AGENCY FOR INTERNATIONAL DEVELOPMENT
WASHINGTON, D. C. 20523

BIBLIOGRAPHIC INPUT SHEET

FOR AID USE ONLY

1. SUBJECT CLASSIFICATION

A. PRIMARY: Agriculture

B. SECONDARY: Fertilizers

2. TITLE AND SUBTITLE

Specification errors and empirical estimation of generalized crop-fertilizer production functions

3. AUTHOR(S)

Batavia, Baia Nandlal

4. DOCUMENT DATE	5. NUMBER OF PAGES	6. ARC NUMBER
1974	115 p.	ARC

7. REFERENCE ORGANIZATION NAME AND ADDRESS

North Carolina State University, Agricultural Experiment Station, Raleigh, North Carolina 27607

8. SUPPLEMENTARY NOTES (Sponsoring Organization, Publishers, Availability)

Doctoral dissertation - North Carolina State

9. ABSTRACT

A report on the process of making the optimum fertilizer recommendations to potato farmers in Peru. The data comes from potato fertilizer experiments carried out in 73 different locations in the Peruvian Sierra region from 1967 to 1971. Models at two levels of aggregation were considered, group models with five data groups leased on Buol's Soil Management Classification System and aggregate models with one data group. The group models were assumed to be true and the aggregate models to be wrong. Fertilizer recommendations and profits using each of the estimated models were made for a number of soil classes using these procedures: 1) for soils not known to belong to a given group, the La Molina soils data and the fixed efficients model were used, and 2) for soils known to belong to a given group, the experimental soils data and the group models were used. Results indicated the importance of considering all the factors in the response relation.

10. CONTROL NUMBER: PN-AAB-480

11. PRICE OF DOCUMENT

12. DESCRIPTORS

13. PROJECT NUMBER

14. CONTRACT NUMBER: CSD-2806 Res.

15. TYPE OF DOCUMENT

AID 590-1 (4-74)

22

SPECIFICATION ERRORS AND EMPIRICAL ESTIMATION OF GENERALIZED CROP-FERTILIZER PRODUCTION FUNCTIONS

by

BALA NANDLAL BATAVIA

A thesis submitted to the Graduate Faculty of
North Carolina State University at Raleigh
in partial fulfillment of the
requirements for the Degree of
Doctor of Philosophy

DEPARTMENT OF ECONOMICS

RALEIGH

1 9 7 4

APPROVED BY:

_____	_____
Cochairman of Advisory Committee	Cochairman of Advisory Committee

ABSTRACT

BATAVIA, BALA NANDLAL. Specification Errors and Empirical Estimation of Generalized Crop-Fertilizer Production Functions. (Under the direction of RICHARD KIDD PERRIN and PAUL REYNOLDS JOHNSON).

This study is concerned with generating optimal fertilizer recommendations to individual potato-growing farmers in Peru using estimated generalized crop-fertilizer production functions. The experimental data come from 73 locations of potato fertilizer experiments conducted in the Sierra region of Peru during the period 1967-1971. A generalized crop-fertilizer function describes the response relation between measured soil characteristics, applied nutrients, weather and crop yield. Models at two levels of aggregation were considered, the group models based on pooling data from locations into five groups on the basis of Buol's Soil Management Classification System and the aggregate models based on pooling data from all the locations. The group models were assumed to be the true structure representing the response relation.

Three specifications of the aggregate model, the fixed coefficient (FC) model, the generalized least squares (GLS) model, and the random coefficients (RC) model, were considered. The assumptions and the estimation procedures of the different specifications of the aggregate model were described. The relationship between the different specifications of the aggregate model and the group models were considered in detail.

The validity of Buol's criteria for grouping the experimental soils into five groups was established by F tests, which resulted in a

rejection of the hypothesis that the coefficients between any two groups are equal.

Many of the estimated models did not satisfy the second order conditions for a maximum and/or generated fertilizer optima that were beyond the experimental range. Hence, the steepest ascent method with constraints was used to generate the fertilizer optima.

Fertilizer recommendations for a number of soil classes were generated using each of the estimated models. Assuming the group models to be true and the aggregate model to be wrong, the expected losses in using the different specifications of the aggregate model were found to be U. S. $45.28/acre for the fixed coefficients model, U. S. $56.26/acre for the generalized least squares model, and U. S. $78.85/acre for the random coefficients model. Using the "cost of the wrong decision" criterion, the fixed coefficients model was chosen as the "best" model for generating fertilizer recommendations for those soils not known to belong to any of the five groups.

Fertilizer recommendations and profits for the different soil classes were estimated using the following procedures: (1) For soils not known to belong to a given group, the La Molina soils data and the fixed coefficients model were used and (2) For soils known to belong to a given group, the experimental soils data and the group models were used. The results show that it is important to take into account simultaneously all the factors in the response relation. The value of soil testing for the different group models and the fixed coefficients model ranged from U. S. $4.22/acre to U. S. $70.91/acre. The value of Buol's classification for each of the groups ranged from U. S. $27.17

to $94.43/acre and was more than two times the value of soil testing. Assuming the fixed coefficients model to be true, the cost of using the recommendations given by Ryan and the La Molina Laboratory was substantial and ranged from 27.9 to 37.5 per cent of the expected profits for the fixed coefficients model.

BIOGRAPHY

The author was born July 19, 1945, in Madras, India. He received his Bachelor's degree in textile technology from the University of Madras in June 1967. In September 1967, he was awarded a graduate research assistantship at the School of Textiles, North Carolina State University, which enabled him to come to the United States to pursue his Masters degree in textile technology. In January 1970, after completing his M.S., he transferred to the Department of Economics to work towards the Doctor of Philosophy degree in economics.

ACKNOWLEDGMENTS

The author expresses his gratitude to the cochairmen of his advisory committee, Dr. R. K. Perrin and Dr. P. R. Johnson, for their guidance and advice throughout the course of this study. The assistance of Dr. T. D. Wallace and Dr. E. C. Pasour in offering criticism and helpful comments is also acknowledged.

The author would also like to thank Dr. W. D. Toussaint for giving the opportunity to undertake this work and to the Agency for International Development for the financial support.

Dr. S. W. Buol, Dr. R. E. McCollum, and Dr. S. Villagarcia, of the Department of Soil Science, North Carolina State University, were most helpful in classifying the soils of the potato-response data from Peru. My thanks to all of them.

Libby Amos, Harriet McLaughlin, and Ann McDermed helped with the numerous computer programming routines used in this work. Ruth Good undertook the final editing and Shirley Isenhour skillfully typed the final draft.

TABLE OF CONTENTS

Page

LIST OF TABLES vi

LIST OF APPENDIX TABLES vii

INTRODUCTION 1

REVIEW OF LITERATURE ON CROP-RESPONSE STUDIES 4

SPECIFICATION ERRORS AND THE ESTIMATION OF DIFFERENT MODELS 10

Functional Form of the Models and the Generation of Optimal Fertilizer Recommendations 12

Structural Model 20

"Cost of the Wrong Decision" Analysis 23

Aggregate Model Specifications and Estimation 25

The Fixed Coefficients (FC) Model 26
The Generalized Least Squares (GLS) Model 27
The Random Coefficients Regression (RC) Model 30

Statistical Tests and the Relationship between the Different Specifications of the Aggregate Model and the Group Models. 40

EMPIRICAL ESTIMATION AND RESULTS 46

The Data 46

Results of the Group Models 47

Results of the Aggregate Models 61

"Cost of the Wrong Decision" Criteria 69

Sensitivity of Optimum Fertilizer Recommendations to Changes in the Soil Variables 77

Value of Soil Testing and Buol's Soil Classification System 85

SUMMARY AND CONCLUSIONS 92

LIST OF REFERENCES 96

TABLE OF CONTENTS (continued)

Page

APPENDICES. 100

Appendix A. Buol's Soil Management Classification System. 101

Appendix B. Simple Statistics of the Variables in the Experimental Data 104

LIST OF TABLES

Page

1. Characteristics of the group data. 48

2. The estimated coefficients and their 't' values for the group models with all the variables included 49

3. The estimated coefficients and their 't' values for the group models with some of the variables (non-significant coefficients at $\alpha = 0.25$) excluded 55

4. Test for equality of coefficients for the group models . . . 61

5. The estimated coefficients and their 't' values for the different specifications of the aggregate model. 63

6. Optimal fertilizer recommendations, yields and profits derived from using the group models for the different soil classes. 72

7. Optimal fertilizer recommendations, yields and profits derived from using the fixed coefficients model for the different soil classes 79

8. Value of soil testing for the group models and the fixed coefficients model, using experimental data. 87

9. Value of Buol's soil classification system 90

LIST OF APPENDIX TABLES

Page

Appendix A

1. Buol's criteria for the fertility capability soil classification system. 102

Appendix B

1. Means and standard deviations of the variables from the experimental data for the group and the overall data 105

CHAPTER 1
INTRODUCTION

In economic analysis of fertilizer responses, a decision is often made to pool all the available data and estimate a single response relation. The purpose of obtaining a single relation for the pooled data is to achieve generality for the derivation of optimum fertilizer recommendations from the relation. The relevant questions raised by such a procedure are:

(1) Assuming the micro-relations for subsets of the complete data to be true, what are the different specifications of the aggregate model that can be considered? What is the relationship between the different specifications of the aggregate model and the micro-relations? By different specifications we mean the different assumptions regarding the stochastic structure of the model.

(2) Assuming the estimated micro-relations to be true, what are the losses, as measured by the fall in the earnings of farmers, incurred by using the different specifications of the aggregate model for deriving optimum fertilizer recommendations?

The main aim of this work is to generate optimal fertilizer recommendations to individual potato growing farmers in Peru. The experimental data come from 73 locations of potato-fertilizer experiments conducted in the Sierra region of Peru during the period 1967-1971. Two aggregation levels are considered, the combining of locations into groups based on Buol's (1972) soil management classification system, and the completely pooled data. The models based on group data are considered as group models and are assumed to be the true structure. The

basic assumptions underlying the structural relationship are that each group model has a different coefficient vector and a different error variance. The models based on completely pooled data are termed as aggregate models. To the data from each of the aggregation levels, a 'generalized' crop-fertilizer is fit. A 'generalized' crop-fertilizer function refers to one which can be utilized under a wide range of soil conditions to generate fertilizer optima based on soil analysis of a site or location. In the words of the National Academy of Science (1961, p. 49):

> One of the most challenging problems in statistical methodology relating to agronomic-economic research in response surface studies is that of developing theory and methods of combining results from a series of experiments to form a generalized production function. To be of practical usefulness, this function must be such that it may be employed over a comparatively large population of conditions by having at hand values for independent variables, such as soil test results, to insert to account for local variations in conditions.

Chapter 2 briefly reviews the literature of the problems concerning the derivation of a generalized crop-fertilizer production function from pooled data.

Chapter 3 deals with the models to be estimated at the two different levels of aggregation. Three different specifications of the aggregate models are considered in this work: the fixed coefficients model, the generalized least squares model, and the random coefficients regression model, recently proposed by Swamy (1970). The relationship between the various specifications of the aggregate model and the group models is considered. The properties of the estimators of the coefficients of each of the specifications of the aggregate model under

various assumptions will also be described in detail. Statistical tests for testing some of these assumptions will be presented. Two ways of generating optimal fertilizer recommendations are considered:

(1) For soils known to belong to any particular group, that estimated group model is used for generating fertilizer recommendations.

(2) For soils not known to belong to any given group, one of the specifications of the aggregate model is chosen for generating fertilizer recommendations to individual farmers.

The choice of the specification is based on the 'cost of the wrong decision' criterion as proposed by Havlicek and Seagraves (1962). This method assumes the estimated group models to be true and estimates the average reduction in net revenues obtained by using the fertilizer recommendations from each of the three specifications of the aggregate model.

