AN ECONOMETRIC ANALYSIS OF FACTORS AFFECTING LAND VALUES IN WESTERN OKLAHOMA

By

FARID EL-HUSSINY ABDEL-BADIE

Bachelor of Science
Ein-shams University
Cairo, Egypt
1960

Master of Science
University of California
Berkeley, California
1964

Submitted to the faculty of the Graduate College
of the Oklahoma State University
in partial fulfillment of the requirements
for the degree of
DOCTOR OF PHILOSOPHY
July, 1967

AN ECONOMETRIC ANALYSIS OF FACTORS AFFECTING

LAND VALUES IN WESTERN OKLAHOMA

Thesis Approved:

L. A. Parcher
Thesis Adviser

Verner R. Eidman

Robert D. Morrison

Daniel D. Badger

D. D. Durham
Dean of the Graduate College

ACKNOWLEDGEMENTS

The author wishes to express his appreciation to the Department of Agricultural Economics of Oklahoma State University at Stillwater for making this study possible. The author is indebted to the Egyptian government for the financial aid he received while pursuing his graduate study.

Sincere gratitude is extended to Dr. Loris A. Parcher, Graduate Committee chairman, for his patience, assistance, and cooperativeness in the graduate program and in the writing of this thesis. Special thanks are due to Dr. James E. Martin for his help and suggestions in developing the econometric models used in this study. Appreciation is extended to Dr. Daniel D. Badger and Dr. Vernon R. Eidman for their suggestions and criticisms.

Thanks are also due to others who helped at various stages of the study including Dr. Charles H. Little and Dr. Robert D. Morrison.

Special appreciation is extended to Dr. James S. Plaxico for his encouragement and assistance during this study.

The author wishes to acknowledge the assistance of the statistical employees of the Department of Agricultural Economics. Miss Pat Cundiff, Mrs. Shireen Bee and Mrs. Martha Hurst were particularly helpful.

Appreciation is also extended to Mrs. Judy Eley and Mrs. Carolyn Hackett for the typing of the drafts and final copy of this dissertation.

The author is indebted to a host of close friends; among them are P. Thomas Cox, Dale Kalbfleisch, James Driscoll, Fred Tyner, Richard Crowder, John G. McNeely, Jr. and William L. Brant.

Special thanks are due my wife, Nawal, and son, Khalid, for the patience they have shown and the encouragement given to me during my graduate study. Their many sacrifices are here gratefully acknowledged.

Finally, I wish to dedicate this dissertation to the memory of my beloved mother and to my father, sisters and brothers who have been constant in their faith in me.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER I

INTRODUCTION

The steady increase in farmland values over the past quarter century has been the subject for much discussion among farmers and those interested in the economics of agriculture. Since 1940 land values in the United States have risen from an average of about $32.00 per acre to $157.00 in 1966 - more than a four-fold increase, (Table I and Figure 1). In Oklahoma during the same period per acre land prices increased from an average of $24.00 per acre to $126.00 which exceeds a five-fold increase.[1]

Since the land market consists of many individual transactions, and since each transaction is an entity with the price of each tract reached by agreement between individuals, it would appear useful to explore the factors which cause market prices to vary from one tract to another. Explanations of changes in the general level of land prices, while useful and revealing, are perhaps not as important to participants in the market as an explanation of why one tract sells for more than another.

---

[1] Farm Real Estate Market Developments, Economic Research Service, United States Department of Agriculture; Farm Real Estate Values, Oklahoma Crop and Livestock Reporting Service, Oklahoma City, Oklahoma; and, Farm Real Estate Values in the United States by Counties, 1950-1959, Edited by Thomas J. Pressly and William H. Scofield, University of Washington Press, (Seattle, 1966).

TABLE I

AVERAGE VALUE PER ACRE AND INDEX OF LAND VALUES, UNITED STATES AND OKLAHOMA, 1940-1966 (1957-59 = 100)

Year	United States		Oklahoma	
	Value per acre	Index	Value per acre	Index
1940	32	30	24	31
1941	32	31	24	32
1942	34	34	25	33
1943	38	36	28	37
1944	43	42	30	40
1945	47	46	33	43
1946	53	52	39	51
1947	60	59	43	56
1948	64	63	43	56
1949	66	66	53	70
1950	65	65	51	67
1951	75	75	60	80
1952	83	82	64	86
1953	83	83	61	83
1954	82	82	60	82
1955	85	85	65	90
1956	89	89	67	91
1957	94	95	69	94
1958	100	99	73	99
1959	108	106	80	107
1960	116	111	86	115
1961	118	112	86	115
1962	124	118	93	124
1963	130	123	102	137
1964	137	131	109	146
1965	146	139	118	158
1966	157	150	126	169

Source: Farm Real Estate Market Developments, Economic Research Service, U. S. Department of Agriculture, CD-68, July 1966, p. 18. Oklahoma values per acre for 1940-1949 were calculated from the indices of land prices in Oklahoma.

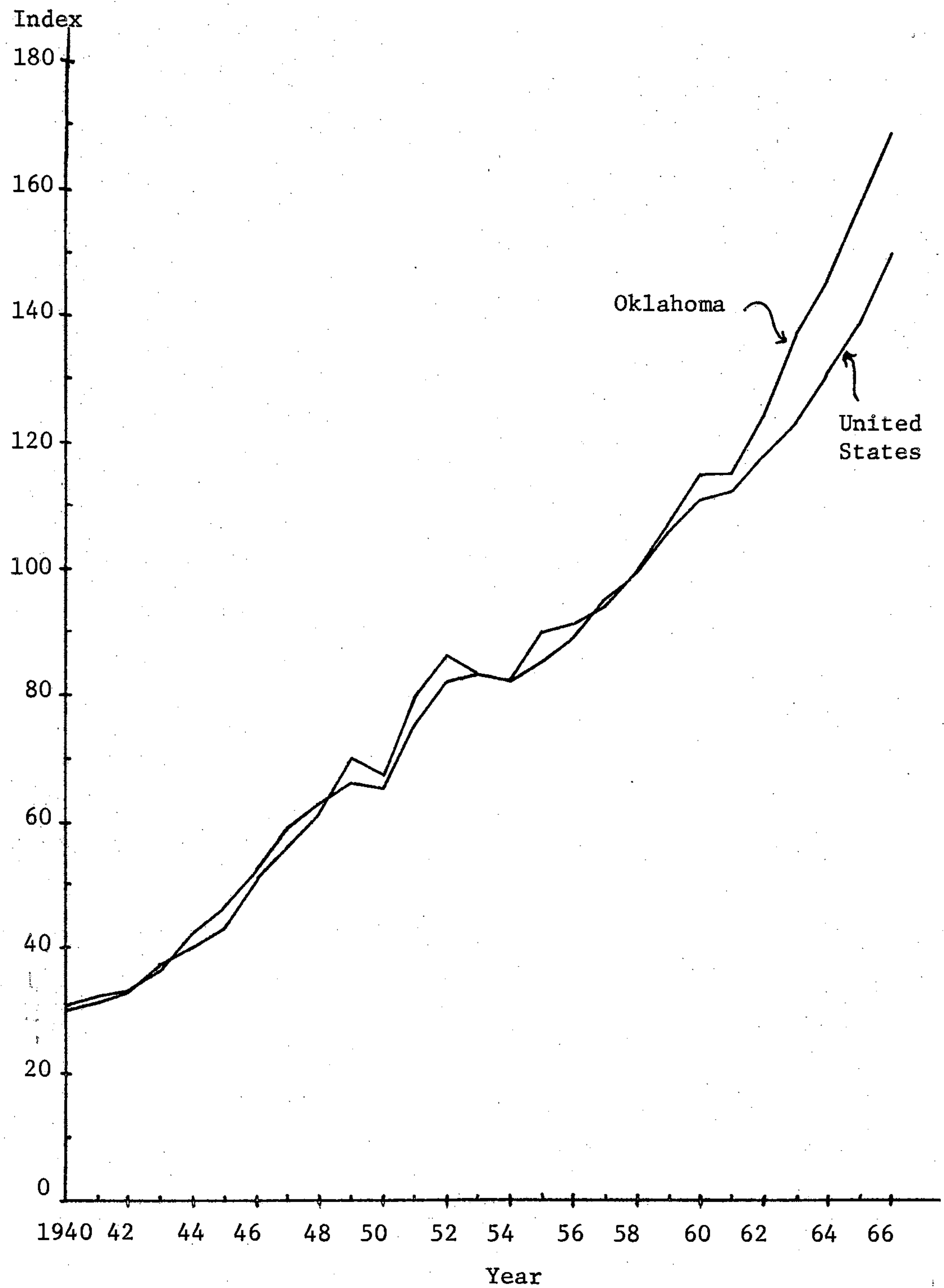

Figure 1. Index of Land Values, United States and Oklahoma, 1940-1966.

(1957-59=100)

Many factors appear to influence the price people pay for farmland. When a buyer and seller enter the land market, each one's subjective price, and consequently his action, is affected by his response to those factors. The same factors are not likely to equally affect all buyers and sellers, nor is it likely that all factors enter into the decision of each party to a transaction. However, it is probably that at least some of the difference in price paid for different tracts of land can be explained by certain factors which observation indicates are important.

The Problem

One of the basic purposes of research in land pricing is to see if procedures might be devised that can be used by buyers, sellers, lenders, and others who need to evaluate land. Land pricing by people in the market usually is based on a "feel" of the market. There is little in the way of precise measurement of value.

It is well known that unlike the market for many other commodities there is a lack of uniformity between tracts of land. There is no widely recognized system of grading upon which land price is based, and individual motivation has much to do with the demand for and the price paid for land. Research on problems of land price and land pricing, therefore, often is concerned with increasing the proficiency of individuals in estimating market values of land. More specifically, land pricing research usually attempts to: (1) determine where imperfections in the farm real estate market lie; (2) devise ways and means of improving the market mechanism and pricing procedures; and (3) determine the elements in the farm land market in addition to value productivity that are reflected in land prices.

The Objectives

The main objective of this study was to investigate the impact of selected price influencing forces on the price of farmland. The specific objectives were to: (1) determine whether hypothesized relationships exist between price per acre and 15 selected independent variables; (2) test the derived coefficients for statistical significance; (3) detect differences between grouped land sales; and (4) discover the underlying cause and effect relationships of these differences.

Source of Data

In this study, 293 bona fide land sales in ten counties in western Oklahoma were compiled and analyzed. These counties are delineated in Figure 2. These sales occurred during the years 1959 through 1964[2] and many of the particulars of each sale were taken from county records. County data were supplemented by data obtained from soils maps and from the Bureau of Indian Affairs on land use and crop allotments. County highway maps were used to determine the type of road adjacent to the property and the location of the property with respect to cities and towns of each tract sold.

In view of the substantial increase in the general level of land prices during the data period - about 6.0 percent per year - land sales prices for earlier years were adjusted to 1964 levels by using the index of land prices as calculated by the Oklahoma Crop and Livestock Reporting Service. The purpose of the adjustment was to minimize, if not eliminate, the time factor.

---

[2]A great majority of the sales occurred after 1960.

