. Each of the functions is discussed in detail below.

Grain Function

The derived grain production function had the following form:

$$X_1 = b_1 + b_2X_2 + b_3X_3 + b_4X_4 \qquad (3\text{-}1)$$

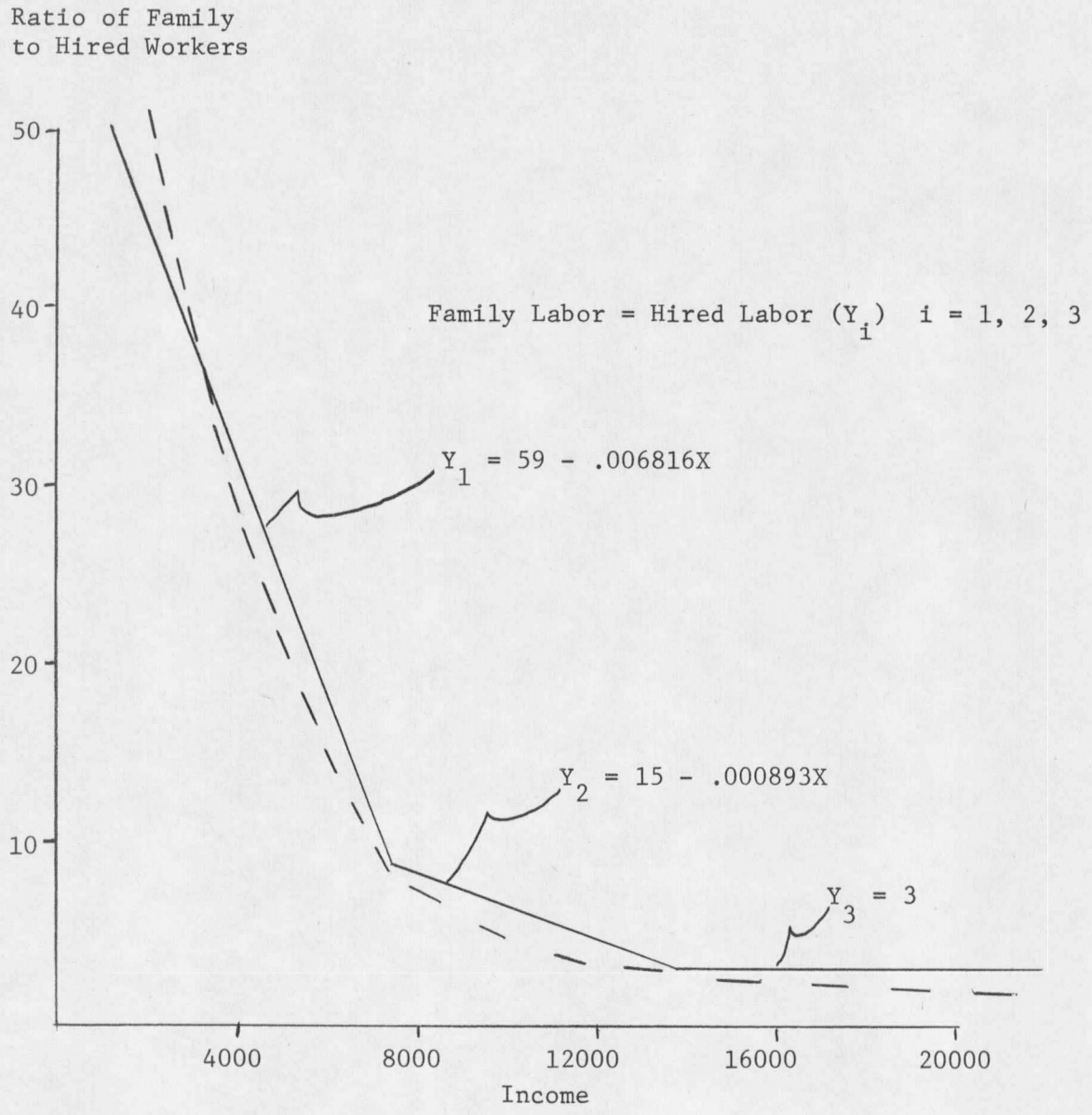


Figure 2. Family Labor for Grain Farms.