Chapter 4 presents the empirical results obtained from the various models. Buol's criteria for classifying soils into homogeneous groups is evaluated. The optimum fertilizer recommendations and the maximum profits for different soil conditions using each of the different models are estimated. The 'cost of the wrong decision' criterion is used to make a choice of the 'best' specification of the aggregate model for generating fertilizer recommendations to farmers. The value of soil test information is evaluated. The recommendations derived from this study are compared with those generated by Ryan (1972) and the La Molina Soils Laboratory in Peru.

The final chapter is devoted to summary and conclusions.

CHAPTER 2
REVIEW OF LITERATURE ON CROP-RESPONSE STUDIES

The major problems that have been encountered in deriving fertilizer recommendations from fertilizer test results can be divided into the following categories.

1. The choice of appropriate functional forms to represent the response relationship.

2. Assumptions concerning the structure of a chosen functional form and the estimation problems when the data are combined over space and time.

3. Economic theories of the behavior of farmers with respect to profits and risks.

4. The underlying nature of the experimental design when using regression analysis to estimate a response function.

The sizable literature in the categories 1, 3 and 4 have been reviewed in detail by Heady and Dillon (1961), Dillon (1968), and Ryan (1972). Here it is proposed to review briefly the second category and discuss some of the limitations of the previous studies in this area.

Up to the mid-1950's, most of the response studies concerned themselves with generating fertilizer optima by estimating location-specific functions. However, as soil testing procedures became advanced, the limitations of the above method were soon realized and more attention was given to the development of a generalized response function.

The early methods developed were designed to estimate crop response to total nutrients available--a combination of soil and applied nutrients. Such research studies have been reported by Anderson (1956), Hurst and Mason (1957), Hildreth (1957), Pesek (1956), and Jensen and Pesek (1959).

In general the procedures estimate the following functional form of the response surface:

$$y = f(X_1, X_2, \ldots, X_n) \tag{2.1}$$

where $X_i = X_{ai} + \lambda_i X_{si}$

X_i is the total ith nutrient available

X_{ai} is the applied ith nutrient

X_{si} is the initial level of ith nutrient in soil

λ_i is the proportionality factor to be determined for the ith nutrient.

Hildreth (1957) recognized that the errors increased when nutrients in the soil are initially ignored. He also pointed out a problem encountered when soil nutrients are included in a yield equation, namely that available soil nutrients are measured in units that differ from those of applied fertilizer. He proposed two methods of estimating the unknown factor λ_i which is used to convert an initial soil measurement into the equivalent value of added nutrient. These methods consisted of the use of a maximum likelihood function and successive iteration of ratios of regression coefficients. The estimated λ_i values are then used to generate ordinary least squares estimates of other parameters.

Anderson (1956) theoretically quantified the biases introduced in yield predictions by failure to recognize and evaluate differences in initial fertility levels between the experimental plot and the field of the farmer. Hurst and Mason (1957) considered the biases outlined by Anderson and presented evidence for a single variable quadratic equation which corroborated the significance of these biases.

Pesek (1956) advanced a procedure which algebraically separated total available nutrients into the added component plus the soil component. He considered the following generalized quadratic function

$$y = c_1 \ (k\alpha + N_f) \ c_2(k\alpha + N_f)^2 + c_3(m\gamma + P_f) + c_4(m\gamma + P_f)^2$$

$$+ c_5(k\alpha + N_f) \ (m\gamma + P_f) \qquad (2.2)$$

where y = yield increase

N_f, P_f = applied nitrogen and phosphorus, respectively

α, γ = soil test values of nitrogen and phosphorus, respectively

k and m are proportionality constants.

Jenson and Pesek (1959) fitted a restrictive quadratic model (with applied nutrients only as variables) to yield data from three lots of soil possessing different initial fertility levels within a given soil type; two soil types were considered. The restrictive equations were then generalized and extended to include soil nutrients similar to equation (2.2). The coefficients of the three generalized equations were then compared statistically; they were able to accept the hypothesis that the coefficients were equal for one soil type and rejected it for the second soil type. Thus, they concluded that the yield equation may be satisfactorily generalized with respect to initial fertility levels for one soil type but not for the other that they considered. They made no further attempts to generalize the yield equations by combining different soil types. Although they assume that the generalized model for each soil type is different, they do not present the estimates

of the error variances from each of the models. However, they said that problems remain in obtaining an adequate estimate of the contribution of soil fertility as a production factor in the generalized equation.

One difficulty of using all the above methods is that estimated conversion factors are different for each nutrient and for each soil. The estimation procedures for the conversion factors are also cumbersome. As pointed out by Ryan (1972), such methods involve using the response data numerous times resulting in biased estimates of the coefficients. A more fruitful approach might be to consider applied and soil nutrients as completely separate factors of production, allowing for an interaction between them. This formulation allows the shape of the response function to applied nutrients to change depending upon the levels of the different soil nutrients.

Fitts et al.[1] fitted a quadratic equation which included soil and weather variables and applied nutrients. They used step wise regression to obtain the equation of the form

$$y = b_o + b_1 \, OM + b_2 \, D + b_3 \, N_A + b_{22} \, N_A^2 + b_{23} \, D \cdot N_A + b_{24} \, D \cdot PH \qquad (2.3)$$

where OM = organic matter

D = drought index

---

[1]See J. W. Fitts, D. D. Mason, D. Cooper, W. W. McPherson and J. Havlicek, Jr., Determining yield response surfaces and economically optimum rates for corn under various soil and climates conditions in North Carolina, unpublished paper, North Carolina State University, Raleigh, N. C., 1959.

NA = applied nitrogen

pH = soil acidity - alkalinity.

They compared this generalized function to an individual location function in a "cost of wrong decision" analysis. They found that the cost of using the generalized equation when the location functions were true was low.

For similar North Carolina data, Havlicek and Seagraves (1962) combined the results of 37 corn fertilization experiments on Norfolk sandy loam soils into one production function. The function was of the form

$$y = b_o + b_1 P_s + b_2 \sqrt{D} + b_3 \sqrt{N} + b_4 N + b_5 \sqrt{ND} \quad . \quad (2.4)$$

Assuming that the location models were correct, they found that the cost of aggregation in using the general model was low at $1.02 per acre. In both these studies, the treatments in each experiment were the same. Under these conditions, Theil (1954) has shown that there is no "aggregation bias" of the estimators of coefficients in the general equations. If the treatments were different between experiments, then the general model would have estimators that contain these "aggregation biases" and the cost may be much higher. This is discussed more fully in the next chapter.

Brown and Oveson (1958) also compared quadratic equations by combining data from similar response years to estimate a single equation. They compared the general response equations to individual equations in a "cost of wrong decision analysis" and found the cost was low at $1.58

per acre. However, they did not incorporate any soil variables into the general equation.

One of the early suggestions of Heady and Pesek (1957) was to group soils on the basis of fertility and fit different functions to each of the groups. They did not pursue the idea empirically and do not present any criteria for separating different groups. Rust and Odell (1957) fit a quadratic model to seven soil types to determine the relative productivity of major Illinois soils. They were attempting to classify soils empirically as opposed to the soil mapping techniques which rely primarily on visual and physical characteristics. However, they were interested in standardizing yields for management levels which influence the amount of fertilizer applications rather than generating fertilizer recommendations. Laird and Cady (1969) found that an "agronomic" model postulated by soil scientists and agronomists had better predictive power compared to a model when all possible variables were selected. This is because in the latter case multicollinearity may seriously affect the "t" values of the estimated coefficients. Ryan (1972) built up a generalized quadratic model based on variables selected agronomically. Studies by Johnsson (1963) and Walker and Long (1966) have noted that when pooled data are used, the R^2 falls even when soil and other variables were included in the response equation. However, Zellner (1962b) has shown that when we consider cross section data, in general, the estimate of R^2 has a downward specification bias.

While by no means complete, the above review points out some of the problems faced in postulating and estimating a general response function with combined data.

CHAPTER 3
SPECIFICATION ERRORS AND THE ESTIMATION OF DIFFERENT MODELS

So far we have considered a review of crop-response function studies and some of the problems in pooling cross section data. In this study we will not be concerned with the choice of the functional form and other related problems. Instead we will assume that a generalized quadratic function incorporating both the soil and the applied nutrients is the appropriate functional form for the response relationship.[2] Using this functional form, the optimal fertilizer combinations for any given soil values of a particular form can be derived.

Two levels of aggregation of soils are considered in this work. At the first level of aggregation, the data from all the locations are pooled into groups on the basis of Buol's (1972) soil management classification system.[3] Briefly, he has proposed a classification by those soil characteristics which can be expected to affect fertility management practices. Management practices as seen by him are not only fertilizer practices but also include the interpretation or calibration of soil test results which serve as a basis for fertilizer recommendations. The goal is to define a reasonable number of soil groups, within which the same or similar soil test correlations and fertilizer application practices can be made. For any given set of locations, n groups can be formed by combining the locations on the basis of the above

---

[2]For similar set of data, Ryan (1972) tried different functional forms and chose the quadratic form as the best for describing the response relation between crop-yield and soil and applied nutrients.

[3]The details of the classification system are given in Appendix A.

criteria. We then define the true structural relationship by assuming that a generalized quadratic model similar to the one used by Ryan (1972) holds for each of the groups. The basic assumptions underlying the structural relationship are that each group has a different coefficient vector and a different error variance.

Two cases for generating the optimal fertilizer recommendations can be distinguished. If we have knowledge regarding a soil's group membership, then that group model can be used to generate the fertilizer recommendations for the given soil. However, if we do not know the soil's group membership, then an aggregate model which in some way represents the group structure must be used. For this we consider the second level of aggregation. In this case, we employ a single generalized quadratic function for all the groups. This model is estimated by using the data obtained by pooling all the locations. Three specifications of the aggregate model, the fixed coefficients (FC) model, the generalized least squares (GLS) model, and the random coefficients (RC) model are considered in this work. Which of these three specifications is to be chosen to generate the optimal fertilizer recommendations for any given soil? The criterion that is to be used here is the "cost of the wrong decision" analysis as proposed by Havlicek and Seagraves (1962). Assuming the group models to be true, the "cost of the wrong decision" analysis calculates the loss in profits in using the different specifications of the aggregate model. The specification with the least monetary loss is chosen as the 'best' aggregate model for generating fertilizer recommendations to farmers. Using the same "cost of the wrong decision" principle, the value of Buol's soil classification system and the value of soil testing will also be evaluated.

In considering the three specifications, a relevant question that can be raised is how well do they represent the true structural models? What criteria do we use to find the relationship between the different specifications of the aggregate model and the group models? We will here consider the criteria of the aggregate model as estimating an average of the group models. In aggregation literature, an average of the group models is the best single relation that represents the true group models for prediction purposes.[4] The conditions under which each of the specifications represent an average of the group models will be considered in detail. The properties of the estimators of the coefficients of each of the specifications of the aggregate model under various assumptions will also be described. Statistical tests for testing some of these assumptions will be presented. It may be noted that even though a specification may represent the average of the group models, it may not necessarily follow that it will give the least loss using the "cost of the wrong decision" criterion.

Functional Form of the Models and the Generation of Optimal Fertilizer Recommendations

The functional form for all the models is a generalized quadratic crop-response function used by Ryan (1972). The relationship represents the response of potatoes to soil nutrients, soil pH, applied nutrients and weather and is based on agronomic considerations. The functional form is given by the relationship:[5]

---

[4]See Theil (1954).