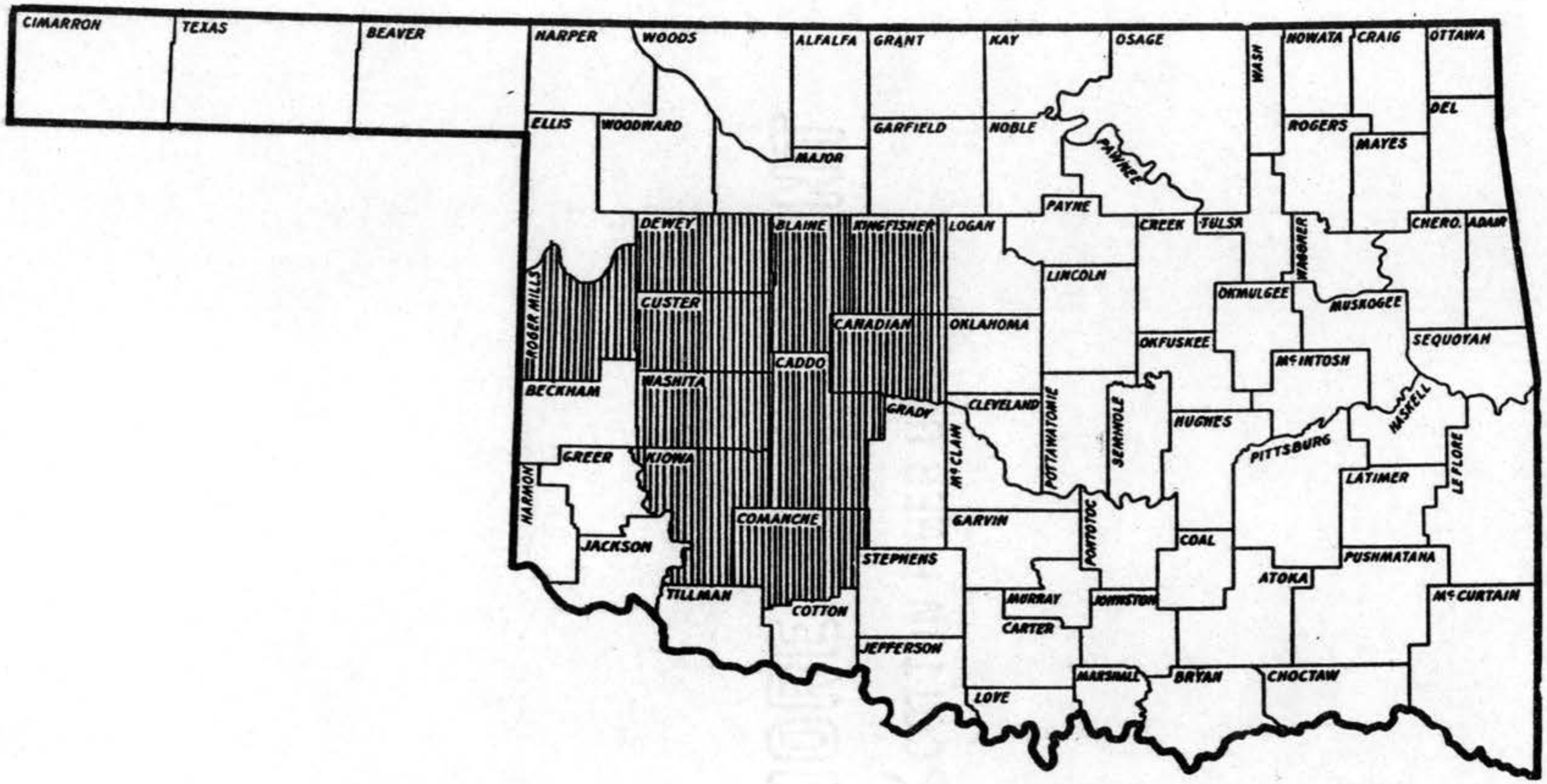

Figure 2. Ten County Area of Western Oklahoma from Which Land Sales Data Were Obtained

Techniques of Analyses

Two different techniques were used to make the analysis. Multiple regression techniques,[3] were considered in Chapter III, in which least squares regression techniques were applied to the data. It was hypothesized that the market price per acre of farmland depended upon 15 explanatory variables (X_1 through X_{15}). To examine this hypothesis by multiple regression, a way had to be found for expressing the form of the functional relationship. That is, not only does one seek a mathematical function which tells how the variables are interrelated, but also one which tells how precisely the value of one variable can be predicted if the values of associated variables are known. This technique is discussed further in Chapter III.

Although the analysis is applicable to the area from which the data were obtained, there is no reason for believing that the influences of the independent variables will not be about the same in other areas. In any event the methodology followed here should be applicable even though a different set of independent variables is chosen.

The second approach used in studying the relationship of land prices and the independent variables was multiple discriminant analysis.[4]

---

[3]For more details on regression analysis see: Bernard Ostle, Statistics in Research, (Ames: The Iowa State University Press, 1963).

[4]For more discussion on discriminant analysis see: Ardie Lubin, "Linear and Non-Linear Discriminating Functions," British Journal of Psychology, Statistical Section 3 (June, 1950), pp. 90-104; Ronald A. Fisher, "The Statistical Utilization of Multiple Measurements," Annals of Eugenics, Vol. 8 (August, 1938), pp. 376-386; Maurice M. Tatsuoka and David V. Tiedeman, "Discriminant Analysis," Review of Educational Research, Washington, D. C., 1954.

The analysis is concerned with the discrimination between three or more groups and is merely an extension of the two group classification. This technique allows one to ask not only: "Are the differences between groups statistically significant?"; but also "Are these differences of practical use? Can they be used to allocate individuals to their proper classification?"

Classification function coefficients developed by the discriminant analysis program[5] may be utilized to show the probabilities that a given land sale falls into one of the price classification groups. The summation of the probabilities that an individual sale will fall within the entire range of prices will equal one. However, a perfect discriminatory function will show this total concentrated in only one group with all other groups showing zero. That is, if the off-diagonal elements show a zero, then there has been no mis-classification. Therefore, the procedure is to minimize the percentage of mis-classified individual sales. This performs the same role as the least squares concept of minimizing the squared error terms.

The two techniques reviewed briefly above are discussed in later chapters showing the procedures which permit an evaluation of land price differentials.

---

[5]Fishers' development of the discriminant function in 1930's and its generalization by Ardie Lubin and Maurice M. Tatsouka in the 1950's has only recently been followed by practical application. For example see: P. Thomas Cox, "A Socio-Economic Analysis of Upstream Watershed Development in Oklahoma" (Unpublished Ph.D. thesis, Oklahoma State University, 1967).

Factors Affecting Land Values
in Western Oklahoma

Previous studies to which later reference will be made, as well as empirical observations have indicated that certain factors are relevant in the land price setting mechanism. On the basis of these studies and observation it appears that the following factors would be of primary importance in explaining differences in price per acre (Y) of land in the area of study.[6]

X_1 = Size of the tract. This refers to the number of acres in the tract sold.

X_2 = Productive quality of farm. This refers to the farm's potential to produce agricultural commodities based on its relative fertility as indicated by soils maps.

X_3 = Type-use combination. The use made of a given type of land. For example: "Bottomland cropland" or "Upland pasture".

X_4 = Quality of predominant land in tract. Usually a tract of land will be mixed in quality, this independent variable refers to the productive potential of the type of land which predominates.

X_5 = Quantity of Mineral rights conveyed. This is the percentage of mineral rights conveyed as shown in the deed. The quantity may vary from zero to 100 percent and in most areas the mineral interest is a valuable property right.

---

[6]Not measurable because of lack of data but also important, are the factors of individual preference, proximity to presently owned land, and the additional value to an operator of adding land to an existing unit.

X_6 = Type of road. This refers to the type of road adjacent to the farmstead or, if the farm is unimproved, to the best type of road touching the farm.

X_7 = Distance to a paved road. This refers to the distance in miles from the tract to the nearest paved road.

X_8 = Distance to nearest town. This is measured in miles to the nearest town.

X_9 = Size of nearest town. This refers to the population of the town as shown in county highway maps and is based on 1960 U. S. census data.

X_{10} = Distance to a principal city. This is the number of miles the tract is located from the nearest city. It generally will be the county seat.

X_{11} = Distance to a metropolitan center. This refers to the distance in miles from the nearest metropolitan area. A metropolitan area is deemed to be any city with a population of 50,000 or more.

X_{12} = Distance to Oklahoma City. In this area of the state the proximity of Oklahoma City is believed to affect land value.

X_{13} = Wheat allotment (acres)

X_{14} = Cotton allotment (acres)

X_{15} = Peanut allotment (acres)

While the quality of farmland often is an important determinant of its value, the size of the tract (X_1) may also influence the price for which it sells. For example, other things being equal, small tracts usually sell for more per acre than large tracts because the total amount of money required for purchase may be smaller and more people are financially able to buy them. Therefore, competition for smaller tracts tends to enhance the per acre price.

The per acre price of farmland is expected to be related to the quality of the soil and the productivity of the land. Therefore, a buyer usually will pay more for land of a higher productive capacity than for land of a lower capacity, because one of the things he is buying is an expected flow of income discounted to the present. The agricultural productivity potential is reflected in X_2, X_3 and X_4.

Rights in minerals (X_5) are included as a factor because whether all mineral rights are included in the transfer may have an important bearing on land values. This was particularly true in the area of study during the early 1960's. The widespread exploration for oil and gas in western Oklahoma kept people in the area conscious of mineral values during the period studied.

Location of a farm as reflected by variables X_6 to X_{12} is deemed to be an important factor in the price per acre for which land sells. Previous studies on the influence of location on farmland prices indicate that:[7]

> Farms on pavement sold for more than farms located on improved dirt roads.
> Farms on unimproved dirt roads sold for less than farms on all-weather roads.
> Farms within a half-mile of an all-weather road sold for more than those two to four miles off such a road.
> Farms near a market sold for more than those farther away.
> Farms within five miles of a principal city sold for more than farms 10 to 15 miles away.

---

[7] Loris A. Parcher, The Influence of Location on Farmland Price, Oklahoma Agricultural Experiment Station, Bul. No. B-417, (March, 1954).

People generally value more highly land that is located on a paved road near a city because of the possibility of converting the land to a higher use. In addition costs of inputs may be lower because of reduced transportation costs.

Observation and studies have shown that the right to produce certain crops on a given farm enhances the price of that farm.[8] Variables X_{13}, X_{14}, and X_{15}, (crop allotments) are included as price influencing factors important in the area of study.

Organization of Remainder of Dissertation

The remainder of this dissertation is organized in the following manner. The next chapter will outline the historic developments of the theory of land value. Chapter III presents and analyzes the results of multiple regression and Chapter IV presents the use of discriminant analysis in the classification of land prices. The study is summarized and conclusions presented in Chapter V. Chapter V is concluded with two sections concerning the need for further research and weakness of the study.

---

[8]Robert F. Bowley, Jr., and W. L. Gibson, Peanut Acreage Allotments and Farm Land Value, Virginia Polytechnic Institute, Blacksburg, Virginia, Technical Bul. 175 (September, 1964).

CHAPTER II

REVIEW OF LITERATURE

The purpose of this chapter is to: (1) outline the historic development of the theory of land value; (2) review and analyze the factors affecting land prices; and (3) analyze the process of supply and demand in the determination of price in the land market.

Land As A Factor of Production

The characteristics of land as a factor of production differ in a number of ways from the other factors of production, labor and capital.[1] These characteristics influence the way in which it is priced, as compared to freely reproducible goods or to capital and labor:

(1) Land is more durable;

(2) Land is not homogenous;

(3) Land is immobile and fixed in location;

(4) Land provides many services which have only a subjective value; and

(5) Land (as space) is indestructible.

Moreover, there are no central markets where land is freely traded and prices quoted. Each transaction is surrounded by all the implications of value theory.

---

[1]Roland R. Renne, Land Economics (New York: Harper and Brothers Publishers, 1947), p. 519.

The Historical Background

The main foundation of land value theory is credited to the mercantilists and in particular to Sir William Petty (1623-1687). While he was not clear in his reasoning concerning the relationship of rents to prices and land values, it appears that he may have seen the differential surplus element in rents. Petty stated the relationship of product prices to land values as follows:[2]

> For as a great need of money heightens exchange, so doth great need of corn raise the price of that likewise, and consequently of the rent of the land that bears corn, and lastly of the land itself.

The Physiocrats, whom Petty preceded, talked much of the surplus from agricultural production or the "net product". The Physiocrats believed that this "net product" was confined to one class of production only, namely, agriculture. They held that no other industry, such as trade or manufacturing, was able to create a surplus of goods. This tenet lends historical support to the application of the net income, or earnings, approach to land used by professional land evaluators.

Classical economics and the development of early value theory were chiefly founded upon the lectures and writings of Adam Smith and David Ricardo. Adam Smith, often called the founder of the classical school, saw the impact of fertility and location on rent which will add to the value of the land. Smith stated:[3]

---

[2]Eric Roll, A History of Economic Thought (New York: Prentice-Hall, Inc., 1949), p. 108.

[3]Adam Smith, The Wealth of Nations, (New York: The Modern Library, Random House, Inc., 1937), p. 147.

> The rent of land not only varies with its fertility, but with its situation, whatever be its fertility. Land in the neighborhood of a town gives a greater rent than land equally fertile in a distant part of the country.

The residual theory of land value as taught today has its documentary beginning in the works of Ricardo. He started his analysis by assuming a newly settled country with an abundance of rich and fertile land, a very small proportion of which is required to be cultivated for the support of the actual population. He then assumed that only the most fertile lands would be brought into cultivation until increases in population numbers and the demand for land made it necessary for society to bring less fertile lands into use.