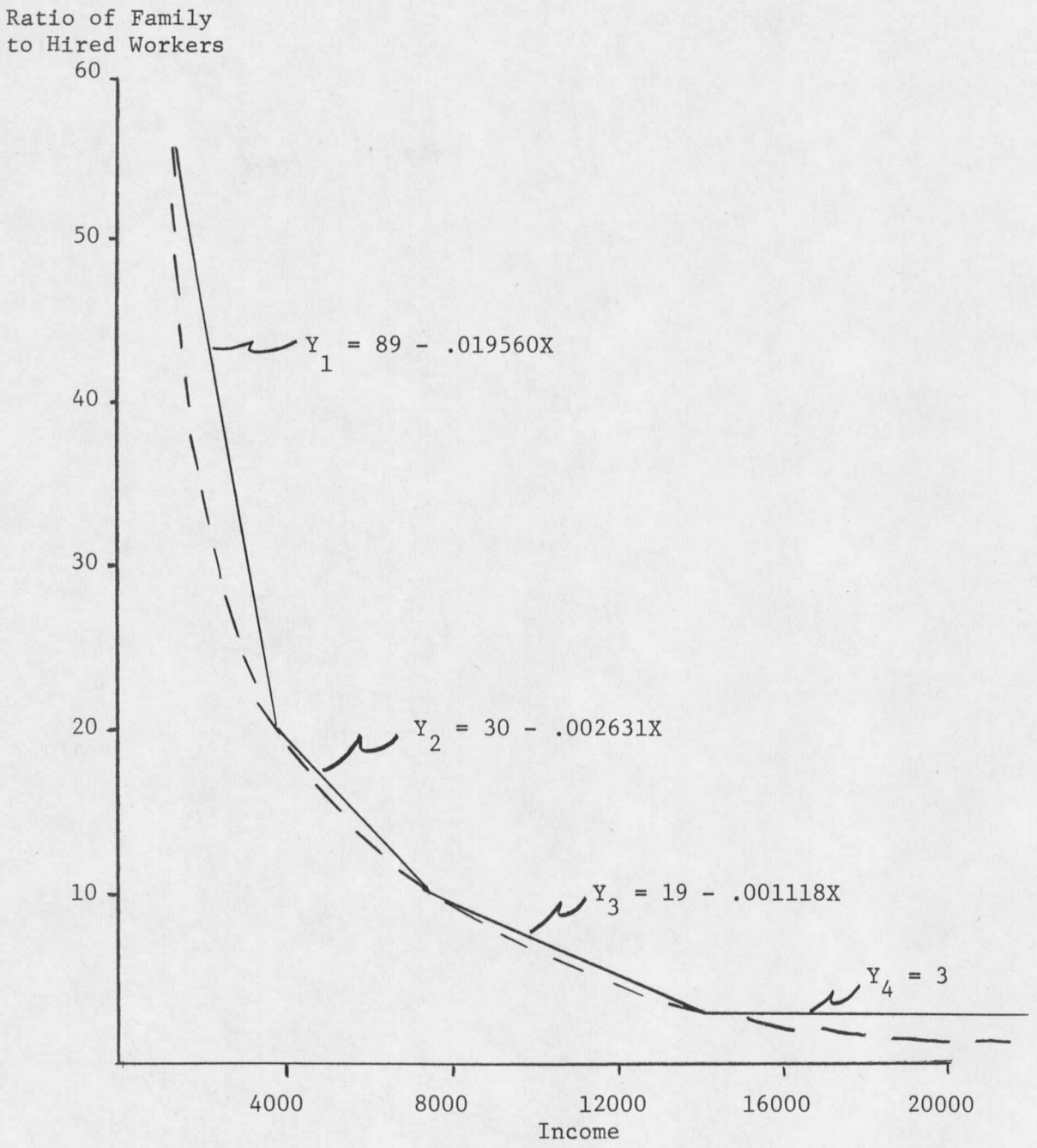


Figure 3. Family Labor for Livestock Farms.

where: X_1 = Total Income

X_2 = Land

X_3 = Labor

X_4 = Capital

All these variables conform to their previous definitions. This function is the general quadratic form under the hypothesis that the regression coefficients on the squared and cross-product terms are zero. Using the general quadratic model does not significantly improve the results obtained by the reduced model,(3-1). That is, the coefficient of multiple determination in the reduced form was not significantly different from that derived with the general quadratic form.

The equation derived from (3-1) is as follows:

$$X_1 = 2171.13 + 11.42X_2 + .47X_3 + .81X_4 \qquad (3\text{-}2)$$

All coefficients were significant at the 1 percent level of probability. This indicates that each included variable does make a significant contribution to production. Statistically, equation (3-2) explains 78 percent of the variation in production on farms which produce primarily grain.

The elasticities of production calcualted by

$$\frac{\partial X_1}{\partial X_i}\left(\frac{\bar{X}_i}{\bar{X}_1}\right) = E_{X_i} \qquad i = 2, 3, 4 \qquad (3\text{-}3)$$

and the returns to scale by

$$\sum_{i=2}^{4} \frac{\partial X_1}{\partial X_i}\left(\frac{\bar{X}_i}{\bar{X}_1}\right) \qquad i = 2, 3, 4 \qquad (3\text{-}4)$$

The elasticity and marginal productivity estimates appear in Table III. The marginal productivities were calculated by a simple partial derivative with respect to each input.

TABLE III. ELASTICITIES OF PRODUCTION AND MARGINAL PRODUCTIVITIES FOR GRAIN FARMS.

	Elasticity a/	Marginal Product	Mean
Land	.1945	11.42	349.66
Labor	.1560	.47	6815.26
Capital	.5450	.81	13816.88
Total	.8955		20534.84 b/

a/ Evaluated at the arithmetic mean.

b/ Arithmetic mean of total income.

It would appear that grain farms in Montana are in a stage of decreasing returns to scale, but a method of statistically testing this inference from the quadratic function is not known.

Theoretically, each of the inputs is being used in something other than optimum quantities. In equilibrium, assuming the existence of competition, marginal productivities for each input should approach unity.

It appears that on the average, grain farms should use less labor and capital and more land. 5/

As a practical matter, the type of sample used should be considered. In a random sample, both efficient and inefficient farms are included. If the rate of decrease in number of farms, in an absolute sense, is correlated with inefficiency in farming, a random sample would be biased toward inefficiency. It must be remembered that many farmers in the sample depend on non-farm income as a supplement to farm income.

With respect to the decreasing returns to scale, it will be recalled that the absence of a measure for the management input will probably bias the scale estimate downward. It is also probable that the marginal product of capital is slightly overstated assuming capital and management are correlated.

Multicollinearity poses no problem in the grain function. The largest coefficient of variation is .8 percent. The standard error for the land coefficient is 2.97, for labor and capital they are .10 and .09 respectively.