[5]For details of the reasons for including the relevant variables, see Ryan (1972).

$$
\begin{aligned}
y = {} & b_o + b_1 N_S + b_2 P_S + b_3 K_S + b_4 pH + b_5 N_A + b_6 P_A \\
& + b_7 K_A + b_8 R + b_9 T + b_{10} N_S^2 + b_{11} P_S^2 + b_{12} K_S^2 + b_{13} pH^2 \\
& + b_{14} N_A^2 + b_{15} P_A^2 + b_{16} K_A^2 + b_{17} R^2 + b_{18} T^2 \\
& + b_{19} N_S pH + b_{20} N_S N_A + b_{21} N_S R + b_{22} N_S T \\
& + b_{23} P_S P_A + b_{24} K_S K_A + b_{25} pH P_A + b_{26} N_A P_A \\
& + b_{27} N_A K_A + b_{28} N_A R + b_{29} N_A T + b_{30} P_A K_A + b_{31} RT \\
& + b_{32} N_A P_S + \varepsilon \qquad (3.1)
\end{aligned}
$$

where N_S = soil organic matter percentage

P_S = soil phosphorus in p.p.m. of P

K_S = soil potassium in p.p.m. of K

pH = soil pH

N_A = applied nitrogen fertilizer in kg/hectare

P_A = applied phosphorous fertilizer in kg/hectare

K_A = applied potassium fertilizer in kg/hectare

y = yield of potatoes in metric tons/hectare

b_o = a constant term

R = total rainfall in the six months October-March, in millimeters

T = means of daily temperatures in January and February in degrees centigrade

ε = the error term.

Using the above model, the optimal fertilizer recommendations can be obtained by setting their marginal products equal to the respective

price ratios and obtaining a simultaneous solution. Taking the first partial derivates of eq. (3.1) with respect to the applied nutrients and setting each equation equal to their respective price ratios, we obtain the system of equations

$$\frac{\partial Y}{\partial N_A} = b_5 + 2b_{14}\, N_A + b_{20}\, N_S + b_{26}\, P_A + b_{23}\, R + b_{29}\, T$$

$$+ b_{32}\, P_S = P_1/P_y$$

$$\frac{\partial y}{\partial P_A} = b_6 + 2b_{15}\, P_A + b_{23}\, P_S + b_{25}\, pH + b_{26}\, N_A$$

$$+ b_{30}\, K_A = P_2/P_y$$

$$\frac{\partial y}{\partial K_A} = b_7 + 2b_{16}\, K_A + b_{24}\, K_S + b_{27}\, N_A + b_{30}\, P_A$$

$$= P_3/P_y \qquad (3.2)$$

where P_1 = price of nitrogen fertilizer in soles/kg

P_2 = price of phosphorous fertilizer in soles/kg

P_3 = price of potassium fertilizer in soles/kg

P_y = price of potatoes in soles/m. ton.

Simplifying and rearranging terms in eq. (3.2), we have the solution for the optimal fertilizer levels (denoted by stars for N_A, P_A, and K_A):

$$\begin{Bmatrix} N_A^* \\ P_A^* \\ K_A^* \end{Bmatrix} = \begin{Bmatrix} 2b_{14} & b_{26} & b_{27} \\ b_{26} & 2b_{15} & b_{30} \\ b_{27} & b_{30} & 2b_{16} \end{Bmatrix}^{-1} \begin{Bmatrix} C_1 \\ C_2 \\ C_3 \end{Bmatrix} \tag{3.3}$$

where $C_1 = P_1/P_y - (b_5 + b_{20} N_S + b_{28} R + b_{29} T + b_{32} P_S)$

$C_2 = P_2/P_y - (b_6 + b_{23} P_s + b_{25} pH)$

$C_3 = P_3/P_y - (b_7 + b_{24} K_s)$.

In matrix form, the eq. (3.3) can be written as

$$\underset{3x1}{X^*} = \underset{3x3}{A^{-1}} \; \underset{3x1}{C} \tag{3.4}$$

when X* is the solution vector, A is a symmetric matrix involving coefficients of the second order terms of the applied nutrients, and C is a vector containing all the soil and weather variables.

However, before the procedure outlined in eq. (3.3) can be used, the A matrix of eq. (3.4) has to satisfy the second order conditions for a maximum for the model under consideration. The classical theory of unconstrained optima requires the following to hold for a maximum:

(a) Given a function $f(\underline{x}_i)$,[6] for which the first partial derivatives $\frac{\partial f}{\partial x_i}$, i = 1, 2, 3, ... exist at all points, a necessary condition for a maximum of $f(\underline{x}_i)$ is that

---

[6]Here x_i, (i = 1, 2, 3) refers to applied nitrogen, phosphorous, and potassium fertilizers, respectively.

$$\frac{\partial f}{\partial x_1} = \frac{\partial f}{\partial x_2} = \frac{\partial f}{\partial x_3} = 0 \tag{3.5}$$

b) A sufficient condition for a point satisfying these conditions to be a maximum is that all second partial derivatives $\frac{\partial^2 f}{\partial x_j \, \partial x_k}$ ($j, k = 1, 2, 3$) exist at this point, and that the principal minors of the matrix of second partial derivatives must alternate in signs. For example, for the A matrix of eq. (3.4), this condition implies that $2b_{14} < 0$,

$$\begin{vmatrix} 2b_{14} & b_{26} \\ b_{26} & 2b_{15} \end{vmatrix} > 0$$

and

$$\begin{vmatrix} 2b_{14} & b_{26} & b_{27} \\ b_{26} & 2b_{15} & b_{30} \\ b_{27} & b_{30} & 2b_{16} \end{vmatrix} < 0 \; . \tag{3.6}$$

The above conditions are necessary and sufficient conditions for a maximum to exist. Even if these conditions are satisfied, the formula (3.3) may not be satisfactory in obtaining a recommendation in those cases where the estimated response curve reached a maximum at fertilizer levels far in excess of the experimental rates. This is likely to occur when the experimental rates applied are not large enough to expose the region of zero or negative marginal product. While it is acceptable practice to predict beyond the range of experimental data, due to large standard error of estimates of the optimum in these cases, such recommendations would impose very large risks on farmers. We therefore

arbitrarily confine all recommendations within the approximate levels of maximum treatments of applied nitrogen, phosphorus and potassium, respectively. Ryan (1972) was fortunate to have the second order conditions for a maximum satisfied for his model with a similar set of data. Thus, he could use the simple formula (3.3) to calculate the fertilizer recommendations. However, he obtained optimal recommendations for many soil classes which were beyond the experimental range.

If the models under consideration here do not satisfy the second order conditions for a maximum and/or have optima which are beyond the experimental range, then alternative iterative methods which take into account both these conditions must be applied. These methods come within the realm of techniques of optimization of nonlinear functions with constraints. In this study, for such cases, the steepest ascent method with constraints as described by Box, et al. (1969) is used to generate the optimal fertilizer recommendations.[7] A brief description of the steepest ascent method will now be given here.

The iterative techniques require an initial point $\underline{x}_o$ to be specified, and proceed by generating a sequence of points $\underline{x}_i$, $i = 1, 2, 3, \ldots$, which represent improved approximation to the solution, that is,

$$f(\underline{x}_{i+1}) \geq f(\underline{x}_i) \tag{3.7}$$

Iterative techniques can be conveniently studied by referring to the following equation:

---

[7]There are a number of other methods also available for constrained optimization of nonlinear functions. For a good review of these methods, see Box, et al. (1969).

$$\underline{x}_{i+1} = \underline{x}_i + \underline{h}_i \, \underline{d}_i' \tag{3.8}$$

where $\underline{d}_i$ is a 3-dimensional vector, and $\underline{h}_i$ is a distance moved along it.

The steepest ascent approach involves the use of the gradient method. Gradient methods are those that select the direction $\underline{d}_i$ in eq. (3.8) using the partial derivatives of the objective function f with respect to the independent decision variables $\underline{x}_i$, as well as the values of the function itself, together with the information gained from previous iterations.

The gradient direction at any point is the direction whose components are proportional to the first partial derivatives of the objective function at the point in question. The role played by the gradient direction in maximization is now explained.

Consider a small perturbation with components $(\delta x_1, \delta x_2, \delta x_3)$ from the current point $\underline{x}_i$. The objective function $f(\underline{x}_i)$ will, to a first order approximation, be changed by an amount

$$df = \sum_{j=1}^{3} \frac{\partial f}{\partial x_j} \cdot \delta x_j \tag{3.9}$$

where the derivatives $\frac{\partial f}{\partial x_j}$ are evaluated at $\underline{x}_i$. Of all possible perturbations from $\underline{x}_i$ of magnitude

$$\Delta = \left[\sum_{j=1}^{3} \delta x_j^2 \right]^{1/2}$$

to obtain the perturbations giving the greatest change in the function, we need to consider the Lagrangian form

$$\phi(x, \lambda) = df + \lambda\left(\sum_{j=1}^{3} \delta x_j^2 - \Delta^2\right) . \tag{3.10}$$

Substituting (3.9) into eq. (3.10),

$$\phi(x, \lambda) = \sum_{j=1}^{3} \frac{\partial f}{\partial x_j} \cdot \delta x_j + \lambda\left(\sum_{j=1}^{3} \delta x_j^2 - \Delta^2\right) . \tag{3.11}$$

Upon differentiation with respect to x_j, we obtain

$$\frac{\partial f}{\partial x_j} + 2\lambda\ \delta x_j = 0 \tag{3.12}$$

so that the required perturbations must satisfy

$$\frac{\delta x_1}{\frac{\partial f}{\partial x_1}} = \frac{\delta x_2}{\frac{\partial f}{\partial x_2}} = \frac{\delta x_3}{\frac{\partial f}{\partial x_3}} \tag{3.13}$$

i.e., of all possible perturbations of magnitude Δ, the greatest change in the function occurs when the δx_j are chosen to be proportional to the corresponding $\frac{\partial f}{\partial x_j}$. Also it can be seen that, for the search to approach a maximum, the constant of proportionality must be positive. This direction is known as the direction of steepest ascent. The gradient direction is orthogonal to the contours of the constant function value (Box, et al. 1969).

The normalized gradient vector at the current point is found by using the formula

$$\underline{d}_1 = \frac{\left(\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2}, \frac{\partial f}{\partial x_3}\right)}{\left[\sum_{j=1}^{3} \left(\frac{\partial f}{\partial x_j}\right)^2\right]^{1/2}} \tag{3.14}$$

The steepest ascent method determines $\underline{d}_i$ and using a specified step length $\underline{h}_i$ obtains a new point by the formula (3.8). The procedure is repeated using a constant step length until a step which does not increase the function value is taken. In this case, the step length is reduced and the procedure continued from the best point.

Whenever a solution is within a specified constraint, the above method works well. If a constraint is encountered in the search procedure, then a method due to Roberts and Lyvers (1961) which "rides" the constraint is used. Thus whenever a step along the search direction results in a constraint violation, the previous point is taken and the step length reduced. When this search direction with reduced step length results in a constraint violation of the x_j variable, x_j is advanced up to the constraint. The x_j variable has the maximum values in the subsequent search and zero step lengths. The remaining two variables have the same initial step lengths as before and the search is continued.

Structural Model

Assume that the data can be divided into n groups on the basis of Buol's soil classification criteria. To each set of group data, a group model of the form given in eq. (3.1) can be fitted. Let the n group models be represented in matrix form as:

$$\underset{T_i x 1}{\underline{Y}_i} = \underset{T_i x \Lambda}{X_i} \; \underset{\Lambda x 1}{\underline{\beta}_i} + \underset{T_i x 1}{\underline{U}_i} \tag{3.15}$$

$\forall\ i = 1, 2, 3, \ldots, n$

where $\underline{Y}_i$ = a T_ix1 vector of observations on the dependent variable of the ith group

X = a T_ixΛ non-stochastic matrix of observations with rank Λ on the Λ independent variables of the ith group. The X_i matrix contains a column of ones which is used to generate an estimate of the intercept terms

$\underline{\beta}_i$ = a Λx1 vector of coefficients for the ith unit which is fixed and same for all observations within the ith group

$\underline{U}_i$ = a T_ix1 vector of random disturbances.