Ricardo reasoned that as the demand for products of the land increased, man would be forced to resort to lower and lower grades of soil to supply this demand. The lowest grade of soil used would be that grade which yielded just enough value product to pay for the labor and capital expended in the effort. The same effort expended on a better soil would yield a surplus which must be attributed to the superior quality of the soil. This surplus he called rent. Purchasers of the superior land would expect to pay for the value of this surplus. A graphic illustration of Ricardo's theory of rent is shown in Figure 3. The Ricardian theory of rent thus reduces itself to the comparatively obvious statement that rent is:[4]

> that portion of the produce of the earth which is paid to the landlord for the use of the original and indestructible powers of the soil.

---

[4] David Ricardo, <u>The Principles of Political Economy and Taxation</u> (London: J. M. Dent and Sons, Ltd., 1911), p. 33.

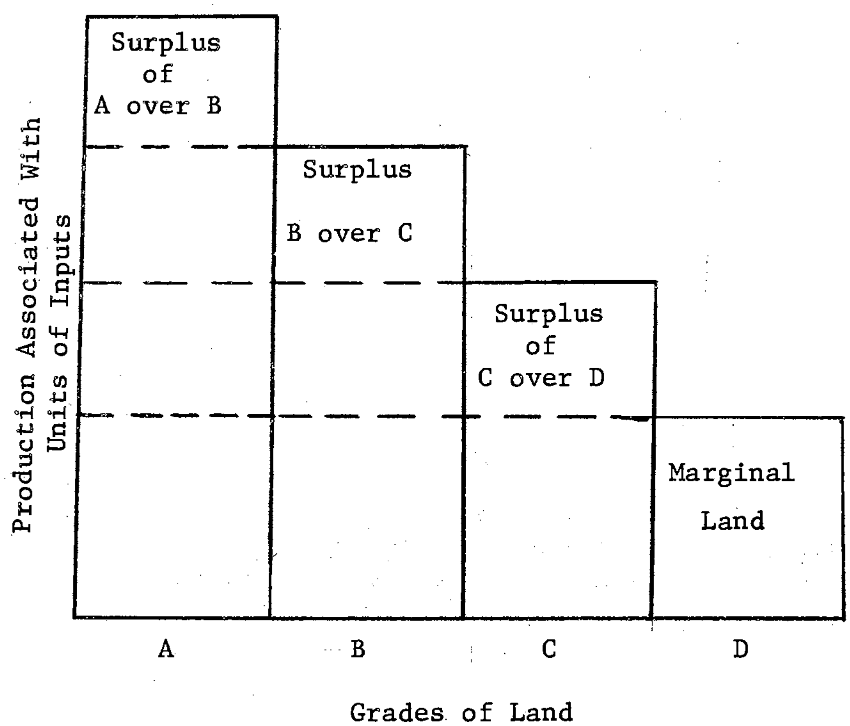

Figure 3. Illustration of Ricardian Theory of Rent

Ricardian theory assumes that prices are determined by the production costs at the intensive and extensive margins of cultivation. He recognized that product prices must rise with the outward shift of the extensive margin of cultivation and that these higher prices at the same time raise the intensive margin on the more fertile lands and thus favor their more intensive use. Ricardo stated:[5]

> It often, and indeed commonly, happens that before . . ., the inferior lands are cultivated, capital can be employed more productively on those lands which are already in cultivation . . .
>
> In such case, capital will be preferably employed on the old land, and will equally create a rent; for rent is always the difference between the produce obtained by the employment of two equal quantities of capital and labour.

Figure 4 illustrates the case of differential rent and involves consideration of both the intensive and extensive margins of production.[6] Assume that there are three different grades of land, with Grade A land the most fertile, Grade B land the next most fertile, and Grade C land the least fertile. Grade C land illustrates the case of marginal land. The price (OPc) is just equal to the average cost of inputs on Grade C land, and there is no economic rent. However, a price of OPc will produce economic rent on land of Grades A and B, since this price is well above the average cost of production on these more fertile lands, and the rent is greater on A than on B because A is better grade land than B.

Thus Figure 4 illustrates the proposition that production will be pushed both intensively and extensively up to the point at which the

---

[5]Ibid., p. 36.

[6]This analysis is based upon: H. H. Liebhafsky, The Nature of Price Theory (Homewood, Illinois: The Dorsey Press, Inc., 1963), Chapter 14.

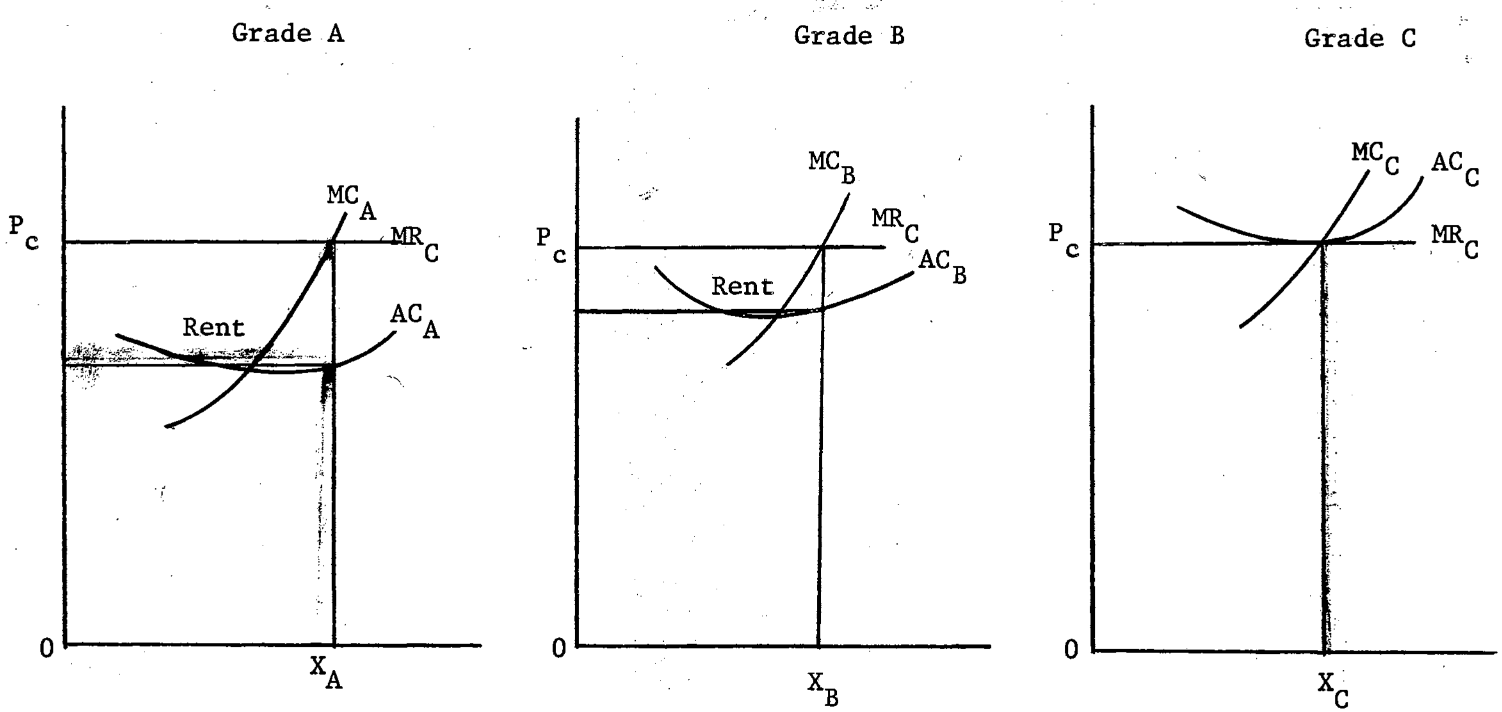

Figure 4. Economic Rent at the Intensive and Extensive Margins of Production on Three Grades of Land

marginal cost of the output is equal to the marginal revenue of price. And so, as poorer grades of land are brought into production, and as better grades of land are used more intensively, economic rent arises as a surplus which give rise to a price for land.

Ricardo's explanation of rent in terms of differences in land quality deals only with one factor that affects rent-paying capacity. Location is another important rent-determinant. The importance of this factor was stressed by Petty and Von Thunen but received little more than passing attention from Ricardo and his contemporaries.

Petty and Von Thunen both observed that when crops produced for a central city market are grown on lands of like fertility, the lands located nearest the city enjoy a definite rent advantage over those located at greater distance. Many differences in rent-paying capacity may be explained in terms of differences in land quality and location.

The important and lasting impact on value theory made by the Austrian School was based upon the importance placed on the human or demand concept of value. Proponents of the School argued that in the final analysis demand determined value. Thus, the Austrian School's theory became the cornerstone of the present utility theory as a measure of value. The Austrian School brought a new dimension into land value theory. Whereas the classical school appeared to attribute land value to its ability to yield surplus income, the Austrian School seemed to feel that land would have value because, like any other commodity, its utility is such that consumers (users) want it and will pay a price to get it.

A reappraisal of economic principles was made by Alfred Marshall. While recognizing the importance and validity of the utility concept of value, Marshall reintroduced the importance of production costs in affecting an equilibrium in the interplay of the forces of supply and demand.

The theories briefly outlined here each have contributed to the general knowledge of value. Value theory and approaches to value advocated by leading appraisal societies today is largely a synthesis of the important ideas and concepts developed by the several schools of thought.

Practicing professional land evaluators, however, still are concerned with the problem of applying value theory in their estimation to the value of individual tracts of land. The factors influencing value are so complex and inter-related that it is extremely difficult to determine what forces set the value on a given piece of land. If students of value theory have difficulty in determining value, the problem of the layman must be infinitely greater.

Heady stated:[7]

> The problem of resource valuation is basically and fundamentally one of allocating or imputing the total product forthcoming in a single production process to each of the several resources involved. The product or reward to one factor of production cannot be established accurately except as the reward for other factors are accurately reflected. Problems of valuation are first those of marginal productivity analysis, and only second those of "placing a price tag" on specific factors. The appraiser does not accept the market price for land, but instead formulates his own expectation of the physical and value productivity of the resource. Yet in doing so he accepts the market estimates of productivity and value (price) for labor, feed, tractor fuel, and so forth; he simply subtracts the market price (expense) of these resources from the total product and imputes the residual to land.

---

[7]Earl O. Heady, Economics of Agricultural Production and Resource Use (Englewood Cliffs, N.J.: Prentice-Hall, Inc., 1964), pp. 402-403.

Recent Research Related to This Study

A recent land market study by Ted R. Nelson[8] pertained to all land and explained changes in the general level of land prices. He made no effort, however, to explore factors which influence the market value of individual tracts of land.

Other studies[9] have indicated that a relationship probably exists between the price people are willing to pay for farmland and certain locational and physical variables such as distance to pavement, distance to towns of various sizes, size of tract, quality and use of the land, and institutional variables such as crops allotment acres.

The relationship of the price paid for land and these kinds of explanatory variables will be examined in the next chapter, where we will seek an answer to the question: Is it possible to measure the impact of selected value influencing forces on the subjective values of buyers and sellers?

---

[8]Ted R. Nelson, An Econometric Model of the Land Market Stressing Effects of Government Programs on Land Values, Unpublished Ph. D. Thesis (Stillwater: Oklahoma State University, 1964).

[9]For example see Mohammed A. Ahmed, An Economic Evaluation of Farmland for Tax Assessment, Tulsa County, Oklahoma, Unpublished Ph. D. Thesis (Stillwater: Oklahoma State University, 1964).

CHAPTER III

MULTIPLE REGRESSION ANALYSIS

The present chapter outlines the more commonly used method of analyzing the relationship of land price to selected independent variables. The multiple regression and correlation approach analyzes the functional relationship of one dependent variable to several independent variables in the form of a mathematical equation.

When information is available on more than one variable, a form may be found to express this relationship if a relationship is hypothesized. It is also possible to measure the strength of the relationship. That is, not only do we seek a mathematical function which tells us how the variables are interrelated, but also how precisely the value of one variable can be predicted if the values of the associated variables are known. The techniques used to accomplish these two objectives are known as correlation analyses and regression analyses. Multiple regression is designed for the purpose of predicting Y (the unknown) if $X_1, \ldots\ldots, X_n$ are known and to explain some of the variation in the unknown given certain explanatory variables.