Livestock Function

The livestock function derived was a Cobb-Douglas type function as expressed by

---

5/ This inference is tentative. A bias growing out of data limitation is contained in the function. It was not possible to obtain all of the land leased by the grain farms in the sample. For this reason, the marginal products for land would seem to be too high and the marginal products of labor and capital would be too low. A method for eliminating this bias was not available.

$$X_1 = aX_2^{b_2}X_3^{b_3}X_4^{b_4} \qquad (3\text{-}5)$$

where: X_1 = Income

X_2 = Land

X_3 = Labor

X_4 = Capital

The quadratic form of production function was applied to the livestock data, but the results were inferior to those using the Cobb-Douglas function. 6/ The difference between the coefficients of determination as calculated from the two forms was substantial and significant.

The actual function derived was

$$X_1 = .80X_2^{.4848}X_3^{.2704}X_4^{.4902} \qquad (3\text{-}6)$$

The coefficients for land and capital are significant at the 1 percent level and the coefficient on labor is significant at the 5 percent level. The elasticities of production and the marginal productivities for this function are shown in Table IV.

Returns to scale were tested and found to be increasing. That is, the sum of the coefficients of production were found to be significantly different from one at the 1 percent level of probability. See Appendix C for a brief outline of this test.

The marginal productivities indicate relatively inefficient use of inputs. The high marginal productivity of land may be partly due to the

---

6/ The quadtratic equation applied to the same data yielded an R^2 of .56, as contrasted to .74 for the Cobb-Douglas form.

TABLE IV. ELASTICITIES OF PRODUCTION AND MARGINAL PRODUCTS FOR LIVESTOCK.

	Elasticity	Marginal Product a/	Mean
Land	.4848	19.18	163
Labor	.2707	.29	6070
Capital	.4902	.63	4040
Total	1.2457		6450 b/

a/ All evaluations at the geometric mean.

b/ Geometric mean of total income.

normalization of the land input. The elasticity of capital is probably overstated because of the bias that results from the absence of a mangement variable in the function. Since the highest coefficient of variation is 1.6 percent, multicollinearity is not a problem.

Comparisons

It was pointed out earlier that the present study is more or less comparable to earlier production function studies made by Heady and Shaw 7/ and by Fienup. 8/ In the discussion that follows these studies will be referred to as: (1) the Haley study (or function); (2) the Heady-Shaw study (or function); and (3) the Fienup study (or function).

7/ Heady and Shaw, op. cit.

8/ Fienup, op. cit.

In the Heady-Shaw study, the Cobb-Douglas type functions derived were as follows:

Crops: $Y_c = 4.85 D_c^{.5032} L_c^{.0394} C_c^{.5804}$ (3-7)

Livestock: $Y_L = 1.89 L_L^{.0839} C_L^{.9370}$ (3-8)

where: Y_c = Value of Crop Production

D_c = Cropland Services in Acres

L_c = Labor Services in Months including Family Labor

C_c = Capital Services used on Crops

Y_L = Value of Livestock Production

L_L = Labor Services in Months including Family Labor

C_L = Capital Services used on Livestock 9/

Land was measured in acres in both studies. The Heady-Shaw "capital services" is about the same as "capital" in the Haley study. Heady and Shaw derived two functions, one for crops and another for livestock. They attempted to relate the value of crop and/or livestock production to the inputs directed specifically toward that production. In the present study the data were of such a nature that inputs could not be so segregated. Heady and Shaw measured labor in man months, while the current study measured labor in dollars as it was reported on the tax returns. Nevertheless, the Heady-Shaw crop function has about the same variables as those used in this study. The Heady-Shaw livestock function included two

---

9/ Heady and Shaw, op. cit., p. 341.

independent variables, "labor services" and "capital services". Land expense was included in "capital services", rather than made a separate variable as in the Haley study.

In the Fienup study, the functions derived were

$$\text{Crops: } Y_c = 2.45X_{c1}^{.2242}X_{c2}^{.3384}X_{c3}^{.4690} \qquad (3\text{-}9)$$

$$\text{Livestock: } Y_L = 1.97X_{L1}^{.2873}X_{L2}^{.5822}X_{L3}^{.1048} \qquad (3\text{-}10)$$

where: Y_c = Value of Crop Output

X_{c1} = Total Crop Acres

X_{c2} = Value of Machine Services

X_{c3} = Total Cash Crop Expense

Y_L = Value of Livestock Output

X_{L1} = Value of Feed

X_{L2} = Value of Livestock Input

X_{L3} = Capital Expense on Livestock 10/

In both (3-9) and (3-10) labor was included as a variable but in neither was it significant.

There are considerable differences in Fienup's variables and those of this study. The most essential differences between the Fienup variables for crops and those for grain in the present study are: (1) Fienup included only hired labor and included such expense in "total cash crop

---

10/ Fienup, op. cit., pp. 44-50.

expense". In the present study an estimate of total labor input was made a separate variable; (2) Machine services in the present study are included in the capital input while they constituted a separate variable in the Fienup function. Otherwise the two functions are more or less comparable. In (3-9) the X_{c1} variable is comparable to the land variable of the present study and variables X_{c2} and X_{c3} combined include about the same things as the capital variable in this thesis. In his livestock function Fienup uses feed and capital expense. Labor and land do not appear as such in the function. In the present study, feed is included partly in the capital input (purchased feed) and partly in the land input. In this thesis, cropland was considered a proxy for purchased feed and hay. It should be borne in mind that the variables in the Fienup and Heady-Shaw studies pertain only to the particular output to which they were regressed. That is, the variables in the crop function pertain only to the value of crop production. The same is true for livestock. In the present study, the variables are intended to explain the value of total farm production, not just a particular output such as grain.