The complete system of eq. (3.15) for the group models is assumed to be the true structural response relationship and can be written as:

$$\begin{bmatrix} \underline{Y}_1 \\ \underline{Y}_2 \\ \underline{Y}_3 \\ \cdot \\ \cdot \\ \cdot \\ \underline{Y}_n \end{bmatrix} = \begin{bmatrix} X_1 & 0 & \cdot & \cdot & \cdot & \cdot & 0 \\ 0 & X_2 & & & & & \cdot \\ \cdot & & X_3 & & & & \cdot \\ \cdot & & & \cdot & & & \cdot \\ \cdot & & & & \cdot & & \cdot \\ \cdot & & & & & \cdot & \cdot \\ 0 & \cdot & \cdot & \cdot & \cdot & \cdot & X_n \end{bmatrix} \begin{bmatrix} \underline{\beta}_1 \\ \underline{\beta}_2 \\ \underline{\beta}_3 \\ \cdot \\ \cdot \\ \cdot \\ \underline{\beta}_n \end{bmatrix} + \begin{bmatrix} \underline{U}_1 \\ \underline{U}_2 \\ \underline{U}_3 \\ \cdot \\ \cdot \\ \cdot \\ \underline{U}_n \end{bmatrix} \qquad (3.16)$$

or also as:

$$\underset{T\text{x}1}{\underline{Y}} = \underset{T\text{x}n\Lambda}{X} \; \underset{n\Lambda\text{x}1}{\underline{\beta}} + \underset{T\text{x}1}{\underline{U}} \qquad (3.17)$$

where $T = \sum_{i=1}^{n} T_i$.

We assume that

$$\underline{\beta}_i = \underline{\beta}_j \quad \text{for } i \neq j \tag{3.18a}$$

$$E(\underline{U}_i) = 0 \tag{3.18b}$$

$$E(\underline{U}_i \, \underline{U}_j') \begin{cases} = \sigma_i^2 \, I_i & \forall \; i = j \\ = \; 0 & \forall \; i \neq j \end{cases} \tag{3.18c}$$

and the $\underline{U}_i$'s are serially uncorrelated.

When we consider the different specifications of the aggregate model as representing this structure, we will also consider the following assumptions:

$$\underline{\beta}_i = \underline{\beta}_j \quad \text{for } i \neq j \tag{3.18d}$$

$$E(\underline{U}_i \, \underline{U}_j') \begin{cases} = \sigma^2 \, I & \forall \; i = j \\ = \; 0 & \forall \; i \neq j \end{cases} \tag{3.18e}$$

$$\underline{\beta}_i = \overline{\underline{\beta}}_R + \underline{\delta}_i \qquad \forall \; i = 1, 2, \ldots, n \tag{3.18f}$$

where $\underline{\delta}_i$ is a Λx1 vector of random elements with $\underline{\delta}_i \sim (0, \Delta)$. This will be discussed in greater detail later in this chapter.

The least squares estimators of the group models, $\underline{b}_i$, are BLUE (best linear unbiased estimators) of $\underline{\beta}_i$ and given by:

$$\underline{b}_i = (X_i' \, X_i)^{-1} \, X_i' \, \underline{Y}_i \qquad i = 1, 2, \ldots, n \tag{3.19}$$

The variance-covariance matrix of the estimators $\underline{b}_i$ is given by:

$$\Sigma_i = (X_i' X_i)^{-1} \sigma_i^2 \quad . \tag{3.20}$$

"Cost of the Wrong Decision" Analysis

We shall now distinguish two different cases of obtaining the profit-maximizing fertilizer recommendations for individual farmers. In the first case, if we have a soil sample and know to which group it belongs, then we can use the estimated coefficients of that group model to obtain the optimum fertilizer combinations as described in the previous section. The second case, which is more important, considers a soil sample for which we have no knowledge regarding its group membership. In this case, we have to use an aggregate model based on all the groups' data to generate the optimum fertilizer combinations. We will postulate three possible specifications of the aggregate model, the fixed coefficients (FC) model, the generalized least squares (GLS) model, and the random coefficients (RC) model. Which of these three specifications is to be chosen for generating the fertilizer recommendations? What criteria should we use to make a choice? The criteria that we consider here is the "cost of the wrong decision" analysis as proposed by Havlicek and Seagraves (1962).

There have been attempts by a number of writers to calculate the "cost of wrong decision" after deriving response functions. Havlicek and Seagraves (1962) assumed that 37 estimated location-year functions for corn in North Carolina were "true." They calculated the optimum levels of fertilizer for each location using a generalized function which included soil variables. These optimum levels were then used in each of

the estimated location-year functions and the net revenue calculated. They compared this net revenue with that generated from using the optimum levels of fertilizer derived from the location-year functions. If the location-year functions are true, they found that the average cost of using the generalized function was quite low at 1.02 dollars per acre. Brown and Oveson (1958) in a similar kind of analysis found the cost of using the general functions to be low at 1.58 dollars per acre. However, they did not incorporate any soil variables into the general function.

The same underlying principle as used by Havlicek and Seagraves (1962) can be applied here to choose the 'best' specification for deriving fertilizer recommendations. This can be done as follows. First, the profit-maximizing fertilizer levels and the corresponding net revenues are obtained using the estimated coefficients of the group models. Next, the optimum fertilizer levels using the estimated coefficients from each of the specifications of the aggregate model are obtained, and these values are used in each of the group models to calculate the corresponding net revenue. This net revenue is compared with that obtained from using the appropriate group model and the difference gives the loss from using each of the specifications of the aggregate model, when the group models are true. The specification with the least monetary loss is then chosen as the 'best' specification for generating fertilizer recommendations to farmers, for those soils not known to belong to any of the five groups. The exact formula and computation procedures are described in greater detail in the next chapter.

Aggregate Model Specifications and Estimation

The fixed coefficients model is the most common form of specification that is considered in crop-response literature. A modification of the fixed coefficients model is to consider a model with heteroscedastic group error variances, the generalized least squares model. Both the FC and the GLS models assume that the coefficients between all the groups are the same. A third possibility is to consider a random coefficients model recently introduced in the literature by Swamy (1970) which does not make this assumption. For the RC model, we can make the assumption that the coefficient vector of each of the group models has a common distribution with the same mean and the same variance-covariance matrix but that the individual group coefficient vectors can be different. In this work, we will estimate the above three models, the FC, GLS, and RC models, using all the experimental data. A relevant question that can be raised in postulating these three specifications is how well do they represent the true structural models given by eq. (3.16)? What criteria should be used to find the relationship between the different specifications of the aggregate model and group structural models? We will here consider the criteria of the aggregate model as representing the average of the group models. An average of the group models is the best single relation that represents the true group models for prediction purposes (Theil, 1954). The properties of the estimators of the coefficients of each of the specifications of the aggregate model as estimating the average of the coefficients of the group models will be discussed in detail. It may be noted that even though a specification of the

aggregate model may represent the average of the group models, it does not necessarily follow that this specification will give the least monetary loss when we apply the "cost of wrong decision" criteria.

The Fixed Coefficients (FC) Model

In postulating the FC model, the following assumptions are made:

(1) The coefficient vector $\underline{\beta}_F$ for the FC model is fixed and same for all the observations.

(2) The coefficient vector of all the n groups are equal, i.e., $\underline{\beta}_1 = \underline{\beta}_2 = \underline{\beta}_3 \ldots = \underline{\beta}_{n-1} = \underline{\beta}_n = \underline{\beta}_F$. This is given by eq. (3.18d).

(3) The error variances assumed for each of the groups are equal, i.e., $\sigma_1^2 = \sigma_2^2 = \sigma_3^2 \ldots = \sigma_{n-1}^2 = \sigma_n^2 = \sigma^2$. This is given by eq. (3.18e).

When the above assumptions are made, the structural system given by eq. (3.16) reduces to the form:

$$\begin{bmatrix} \underline{Y}_1 \\ \underline{Y}_2 \\ \underline{Y}_3 \\ \cdot \\ \cdot \\ \cdot \\ \underline{Y}_n \end{bmatrix} = \begin{bmatrix} X_1 \\ X_2 \\ X_3 \\ \cdot \\ \cdot \\ \cdot \\ X_n \end{bmatrix} \underline{\beta}_F + \begin{bmatrix} \underline{V}_1 \\ \underline{V}_2 \\ \underline{V}_3 \\ \cdot \\ \cdot \\ \cdot \\ \underline{V}_n \end{bmatrix} \tag{3.21}$$

or

$$\underset{T\times 1}{\underline{Y}} = \underset{T\times\Lambda}{Z} \; \underset{\Lambda\times 1}{\underline{\beta}_F} + \underset{T\times 1}{\underline{V}} \quad . \tag{3.22}$$

It is assumed that

$$E(\underline{V}) = 0 \tag{3.23a}$$

$$E(\underline{V}^2) = \sigma^2 \, I \tag{3.23b}$$

and the $\underline{V}$'s are serially uncorrelated. The least squares estimator of the coefficient vector in the above model is given by:

$$\underline{b}_F = (Z'Z)^{-1} Z' \underline{Y} \tag{3.24}$$

$$= \left[\sum_{i=1}^{n} X_i' X_i\right]^{-1} \left[\sum_{i=1}^{n} X_i' \underline{Y}_i\right]$$

$$= \sum_{i=1}^{n} W_{Fi} \underline{b}_i \tag{3.25}$$

where $$W_{Fi} = \left[\sum_{i=1}^{n} (X_i' X_i)\right]^{-1} (X_i' X_i) \ . \tag{3.26}$$

The estimator $\underline{b}_F$ is thus a weighted average of the group models' least squares estimators $\underline{b}_i$, where the weights are given by W_{Fi}. If the assumptions described above are true, then the least squares estimators are BLUE of $\underline{\beta}_F$. The variance of the estimator $\underline{b}_F$ is given by

$$\Sigma_F = (Z'Z)^{-1} \sigma^2 \ . \tag{3.27}$$

The Generalized Least Squares (GLS) Model

To take into account the heteroscedasticity of error variances between the group models, we now turn to an analysis of the Zellner-Aitken's generalized least squares model.

The assumptions of the generalized least squares (GLS) model can be written as:

(1) The coefficient vector for the GLS model is fixed and same for all the observations.

(2) The coefficient vector for all the group models are equal, *i.e.*, $\underline{\beta}_1 = \underline{\beta}_2 = \underline{\beta}_3 \ldots = \underline{\beta}_n = \underline{\beta}_G$, also given by eq. (3.18d).

(3) The error variances between the n group models are not equal, *i.e.*, $\sigma_1^2 \neq \sigma_2^2 \neq \sigma_3^2 \neq \ldots \neq \sigma_n^2$. This is also given by eq. (3.18c).

The GLS model is then of the form:

$$\begin{bmatrix} \underline{Y}_1 \\ \underline{Y}_2 \\ \underline{Y}_3 \\ \cdot \\ \cdot \\ \cdot \\ \underline{Y}_n \end{bmatrix} = \begin{bmatrix} X_1 \\ X_2 \\ X_3 \\ \cdot \\ \cdot \\ \cdot \\ X_n \end{bmatrix} \underline{\beta}_G + \begin{bmatrix} \underline{V}_1' \\ \underline{V}_2' \\ \underline{V}_3' \\ \cdot \\ \cdot \\ \cdot \\ \underline{V}_n' \end{bmatrix} \tag{3.28}$$

or

$$\underset{T\times 1}{\underline{Y}} = \underset{T\times \Lambda}{Z} \; \underset{\Lambda\times 1}{\underline{\beta}_G} + \underset{T\times 1}{\underline{V}'} \tag{3.29}$$

We assume that

$$E(\underline{V}_i') = 0 \tag{3.30a}$$

$$E(\underline{V}_i'^2) = \sigma_i^2 \, I_i \qquad \forall \; i = 1, 2, \ldots, n \; . \tag{3.30b}$$

This means that the disturbances between the groups are heteroscedastic and each group has a different error variance. It is also assumed that the disturbances are not serially correlated.