By use of multiple regression analysis one can measure by means of empirical data whether a relatioship exists between one factor and certain selected variables.

A multiple regression[1] analysis was made of the relationship between the dependent variable (Y = price per acre of farm land) and of the 15 independent variables set forth in Chapter I. Under certain assumptions the method of least squares gives the best, unbaised linear estimate of Y according to the Markoff theorem.[2] "Best" means that the least squares estimate has the smallest variance or standard error among all linear unbiased estimates. The linear estimate is unbiased in the sense that the mean value of the estimate is equal to the population value. It is well to refer to the assumptions necessary for the estimate to give optimum statistical properties. The least squares statistical model is in the form:

$$Y_t = \beta_o + \beta_1 X_{1t} + \beta_2 X_{2t} + \ldots + \beta_n X_{nt} + \varepsilon_t$$

The assumptions of the model are:

(1) The parameters are constant and enter the model linearly;

(2) The expected value of the error term ε_t is zero i.e. $E(\varepsilon_t) = 0$ for $t = 1, 2, \ldots, m$

(3) The independent variables X_{it} are fixed (non-stochastic) and measured without error; $i = 1, 2, \ldots n$

(4) The covariance between the error ε_t and the independent variables X_{it} is zero, i.e. $E(\varepsilon_t X_{it}) = 0$ for all i and t.

---

[1] For more information about multiple regression see:
(a) J. Johnson, Econometric Methods, (New York, McGraw-Hill Book Company, Inc., 1960), pp. 108-115.
(b) M. G. Kendall, The Advanced Theory of Statistics, Vol. I., (London, 1945) p. 368.
(c) Karl A. Fox and James F. Coonery, Jr., Effects of Intercorrelations Upon Multiple Correlation and Regression Measures, United States Department of Agriculture, AMS-341, 1959.

[2] F. N. David and J. Neyman, Extension of the Markoff Theoreum on Least Squares, Statistical Research Memoris, Vol. 2, (1938), p. 105.

(5) The error ε_t is not autocorrelated, i.e. $E(\varepsilon_t \varepsilon_{t+i}) = 0$ for $i \neq 0$.

(6) The variance of the error is homogenous over time, i.e. $E(\varepsilon_t^2) = \sigma^2$ for $t = 1, 2, \ldots.m$.

(7) The error ε_t is normally distributed.

(8) The matrix of the independent variables X_{it} is not singular.

The β_i coefficients and the parameters of the ε_t distribution are unknown, and the problem is to obtain estimates of these unknowns. If the above eight assumptions hold, least squares can be used to estimate the parameters. Let $\hat{\beta}'$ denote a vector of estimates of β as

$$\hat{\beta}'=\{\hat{\beta}_1,\hat{\beta}_2 \ldots.\hat{\beta}_n\}$$

The n equations can be expressed in matrix notation as $Y = X\hat{\beta} + \varepsilon$, where ε denotes the column vector of n residuals $(Y-X\hat{\beta})$. The sum of squared residuals is: $\sum_{t=1}^{m} \varepsilon_t^2 = \varepsilon'\varepsilon$

$$= (Y-X\hat{\beta})' (Y-X\hat{\beta})$$

$$= Y'Y - 2\hat{\beta}' X' Y + \hat{\beta}' X' X\hat{\beta}$$

Now, to get the value of $\hat{\beta}$ which minimizes the sum of squared residuals we differentiate $\frac{\partial \varepsilon'\varepsilon}{\partial \hat{\beta}} = -2X'Y + 2X'X\hat{\beta}$, set the derrivative equal to zero and solve for $\hat{\beta}$.

$$-2X'Y+2X'X \hat{\beta} = 0$$

$$2 X'X \hat{\beta} = 2 X'Y$$

$$X'X \hat{\beta} = X'Y$$

$$\hat{\beta} = (X'X)^{-1} X'Y$$

If the above eight conditions are met, least squares estimates have the desirable statistical properties listed earlier.

The Analysis

The independent variables of the system were chosen on the basis of their hypothesized effect on the dependent variable land price. The independent variables and the codes used are as follows:

(1) Size of tract sold in acres.

(2) Productive quality of farm

very high, 1; high, 2; high medium, 3; medium, 4; low medium 5; poor to fair, 6; poor, 7; very poor, 8; mostly wasteland, 9.

(3) Type-use combination

Mostly good bottom land crop, 0;

Mostly good upland crop, 1;

Combination of above, 2;

Good mixed farm (crop & pasture), 3;

Medium quality mostly crop, 4;

Medium quality combination crop and pasture 5;

Medium quality mostly pasture, 6;

Low quality, mostly crop, 7;

Low quality, combination, 8; and

Low quality, mostly pasture, 9.

(4) Quality of predominant land in tract: Best bottom, 0; Best upland, 1; Good bottom, 2; Good upland, 3; Best pasture, 4; Good pasture, 5; Poor bottom, 6; Poor upland, 7; Poor pasture, 8; Miscellaneous, Waste and other, 9.

(5) Quantity of mineral right conveyed, percent conveyed.

None, 0; 1-10% 1; 11-24%, 2; 25%, 3; 26-49%, 4; 50%, 5; 51-74%, 6; 75%, 7; 76-99%, 8; 100%, 9.

(6) Type of road on which farm is located

Federal highway,0; state highway, 1; other paved or bituminous surfaced road, 2; gravel, 3; graded dirt, 4; ungraded dirt, 5; no road passes land, 6.

(7) Distance to a paved road.

(0-0.2 mile), 0; (0.3-0.5 mile), 1; (0.6-1.0 mile), 2; (1.1-2.0 miles), 3; (2.1-3.5 miles), 4; (3.6-5.0 miles), 5; (5.1-7.5 miles), 6; (7.6-10.0 miles), 7; (10.1 miles and up, 8.

(8) Distance to nearest town.

(0-.5 mile), 0; (0.6-1.5 miles), 1; (1.6-2.5 miles), 2; (2.6-5.0 miles), 3; (5.1-7.5 miles), 4; (7.6-12.0 miles), 5; (12.1 miles and up), 6.

(9) Population of nearest town.

(less than 100), 0; (101-250), 1; (251-500), 2; (501-1,000), 3; (1,001-1,750), 4; (1,751-3,000), 5; (3,001-5,000), 6; (5,001-10,000) 7; (10,001-25,000), 8; (Over 25,000), 9;.

(10) Distance in miles to a principal city or to the county seat if county seat has population of at least 1,500.

(up to 2.5), 0; (2.6-5.0), 1; (5.1-10.0), 2; (10.1-15.0), 3; (15.1-25.0), 4; (25.1-40.0), 5; (Over 40), 6.

(11) Distance to a metropolitan area (50,000 population or more)

(less than 10 miles), 0; (10-20 miles), 1; (21-40 miles), 2; (41-65 miles), 3; (66-90 miles), 4; (over 90 miles), 5.

(12) Distance to Oklahoma City (state highway distances)

(less than 10 miles), 0; (10-25 miles), 1; (26-40 miles), 2; (41-65 miles), 3; (66-100 miles), 4; (over 100 miles), 5.

(13) Wheat allotment acres

(14) Cotton allotment acres

(15) Peanut allotment acres

Analysis of the Results

In the analysis 15 independent variables were regressed on the dependent variable. The purpose of this process was to seek the best possible fit of 15 explanatory variables which may explain variation in land price in Western Oklahoma. In order to apply ordinary least squares regression techniques, the equation was made linear in the variables and the parameters. It is convenient to begin this discussion by showing the economic and statistical analyses made through use of the equation. The estimated equation appears below.

$$\hat{Y} = 235.02485 - .07917X_1 - 5.62210X_2 - 5.26324X_3 - 4.35267X_4 + 1.56240X_5 + 5.07328X_6 - 1.15358X_7 + 4.43975X_8 + .47418X_9 + .02321X_{10} - 3.72834X_{11} - 11.57324X_{12} + .77419X_{13} + .12393X_{14} + .95618X_{15}. \qquad (R^2 = .51)$$

The simple correlation technique was utilized in depicting qualitative types of relationships among the variables under study. Knowledge of intercorrelations among the independent variables may assist an evaluator in eliminating one or more of the highly correlated independent variables in an equation. For example, the prevalence of high intercorrelation between two independent variables means that the inclusion of one of them in an equation may explain as much variation in the dependent variable as can be explained by both; they are actually the same for all practical purposes. Therefore, the elimination of one may save a great deal of time and calculation effort. The attainment

TABLE II

THE SIGN, SIZE, STANDARD ERROR OF COEFFICIENTS, AND THE t-VALUES: FOR 15 INDEPENDENT VARIABLES USED TO PREDICT PRICE OF FARMLAND

Variable	Regression Coefficient	Standard Error of Coefficient	t-Value
X_1	-0.07917	0.03921	-2.01905**
X_2	-5.62210	2.95856	-1.90028*
X_3	-5.26324	2.02581	-2.59810***
X_4	-4.35267	1.33513	-3.26011***
X_5	1.56420	0.89152	1.75252*
X_6	5.07328	2.35686	2.15256**
X_7	-1.15358	1.88989	-0.61040
X_8	4.43975	2.34064	1.89681*
X_9	0.57518	1.84786	0.25661
X_{10}	0.02321	1.98413	0.01170
X_{11}	-3.72834	4.01686	-0.92817
X_{12}	-11.57324	5.45884	-2.12009**
X_{13}	0.77419	0.19194	4.03355***
X_{14}	0.12393	0.40636	0.30497
X_{15}	0.95618	1.52794	0.62580

*Statistically significant at the 10% probability level.

**Statistically significant at the 5% probability level.

***Statistically significant at the 1% probability level.

of low intercorrelations among the independent variables is advantageous. The prevalence of low intercorrelations means that when an equation is fitted to the data, the regression coefficients tend to be stable and each of the independent variables shows its impact separately on land price. Thus, a simple correlation technique provides valuable information in establishing functional relationships between the dependent variable (Y = per acre price of land) and the selected independent variables X_1 through X_{15} (Table III).

The explanatory variables are defined in Chapter I. The coefficient of determinations (R^2) shows that about 51 percent of the price variation between tracts of land in these ten counties in Western Oklahoma is explained by the fifteen independent variables included in the equation. Coefficients on X_3 (type use combinations), X_4 (quality of predominant land), and X_{13} (wheat allotment acres) are statistically significant at the 1% probability level. The coefficients X_1 (size of the tract sold) X_6 (type of road), and X_{12} (distance to Oklahoma City) are significant at the 5% probability level. Coefficients X_2 (productive quality of the whole farm), X_5 (quantity of mineral rights conveyed), and X_8 (distance to a nearest town) are statistically significant at the 10% probability level. The coefficients X_7 (distance to a paved road), X_9 (size of nearest town), X_{10} (distance to a principal city), and X_{11} (distance to a metropolitan center), X_{14} (cotton allotment acres) and X_{15} (peanut allotment acres) are not significant at the 10% probability level.

Peanut allotment acres (X_{15}) were not a significant factor in land price in the area of study probably for two reasons: (1) lack of sales

TABLE III

SIMPLE CORRELATION MATRIX OF THE VARIABLES

		X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}	X_{11}	X_{12}	X_{13}	X_{14}	X_{15}	Y
Size of the tract sold	X_1	1.000	.066287	.073704	.087833	-.225060	.063675	.135950	.048507	-.190229	.153868	.120520	.202557	.105557	.072100	.047850	-.162771
Productive Quality of farm	X_2		1.000000	.861702	0.673499	-.091320	.184806	.224088	-.010257	-.178632	.016021	.299114	.247108	-.205156	-.041263	.047759	-.589447
Type-use combination	X_3			1.000000	.721936	.009770	.169910	.205153	.010715	-.174579	.049536	.240424	.192755	-.163094	-.018732	.066424	-.591260
Quality of Predominant land in tract	X_4				1.000000	-.029836	.095615	.133622	-.033202	-.146081	-.009138	.255402	.144743	-.129179	-.044832	.126960	-.551463
Quantity of mineral rights	X_5					1.000000	-.023289	-.096437	-.005603	.297764	.006398	-.147592	-.279093	.034008	.129592	.063771	.179296
Type of road	X_6						1.000000	.522525	.199181	-.076255	.090588	-.039872	-.012518	-.082653	.013915	.025254	-.001291
Distance to a paved road	X_7							1.000000	.429633	-.323211	.249878	.032697	.202288	-.043906	.034669	-.015335	-.112069
Distance to nearest town	X_8								1.000000	.022848	.255139	-.022092	.255139	-.054733	-.007292	-.039390	.083313
Size of nearest town	X_9									1.000000	-.411314	-.254485	-.471975	.036175	.044849	.092487	.244798
Distance to a principal city	X_{10}										1.000000	-.006426	.242812	-.064042	-.008364	-.010120	-.048140
Distance to a metropolitan center	X_{11}											1.000000	.552253	-.122878	.008585	.161858	-.330950
Distance to Oklahoma City	X_{12}												1.000000	-.091893	-.087057	-.068233	-.338934
Wheat Allotment acres	X_{13}													1.000000	.081345	-.023880	.276534
Cotton Allotment acres	X_{14}														1.000000	.310514	.072651
Peanut Allotment acres	X_{15}															1.000000	-.018721
Price per acre	Y																1.000000

having peanut allotments, and (2) the area in general is not a peanut producing sector.