It may be said that Heady-Shaw and Fienup faced a "segregation problem" while the present analyst faced an "aggregation problem". The former, in their attempt to explain crop production and livestock production <u>per se</u>, were required to segregate inputs according to whether the inputs were directed toward crop production or livestock production. Some inputs are "joint inputs" in the sense that they have a dual function and other inputs are inseparable by nature. Nevertheless, such inputs

have to be separated when any "crop farms" produce livestock and when any "livestock farms" produce crops. A mixture of crop and livestock production on a given farm is common.

The nature of the data in the present study was such that it was not possible to segregate inputs according to whether the input was directed toward one output or the other. So both output and the inputs were aggregated. The resulting functions apply to either farms which are classified as "primarily grain" or "primarily livestock", thus avoiding the problem of joint or common inputs. The Haley functions are thereby more general than the Heady-Shaw, Fienup functions, but some comparison can be made.

Table V contains the various comparable values from each of the functions. All values in Table V except the Haley grain function are evaluated at the geometric mean of the sample from which they were derived. The grain function for the present study was evaluated at the arithmetic mean. The reason for evaluating a function at its geometric mean is because it is calculated in logarithms as is the case in a Cobb-Douglas type function. For purposes of comparison, the marginal product of labor for the Haley functions were converted to the dollar addition to total output per month of labor input. This was accomplished by multiplying the average monthly labor expense by the marginal product of labor for grain and livestock. The average monthly mean was calculated by dividing the sample mean by 12.

TABLE V. SOME COMPARABLE VALUES FROM THE FIENUP AND HEADY-SHAW STUDIES.

Input	Heady-Shaw		Fienup		Haley	
	MP	Elasticity	MP	Elasticity	MP	Elasticity
Crop:						
Land	10.29	.5032	4.68		11.42	.1945
Labor	56.48 a/	.0394 b/			266.93 d/	.1560
Capital	2.13	.5804			.81	.5450
Sum		1.1230 c/		1.0316		.8955
Livestock:						
Labor	74.27 a/	.0839 b/			126.69 d/	.2707
Land					19.18	.4848
Capital	1.21 e/	.9370 e/			.63	.4902
Sum		1.0209		.9743		1.2457

a/ Measured in (dollars/month).

b/ Significantly different from zero at the 8 percent level.

c/ Significantly different from one at the 8 percent level.

d/ Converted to (dollars/month) from (dollars/dollars).

e/ Land was considered a capital input.

The grain function results derived in this study are about what one would expect in light of 1950-64 growth. The marginal productivity of capital has decreased from 2.13 to .81 and the marginal productivity of labor increased from 56.48 to 266.93. The increase in the value of the marginal product of labor was almost fivefold. Farming has become more heavily mechanized since 1950, the year from which the Heady-Shaw data were taken, and less labor is required. Under these circumstances one would expect that the marginal product of capital would decline and that the marginal product of labor would increase. Some of the increase in the value of the marginal product of labor is attributable to increases in the wage rate between 1950 and 1964. Montana farm wage rates increased about 25 percent between 1950 and 1964. 11/ But a large part of the increase is no doubt a result of improvement in the basic productivity of labor. The difference in total elasticities indicates farms have become more efficient. This difference, however, may be in some part attributable to the difference in types of functions derived. The Heady-Shaw crop function was homogeneous although not linear, whereas the grain function for this study is linear but not homogeneous. It would appear the income tax data have provided results for grain farms that make economic sense and that compare well with results from the previous studies.

11/ Maurice C. Taylor, Farm Price Index Trends, Montana--1948-1962, (Montana Agricultural Experiment Station, Montana State College, Bozeman, Montana, Bulletin 588, June 1964), p. 16, and an unpublished manuscript, Prices Paid and Prices Received by Montana Farmers and Ranchers, 1957-1966, (Montana Agricultural Experiment Station, Montana State University, Bozeman, Montana).

Comparing the derived livestock function in this thesis to those from the Heady-Shaw, Fienup studies is more tenuous. In the Heady-Shaw and Fienup studies, the livestock enterprise was a secondary enterprise and the farmers in the sample had no plans to increase the livestock operation. In this study, livestock is the primary enterprise in the sample of livestock farms. Labor is the only input that seems comparable to the Heady-Shaw work. Heady and Shaw aggregated land and capital into a single independent variable and the Fienup input categories did not include a significant labor variable. The marginal product of labor derived in the present study is about 70 percent higher than the one from the Heady-Shaw analysis. This change seems reasonable in light of the 1950-64 changes in wage rates. The 70 percent change in the value of the marginal product of labor compensates for a 25 percent 12/ change in money wage rates since 1950 and leaves considerable room for any increase in basic labor productivity. The sum of the livestock coefficients indicates that Montana ranches face increasing returns to scale. If so, livestock farms in Montana have room to grow in an efficient way. It appears that labor is still an inefficient input and that land is the most efficient input. The marginal product of capital is difficult to evaluate. It is plausible that livestock farming has become more capital intensive but how much is still a question.

12/ Ibid., p. 16.

PART IV

SUMMARY AND CONCLUSIONS

A. Summary

The first part of this paper was devoted primarily to a discussion of why income tax data might be an important source of information for economic research. It was pointed out that in addition to being a relatively cheap means of gathering data, each tax return is prepared according to a set of rules. When the penalty for breaking rules may be severe, as it is in the income tax case, there is an incentive for the return to be filled out correctly. The existence of such rules, even though the rule framework may not always conform to economic concepts and definitions, tends to temper the understatement of incomings and the overstatement of outgoings. And if taxpayers are more or less economically rational, they would not generally overstate incomings and understate outgoings. For this reason, income tax data should be at least as consistent as data gathered through interviews and questionnaires.