Under these assumptions the BLUE estimator of the coefficient vector is given by the Aitken's generalized least squares vector (Goldberger, 1964, p. 233);

$$\underline{b}_G = (Z'\Omega^{-1} Z)^{-1} Z' \, \Omega^{-1} \, \underline{Y} \tag{3.31}$$

where $$\Omega = \begin{bmatrix} \sigma_1^2 I_1 & 0 & \cdots & 0 \\ 0 & \sigma_2^2 I_2 & & \vdots \\ \vdots & & \ddots & \vdots \\ 0 & \cdots & \cdots & \sigma_n^2 I_n \end{bmatrix}$$

Eq. (3.31) can also be written as

$$\underline{b}_G = \left[\sum_{i=1}^{n} \frac{X_i' X_i}{\sigma_i^2}\right]^{-1} \left[\sum_{i=1}^{n} \frac{X_i Y_i}{\sigma_i^2}\right]$$

$$= \sum_{i=1}^{n} W_{Gi} \underline{b}_i \tag{3.32}$$

where $$W_{Gi} = \left[\sum_{i=1}^{n} \frac{X_i' X_i}{\sigma_i^2}\right]^{-1} \left[\frac{X_i' X_i}{\sigma_i^2}\right] , \tag{3.33}$$

The estimators $\underline{b}_G$ are weighted averages of the microunits' least squares estimators $\underline{b}_i$, with the weights equal to W_{Gi}. The variance of $\underline{b}_G$ is given by

$$\Sigma_G = (Z' \Omega^{-1} Z)^{-1} . \tag{3.34}$$

The above estimator can be used only if we have knowledge of Ω, i.e., the true values of σ_i^2 for all i. In practice we do not have such knowledge. However, Zellner (1962a) has suggested a method where we can employ estimates of the disturbance terms' variances based on the residuals derived from a least squares application of each of the group models. When these estimates are substituted in eq. (3.31), he shows that the resulting estimator possesses good large sample properties,

i.e., it is consistent and asymptotically efficient. Thus the Zellner-Aitken's GLS estimator is given by:

$$\hat{\underline{b}}_G = (Z' \hat{\Omega}^{-1} Z)^{-1} Z' \hat{\Omega}^{-1} \underline{Y} \tag{3.35}$$

where

$$\hat{\Omega} = \begin{bmatrix} s_1^2 I_1 & 0 & \cdots & 0 \\ 0 & s_2^2 I_2 & & \vdots \\ \vdots & & \ddots & \vdots \\ 0 & \cdots & \cdots & s_n^2 I_n \end{bmatrix} . \tag{3.36}$$

The Random Coefficients Regression (RC) Model

So far, for both the FC and the GLS models, we have assumed that the coefficient vectors for each of the groups are the same. We can now propose a third specification which takes into account the non-equality of the coefficients between the different groups. This leads to a consideration of the random coefficients regression model. The properties of the estimators of such a model have been fully dealt with by Swamy (1970). We will borrow heavily from his work in specifying a suitable model for application to estimating a generalized aggregate crop-response function.

The basic assumptions of the random coefficients model in comparison with the FC model and GLS model are:

(1) The coefficient vector for any given group is fixed and same for all sample observations within that group.

(2) The coefficient vector over all the groups is assumed to be distributed with the same mean and the same variance-covariance matrix, *i.e.*, $\beta_i \sim (\bar{\underline{\beta}}_R, \Delta)$. This is also given by eq. (3.18f).

(3) The error variances between all the groups are not equal, i.e., $\sigma_1^2 \neq \sigma_2^2 \neq \sigma_3^2 \neq \ldots \neq \sigma_n^2$. This is also given by eq. (3.18c).

The most important difference between the RC and the FC model is due to changes in the second assumption. Klein's (1953) brief but illuminating discussion on estimation from cross section data explains more clearly the rationale for the underlying nature of this assumption of the RC model. He uses the following simple example for illustrative purposes.

$$y_{it} = \beta_o + \beta_1 x_{it} + u_{it} \tag{3.37}$$

where t = 1, 2, ..., T the number of data observations for a given year for the individual i

i = 1, 2, ..., n the number of individual units

y_{it} = dependent variable for individual i

x_{it} = independent variable for individual i

and u_{it} = error term for individual i.

If we compare any two individuals i, j ($i \neq j$), it is likely that the same relationship (3.37) may not hold. The most plausible reason why different relationships between x and y underlie different individual's behavior is that some important explanatory variables are missing or are not measurable. For example, if y is household saving, and x is household income, it may be that a young household adjusts its saving to its income along the relation,

$$y_{it} = \beta_{oi} + \beta_{1i} x_{it} + u_{it}$$

and an old household along another relation,

$$y_{it} = \beta_{oj} + \beta_{1j} x_{it} + u_{it} \quad .$$

Thus we find that $\beta_{oi} \neq \beta_{oj}$ and $\beta_{1i} \neq \beta_{1j}$. In this case age is one of the missing variables which accounts for the difference in their behavior. It is possible that if age is included in the relationship, we may get identical parameters for each individual. When we are compelled to work with fewer variables than required, it may be desirable to specify that each individual has his own coefficient vector. Once different coefficient vectors for each individual are allowed for, it is more reasonable to assume that the remaining differences between their response patterns are attributable to just the u_i term. Thus, it might be useful to specify a random coefficients model in such cases. The analogy of the above example to the case considered here is straightforward.

It may be noted that the second assumption of the random coefficients model covers the fixed coefficient model as a special case. For the latter, the assumption becomes that the coefficient vector of each group is distributed with the same mean and zero variances. The assumptions for the RC model allows for heterogeneity among the different groups, at the same time enabling one to estimate efficiently a single relation for the cross section data. The difference in the assumptions underlying the group models and the RC models may also be noted. In the group models we assume that none of the n groups are related to each other, i.e., we assume that each group model is separate and represents a response relationship for the data within that group only. For the RC model, we assume that each of the n group models can

be considered as a sample from a population which has the same underlying response relation.

An example of the possible use of RC model in crop-response functions is given by Hildreth and Houck (1968). If a particular coefficient represents the response of a plant to nitrogen fertilizer, it is well known that this response is strongly influenced by temperature, rainfall, soil characteristics and many other factors. If these factors are held constant, the coefficient can be expected to be constant to a tolerable approximation. If these factors vary and are observed, it is of course desirable to incorporate their influences into the model. If they vary and are unobserved, regarding the coefficient as a mean of a random response may be better than assuming it to be constant. However, unlike the present case, Hildreth and Houck (1968) considered a single regression equation in which the coefficient vector is random from observation to observation.[8] Rao (1965b) also discussed the estimation of a random coefficients model with cross section data; however, he confined his attention to the experimental situation in which the "design matrix" X_t is kept constant for all microunits, i.e., each individual unit has the same matrix of observations on independent variables.

---

[8]They have considered a number of estimators for the coefficient vector and their properties; however, further investigation of the sampling properties of these estimators are needed before they can be used for empirical applications.

The basic RC model can now be written as

$$\begin{bmatrix} \underline{Y}_1 \\ \underline{Y}_2 \\ \underline{Y}_3 \\ \vdots \\ \underline{Y}_n \end{bmatrix} = \begin{bmatrix} X_1 & 0 & \cdot & \cdots & 0 \\ 0 & X_2 & & & \cdot \\ \cdot & & X_3 & & \cdot \\ \vdots & & & \ddots & \vdots \\ 0 & \cdot & \cdot & \cdots & X_n \end{bmatrix} \begin{bmatrix} \underline{\beta}_1 \\ \underline{\beta}_2 \\ \underline{\beta}_3 \\ \vdots \\ \underline{\beta}_n \end{bmatrix} + \begin{bmatrix} \underline{W}_1 \\ \underline{W}_2 \\ \underline{W}_3 \\ \vdots \\ \underline{W}_n \end{bmatrix} . \quad (3.38)$$

Let $\underline{\beta}_i = \overline{\underline{\beta}}_R + \underline{\delta}_i \qquad i = 1, 2, 3, \ldots, n$ (3.39a)

where $\underline{\delta}_i$ is a Λx1 vector of random elements with

$$\underline{\delta}_i \sim (0, \Delta) . \quad (3.39b)$$

Assume that

$$E(\underline{W}_i) = 0 \quad (3.39c)$$

$$E(\underline{W}_i \, \underline{W}_j') = \begin{cases} \sigma_i^2 \, I_i & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases} . \quad (3.39d)$$

This implies that the disturbances have different variances for different microunits

$$E(\underline{\beta}_i) = \overline{\underline{\beta}}_R \quad (3.39e)$$

$$E(\underline{\beta}_i - \overline{\underline{\beta}}_R)\,(\underline{\beta}_j - \overline{\underline{\beta}}_R)' = \begin{cases} \Delta & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases} \quad (3.39f)$$

$\underline{\beta}_i$ and $\underline{u}_j$ are independent

$\underline{\beta}_i$ and $\underline{\beta}_j$ are independent for $i \neq j$.

These assumptions imply that the $\underline{\beta}_i$ are random and uncorrelated across microunits, but follow the same distribution with mean $\overline{\underline{\beta}}_R$ and variance-covariance matrix Δ. This distribution is assumed to be stable over the observations.

Substituting eq. (3.39a) in (3.38) and simplifying, we get

$$\begin{bmatrix} \underline{Y}_1 \\ \underline{Y}_2 \\ \underline{Y}_3 \\ \cdot \\ \cdot \\ \cdot \\ \underline{Y}_n \end{bmatrix} = \begin{bmatrix} X_1 \\ X_2 \\ X_3 \\ \cdot \\ \cdot \\ \cdot \\ X_n \end{bmatrix} \overline{\underline{\beta}}_R + \begin{bmatrix} X_1 & 0 & \cdot & \cdots & 0 \\ 0 & X_2 & & & \cdot \\ \cdot & & X_3 & & \cdot \\ \cdot & & & \cdot & \cdot \\ \cdot & & & \cdot & \cdot \\ \cdot & & & \cdot & \cdot \\ 0 & \cdot & \cdot & \cdots & X_n \end{bmatrix} \begin{bmatrix} \underline{\delta}_1 \\ \underline{\delta}_2 \\ \underline{\delta}_3 \\ \cdot \\ \cdot \\ \cdot \\ \underline{\delta}_n \end{bmatrix}$$

$$+ \begin{bmatrix} \underline{W}_1 \\ \underline{W}_2 \\ \underline{W}_3 \\ \cdot \\ \cdot \\ \cdot \\ \underline{W}_n \end{bmatrix} \tag{3.40}$$

or, more simply,

$$\underset{T\times 1}{\underline{Y}} = \underset{T\times\Lambda}{Z} \; \underset{\Lambda\times 1}{\overline{\underline{\beta}}_R} + \underset{T\times n\Lambda}{X} \; \underset{n\Lambda\times 1}{\underline{\delta}} + \underset{T\times 1}{\underline{W}} \; . \tag{3.41}$$

Let $\underset{T\times 1}{\theta} = X\,\underline{\delta} + \underline{W}$. Then the variance-covariance matrix of θ is given by

$$V(\theta) = \begin{bmatrix} X_1 \Delta X_1' + \sigma_1^2 I_1 & 0 & \cdots & 0 \\ 0 & X_2 \Delta X_2' + \sigma_2^2 I_2 & & \vdots \\ \vdots & & \ddots & \vdots \\ 0 & \cdots & \cdots & X_n \Delta X_n' + \sigma_n^2 I_n \end{bmatrix} . \tag{3.42}$$

The matrix $V(\theta)$ is of dimension (TxT) and is a function of the X_i's and an unknown $1/2\ \Lambda(\Lambda + 1) + n$ elements of Δ and σ_i^2.