The coefficient for the explanatory variable X_1 has a negative sign and is a relatively small value. The sign is consistent with the economic interpretation, that there often is a greater demand for smaller tracts since there usually will be more people in a financial position to buy such tracts because the total outlay is less. This may be interpretated in the following manner: As the size of the tract increased one acre the price per acre decreased by $.07917.

The coefficient for variable X_2 (productive quality of farm) has a negative sign and a relatively large value. The negative sign is logical because of the manner in which productive quality was coded. That is, the lower code numbers represented the higher quality soils, thus the lower the numerical value of this variable the higher the price. Therefore, as the productive quality of a farm changed by one numerical unit the price per acre changed in an opposite direction by $5.62210 assuming other things equal.

Variable X_3 has a coefficient with a negative sign and is relatively large in size. Again the sign is logical. The scheme of coding was such that the lower the code number the greater the income potential of the land. That is, as we move from one class to the other price per acre decreased by $5.25324.

Variable X_4 was coded in such a manner that lower numbers were assigned to better qualities of land; the numerical value rising as land quality decreased. It is for this reason that the negative coefficient of X_4 is logical. The coefficient may be interpretated in the following manner: as quality of the predominant land in the tract increased by one class, that is the numerical value falls by one, the price per acre increases by \$4.36267.

Studies[3] have shown a rather clearly defined relationship between price paid and the proportion of minerals conveyed in the sale. The coefficient for variable X_5 (the quantity of mineral right conveyed) agrees in sign with a priori reasoning. That is, as the proportion of mineral rights conveyed increases one class interval[4] the price per acre increases by \$1.56240.

The coefficient for variable X_6 (type of road on which the tract is located) is significant at the 10% probability level, but the sign of the coefficient does not agree with a priori reasoning. Logic would suggest that farms on good roads will sell for more than farms on poor roads and empirical studies have born out the truth of this relationship. However, one study[5] showed that the relationship might vary from the expected. In the study cited it was pointed out that other factors may outweigh road type and as a result farms on paved roads sometimes sell for less than farms on gravel roads. Here the equation indicates that the price increases as the road becomes lower in quality.

The distance one must travel on a lower quality road to reach pavement also influenced price. The negative sign of the variable X_7,

---

[3]L. A. Parcher, Some Factors Influencing Mineral Rights Separation in Land Sales, Bulletin B-431, Oklahoma Agricultural Experiment Station, Stillwater, Oklahoma, (July 1954).

[4]See p. 26 for distribution of class intervals. There are 10 class intervals which although not equal, correspond quite well to the actual proportions of minerals transferred in land market transactions. The change in value between classes, incidentally, indicates a very close correspondence with U. S. Geological Survey estimates of per acre values of mineral rights in the area.

[5]L. A. Parcher, The Influence of Location on Farmland Prices, Bulletin No. B-417, Oklahoma Agri. Experiment Station, (March 1954).

distance to a paved road, is logical and can be interpreted as follows: as distance to a paved road increases, the price per acre decreases.

The proximity of any town was also thought to positively influence the price paid for land. The coefficient for the variable X_8 (Distance to nearest town) is positive, but in the scheme used for coding this positive sign means that as distance increases the price per acre increases. This is not as would be expected and is a probable consequence of many inter-relationships.

The coefficients for variables X_9 and X_{10} are not significant, but show a positive relationship with land price. The magnitudes are also small.

The positive value of X_9 agrees with a priori reasoning; the positive value of X_{10} does not. The size of the nearest town and distance to a principal city may or may not affect land prices, but only slightly in any case.

Variable X_{11}, relates to the distance to a metropolitan area, because the numerical value of the code is lowest when the tract is nearest a metropolitan area, the sign of the coefficient is in agreement with a priori reasoning, that is, as distance to a metropolitan area decreases, land value increases.

Variable X_{12} (distance to Oklahoma City) is so coded that as the number of miles increases, the higher is the code number. The coefficient is economically consistent with a priori reasoning and statistically significant. The beta value is relatively large which indicates the importance of Oklahoma City in explaining the differences in price per acre between tracts.

Of the coefficients for the crop allotment variables, X_{13}, X_{14}, and X_{15} only X_{13} was statistically significant. The variable X_{13} is the number of acres of wheat allotment on the land sold. In most of the area included in this study, wheat is a more important crop than either cotton (X_{14}) or peanuts (X_{15}). Suitability of resources and grower preferences have caused this emphasis to be placed on small grain. It probably is for these reasons that we find the wheat acreage allotment to be a highly significant explanatory variable.

Summary

One of the main objectives of this study is to investigate the influence of market forces on the per acre price of farmland in an agriculture area. Therefore, the study was designed to use actual farmland prices of recently sold tracts of land in developing a systematic technique for estimating the per acre price of unsold land.

Previous studies have indicated that certain factors are associated with the price paid per acre (Y). These factors plus others were combined into a regression equation. The Regression technique was then utilized to estimate by an equation the price per acre. Such an equation can benefit real estate appraisers and assessors, but its usefulness will be limited to western Oklahoma.

Variables X_1 (size of the tract sold), X_2 (productive quality of farm), X_3 (type-use combination), X_4 (quality of predominant land in tract), X_5 (quantity of mineral rights) X_{12} (distance to Oklahoma City) and X_{13} (wheat allotment acres) are statistically significant and consistent with _a priori_ reasoning, which indicate their influence on price per acre of farmland in Western Oklahoma. Distance to Oklahoma City has a relatively large coefficient in explaining the difference in price per acre of farmland. Wheat allotment has a highly significant explanatory variable, since in most of the area of this study wheat is a more important crop than either cotton or peanut.

In the estimated regression equation, several of the variables were not significant and some had signs not in accord with a priori reasoning. But this does not necessarily mean that factors which were neither economically nor statistically significant are not important in explaining the difference in price per acre. Moreover, it was found that the independent variables used explained only about half of the difference in price per acre between tracts. This raised the question of alternative approaches to the problem. Are there alternative approaches which might lead to more efficient land pricing by individuals?

The next chapter discusses in some detail approaches to value which have been used and presents a new alternative approach, discriminant analysis, to see whether it will be useful in solving the problem of evaluation.

CHAPTER IV

DISCRIMINANT ANALYSIS IN THE CLASSIFICATION OF LAND PRICES

There are many ways one can study relationships which exist between the selling price of land and certain price influencing factors. One of the simpler approaches used in such studies is to determine how the land price varies when all factors but one are assumed to be the same. This has been the approach used in several studies in Oklahoma as well as elsewhere.

In studies of this nature some factor such as type of road or distance to market is assumed to be the only variable and sales are classified with respect to this single factor and its effect on price is determined. Variations from this approach have sometimes had sales being sorted into land quality and further sub-divided into tract size and then into type of road or distance to market. Such studies have been useful in that they have indicated that certain relationships do exist between land price and certain variable factors. This knowledge has helped to sharpen skills of land evaluators in estimating market values. But in this simplified approach, one can never be sure of the extent to which factors not considered have influenced results.

The many complex inter-relationships between variables affecting land values confound the researcher in his attempts to analyze cause and effect. Multiple regression analysis has been used in at least one other study which had seven variables deemed to affect land price. This study showed an R^2 value of .80[1] and Ahmed's study provided valuable insights as to the direction of the effect and relative importance or contribution of the various independent variables of land price determination in one county. His analysis was accomplished through interpretation of sign and size of the structural coefficients. However, many of the variables he used had coefficients which were not significantly different from zero, and while land prices could be predicted with fair accuracy for many sales there still existed fairly wide variations between estimated price and actual price in other sales when the formula was applied to tracts for which the price was known.

There is an apparent widespread interest[2] in a method which would help in predicting the market price of unsold tracts. In the approach, presented in this chapter, instead of a formula which attempts to estimate exact value, the formula will attempt to estimate on the basis of the independent variables a range of values into which the price of a given tract could be expected to fall. This approach is called discriminant analysis.

It is the purpose here to investigate the merits of this alternative predictive procedure and compare the results so obtained with those obtained by the regression technique.

---

[1] Mohammed A. Ahmed, p. 114.

[2] Particularly among tax assessors who are faced with the current evaluation of every ownership tract and among appraisers who spend much time in analyzing sales to arrive at an estimate of value of a given tract.

Very often the purpose of a study is to detect differences between groups of observations in order to discover the underlying cause and effect relationships. For instance, it is the purpose here to determine what differences may exist between qualities associated with high versus low land values. If it is possible to detect differences between these groups, then further observation and analysis may permit an explanation of why one tract of land sells for more than another.

Discriminant analysis will show whether it is possible to discriminate between several groups of land sales which range from low to high in value. Several independent variables are used in the discriminating equations in an effort to predict the group into which a given land sale would be most likely to fall. This allows each individual case to be analyzed separately by providing the probability of a case falling within one of the groups. The probability of not appearing in the group (1-P) is the error term of the individual case, whereas in the regression technique the error term is given only for the entire group. The advantage of the above model is that its prediction in terms of a probability gives some indication of cause of error. A discussion of the discriminant technique and its application may be useful at this point.

Discriminant analysis is a procedure for separating individual sales into discriminant groups, given specific independent variables.[3] Discriminant functions represent a fairly new addition to the statistical techniques that can be used by economists. During the past few

---

[3]W. W. Cooley and Paul R. Cohnes, Multivariate Procedures for the Behavioral Sciences (New York: John Wiley and Sons, Inc., 1965).

years, however, the potential value of discriminant functions in agricultural economics research has received increased attention. The discriminant function is useful whenever there are a set of "g" mutually exclusive classifications or groups and an individual must be assigned to one of these classes on the basis of a standard set of quantitative scores.[4] We first ascertain whether a significant difference exists between the groups, and then interest turns to:

(1) The distances separating the g groups,

(2) The directions in which the g groups differ, and

(3) The assignment to one of the g groups of an unclassified individual known to belong to one of the g groups. Significance, distance and direction, and assignment are the issues of discriminant analysis.[5]

The Technique

Discriminant analysis involves the computation of a set of linear functions for the purpose of classifying an individual into one of several groups. An individual is classified as belonging to that group for which the computed linear function corresponding to each of the groups has the largest probability (a probability ≤ 1). The group assignment procedure is derived from a model of a multivariate normal distribution of observations within groups such that the variance matrix is the same for all groups.

---

[4]Ardie Lubin, "Linear and Non-Linear Functions," British Journal of Psychology, Statistical Section 3 (June, 1950), p. 91.

[5]Maurice M. Tatsuoka and Davis V. Tiedeman, "Discriminant Analysis," Review of Educational Research (Washington, D. C.: American Educational Research Association, 1954).

The data are symoblized by:

$$X_{ijk}$$

Where

$i = 1, \ldots, g$, number of the groups,

$j = 1, \ldots, n_i$, sample size of the i^{th} group,

$k = 1, \ldots, m$, number of variables.

The mean of variables considered within each group can be denoted by:

$$X_{i \cdot k}$$

i.e., $(X_{i \cdot 1}, X_{i \cdot 2}, \ldots, X_{i \cdot m})$

$i = 1, \ldots, g$, number of the groups.

The sample size normally differs from one group to another, but the number of independent variables must be the same for all groups.