In any case these data can be used to supplement or verify data gathered from the schedule method and such use should improve the quality of schedule data. The use of income tax data in this manner should reduce the number of questions involving intimate financial detail and questions considered to be personal in nature.

In order to determine whether income tax data does yield consistent and reasonable results, it was hypothesized that these data would fit into a production function and yield significant coefficients. The derived

function could then be compared to results obtained using Montana data from interview schedules.

The second part of the report dealt with three problems usually associated with statistically derived production functions. It was shown that the omission of the management variable tends to bias returns to scale downward and the elasticity of capital upward. The general problem of multicollinearity was discussed and the means of discovering the magnitude of the multicollinearity problem were pointed out. The "regression fallacy" was shown to have little influence on the Cobb-Douglas type production function, and it was suggested that if the random occurrences that create the "regression fallacy" could be assumed to be normally distributed, no bias would exist.

Simultaneous equations and multiple regression analysis were discussed as possible methods of deriving input-output relationships. Four alternative production function forms were discussed. These were: the Spillman function, the Cobb-Douglas function, the Quadratic function, and the Square Root function.

Part III contains the analysis of the income tax data and results obtained along with a comparison with the results of previous research efforts in the same problem area. An explanation of the sample procedure and a description of the methods used to aggregate the income tax data and adjust or eliminate abnormal observations preceeded the presentation of the acutal analysis. Each type of exclusion was explained and justified. The origin and the makeup of the land data was described in detail, along with the process of merging the land data with the income tax data.

Each of the four variables used in the analysis was defined and described in the next section of Part III. Differences between the variables used in the grain function and those used in the livestock function were noted and justified. The labor variable was the most difficult to handle because the income tax returns contained only cash expenditures for labor. It was necessary to find a reasonable means of evaluating family labor. The process used to evaluate family labor was described and discussed in detail. The aggregation of each variable was described and possible shortcomings were noted.

The estimated marginal value productivity of land for the livestock farms was $19.18. This productivity can be interpreted as applying to the weighted "average acre of land" on livestock farms. The average livestock farm has 64 acres of irrigated cropland, 256 acres of non-irrigated cropland and 2,755 acres of grazing land. Weighted according to the 1965 average market prices of the three types of land, the value per acre turns out to be about $40. The marginal return of $19.18 can be set against an acre of land worth about $40. This high marginal value productivity would seem to indicate that land is not being used at its economic optimum.

Each function was then presented and the various statistical and economic characteristics were detailed. The results of this study was then compared to previous results of research done by Heady and Shaw and Fienup. It was difficult to compare the livestock function from this study to those of the Heady-Shaw and Fienup studies because of differences in samples and in the variables used. The grain function from the present

study was found to be quite comparable to the crop function developed by Heady and Shaw. Differences and similarities between this study and the other studies were explained and some possible economic reasons for the differences were presented.

B. Conclusions

Put in a negative way, it can be said that the hypothesis of this thesis cannot be rejected. At least it cannot be said that income tax data are not appropriate for production function research. The regression coefficients were significant for all variables used in both the grain function and the livestock function.

The derived functions were satisfactory in a statistical sense and appeared to be economically plausible. The marginal productivity of capital was found to have decreased in the 14 years between the Heady-Shaw study and the current study. This is explained by the large increase in the consumption of capital by Montana farms during that period. The marginal product of labor was found to have increased during the period, but is still low relative to farm wage rates. The marginal product of cropland was found to be about the same as in 1950 and this study indicated that grain land still needs to be more intensively used.

The livestock function is more difficult to evaluate. Increasing returns to scale were found at a significant level of statistical probability. This indicates that livestock farms are too small in relation to the economic optimum.

C. Recommendations for Future Research

The results of this study could be improved upon by stratifying the farm sample into income groups. This would serve to add more meaning to the interpretation. With such a sample, efficiencies of size could be isolated and compared. If the stratification were narrow enough, recommendations to specific farm groups and to the individual farmer might be made.

The use of income tax data in analyzing other important problems comes to mind. One such problem has to do with the effect of governmental control programs on farm size. Little, if any, quantitative research has been done on this problem. Another such problem has to do with the effect of governmental price support and control programs on the distribution of income among farms of various sizes. It has been suggested that government farm programs, which were designed to help the small farmer, have the effect of putting them at a long-run disadvantage.

Income tax data seem well-suited for testing a hypothesis of this sort. It now appears that arrangements for a continuous sampling of Montana farm income tax returns may materialize. In this event, it will be possible to analyze the temporal aspects of farm economics as well as the cross-sectional aspects. This would seem to put us at the threshold of a bonanza of information for analyzing the problems of Montana agriculture.