The coefficients $\underline{\bar{\beta}}_R$ can be estimated by applying Aitkin's generalized least squares to eq. (3.41). Thus a BLUE of $\underline{\bar{\beta}}_R$ is given by:

$$\underline{\bar{b}}_R\ (\theta) = (Z'\ V(\underline{\theta})^{-1}\ Z)^{-1}\ Z'\ V(\underline{\theta})^{-1}\ \underline{Y} \quad . \tag{3.43}$$

Since $V(\underline{\theta})$ is of dimension (TxT), it has to be simplifed before it can be used practically. Swamy (1970) has shown that the above expression reduces to[9]

$$\underline{\bar{b}}_R\ (\underline{\theta}) = \sum_{i=1}^{n} \left[\left\{ \sum_{j=1}^{n} (\Delta + \sigma_j^2\ (X_j'\ X_j)^{-1})^{-1} \right\}^{-1} \right.$$

$$\left. \cdot \left\{ \Delta + \sigma_i^2\ (X_i'\ X_i)^{-1} \right\}^{-1} \right] \underline{b}_i = \sum_{i=1}^{n} W_{Ri}\ \underline{b}_i \tag{3.44}$$

where the weights are given by the terms in the squared brackets.

---

[9] He uses the following result obtained by Rao (1965a, p. 29): If A and D are nonsingular matrices of orders m and n and B a mxn matrix, then

$$(A + BDB')^{-1} = A^{-1} - A^{-1}\ BEB'\ A^{-1} + A^{-1}\ BE\ (E + D)^{-1}\ EB'\ A^{-1}$$

where $E = (B'\ A^{-1}\ B)^{-1}$.

Thus, it can be seen that $\bar{\underline{b}}_R(\underline{\theta})$ is a weighted average of the microestimators $\underline{b}_i$ with weights inversely proportional to their covariance matrices. It is interesting to compare the weighting scheme used to combine the $\underline{b}_i$'s for the RC model in eq. (3.44), the GLS model in eq. (3.33) and the FC model in eq. (3.25). The weighting scheme in eq. (3.25) takes into account only the moment matrix of the regressors, while that in equation (3.33) takes into account the variances of the disturbances also. The weighting scheme in eq. (3.44) takes into account the variances and covariances of the coefficients and the moment matrix of the regressors besides the variances of disturbances.

The variance-covariance matrix of the estimator is given by

$$C(\underline{\theta}) = [Z' \; V(\theta)^{-1} \; Z]^{-1} \quad . \tag{3.45}$$

So far it was assumed that we can use eq. (3.44) to obtain estimates of the RC model coefficients. In general, Δ and σ_i^2 are unknown and so we cannot use these equations in practice. This is one of the major drawbacks of the RC model. However, we can employ estimates of Δ and σ_i^2 in constructing the Aitken estimator of the RC model.

An unbiased estimate of σ_i^2 is given by (Goldberger, 1964, p. 166)

$$s_i^2 = \frac{\underline{Y}_i' \; [I_i - X_i' \; (X_i' \; X_i) \; X_i] \; \underline{Y}_i}{T_i - \Lambda} \quad . \tag{3.46}$$

To estimate Δ we treat the ordinary least squares estimators $\underline{b}_i$ ($i = 1, 2, \ldots, n$) as a random sample of size n. Let the sample covariance matrix of the b_i^s be given by

$$\frac{S_b}{n-1} = \frac{1}{n-1}\left[\sum_{i=1}^{n} \underline{b}_i \underline{b}_i' - \frac{1}{n} \sum_{i=1}^{n} \underline{b}_i \sum_{i=1}^{n} \underline{b}_i'\right] . \tag{3.47}$$

Substituting eq. (3.15) in (3.19), we obtain

$$\underline{b}_i = \underline{\beta}_i + (X_i' X_i)^{-1} X_i' \underline{U}_i = \underline{\beta}_i + P_i \underline{U}_i \tag{3.48}$$

where $\quad P_i = (X_i' X_i)^{-1} X_i' \underline{U}_i$.

Using eq. (3.48), we get for eq. (3.47),

$$S_b = \sum_{i=1}^{n} (\underline{\beta}_i + P_i \underline{U}_i)(\underline{\beta}_i' + U_i' P_i') - \frac{1}{n} \sum_{i=1}^{n} (\underline{\beta}_i + P_i \underline{U}_i)$$

$$\cdot \sum_{j=1}^{n} (\underline{\beta}_j' + \underline{U}_j' P_j') . \tag{3.49}$$

Using assumption (3.39b) and taking expectations in (3.49),

$$E(S_b) = n(\overline{\underline{\beta}}_R \overline{\underline{\beta}}_R' + \Delta) + E\left(\sum_{i=1}^{n} P_i \underline{U}_i \underline{U}_i' P_i'\right)$$

$$- \frac{1}{n}\left[n^2 \overline{\underline{\beta}}_R \overline{\underline{\beta}}_R' + n\,\Delta + E\left(\sum_{i=1}^{n} P_i \underline{U}_i\right)\left(\sum_{j=1}^{n} \underline{U}_j' P_j'\right)\right] . \tag{3.50}$$

Making use of the assumption (3.18c), the eq. (3.50) can be simplified to

$$E(S_b) = (n-1)\,\Delta + \frac{n-1}{n}\left(\sum_{i=1}^{n} (X_i' X_i)^{-1} \sigma_i^2\right) . \tag{3.51}$$

The estimator of Δ is given by

$$\hat{\Delta} = \frac{S_b}{n-1} - \frac{n-1}{n}\left[\sum_{i=1}^{n} (X_i' X_i)^{-1} s_i^2\right]. \tag{3.52}$$

These estimators s_i^2 and $\hat{\Delta}$ can be substituted for σ_i^2 and Δ, respectively, in eq. (3.44) to obtain the estimator $\underline{b}_R(\hat{\underline{\theta}})$. Swamy (1970) has shown that $\underline{b}_R(\hat{\underline{\theta}})$ is a consistent, asymptotically efficient estimator with asymptotic variance-covariance matrix

$$\hat{\Delta}/n \qquad \text{as } T \longrightarrow \infty \qquad \text{and } n \text{ is fixed.}$$

However, some fundamental problems associated with the estimator $\hat{\Delta}$ need to be pointed out. Since Δ is a variance-covariance matrix and assumed to be nonsingular, all its diagonal elements should be positive and its i-jth off diagonal element, when squared, should be equal to or less than the product of its ith and jth diagonal elements. These restrictions are neglected in the above procedure, and thus it is possible in empirical applications for $\hat{\Delta}$ to yield negative estimates for the variances of some coefficients.[10] Thompson (1962) has suggested two possible explanations of a negative estimate for the variance: (1) the basic model is misspecified, i.e., some of the various assumptions listed earlier for the RC model may not hold and (2) statistical noise may have obscured the underlying physical situation. The most common approach to the solution of this problem is to set the negative estimates equal to zero. However, the estimates thus obtained are biased (Scheffe, 1959). Thus if negative estimates for the variance are

---

[10]This difficulty is familiar in discussions of random-effect models in the analysis of variance. See Scheffe (1959).

obtained, we prefer to conclude that the basic model is misspecified and inconsistent with the data.

Summarizing, we have discussed in detail the underlying nature of the assumptions for each of the specifications of the aggregate model, the FC model, the GLS model, and the RC model. The estimation problems and the properties of the estimators of the above models were discussed in considerable detail. In the next chapter, it is proposed to try out the various models for the estimation of a response relationship for the Peruvian potato-fertilizer data.

In the next section, we will outline the statistical criteria for testing the validity of some of the assumptions made for each of the specifications of the aggregate model. If the assumptions are not true, what can be said about the properties of the estimators of each of the specifications of the aggregate model? What is the relationship between each of the specifications and the true group models under different assumptions? These questions will be taken up in the next section.

Statistical Tests and the Relationship between the Different Specifications of the Aggregate Model and the Group Models

For the FC model, we make the assumption that the error variances for all the n group models are the same. This can be tested by using Bartlett's test for the homogeneity of variances.[11] An outline procedure of the test is as follows:

---

[11]The detailed derivation of the test is given in Anderson and Bancroft (1952, pp. 141-144).

Let there be n independent group sample variances s_i^2 with m_i degrees of freedom each ($i = 1, 2, 3, \ldots, n$). The null hypothesis is given by:

$$H_o: \quad \sigma_1^2 = \sigma_2^2 = \ldots = \sigma_n^2 = \sigma^2$$

i.e., each s_i^2 is an estimate of the same population variance, σ^2. The null hypothesis can be tested by the criterion Q/ℓ, which can be shown to be distributed approximately as χ^2 with (n-1) degrees of freedom. In this expression

$$Q = m \log \bar{\sigma}^2 - \sum_{i=1}^{n} m_i \log s_i^2 \tag{3.53}$$

and

$$\ell = 1 + \frac{1}{3(n-1)} \left[\sum_{i=1}^{n} \frac{1}{m_i} - \frac{1}{m} \right] \tag{3.54}$$

where $m = \sum_{i=1}^{n} m_i$ and $\bar{\sigma}^2 = \sum_{i=1}^{n} \left(\frac{m_i \, s_i^2}{m} \right)$.

If the error variances of all the group models are not the same, then it can be shown that the least squares estimators of the fixed coefficient model are inefficient, although still unbiased (Goldberger, 1964, p. 238). Also the classical estimator of σ^2 from using the fixed coefficient model is no longer unbiased (Goldberger, 1964, pp. 238-243).

So far for both the FC and the GLS models, we have assumed that the coefficient vectors of the group models are equal. However, this assumption may not be true and in that case both the FC and GLS models

have estimators that contain aggregation biases in estimating an average vector of coefficients (Theil, 1954).[12] While Theil's work is confined to aggregation biases in the macromodel estimators, similar results have been extended to analyzing such errors in specification in combining cross section data (Zellner, 1962b, Theil, 1957, Griliches, 1957). These results can be shown for the FC model as follows:

Substituting the value of $\underline{Y}$ in eq. (3.17) into (3.24) we have

$$\underline{b}_F = (Z'Z)^{-1} Z'[X \underline{\beta} + \underline{U}] \tag{3.55}$$

Eq. (3.55) can be rewritten as,

$$\underline{b}_F = \underline{\beta}_A + (P - H) \underline{\beta} + (Z'Z)^{-1} Z'\underline{U} \tag{3.56}$$

where $P = (Z'Z)^{-1} Z'X$, $H\underline{\beta} = \underline{\beta}_A$, and $\underline{\beta}_A = \frac{1}{n} \sum_{i=1}^{n} \underline{\beta}_1$

$$\underline{b}_F = \underline{\beta}_A + \underline{\delta}_A \tag{3.57}$$

where $\underline{\delta}_A = (P - H) \underline{\beta} + (Z'Z)^{-1} Z'\underline{U}$.

If we take expectations on both sides,

$$E(\underline{b}_F) = \underline{\beta}_A + E(\underline{\delta}_A) \quad . \tag{3.58}$$

The term given by $E(\underline{\delta}_A)$ can be considered as the bias term when using $\underline{b}_F$ as an estimator of the average of the group coefficients. It can readily be seen that the expected value of each element of the FC model

---

[12] Allen (1956) gives a useful summary of Theil's work. Boot and de Wit (1960) have empirically quantified such biases in an analysis of investment demand of ten U. S. corporations.

least squares estimator will depend on both the corresponding as well as non-corresponding parameters of the group models. Similar results can easily be shown to hold for the GLS model estimators, although it is not done here. The standard well-known procedure for testing the assumption of equality of coefficient vectors between the groups has been the F test.[13] Wallace (1972) claims that the F test is an overly strong criterion. The FC model estimators, although biased, have smaller variances than the group estimators under the restriction of equality of coefficients between groups. In such cases he shows that a more appropriate test can be based on comparison of the mean square errors (MSE) of the FC model estimators and the estimators of the group models.[14] For the FC model, we can thus use the F test to test the hypothesis that the coefficient vectors of the group models are equal, or use the MSE criteria to judge if the restricted estimator is better than the unrestricted one. For the GLS model, we can test the hypothesis of equality of coefficients between groups assuming they have different error variances using a test procedure proposed by Rao (1965a, p. 323). Briefly, the test procedure involves a consideration of the following statistic:

$$H_\beta = \sum_{i=1}^{n} \frac{(\underline{b}_i - \underline{b}_G)' X_i' X_i (\underline{b}_i - \underline{b}_G)}{s_i^2} \tag{3.59}$$

---

[13] A clear exposition of the test is given by Goldberger (1964, pp. 256-265).