To develop such a model, we first calculate the means of the variables within each group. We next calculate the matrix S_i, which represents the sum of cross products of deviation from the means denoted by:

$$S_i = (S^i_{kk'})$$

$k' = 1, \ldots, m$, number of variables

$i = 1, \ldots, g$, number of groups

where

$$S^i_{kk'} = \sum_{j=1}^{n_i} (X_{ijk} - X_{i \cdot k})(X_{ijk'} - X_{i \cdot k'})$$

Then compute the pooled dispersion matrix based upon S_i, which is described by:

$$D = \frac{\sum_{i=1}^{g} S_i}{\sum_{i=1}^{g} n_i - g}$$

with the common mean:

$$X_{\cdot\cdot k} = \frac{\sum_{i=1}^{g} n_i X_{i\cdot k}}{\sum_{i=1}^{g} n_i}$$

To calculate the generalized Mahalanobis D^2 statistic, V, we have to invert the dispersion matrix D, denoted by $D^{-1}_{kk'}$:

$$V = \sum_{k=1}^{m} \sum_{k'=1}^{m} (D^{-1}_{kk'}) \sum_{i=1}^{g} n_i (X_{i\cdot k} - X_{\cdot\cdot k}) (X_{i\cdot k'} - X_{\cdot\cdot k'})$$

V is distributed as chi-square with m(g-1) degrees of freedom. The chi-square distribution can be used to test the hypothesis that the mean values are the same in each of the g groups for these m explanatory variables. The independent variables are capable of discriminating among groups if the hypothesis of no difference is rejected. If not rejected, there are no significant group differences, alternative variables should be selected and the above processes repeated. If the difference existed among the groups for the independent variables, the second step is to calculate the $(i^*)^{th}$ discriminating function:

$$f^{mx1}_{i^*} = \sum_{\ell=1}^{m} Z_\ell \ C_{i^*\ell} + C_{i^*o}$$

where:

$i^* = 1, \ldots, g$, the number of functions

$\ell = 1, \ldots, m$, the number of the variables

Z_ℓ = observation for each variable

$C_{i^*\ell}$ = classification function coefficient

C_{i^*o} = constant

and

$$C_{i*\ell} = \sum_{k=1}^{m} d_{\ell k} X_{i\cdot k}$$

$$C_{i*o} = -\tfrac{1}{2} \sum_{k=1}^{m} \sum_{k'=1}^{m} d_{kk'} X_{i\cdot k} X_{i\cdot k'}$$

where

$$(d_{k1}, d_{k2}, \ldots, d_{km}) = k^{th} \text{ row of } D^{-1}$$

D^{-1} = inverse of pooled dispersion matrix.

The $(i*)^{th}$ discriminating function is then used to evaluate each data point such that for each observation:[6]

$$P_i = \frac{e^{(f_{i*} - \max f_{i*})}}{\sum_{i*} e^{(f_{i*} - \max f_{i*})}}$$

resulting in a probability P_i that a single observation will fall in one of the groups. If the experimental groups are widely separated, then the diagonal of the frequency matrix will contain a large number compared to the off-diagonal elements (frequency of lesser probabilities).

Thus, it can be seen that discriminant analysis can be used as a unified approach involving multivariate comparison of several groups, which has as its three phases: (a) the establishment of significant group differences; (b) the study and "explanation" of these differences; and (c) the utilization of multivariates from the samples

---

[6]The above formulas are based upon: Biomedical Computer Programs, BMDO5M, "Discriminant Analysis for Several Groups," Health Services Computing Facility, Dept. of Preventive Medicine and Public Health, School of Medicine, University of California, Los Angeles (January, 1964).

studied in classifying a future individual known to belong to one of the groups represented.[7]

Classification and Analysis of the Data

The land sales were classified and analysis of the data arbitrarily divided into several price groups. These data were divided in two ways. First, the land sales were divided into ten groups, which had to be subdivided due to the restrictions of the program which limits the number in any one problem to five groups. The ranges of these ten groups are as follows:

Problem 1	25 - 50 dollars per acre
	51 - 75 dollars per acre
	76 - 100 dollars per acre
	101 - 125 dollars per acre
	126 - 150 dollars per acre
Problem 2	151 - 175 dollars per acre
	176 - 200 dollars per acre
	201 - 225 dollars per acre
	226 - 250 dollars per acre
	251 - 275 dollars per acre

Secondly, the data were divided into five groups as follows:

25 - 100 dollars per acre

101 - 150 dollars per acre

151 - 175 dollars per acre

176 - 250 dollars per acre

251 - 275 dollars per acre

---

[7] Maurice M. Tatsuoka and David M. Tiedeman, p. 414.

The data were tabulated and then punched on IBM cards for statistical analysis. The discriminant analysis program utilized the data without further transformation.

The first step in the program was the calculation of the mean score of each variable by group. These scores are given in Table IV for Problem 1. Analysis of the data at this stage is impossible since the independent variables are highly inter-related. The Generalized Mahalanobis D^2 statistics was calculated.[8] The value is 164.62434. A chi-square distribution was used to test the hypothesis that the mean values are the same in all groups. In this case the hypothesis was rejected at the .05 level of significance.

The second step of the discriminant program was to compute the discriminating functions, and the relative size and sign of these function coefficients. It is of some analytical value to observe the relative change in importance between variables, as shown in Table V.

In problem 1, in which the lower price groupings were analyzed, variables 7 and 8 have relatively small functional coefficients. Table V indicates that the explanatory variables 6, 9, 10, and 11 have relatively large functional coefficients, which indicate their importance in discriminating between the groups. Variable 12 (distance to Oklahoma City) has a very large functional coefficient thus contributing much of the explanation in price variation.

---

[8] A chi-square distribution with m(g-1) degrees of freedom can be used to test the hypothesis that there is no difference between means of the groups.

TABLE IV

MEAN SCORES OF 15 SELECTED INDEPENDENT VARIABLES BY PRICE GROUPS FOR 194 LAND SALES

Independent Variables	Price Classification Groups* 1 $25-50	2 $51-75	3 $76-100	4 $101-125	5 $126-150
	Mean Scores				
1	155.15789	144.78049	116.86274	124.97500	117.60465
2	6.00000	5.41463	5.58824	4.30000	3.90698
3	7.57895	7.07317	6.80392	5.37500	4.69767
4	6.78947	5.97561	6.00000	4.45000	5.00000
5	2.78947	4.07317	4.27451	4.80000	5.20930
6	3.68421	3.58537	3.47059	3.62500	3.67442
7	3.21053	2.90244	2.52941	2.30000	2.55814
8	3.52632	3.65854	3.74510	3.85000	4.20930
9	1.89474	2.51220	2.80392	2.92500	3.46512
10	3.63158	4.14634	3.45098	3.40000	3.51163
11	5.15789	4.90244	4.96078	4.67500	4.88372
12	4.84211	4.65854	4.64706	4.30000	4.44186
13	0.00000	0.00000	0.37255	0.50000	1.65116
14	0.00000	0.00000	0.72549	2.92500	0.93023
15	0.00000	0.00000	0.62745	0.05000	0.00000

*Sample size for groups 1 through 5 are 19, 41, 51, 40, and 43 respectively.

TABLE V

CLASSIFICATION FUNCTION COEFFICIENTS FOR 15 SELECTED
INDEPENDENT VARIABLES BY PRICE GROUPS
FOR 194 LAND SALES

Independent Variables	Price Classification Groups*				
	1 $25-50	2 $51-75	3 $76-100	4 $101-125	5 $126-150
	Function Coefficients				
1	0.01242	0.01089	0.00700	0.00940	0.00752
2	0.64978	0.38530	0.90089	0.37536	0.16852
3	0.54070	0.63432	0.23855	0.24424	-0.00156
4	1.37240	1.22711	1.23875	1.05756	1.37891
5	0.18112	0.26105	0.32705	0.39128	0.43160
6	1.66192	1.65957	1.75885	2.06619	2.09214
7	-0.21833	-0.31114	-0.54607	-0.73776	-0.58229
8	-0.41380	-0.29492	0.05222	0.34797	0.39533
9	3.79814	4.01301	3.87007	3.73543	4.08880
10	2.24060	2.60015	2.17561	2.02584	2.22335
11	5.44671	5.40140	5.11049	5.18851	5.56971
12	15.45845	14.99103	15.28560	13.99601	14.39953
13	-0.24461	-0.23131	-0.22051	-0.24731	-0.13529
14	0.21212	0.19182	0.21111	0.28450	0.22100
15	-0.94887	-0.94728	-0.63732	-0.94972	-0.96486
Constant	-70.99370	-68.84626	-57.82028	-60.36642	-65.82982

*Sample size for groups 1 through 5 are 19, 41, 51, 40, and 43 respectively.

Generalized Mahalanobis D-square 164.62434. The value 164.62434 can be used as chi-square with 60 degrees of freedom to test the hypothesis that the mean values are the same in each of the 5 groups for these 15 variables.

The discriminating functions were then utilized for determining the probabilities of each original land sale falling within each of the five groups depicting a range of low to high price land sales.

The classification of each case by probabilities was computed and is summarized in Table VI. These probabilities show the chance of the individual case appearing in each of the 5 groups. It is seen that not all individual land sales fall into the same group as ranked in the original sales. If all values fell on the main diagonal, we would have perfect classification. Thus the misclassifications fall in the off-diagonal elements. Inspection of Table VI reveals that slightly more than 44 percent of the cases were classified in the same way as the original ranking procedure.

In Problem 2, the land sales were divided as shown earlier with the means of the independent variables given in Table VII. The generalized Mahalanobis D^2 Statistic had a value of 127.85541 with 60 degrees of freedom (Table VIII). This test, when utilized as a chi-square is the basis for the rejection of the null hypothesis that the means were the same in all groups at the .05 significance level.

The classification of each case for Problem 2 was computed and is summarized in Table IX. Table IX indicates the classification matrix with 59.4 percent of the cases classified in the same manner as the original ranking procedure.

Whereas in Problem 1 the extreme groups had the higher percentage of classification agreement, in Problem 2 the classificaion agreement was more equally distributed except in Group 5 where 8 out of 10 were in agreement.

TABLE VI

CLASSIFICATION MATRIX FOR 194 LAND SALES

	Price Classification Groups						
	1	2	3	4	5		
Group	$24-50	$51-75	$76-100	$101-125	$126-150	Total	Percentage
1	(9)	5	3	1	1	19	47%
2	7	(18)	7	3	6	41	44%
3	9	10	(17)	9	6	51	33%
4	2	5	5	(18)	10	40	45%
5	2	5	5	9	(22)	43	51%
					Grand Total	194	

TABLE VII

MEAN SCORES OF 15 SELECTED INDEPENDENT VARIABLES
BY GROUPS FOR 99 LAND SALES

Independent Variables	Price Classification Groups* 1 $151-175	2 $176-200	3 $201-225	4 $226-250	5 $251-275
	Mean Scores				
1	118.59375	115.52632	96.45833	100.00000	119.90000
2	3.53125	2.78947	3.41667	2.64286	1.70000
3	4.40625	3.15789	3.62500	2.42857	1.10000
4	3.65625	2.00000	2.70833	1.64286	0.70000
5	6.12500	4.47368	5.16667	5.50000	5.20000
6	3.75000	2.84211	3.66667	3.64286	3.80000
7	2.71875	1.68421	1.66667	2.35714	3.30000
8	3.87500	4.00000	3.70833	3.50000	4.60000
9	3.68750	4.21053	3.66667	3.64286	2.70000
10	3.71875	3.84211	3.20833	3.57143	3.80000
11	4.62500	4.26316	4.41667	3.92857	4.20000
12	4.31250	4.05263	4.41667	3.78571	4.00000
13	4.96875	11.26316	5.45833	10.07143	15.60000
14	1.50000	0.73684	0.16667	0.00000	6.80000
15	0.68750	0.00000	0.00000	0.00000	0.00000

*Sample size for groups 1 through 5 are 32, 19, 24, 14, and 10 respectively.