APPENDICES

APPENDIX A

TABULATION SHEETS FOR 1964 FARM INCOME TAX RETURNS

RETURN NUMBER (1 thru 6)				
INCOME:				
NON AGRICULTURAL (1 in Col. 80)				
Dividends (7 thru 12)				
Interest (13 thru 18)				
Rent & Royalty (19 thru 24)				
Other Business (25 thru 30)				
Gain or Loss (31 thru 36)				
Wages & Salaries (37 thru 42)				
Other N-F Income (43 thru 48)				
TOTAL NONFARM (49 thru 55)				
FARM (2 in Col. 80)				
Cattle (7 thru 12)				
Sheep (13 thru 18)				
Swine (19 thru 24)				
Poultry (25 thru 30)				
Grain (31 thru 36)				
Hay (37 thru 42)				
Dairy (43 thru 48)				
Eggs (49 thru 54)				
Wool (55 thru 60)				
Other Products (61 thru 66)				
OTHER FARM (3 in Col. 80)				
Barter (7 thru 12)				
Machine Work (13 thru 18)				
Breeding Fees (19 thru 24)				
Patronage Div. (25 thru 30)				
Agric. Prog. (31 thru 36)				
Gas Refund (37 thru 42)				
Hail Ins. Payts. (43 thru 48)				
Pasture Rent (49 thru 54)				
Gain or Loss (55 thru 60)				
TOTAL FARM (61 thru 69)				
PROFIT (OR LOSS) ON (4 in Col. 80) SALE OF PURCHASED LIVESTOCK OR OTHER PURCHASED ITEMS (7 thru 13)				
GROSS PROFITS (14 thru 20)				

RETURN NUMBER (1 thru 6)				
EXPENSES:				
FARM (5 in Col. 80)				
Labor Hired (7 thru 12)				
Repairs, Maint. (13 thru 18)				
Interest (19 thru 24)				
Feed Purchased (25 thru 30)				
Seeds, Plants (31 thru 36)				
Fertilizers, Lime (37 thru 42)				
Machine Hire (43 thru 48)				
Supplies Purchased (49 thru 54)				
Breeding Fees (55 thru 60)				
Vet, Medicine (61 thru 66)				
Gas, Fuel, Oil (67 thru 72)				
FARM CONTINUED (6 in Col. 80)				
Storage, Whse. (7 thru 12)				
Taxes (13 thru 18)				
Insurance (19 thru 24)				
Utilities (25 thru 30)				
Rent, Farm or Past. (31 thru 36)				
Freight, Trucking (37 thru 42)				
Conservation Exp. (43 thru 48)				
Retirement Plans (49 thru 54)				
Vehicle Lic. & Tax (55 thru 60)				
Farm Car Exp. (61 thru 66)				
Cl & Treating & Spray (67 thru 72)				
FARM CONTINUED (7 in Col. 80)				
1/2 SS Hired Labor (7 thru 12)				
Miscellaneous (13 thru 18)				
TOTAL FARM EXPENSE (19 thru 25)				
DEPRECIATION (26 thru 31)				
OTHER FARM DEDUCT. (32 thru 37)				
TOTAL DEDUCTIONS (38 thru 44)				
NET PROFIT OR LOSS (45 thru 51)				

APPENDIX B

Taxpayer's Name: _____ RETURN NUMBER: _____
Address: _____

RECORD OF LAND CLASSIFICATION

Tillable Irrigated

Grade	Assessed Value	No. of Acres
1A		
1B		
2		
3		
4		
5		
6		
7		
8		

Wild Hay

Grade	Assessed Value	No. of Acres
1		
2		
3		
4		
5		
6		
7		

Non-Irrigated Farm

Grade	Assessed Value	No. of Acres
1A3		
1A2		
1A1		
1A		
1B		
2A		
2B		
2C		
3A		
3B		
4A		
4B		
5		

Grazing

Grade	Assessed Value	No. of Acres
1A2		
1A1		
1A		
1B		
2A		
2B		
3		
4		
5		
6		

Non-Irrigated Continuously Cropped

Grade	Assessed Value	No. of Acres
1		
2		
3		
4		
5		
6		
7		

Grade	Assessed Value	No. of Acres
8		
9		
10		
11		
12		
13		
14		

Timber Land			Comments
Grade	Assessed Value	No. of Acres	

Please complete this form and return to the State Board of Equalization, Helena, Montana.

Does this taxpayer own land in an adjoining county? Yes____ No____

May 3, 1967

Dear

At the present time we are attempting to determine what kind of relationship there is between agricultural production in Montana and the various inputs that go into making up the production. Of particular interest is utilization of leased land in farming and cattle ranching.

Your information, combined with that of other farmers and ranchers in Montana will be extremely helpful in determining the contribution of leased land to agricultural production.

We sincerely ask your cooperation in filling out the enclosed questionnaire. It is very brief and the information, of course, will be kept in strict confidence.

Return of the quesionnaire at your earliest convenience will be appreciated.

Sincerely yours,

William J. Haley
Project Coordinator

WJH:jdb

The following information is requested for LEASED LAND ONLY. Please do not include land you own.*

Fill in the appropriate blank indicating the amount of land you leased for the years 1963 and 1964.

	1963	1964
Acres Tillable-Irrigated	_____	_____
Acres Non-Irrigated Cropped	_____	_____
Acres Grazing	_____	_____

*We have adequate secondary information about land owned in Montana.

APPENDIX C

The following discussion will outline the process of testing the hypothesis that returns to scale are constant on a derived Cobb-Douglas production function. The discussion will take up the case of three independent variables. Extension to n variables is obvious. 1/

Consider the following set of normal equations for a three variable regression equation.

$$\hat{\beta}'_2\Sigma x_2^2 + \hat{\beta}'_3\Sigma x_2x_3 + \hat{\beta}'_4\Sigma x_2x_4 + \lambda = \Sigma x_1x_2$$

$$\hat{\beta}'_2\Sigma x_2x_3 + \hat{\beta}'_3\Sigma x_3^2 + \hat{\beta}'_4\Sigma x_3x_4 + \lambda = \Sigma x_1x_3 \qquad \text{(C-1)}$$

$$\hat{\beta}'\Sigma x_2x_4 + \hat{\beta}'\Sigma x_3x_4 + \hat{\beta}'_4\Sigma x_4^2 + \lambda = \Sigma x_1x_4$$

$$\hat{\beta}'_2 + \hat{\beta}'_3 + \hat{\beta}'_4 = 1$$

If all the sum of cross-products are available, this system can be solved for the Lagrange multiplier and the betas, under the constraint that the sum of the betas equal one.