[14] The details of the relevant test procedure are given in Toro and Wallace (1968).

where $\underline{b}_i$ and $\underline{b}_G$ are defined in equations (3.19) and (3.32), respectively, and s_i^2 is an estimate of σ_i^2. The statistic H_β is asymptotically distributed as χ^2 with $\Lambda(n-1)$ d.f. as $T \longrightarrow \infty$ and n is fixed. Zellner (1962a) shows that the asymptotic distribution of $H_\beta/\Lambda(n-1)$ can be approximated by F with $\Lambda(n-1)$, $\sum_{i=1}^{n} (T_i - \Lambda)$ d.f.

If the above tests are used and result in a rejection of the hypothesis that the coefficients are equal or that the restricted estimators are not better than the unrestricted ones, then we can accept the hypothesis that the estimators of the FC model and the GLS model are biased estimators of the average coefficient vector of the group models, as shown in eq. (3.58). Since the profit-maximizing fertilizer levels are functions of the estimated coefficients, we can also accept the hypothesis that the fertilizer recommendations derived from the FC and the GLS models are biased estimators of the fertilizer recommendations from the average of the group models.[15]

We can now note one result of the RC model. Zellner (1969) applied a RC model to the aggregation problem and showed that for such models there would be no aggregation bias in the least squares estimators of the coefficients in the aggregate model. In the context of the problem considered here, this can be shown as follows:

Substituting eq. (3.41) into (3.43), we have

---

[15]This result follows from the theorem (see Goldberger, 1964, p. 128) which states, "If $\hat{\theta}$ is a consistent estimator of θ, and if $\Psi = g(\theta)$ is a continuous function of θ, then $\hat{\Psi} = g(\hat{\theta})$ is a consistent estimator of Ψ."

$$\underline{\bar{b}}_R(\theta) = (Z'V(\underline{\theta})^{-1} Z)^{-1} Z'V(\underline{\theta})^{-1} [Z \underline{\bar{\beta}}_R + X \underline{\delta} + \underline{W}] \tag{3.60}$$

$$= \underline{\bar{\beta}}_R + [(Z'V(\underline{\theta})^{-1} Z)^{-1} Z'V(\theta)^{-1} X \underline{\delta}$$

$$+ (Z'V(\underline{\theta})^{-1} Z'V(\underline{\theta})^{-1} \underline{W}] = \underline{\bar{\beta}}_R + \underline{\gamma} \tag{3.61}$$

where $\underline{\gamma}$ refers to the terms within the square brackets. Taking expectations on both sides of (3.61)

$$E(\underline{\bar{b}}_R\ (\underline{\theta})) = \underline{\bar{\beta}}_R$$

since $$E[(Z'V(\underline{\theta})^{-1} Z)^{-1} Z'V(\underline{\theta})^{-1} X \underline{\delta}] = 0 \tag{3.62}$$

and $$E[(Z'V(\underline{\theta})^{-1} Z)^{-1} Z'V(\underline{\theta})^{-1} \underline{W}] = 0$$

Analogous to the arguments made earlier,[16] we can show that the profit-maximizing fertilizer levels derived from the RC model are consistent estimators of the fertilizer recommendations from the average of the group models.

---

[16] See footnote 15.

CHAPTER 4
EMPIRICAL ESTIMATION AND RESULTS

The Data

For this study, the data were chosen from the potato field experiments carried out by Professor R. E. McCollum of the North Carolina State University Mission to Peru. The data consisted of the results of potato field experiments conducted in 73 locations over the period 1967-1971 in the Sierra region of Peru.[17] There were 12 locations used for the year 1967-68, 18 for the year 1968-69, 19 for year 1969-70, and 24 for the year 1970-71. The field trials were made up of central composite and factorial response surface designs.[18] For all the experiments included in the data, each had soil analysis made.[19] Four replications of each treatment were conducted for all the experiments. The total number of observations for the experimental data was 4569.[20] Each observation had eight variables--potato yield, soil nitrogen, soil phosphorus, soil potassium, soil pH, applied nitrogen fertilizer, applied phosphorous fertilizer and applied potassium fertilizer.

---

[17] Part of the data between 1967-1970 was the same as that used by Ryan (1972).

[18] These types of designs have been found to be most appropriate for estimating a quadratic response in applied nutrients.

[19] Some experiments which had magnesium deficiency in the soil and/or had no soil analysis were omitted from the experimental data.

[20] Due to space limitations, the data are not presented here. A computer listing is held by the author.

Climatic data were not available for each of the 73 locations. Data from 10 weather stations located at Cajamarca, Huamachuco, Otuzco, Huancayo, Cuzco, Puno, Asangaro, Junin, Huaraz and Huancavelica were used for all the experiments in their vicinity.

Although the data were obtained over the four-year period 1967-1971, none of the locations were repeated from one year to another. Hence, the set of 73 locations can be considered as a set of cross section data.

Results of the Group Models

For the group models, an application of the Buol's soil management classification scheme to data from 73 locations resulted in the formation of five groups. The details of the characteristics of each of the five groups and the number of locations and observations in each group are presented in Table 1.[21] The means and standard deviations of each of the variables for all the groups are given in Appendix B, Table 1. To each of these five groups of data, the model in eq. (3.1) was fit and the ordinary least squares estimation procedure applied to obtain the estimates of the coefficients. The least squares estimates of the coefficients and their 't' values for each of the five group models are given in Table 2. The R^2 of the group models were generally high and ranged from 0.63 to 0.82. As can be seen from Table 2, there were many (about ten on the average for all the groups) estimates that were not

---

[21] I am indebted to Dr. S. W. Buol, Dr. R. E. McCollum, and Dr. S. Villagarcia of the Soil Science Department at North Carolina State University for their help in classifying the soils according to groups for this data.

Table 1. Characteristics of the group data

Group number	Soil code[a]	Number of locations	Number of observations
1	L Ca d k X	6	314
2	L a d k	23	1538
3	L h d k	27	1578
4	L Ca d k	11	622
5	C Ca d k	6	517
Total		73	4569

[a]The details of the soil codes are explained in Appendix A.

Table 2. The estimated coefficients and their 't' values for the group models with all the variables included

Models	B_o	B_{N_S}	B_{P_S}	B_{K_S}	B_{pH}	B_{N_A}	B_{P_A}	B_{K_A}	B_R
G_1	3193.01 (7.08)	-660.89 (-5.44)	0.6103 (1.91)	0.0652 (2.82)	-133.22 (-1.78)	-5.081 (-4.49)	-0.0625 (-0.43)	0.0462 (2.64)	-5.314 (-7.01)
G_2	-123.33 (-5.08)	-23.84 (-10.55)	-0.4398 (-2.44)	0.01882 (1.73)	-0.0898 (-0.02)	-0.0328 (-1.00)	0.05988 (2.79)	0.02752 (2.09)	-0.06325 (-2.45)
G_3	-462.80 (-7.19)	-16.01 (-4.47)	-0.2706 (-2.10)	0.02228 (3.00)	10.25 (1.51)	0.05369 (1.04)	0.2722 (10.48)	0.001116 (0.09)	0.7179 (5.27)
G_4	100.16 (1.15)	6.002 (0.63)	0.9357 (4.71)	0.1365 (3.38)	-61.02 (-3.75)	-0.005606 (-0.19)	-0.09052 (-2.04)	-0.009480 (-0.57)	-0.1446 (-2.19)
G_5	-317.17 (-0.69)	122.90 (0.91)	-0.2893 (-0.18)	-0.005301 (-0.34)	-30.98 (-0.77)	0.3744 (0.62)	-0.1222 (-2.44)	-0.001097 (-0.08)	0.4559 (1.08)

Table 2 (continued)

Models	B_T	$B_{N_S}^2$	$B_{P_S}^2$	$B_{K_S}^2$	B_{pH}^2	$B_{N_A}^2$	$B_{P_A}^2$	$B_{K_A}^2$	B_R^2
G_1	0.0 (0.0)	1.649 (4.18)	0.01734 (1.09)	-0.0001135 (-2.52)	13.15 (2.15)	0.00003438 (0.54)	-0.00000219 (-0.04)	0.00000158 (0.06)	0.0 (0.0)
G_2	52.49 (15.43)	-0.1350 (-4.92)	0.02493 (5.58)	0.00000504 (0.22)	-0.2193 (-0.65)	-0.00003015 (-0.59)	-0.00005582 (-4.95)	-0.00007711 (-2.03)	0.0002138 (14.05)
G_3	89.70 (7.45)	0.4454 (4.22)	0.005137 (1.63)	-0.0000237 (-3.50)	-1.007 (-1.88)	-0.00003435 (-0.78)	-0.0001280 (-10.57)	-0.00004607 (-1.45)	-0.0002927 (-3.49)
G_4	32.54 (1.51)	0.3149 (1.40)	-0.01869 (-4.77)	-0.0003834 (-3.47)	4.116 (3.87)	-0.0001248 (-2.69)	-0.00002344 (-2.70)	0.00004530 (1.11)	0.00004983 (1.23)
G_5	26.26 (0.92)	1.491 (4.51)	-0.1998 (-1.20)	0.00000965 (0.41)	1.969 (0.80)	-0.00006585 (-0.63)	-0.00005870 (-6.32)	0.00002121 (0.66)	0.0 (0.0)

Table 2 (continued)

Models	B_T^2	$B_{N_S pH}$	$B_{N_S N_A}$	$B_{N_S R}$	$B_{N_S T}$	$B_{P_S P_A}$	$B_{K_S K_A}$	B_{pHP_A}
G_1	0.0 (0.0)	-9.707 (-3.81)	-0.002927 (-0.67)	1.320 (5.92)	0.0 (0.0)	-0.001102 (-2.43)	-0.00006337 (-2.83)	0.004685 (0.26)
G_2	-2.653 (-17.76)	0.6900 (5.67)	0.005131 (3.77)	0.02666 (14.33)	1.300 (10.01)	-0.003374 (-5.93)	-0.00009432 (-2.45)	0.0009362 (0.21)
G_3	-3.353 (-9.56)	0.9263 (2.77)	-0.007598 (-2.72)	0.03834 (8.59)	-1.349 (-6.98)	-0.0009373 (-2.27)	-0.00003822 (-1.04)	-0.03267 (-7.60)
G_4	-1.864 (-1.34)	1.433 (2.90)	0.003264 (1.32)	0.0009959 (0.17)	-2.920 (-2.31)	-0.001421 (-2.69)	-0.00000611 (-0.09)	0.01404 (2.58)
G_5	0.0 (0.0)	0.5320 (0.29)	-0.03068 (-5.58)	-0.1371 (-1.09)	-7.460 (-0.87)	0.005110 (2.08)	-0.00003062 (-0.94)	0.01837 (3.44)

Table 2 (continued)

Models	$B_{N_A P_A}$	$B_{N_A K_A}$	$B_{N_A R}$	$B_{N_A T}$	$B_{P_A K_A}$	B_{RT}	$B_{N_A P_S}$
G_1	0.0001487 (1.51)	-0.0003773 (-2.03)	0.009474 (4.62)	0.0 (0.0)	0.0002033 (1.11)	0.0 (0.0)	0.0009035 (1.71)
G_2	-0.00000922 (-0.33)	0.00003772 (0.58)	-0.00000741 (-0.30)	0.005340 (2.10)	-0.00001536 (-0.49)	-0.03711 (-14.94)	0.001024 (1.28)
G_3	-0.00000426 (-0.12)	0.0002136 (3.80)	-0.00000600 (-0.08)	0.002392 (0.79)	-0.00003696 (-1.04)	-0.08442 (-5.53)	-0.002123 (-4.30)
G_4	0.00003270 (1.81)	0.00000262 (0.06)	-0.00000443 (-0.17)	0.009159 (1.58)	0.00002771 (1.18)	0.01283 (1.54)	0.00005577 (0.12)
G_5	0.00007408 (3.20)	0.00001244 (0.18)	-0.0002495 (-0.44)	-0.01218 (-0.31)	0.00000686 (0.27)	0.0 (0.0)	0.004531 (0.90)

significantly different from zero.[22] All variables that were not significant at the 0.25 level were dropped, and the functions described by eq. (3.1) were refitted with the remaining variables. Although such a procedure entails the well known sequential biases of the estimators, it was considered that such a model with reduced variables would represent the response relation more accurately.[23] Other procedures such as the forward selection procedure, the backward selection procedure, stepwise regression, etc., also involve similar sequential biases and are more time consuming and expensive.[24] No clear result has been established regarding the superiority of any of the listed procedures. Thus the method of dropping all nonsignificant variables at the 0.25 level was chosen for simplicity of computational procedures. Because of the sequential bias of the resultant estimator that is obtained, it has been suggested (Anderson _et al._, 1970, Wallace, 1972) that the $\alpha = 0.05$ level is an overly strong criterion and that a more reasonable level would be much higher. The level of 0.25 was chosen here so that only a reasonable number of variables may be dropped from consideration. The results of fitting the group models to the reduced number of variables is given in Table 3.[25] It can be seen that there is very little

---

[22] Many data variables did not have enough variability to obtain significant estimates of coefficients. See Appendix B, Table 1.