TABLE VIII

CLASSIFICATION FUNCTION COEFFICIENTS FOR 15 SELECTED INDEPENDENT VARIABLES BY GROUPS FOR 99 LAND SALES

Independent Variables	Price Classification Groups* 1 $151-175	2 $176-200	3 $201-225	4 $226-250	5 $251-275
	Function Coefficient				
1	0.02876	0.02501	0.01830	0.02366	0.02333
2	0.85527	0.64781	1.19107	0.92137	0.14121
3	-0.28315	-0.29185	-0.46850	-0.61512	-0.88398
4	-0.05138	-0.21687	-0.13052	-0.27502	-0.27141
5	1.70558	1.31132	1.55324	1.54286	1.53432
6	5.11065	4.42291	5.28805	4.84518	4.73714
7	-0.61411	-0.75715	-1.28847	-0.46946	-0.16042
8	1.83865	2.03542	2.22673	1.53926	2.54932
9	4.33557	4.39587	4.08373	4.05244	3.67780
10	1.18156	1.45863	0.85824	1.30120	1.09615
11	5.38210	4.38100	3.81922	4.60867	4.72588
12	15.95882	15.95048	18.10658	14.84684	15.47064
13	0.08562	0.09855	0.09723	0.09635	0.10856
14	-0.26850	-0.22915	-0.27621	-0.25585	-0.07999
15	-0.10380	-0.12284	-0.06210	-0.12252	-0.28059
Constant	-77.05511	-68.51031	-76.22792	-63.98180	-68.01445

*Sample size for groups 1 through 5 are 32, 19, 24, 14, and 10 respectively.

Generalized Mahalanobis D-square 127.86541. The value 127.86541 can be used as chi-square with 60 degrees of freedom to test the hypothesis that the mean values are the same in all the 5 groups for these 15 variables.

TABLE IX

CLASSIFICATION MATRIX FOR 99 LAND SALES

	Price Classification Groups						
	1	2	3	4	5		
Group	$151-175	$176-200	$201-225	$226-250	$251-275	Total	Percentage
1	(19)	4	5	4	0	32	59%
2	1	(9)	5	2	2	19	47%
3	6	3	(13)	2	0	24	54%
4	0	3	2	(8)	1	14	57%
5	0	1	0	1	(8)	10	80%
					Grand Total	99	

In Problem 2, in which the higher price groupings were analyzed, variables 3, 4, 7, 14 and 15 have relatively small functional coefficients, while variable 12 (distance to Oklahoma City) remains large as in Problem 1. The explanatory variables 6, 9, and 11 have relatively large functional coefficients, which indicate their importance in discriminating between the groups. The magnitude and the sign of the coefficients can not be interpreted in the same manner as regression coefficients. Rather, the coefficients in discriminant analysis allow the calculation of a function which provides a mathematical procedure for classifying land price into one of five different groups.

In Problem 3, where the total data were divided into 5 groups, the Mahalanobis D^2 Statistic of 376.33968 with 60 degrees of freedom is the basis for rejecting the null hypothesis that the group means were the same at the .05 significance level. The means of the independent variables by groups are shown in Table X.

The classification function coefficients and constant terms are shown in Table XI and the classification of observations were calculated and are summarized in Table XII. The classification matrix reveals that 57.6 percent of the cases were classified as the original ranking procedure, indicating a reasonable fit of the model. As in Problem 1, the extreme groups show a better fit while the middle groups are less associated. This procedure may indicate that fewer classifications could lead to an even better classification or grouping. For example, for classifying lower quality[9] land the use of the selected independent

---

[9]"Quality" refers here to all factors which appear to affect price.

TABLE X

MEAN SCORES OF 15 SELECTED INDEPENDENT VARIABLES BY GROUPS FOR 293 LAND SALES

Independent Variables	Price Classification Groups* 1 $25-100	2 $101-150	3 $151-175	4 $176-250	5 $251-275
	Mean Scores				
1	133.72973	121.15663	118.59375	103.68421	119.90000
2	5.59459	4.09639	3.53125	3.01754	1.70000
3	7.03604	5.02410	4.40625	3.17544	1.10000
4	6.12613	4.73494	3.65625	2.21053	0.70000
5	3.94595	5.01205	6.12500	5.01754	5.20000
6	3.54955	3.65060	3.75000	3.38596	3.80000
7	2.78378	2.43373	2.71875	1.84211	3.30000
8	3.67568	4.03614	3.87500	3.75439	4.60000
9	2.54054	3.20482	3.68750	3.84211	2.70000
10	3.73874	3.45783	3.71875	3.50877	3.80000
11	4.97297	4.78313	4.62500	4.24561	4.20000
12	4.68468	4.37349	4.31250	4.14035	4.00000
13	0.17117	1.09639	4.96875	8.94737	15.60000
14	0.33333	1.89157	1.50000	0.31579	6.80000
15	0.28829	0.02410	0.68750	0.00000	0.00000

*Sample size for groups 1 through 5 are 111, 83, 32, 57, and 10 respectively.

TABLE XI

CLASSIFICATION FUNCTION COEFFICIENTS FOR 15 SELECTED INDEPENDENT VARIABLES BY GROUPS FOR 293 LAND SALES

Independent Variables	Price Classification Groups* 1 $25-100	2 $101-150	3 $151-175	4 $176-250	5 $251-275
	Function Coefficient				
1	0.01699	0.01533	0.01467	0.01121	0.01143
2	1.28567	0.70985	0.48687	0.78577	0.47771
3	0.09365	-0.09993	-0.02007	-0.23308	-0.64679
4	0.69603	0.70453	0.49618	0.30618	0.27932
5	0.62408	0.71875	0.81069	0.68940	0.78134
6	2.40307	2.74231	2.75899	2.77180	2.93451
7	-0.64057	-0.78229	-0.52370	-0.82175	-0.55142
8	0.64459	1.05706	0.74875	0.91973	1.52564
9	3.72873	3.68815	3.93784	3.83578	3.23588
10	1.87875	1.76014	1.95957	1.94153	1.71594
11	4.49028	4.78857	4.63384	4.12598	4.57644
12	13.70344	12.89768	13.12264	12.78982	11.41559
13	0.04550	0.05088	0.07565	0.10271	0.13626
14	0.01270	0.06247	0.01679	0.01183	0.16980
15	-0.44272	-0.60355	-0.33032	-0.43914	-0.58886
Constant	-64.43728	-60.47675	-60.98941	-55.59882	-52.75028

*Sample size for groups 1 through 5 are 111, 83, 32, 57, and 10 respectively.

Generalized Mahalanobis D-square 376.33968. The value 376.33968 can be used as chi-square with 60 degrees of freedom to test the hypothesis that the mean values are the same in all the 5 groups for these 15 variables.

TABLE XII

CLASSIFICATION MATRIX FOR 293 LAND SALES

	Price Classification Groups						
	1	2	3	4	5		
Group	$25-100	$101-150	$151-175	$176-250	$251-275	Total	Percentage
1	(80)	14	12	5	0	111	72%
2	16	(36)	14	9	8	83	43%
3	2	7	(11)	10	2	32	34%
4	8	5	6	(28)	10	57	49%
5	0	0	1	0	(9)	10	90%
					Grand Total	293	

variables apparently would result in the proper price range grouping in 72 percent of the cases. For the higher quality land the proper price classification apparently would result in 90 percent of the cases. The $151 to $175 price range appears to be the most difficult to determine on the basis of the selected independent variables.

In Problem 3, in which all the price groupings are analyzed, variables 7 and 15 have negative functional coefficients. Variable 12 remains large and positive as before. The independent variables 6, 9, and 11 have a relatively large function coefficients, which indicate their importance in price variation.

When comparing the size and the sign of the functional coefficients in the three problems, we note that variables 7, distance to a paved road, variable 12, distance to Oklahoma City, and variable 15, peanut allotment acres retain the same sign throughout the process. Variables 7 and 15 retain small coefficients, variable 12 retains a large coefficient.

Summary

The use of multiple discriminant analysis to classify land sales allows one to ascertain whether a tract of land can be assigned to a price range group on the basis of selected value influencing factors. The program permits the computation of discriminating functions. Its usefulness lies in the fact that it permits one to observe the relative change in importance between variables.

Discriminant functions can be used for predicting the probable price range of unsold tract when its physical and locational characteristics are known. The function coefficients can be applied to data for unpriced land for determining the probability that the price of a given tract of land will conform in value with one of the groups. The procedure for doing this is as follows: The physical and locational characteristics of an unpriced tract of farm land can be coded according to the coding outlined in Chapter III. Then the code numbers of the characteristics are multiplied by the function coefficients as shown in Table VIII. The results are added to the constant terms and a total is obtained. This gives a discriminatory function which can be placed in a computer to get the probability that the unpriced tract would fall within one of the five price groups. The nature of the program is such that the sum of the probabilities that an unpriced tract will fall within one of the whole range of prices will equal one. But one fractional portion will be larger in a particular price grouping, and this reveals the probability that this tract will sell within this price range.

A study of the functional coefficients indicated that some of the variables are more important than others in the classification procedure. In Problem 1 in which the lower price groupings were analyzed, variables 6 (type of road), 9 (size of nearest town), 10 (distance to a principal city), and 11 (distance to a metropolitan center) have relatively large functional coefficients, indicating their greater importance in price variation between tracts. In problem 2 where the higher priced land is analyzed, variables 6 (type of road), 9 (size of nearest town), and 11 (distance to a metropolitan center) play a much greater part in land price variation. These same variables were also important in problem 3 where the data were combined.

In examining the relative values of the derived classification function coefficients of the selected explanatory variables, certain variables seemed to stand out in importance. One in particular, distance to Oklahoma City dominated in all three problems. However, variables 6 (type of road), 9 (size of nearest town), and 11 (distance to any metropolitan area) also had high values for the function coefficients in all three problems.

CHAPTER V

SUMMARY AND CONCLUSIONS

Land is becoming an increasingly important component of resource cost to the farm firm. Any market estimate which measures the importance of this component requires a thorough analysis of all factors influencing market price and a weighing of the impact of these factors on the property being evaluated. A method of systematically evaluating the land resource would be important to farmers and to investors. While a study or an analysis of value influencing forces ordinarily do not give value as such, the process may provide a basis upon which judgments can be made or action taken.

This study attempted to explore the factors which influence the market value of individual tracts of land in an endeavor to explain why market values vary from one tract to another. For the purposes of this study, several factors were deemed to affect farmland market value in Western Oklahoma. The market price of real property is considered to be a satisfactory measure of value.

To analyze sales and to study the factors affecting the variation in land prices in the study area, an attempt was made to select all factors which might cause one tract of land to sell for more or less than another. In the analysis those factors were selected which logic would lead one to believe were most likely to influence the price of an individual tract of land.

The variables which were assumed to be related to value as reflected by market price included variables which reflected the quality, type, and use of farmland, its location, and crop acreage allotments. The latter variables have been shown in other studies to be an important factor in selling price of land.

It is well known that differences in market value exist among various tracts of land. Such differences are not unexpected since each tract of land and each sale has unique characteristics. It was hypothesized that this uniqueness is based upon certain variables which can be measured and that these variables will help to explain differences in value.

The final decision with respect to value based upon formula cannot be exact nor can it be expressed easily as a single dogmatic statement. But it can be said that an estimate is the best clue to value, since it is based on observations of the basic economic forces which influence value.

In the analyses of the selected factors, two types of techniques were employed: (1) least squares regression, and (2) discriminant analysis. The discriminant analysis function was utilized to generate the probabilities that a given land sale will fall into one of several price classification groups.

Factors influencing price are so complex and inter-related that it is extremely difficult to determine what forces have been at work to set the value on a given piece of land. Through the use of the two techniques mentioned above, however, the factors influencing price were evaluated with greater success than previous studies permitted.

The data on selected variables were secured and analyzed with price per acre of land set as a linear function of the fifteen variables.

The least squares technique was used to estimate the parameters of the following model:

$$Y_t = \beta_o + \beta_1 X_{1t} + . . ., + \beta_{15} X_{15t} + \varepsilon_t$$

t = 1, 2, ..., m, number of observations.

Price of land is a function of the many independent variables. The 15 independent variables tested in this study explained only a little more than half of the difference in price (R^2 = .51).

$$\hat{Y} = 235.02485 - .07917X_1 - 5.62210X_2 - 5.26324X_3 - 4.35267X_4 + 1.56240X_5 + 5.07328X_6 - 1.15358X_7 + 4.43975X_8 + .47418X_9 + .02321X_{10} - 3.72834X_{11} - 11.57324X_{12} + .77419X_{13} + 12393X_{14} + .95618X_{15}$$

$$R^2 = .51$$

The above equation could be useful in estimating the per acre price of farmland because an estimate so obtained is superior to guesswork. The main limitation of the equation is that it yields an estimate which might be interpreted as being more precise than it actually is.