In order to solve for the betas when the sum of cross-products is not available, the normal equations can be written in the following form where the r's refer to correlation coefficients given in Appendix D.

1/ A comprehensive discussion of this method when the normal equations, in summation form, are available can be found in: Gerhard Tintner, "Production Functions from Farm Records", Econometrics, Vol. 12, (January 1944), pp. 26-34.

$$b'_2 r_{22} + b'_3 r_{23} + b'_4 r_{34} + \lambda = r_{12}$$

$$b'_2 r_{32} + b'_3 r_{33} + b'_4 r_{34} + \lambda = r_{13} \qquad \text{(C-2)}$$

$$b'_2 r_{42} + b'_3 r_{43} + b'_4 r_{44} + \lambda = r_{14}$$

$$b'_2 + b'_3 + b'_4 = 1$$

$$b'_i = \hat{\beta}'_i \frac{\hat{\sigma}_i}{\hat{\sigma}_1} \qquad i = 2, 3, 4 \qquad \text{(C-3)}$$

Once each b'_i and λ have been solved for, $\hat{\beta}'_i$ can then be obtained by

$$\hat{\beta}'_i = b'_i \frac{\hat{\sigma}_1}{\hat{\sigma}_i} \qquad i = 2, 3, 4 \qquad \text{(C-4)}$$

The $\hat{\beta}'_i$'s will probably not add to one but by normalizing them this constraint can be obtained.

Using the normalized betas, a test statistic can now be formed. Let

$$Q_1 = \Sigma x_1^2 - \hat{\beta}_2 \Sigma x_1 x_2 - \hat{\beta}_3 \Sigma x_1 x_3 - \hat{\beta}_4 \Sigma x_1 x_4 \qquad \text{(C-5)}$$

be the residual sum of squares unconstrained. Let

$$Q_2 = \Sigma x_1^2 - \hat{\beta}'_2 \Sigma x_1 x_2 - \hat{\beta}'_3 \Sigma x_1 x_3 - \hat{\beta}'_4 \Sigma x_1 x_4 - \lambda(\hat{\beta}'_2 + \hat{\beta}'_3 + \hat{\beta}'_4) \qquad \text{(C-6)}$$

be the residual sum of squares under the constraint that returns to scale are constant.

The test statistic

$$F = \frac{(Q_2 - Q_1)(N - 3)}{Q_1} \qquad \text{(C–7)}$$

is distributed like Snedecor's F with 1 and N - 3 degrees of freedom.

The numerical solution to the problem of testing the hypothesis that returns to scale are constant in the derived livestock function is as follows. Solving system (C–2) for each b_i' yields

$b_2' = .4541$ $\qquad$ $b_4' = .4245$

$b_3' = .1214$ $\qquad$ $\lambda = .0260$

Substitution of those values into (C–4) and normalizing

$\hat{\beta}_2' = .3960$

$\hat{\beta}_3' = .2005$

$\hat{\beta}_4' = .4036$

Using these values in equation (C–6), $Q_2 = 39.16$ and using the betas from the unconstrained system, $Q_1 = 25.26$. The difference in the calculated Q_1 and the Q_1 as given in Appendix D is due to rounding error.

Therefore, (C–7) is

$$F = \frac{13.90(61)}{39.16} = 15.80$$

which is significant at the one percent level.

APPENDIX D

The Grain Function

The general form of the grain function is as follows:

$$X_1 = b_1 + b_2X_2 + b_3X_3 + b_4X_4$$

The computed function is:

$$X_1 = 2171.13 + 11.42X_2 + .47X_3 + .81X_4$$

The test statistic for each coefficient is:

	T
b_2	3.84**
b_3	4.62**
b_4	9.24**

**Significant at the 1 percent level.

The analysis of variance for the regression equation is:

Source	df	SS	MS	F
Due to Reg.	3	3.41×10^{10}	1.14×10^{10}	177.81**
About Reg.	148	$.95 \times 10^{10}$	$.64 \times 10^{8}$	
Total	151	4.36×10^{10}		

**Significant at the 1 percent level.

The correlations between the variables are:

$r_{12} = .62$ $r_{13} = .76$ $r_{14} = .85$

$r_{23} = .51$ $r_{24} = .55$

$r_{34} = .74$

$R = .88$ $R^2 = .78$

The Livestock Function

The general form of the livestock function is as follows:

$$X_1 = b_1 X_2^{b_2} X_3^{b_3} X_4^{b_4}$$

The computed function is:

$$X_1 = .80 X_2^{.48} X_3^{.27} X_4^{.49}$$

The test statistic for each coefficient is:

	T	
b_2	6.02**	* Significant at the 5 percent level
b_3	1.89*	
b_4	5.25**	**Significant at the 1 percent level

The analysis of variance for the regression equation is:

Source	df	SS	MS	F
Due to Reg.	3	73.807	24.602	57.82**
About Reg.	61	25.955	.425	
Total	64	99.762		

**Significant at the 1 percent level.

The correlations between the variables are:

$r_{12} = .74$ $r_{13} = .47$ $r_{14} = .75$

$r_{23} = .29$ $r_{24} = .54$

$r_{34} = .45$

$R = .86$ $R^2 = .74$

APPENDIX E

Due to the inaccessiblity of 1964 land data, all land data furnished by the counties was for the year 1965. To justify using 1965 land data it must be shown that the size of farms in Montana did not change substantially in one year. Table I shows, by interpolation between 1959 and 1964, the per year change in the number of farms by size. The number of acres farmed increase about .05 percent per year between 1959 and 1964 and the average size farm increased about 1.8 percent. 1/