Three of these variables were statistically significant at the 10% probability level, three of them were statistically significant at 5% probability level, and three of them were statistically significant at 1% probability level. Six of the variables had coefficients which were not significantly different from zero at 10% probability level or less and some had signs which did not agree with <u>a priori</u> reasoning.

The linear regression equation showed a relationship between price per acre of farmland and the explanatory variables of only $R^2 = .51$, and since the estimated value sometimes is different from the actual value per acre of farmland, a discriminant analysis technique was applied to the data to give a range of price into which a tract of land would fall. The use of the technique revealed differences between groups of land sales so that differences between qualities associated with the various groups could be studied.

The data were first divided into ten price range groups. Because only five groups can be handled at a time in a discriminant analysis, the ten price groups were divided into two problems. Finally the entire range of price was divided into five price range groups. In all of these classification schemes it was found that the difference between groups was significant at the .05 level. The discovered difference between the groups explains the importance of the selected explanatory variables as economic forces contributing to price variations of farmland.

In the first problem 194 land sales were grouped into relatively narrow price ranges. This resulted in the model successfully discriminating in 44% of the cases. In the second problem, using the same range in price for 99 higher priced land sales, the model performed more adequately, discriminating successfully in 59.4% of the cases. Then, when all observations were divided into five price ranges (Problem 3), the success rate fell to 57.6%. This resulted because of the higher percentage of misclassification in the middle price ranges. The percentage of correct classification was very high at the extreme ranges. One may conclude from this experience with the discriminating model that wider ranges in the middle price groups might result in a greater rate of success in classification. This is what one would expect, but only through trials such as reported in this study can the best range and number of group classifications be determined.

Development of the classification function coefficients allows the researcher to make two determinations: (1) given similar data on land the price of which is unknown, one can determine the price class range into which the particular land sale would be most likely to fall; and (2) by observing the size and sign of the coefficient, one can determine the importance or significance of the individual variables as discriminators in relation to the other variables of less size and different sign.

Because of the more gross classification and favorable success rate, problem 3 was used for illustrative purposes. In that problem variable 12 (distance to Oklahoma City) was the most important factor affecting land sales value.

This result is very important because of its simplicity in application, i.e., an assessor can easily determine this variable and weigh his estimated values accordingly. Variables 6 (type of road), 9 (size of nearest town), and 11 (distance to a metropolitan area) are also important and, while not as strong a factor in price as distance to Oklahoma City, are easily found by the land evaluator.

The ability to discriminate between land sales provides results which suggest that a more gross classification assures a greater success rate. The model also provides some explanation by evaluation of the classification coefficients although it must be remembered that discriminant analysis is primarily a predictive model.

Weaknesses of the Study

It was proposed that the equation fitted to the data might be used in estimating the per acre price of farmland. Its empirical results, however, are applicable only to Western Oklahoma, unless there are counties identical to Western Oklahoma in all respects. The main weaknesses of this study can be summarized as follows:

(1) Data not statistically selected, that is, all sales obtainable were used.

(2) The conditions surrounding the sales were not known, therefore, only physical and geographic factors were considered.

(3) Sales, perhaps, were too few to really permit a definite analysis of sales.

(4) There was a preponderance of sales of Indian land included in the analysis. Such sales may not be fully representative of all sales.

(5) Sales occurred over a period of time and were adjusted by means of the index of land prices in Oklahoma, which may not be proper for individual sales.

(6) The approaches to estimated market price likely will be of little practical use to the typical land evaluator since the number of variables used is so great that the use of an electronic computer is almost a necessity.

To keep abreast of the changes in the farmland market, it may be necessary to continually re-examine the relationships between price per acre of farmland and the explanatory variables. Perhaps a re-examination every five years would suffice. Any new explanatory variables found to influence the price per acre of farmland can be used in a discriminant analysis predictive model to help land evaluators to estimate land sales prices.

Need for Further Research

Methods were developed which provide a classification and analysis of land sales. These methods also permit the assessment of the variables as to their contribution to the prediction of land sales into one of the price classification groups. The assessment of the importance of contribution of a given variable is still unanswered except in gross terms.

The differences between the contribution of the many variables needs further analysis. Possibly the interactions of several variables is the significant factor in a good prediction or model. The proposed techniques for further research would include factor analysis or principal component analysis. A need for application of these and other untried statistical techniques is evident.

It would seem particularly useful if it were possible to eliminate variables or use new and fewer variables to get the program into a more manageable form.

Other variables and other areas of the study could be analyzed as was done in this study. These results could be compared to these obtained in this study to adequately assess the techniques used.

SELECTED BIBLIOGRAPHY

Alfred, Ring. The Valuation of Real Estate. Englewood Cliffs, New Jersey: Prentice-Hall, Inc., 1963.

Allen, R. G. D. Mathematical Economics. London: Macmillan and Company, 1959.

Anderson, R. L., and Bancroft, T. A. Statistical Theory in Research. New York: McGraw-Hill Book Company, 1952.

Bartlett, Maurice. Multivariate Analysis. Supplement to the Journal of the Royal Statistical Society, g, 1947, pp. 176-190.

________. "The Standard Errors of Discriminant Function Coefficients." Supplement to the Journal of the Royal Statistical Society, 6, 1939, pp. 168-73.

Baten, William D., and Hatcher, Hazel M. "Use of the Discriminant Functions in the Proximate Coal Analyses." Industrial and Chemical Engineering, 16 (January, 1944), pp. 32-34.

Baumol, William J. Economic Theory and Operations Analysis. Englewood Cliffs, New Jersey: Prentice-Hall, Inc., 1965.

Blood, Dwight Mr. "Discriminant Analysis and Farm Management Research." Paper presented at GP-2 Technical Committee Workshop, Lincoln, Nebraska, May 5-7, 1959.

Boulding, Kenneth E. Economic Analysis. New York: Harper and Brothers, 1955. Biomedical Computer Programs (BMD, W. J. Dixon, Editor), Health Sciences Computing Facility, Department of Preventive Medicine and Public Health, School of Medicine, University of California, Los Angeles, January 1, 1964.

Brown, George W. "Discriminant Functions." Annals of Mathematical Statistics, 18 (December, 1947), pp. 514-28.

Carlson, Sune. A Study in the Pure Theory of Production. London: King and Son, 1939.

Chamberlin, E. H. The Theory of Monopolistic Competition. Cambridge: Harvard University Press, 1948.

Cooley, W. W., and Paul R. Lohnes. Multivariate Procedures for the Behavioral Sciences. New York: John Wiley and Sons, Inc., 1962.

Dixon, Wilfred T., and Frank J. Massey, Jr. Introduction to Statistical Analysis. Second Edition. New York: McGraw-Hill, 1957.

Ely, Richard T., and George S. Wehrwein. Land Economics. New York: The Macmillan Company, 1940.

Farm Real Estate Values in the United States by Counties, 1850-1959. Edited by Thomas J. Pressly and William H. Scofield. Seattle: University of Washington Press, 1965.

Fellner, William. Emergence and Content of Modern Economic Analysis. New York: McGraw-Hill, 1960.

Fisher, Ronald A. The Statistical Utilization of Multiple Measurements (August, 1938), pp. 376-86.

Friedman, Milton. Price Theory, A Provisional Text. Chicago: Aldine, 1962.

Heady, Earl O., and others. Dynamics of Land Use - Needed Adjustment. Ames, Iowa: The Iowa State University Press, 1961.

Heady, Earl O., and Luther G. Tweeten. Resource Demand and Structure of the Agricultural Industry. Ames: The Iowa State University Press, 1963.

Henderson, James M., and Richard E. Quandt. Microeconomic Theory. New York: McGraw-Hill, 1958.

Hicks, G. R. Value and Capital. Oxford: The Clarendon Press, 1959.

Hoel, Paul G. Introduction to Mathematical Statistics. New York: John Wiley and Sons, 1947.

Johnson, Webster W., and Raleigh Barlowe. Land Problems and Policies. New York: McGraw-Hill, 1954.

Kendall, M. G. A Course in Multivariate Analysis. London: Charles Griffin and Company, 1957.

Klein, L. R. Introduction to Econometrics. Englewood Cliffs, N.J.: Prentice-Hall, 1962.

Kurnow, Ernest. Theory and Measurement of Rent. Philadelphia and New York: Chilton Company, Book Division, 1961.

Leftwich, Richard H. The Price and Resource Allocation. New York: Holt, Rinehart and Winston, March, 1960.

Liebhafsky, H. H. The Nature of Price Theory. Homewood, Illinois: The Dorsey Press, Inc., 1963.

Lubin, Aride. "Linear and Non-Linear Discriminating Functions." British Journal of Psychology, Statistical Section 3 (June, 1950), pp. 90-104.

Malthus, Robert. "The Principles of Political Economy." The International Economic Circle, Tokyo, 1936.

Mill, L. S. Principles of Political Economy With Some of Their Applications to Social Philosophy. New York: Longmans, Green, and Co., 1929.

Mohammed, M. A., Ahmed. An Economic Evaluation for Tax Assessment. Unpublished Ph.D. Thesis. Stillwater: Oklahoma State University, May, 1964.

Mood, A. M. Introduction to the Theory of Statistics. New York: McGraw-Hill, 1950.

Nelson, Ted Richard. An Econometric Model of the Land Market Stressing Effects of Government Programs on Land Values. Unpublished Ph.D. Dissertation. Stillwater: Oklahoma State University, 1964.

Ostle, Bernard. Statistics in Research. Ames: The Iowa State University Press, 1963.

Parcher, Loris A. The Influence of Location on Farmland Price. Oklahoma Agricultural Experiment Station, Bul. No. 417 (March, 1954).

Renne, Roland R. Land Economics. New York: Harper and Brothers, 1947.

Ricardo, David. The Principles of Political Economy and Taxation. Toronto, Canada: T. H. Best Printing Co., 1948.

Robert, F. Boxley, Jr., and W. L. Gibson. Peanut Acreage Allotments and Farm Land Value. Blacksburg, Virginia: Virginia Polytechnic Institute, Bul. 175, September, 1964.

Robinson, Y. Economics of Imperfect Competition. New York: Macmillan Company, 1954.

Roll, Eric. A History of Economic Thought. New York: Prentice-Hall, 1949.

Samuelson, P. A. Foundations of Economic Analysis. Cambridge, Massachusetts, 1948.

Smith, Adam. The Wealth of Nations. New York: The Modern Library, Random House, Inc., 1937.

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

Tatsuoka, Maurice M., and Davis V. Tiedman. Discriminant Analysis. Review of Educational Research, American Educational Research Association, Washington, D. C., 1954.

Theil, H. Economic Forecasts and Policy. Second Edition. North Holland, 1961.

Tiedeman, David V., and Joseph G. Bryan. "Prediction of College Field of Concentration." Harvard Educational Review, 24 (1954), pp. 122-139.

Tinter, Gerhard. "Some Application of Multivariate Analysis to Economic Data." Journal of the American Statistical Association, 41 (December, 1946), pp. 472-500.

________. Econometrics. New York: John Wiley and Sons, Inc., March, 1959.

Tweeten, Luther Gilbert. An Economic Analysis of the Resource Structure of United States Agriculture. Unpublished Ph.D. Thesis. Ames: Iowa State University, 1962.

VITA

Farid El-Hussiny Abdel-Badie

Candidate for the Degree of

Doctor of Philosophy

Thesis: AN ECONOMETRIC ANALYSIS OF FACTORS AFFECTING LAND VALUES IN WESTERN OKLAHOMA

Major Field: Agricultural Economics

Biographical:

Personal Data: Born in Abo-Kebir, Sharkiyah, Egypt, April 2, 1939, the son of El-Hussiny Abdel-Badie and Z. A. Amro.

Education: Attended elementary school and secondary school at Abo-Kebir, Sharkiyah, Egypt; graduated from secondary school in 1956; received the Bachelor of Science Degree (Agricultural Economics) from Ein-Shams University, College of Agriculture, Cairo, Egypt, in June, 1960; received the Master of Science Degree from University of California, Berkeley, in July, 1964, with a major in Agricultural Economics. Engaged in graduate study toward the Degree of Doctor of Philosophy at Oklahoma State University, Stillwater, Oklahoma from September, 1964, to June, 1967.

Professional Experience: Taught Agricultural Economics at Baniseuf High School, Egypt, from September, 1960 to December, 1961; received a scholarship from the Egyptian government to study for the Ph. D. Degree. At the end of my third year at Ein-Shams University, I was given one of the scholarships awarded to the two best students in Agricultural Economics.