TABLE I.*

Size in Acres	Number		Total Percent Change	Yearly Percent Change
	1964	1959		
Less than 10	704	673	+ 4.6	+ .9
10 - 49	1641	1656	- .9	- .2
50 - 69	385	395	- 2.5	- .5
70 - 99	897	936	- 4.2	- .8
100 - 139	758	885	-14.4	-2.9
140 - 179	1353	1537	-12.0	-2.4
180 - 219	640	683	- 6.3	-1.3
220 - 259	626	666	- 6.1	-1.2
260 - 499	3130	3525	-11.2	-2.2
500 - 999	3954	4665	-15.3	-3.1
1000 - 1999	5101	5664	- 9.9	-1.9
2000 & over	7831	7517	+ 4.2	+ .8
Total	27020	28802	- 6.2	-1.2

*Source: U. S. Bureau of the Census, Census of Agriculture, Statistics for State and Counties, Montana, (U. S. Government Printing Office, Washington, D. C., 1967), p. 8.

It would appear that there is some change in size of farm operations from year to year. The farms of 10 acres or under and 2000 acres or over

1/ U. S. Bureau of the Census, op. cit., p. 8.

are the only categories which increase in number from 1959 to 1964. Since farm size of 10 acres or under is such a small percentage of the total number of farms (2.6 percent) and few if any of this size farm could be classified as a legitimate commercial operation, the increase in acreage from 1964 to 1965 attributable to farms of 10 acres or under is negligible at best.

The farms of 2000 acres or over constitute 29 percent of the total. This would mean that since the sample taken for this study was random, farms 2000 acres or over should constitute about 29 percent of the sample. Therefore, any influence on this size farm will certainly have in influence on the sample of which it is a part. If anything, the amount of land will be overstated but it is doubtful if it is overstated enough to make a difference in the results. One or two percent change in a given farmer's land from one year to the next will have little effect on the relative efficiency of the land input.

LITERATURE CITED

Business Concentration and Price Policy, Princeton, 1952.

Farrell, M. J., "The Measurement of Productive Efficiency", Journal of Royal Statistical Society, Part III, March 1957.

Fienup, Darrell F., Resource Productivity on Montana Dry-Land Crop Farms, Montana State College, Agricultural Experiment Station, Mimeograph Circular 66, June 1952.

Fisher, Franklin M., The Identification Problem in Econometrics, McGraw-Hill Book Co., 1966.

Griliches, Zvi., "Specification Bias in Estimates of Production Functions", Journal of Farm Economics, February 1957.

Heady, Earl O., and Shaw, Russell, Resource Returns and Productivity Coefficients in Selected Farming Areas of Iowa, Montana, and Alabama, Ames, Iowa, Agricultural Experiment Station, Research Bulletin 425, April 1955.

_____, "Technical Considerations in Estimating Production Functions", Resource Productivity, Returns to Scale, and Farm Size, ed. by Earl O. Heady, Glenn L. Johnson, and Lowell S. Hardin, Ames, Iowa, Iowa State College Press, 1956.

_____, and Dillon, John L., Agricultural Production Functions, Ames, Iowa, Iowa State College Press, 1961.

Henderson, J. M., and Quandt, R. E., Microeconomic Theory, New York, McGraw-Hill Book Co., 1958.

Johnston, J., Econometric Methods, New York, McGraw-Hill Book Co., 1960.

_____, "Statistical Cost Functions: A Re-Appraisal", Review of Economics and Statistics, November 1958.

Malinvaud, E., Statistical Methods of Econometrics, Chicago, Rand-McNally, 1966.

Perlis, Sam, Theory of Matrices, Reading, Mass., Addison-Wesley Publishing Co., 1952.

State of Montana Twenty-First Biennial Report of the Montana State Board of Equalization, State of Montana, December 1964.

Stigler, G. J., The Theory of Price, New York, 1952.

Swanson, Earl B., "Determining Optimum Size of Business From Production Functions", Resource Productivity, Returns to Scale, and Farm Size, ed. by Earl O. Heady, Glenn L. Johnson, and Lowell S. Hardin, Ames, Iowa, Iowa State College Press, 1956.

Taylor, Maurice C., Farm Price Index Trends, Montana--1948-1962, Montana Agricultural Experiment Station, Montana State College, Bozeman, Montana, Bulletin 588, June 1964.

_____, Prices Paid and Prices Received by Montana Farmers and Ranchers, 1957-1966, Unpublished Manuscript, Montana Agri-Cultural Experiment Station, Montana State University, Bozeman, Montana.

Tintner, Gerhard, "Production Functions from Farm Records", Econometrica, Vol. 12, January 1944.

U. S. Bureau of the Census, Census of Agriculture, Statistics for the State and Counties, Montana, Washington, D. C., U. S. Government Printing Office, 1967.

Walters, A. A., "Expectations and the Regression Fallacy in Estimating Cost Functions", Review of Economics and Statistics, May 1960.

MONTANA STATE UNIVERSITY LIBRARIES
3 1762 10014175 1

N378
H136
cop.2

Haley, W.J.
Livestock and grain production functions from a new data source

NAME AND ADDRESS
APR 10 191 Arval D. Haroldson
Econ Dept
Ginger Schmid
703 N Grand 7-1014

N378
H136
cop. 2

BOUND BY HM