[23] For a clear exposition of this problem, see Wallace and Ashar (1972).

[24] See Draper and Smith (1966) for a description of these procedures.

[25] In all cases the hypothesis that the subsets of coefficients dropped from each of the regressions are equal to zero was accepted at the 0.05 level (see Goldberger, 1964, p. 177 for the relevant test procedure).

Table 3. The estimated coefficients and their 't' values for the group models with some of the variables (non-significant coefficients at $\alpha = 0.25$) excluded

Models	B_o	B_{N_S}	B_{P_S}	B_{K_S}	B_{pH}	B_{N_A}	B_{P_A}	B_{K_A}	B_R
G_1	3240.84 (7.63)	-654.92 (-5.91)	0.5812 (1.88)	0.06394 (2.89)	-125.06 (-1.71)	-5.761 (-7.47)	0.0 (0.0)	0.03254 (3.28)	-5.451 (-7.72)
G_2	-137.23 (-7.40)	-18.94 (-9.38)	-0.6829 (-4.18)	0.01589 (2.70)	0.0 (0.0)	-0.04324 (-2.42)	0.05614 (9.22)	0.02571 (2.26)	-0.03939 (-1.63)
G_3	-477.11 (-7.43)	-16.57 (-4.65)	-0.2595 (-2.02)	0.02222 (3.22)	8.866 (1.31)	0.06490 (4.08)	0.2679 (10.39)	0.0 (0.0)	0.7483 (5.54)
G_4	135.67 (2.58)	0.0 (0.0)	0.9556 (5.26)	0.1416 (3.63)	-60.37 (-3.91)	0.0 (0.0)	-0.08923 (-2.04)	-0.01048 (-1.13)	-0.1165 (-2.47)
G_5	-27.97 (-2.12)	11.21 (2.65)	0.0 (0.0)	0.0 (0.0)	0.0 (0.0)	0.1762 (5.60)	-0.1342 (-4.53)	0.0 (0.0)	0.06177 (3.35)

Table 3 (continued)

Models	B_T	$B_{N_S}^2$	$B_{P_S}^2$	$B_{K_S}^2$	B_{pH}^2	$B_{N_A}^2$	$B_{P_A}^2$	$B_{K_A}^2$
G_1	0.0 (0.0)	1.624 (4.16)	0.01705 (1.09)	-0.0001124 (-2.55)	12.56 (2.09)	0.0 (0.0)	0.0 (0.0)	0.0 (0.0)
G_2	52.25 (15.53)	-0.1643 (-6.19)	0.02404 (5.58)	0.0 (0.0)	0.0 (0.0)	-0.00005135 (-1.07)	-0.00005916 (-6.24)	-0.00009127 (-2.46)
G_3	93.18 (7.78)	0.4588 (4.36)	0.004733 (1.51)	-0.00002407 (-3.81)	-0.9215 (-1.73)	-0.00002169 (-0.50)	-0.0001251 (-10.57)	-0.00003714 (-1.45)
G_4	21.22 (2.30)	0.4516 (4.73)	-0.01904 (-5.03)	-0.0004010 (-3.83)	4.056 (4.11)	-0.0001234 (-3.13)	-0.00002363 (-2.82)	0.00004753 (1.30)
G_5	0.0 (0.0)	1.412 (5.55)	-0.2439 (-3.31)	0.0 (0.0)	0.0 (0.0)	0.0 (0.0)	-0.00005833 (-6.97)	0.00002506 (1.82)

Table 3 (continued)

Models	B_R2	B_T2	$B_{N_S pH}$	$B_{N_S N_A}$	$B_{N_S R}$	$B_{N_S T}$	$B_{P_S P_A}$	$B_{K_S K_A}$
G_1	0.0 (0.0)	0.0 (0.0)	-9.548 (-3.78)	0.0 (0.0)	1.307 (6.44)	0.0 (0.0)	-0.0008735 (-2.11)	-0.00005750 (-2.77)
G_2	0.0002000 (14.51)	-2.513 (-17.07)	0.2933 (3.64)	0.005532 (4.15)	0.02376 (13.30)	1.079 (8.86)	-0.002311 (-4.38)	-0.0001017 (-2.70)
G_3	-0.0003012 (-3.66)	-3.437 (-9.91)	1.008 (3.02)	-0.007562 (-3.31)	0.03871 (8.71)	-1.389 (-7.22)	-0.0009544 (-2.33)	-0.00003105 (-1.07)
G_4	0.00003375 (1.10)	-1.132 (-1.63)	1.553 (3.76)	0.003060 (2.22)	0.0 (0.0)	-2.1526 (-4.94)	-0.001398 (-2.79)	0.0 (0.0)
G_5	0.0 (0.0)	0.0 (0.0)	0.0 (0.0)	-0.03134 (-6.66)	-0.02610 (-4.19)	0.0 (0.0)	0.005683 (3.59)	-0.00002483 (-1.76)

Table 3 (continued)

Models	B_{pHP_A}	$B_{N_AP_A}$	$B_{N_AK_A}$	B_{N_AR}	B_{N_AT}	$B_{P_AK_A}$	B_{RT}	$B_{N_AP_S}$
G_1	0.0 (0.0)	0.00007390 (1.61)	-0.0001510 (-3.15)	0.01068 (7.51)	0.0 (0.0)	0.00005841 (1.64)	0.0 (0.0)	0.0007986 (1.69)
G_2	0.0 (0.0)	0.0 (0.0)	0.00006582 (1.05)	0.0 (0.0)	0.005970 (3.12)	0.0 (0.0)	-0.03757 (-15.08)	0.001205 (1.57)
G_3	-0.03221 (-7.59)	0.0 (0.0)	0.0001915 (3.64)	0.0 (0.0)	0.0 (0.0)	-0.00003460 (-1.09)	-0.08816 (-5.80)	-0.002101 (-4.33)
G_4	0.01389 (2.59)	0.00003207 (1.85)	0.0 (0.0)	0.0 (0.0)	0.008041 (5.29)	0.00002735 (1.20)	0.01153 (1.71)	0.0 (0.0)
G_5	0.01972 (6.40)	0.00007564 (3.42)	0.0 (0.0)	-0.00007071 (-1.47)	0.0 (0.0)	0.0 (0.0)	0.0 (0.0)	0.004928 (1.16)

Table 3 (continued)

Models	Number of observations	R^2	Error mean square	F value
G_1	314	0.81	7.23	57.93
G_2	1,538	0.62	30.28	99.82
G_3	1,578	0.69	45.27	120.90
G_4	622	0.82	11.46	106.65
G_5	517	0.75	20.91	76.94

reduction in the R^2 and the error mean squares. The estimates of the coefficients from these models were used to generate fertilizer recommendations for given soil conditions. Since many of the variables dropped were soil-applied nutrient interactions, these soil values will not have any influence on the optimum fertilizer levels. For example, the coefficient for the interaction term $pH.P_A$ is not significantly different from zero (a "t" value of 0.26) for the first group model. Hence values of pH will have no influence on the optimum level of fertilizers. There are at least two possible reasons for the non-significance of some of the coefficients: (a) there is an insufficient number of data observations to obtain an estimate for all the coefficients and (b) the underlying response relation for the group may be such that these coefficients are zero.

To test the suitability of Buol's soil management classification system for this set of data, the five groups have to be compared with each other for any differences between them. One way to do this is to test the hypothesis that the coefficient vector between any two groups is the same. This can be done by using the F test referred to in the last chapter.[26] Table 4 gives the F values for all possible combinations of two groups at a time. In all cases, the hypothesis was rejected leading to the conclusion that the Buol's criterion does classify soils into different management groups.

---

[26] See footnote 13.

Table 4. Test for equality of coefficients for the group models

Model	SSE	D.F.	Calculated F value
G_1	2,101.1	282	-
G_2	44,821.7	1,506	-
G_3	69,996.2	1,546	-
G_4	6,823.2	590	-
G_5	10,409.1	485	-
$G_1 + G_2$	56,317.8	1,788	8.59**
$G_1 + G_3$	86,703.0	1,828	9.48**
$G_1 + G_4$	11,631.6	772	5.71**
$G_1 + G_5$	14,830.0	767	3.36**
$G_2 + G_3$	177,869.8	3,052	36.41**
$G_2 + G_4$	75,651.6	2,096	19.21**
$G_2 + G_5$	73,560.7	1,991	14.31**
$G_3 + G_4$	103,126.6	2,136	15.75**
$G_3 + G_5$	119,767.3	2,031	10.57**
$G_4 + G_5$	20,257.0	1,075	4.12**

**Significant at the 0.01 level.

Results of the Aggregate Models

Equation (3.1) was fit to each of the three specifications at the aggregate level: the FC model, the GLS model, and the RC model. The coefficients of the FC model were estimated by ordinary least squares.

For the GLS model, the error variances from each of the group models were used to estimate the coefficients by the use of the formula in eq. (3.31). The $\hat{\Delta}$ matrix for the RC model was obtained from eq. (3.52) using all the estimated coefficients from the 5 group models. The resultant estimates of the coefficients of the RC model was obtained by using estimates of Δ and σ_i^2 in eq. (3.44). The results of the estimates of the coefficients and their 't' values for all the aggregate models are given in Table 5. The R^2 for these models are much lower compared to the group models and ranged from 0.28 to 0.45. However, this is to be expected with the larger set of data used to fit each aggregated model.

In the previous chapter the assumptions of each of these models were outlined. We will now examine the validity of these assumptions of each of the aggregated models for the set of experimental data considered in this study.

The first hypothesis to be tested is the homogeneity of error variances of the group models. This is one of the assumptions made for the estimation of the FC model. Bartlett's test described in the last chapter (p. 41) to test this hypothesis resulted in a Q/ℓ value of 558.91. This is well above the tabulated χ^2 value of 13.3 with 4 d.f. at the 0.01 level. Thus, it can be concluded that the error variances between the groups are not equal. If this is so, it was noted that the least squares estimates of the FC model are inefficient. Also, an F test on the equality of coefficients between the groups resulted in an F value of 48.67 which again was higher than the tabulated F value of 1.22 with (132, 4404) degrees of freedom at the 0.05

Table 5. The estimated coefficients and their 't' values for the different specifications of the aggregate model

Models	B_o	B_{N_S}	B_{P_S}	B_{K_S}	B_{pH}	B_{N_A}	B_{P_A}	B_{K_A}	B_R
FC	-31.06	-1.366	-0.9528	0.03278	15.82	-0.03608	0.05286	0.007867	0.09877
	(-4.15)	(-1.28)	(-10.76)	(7.61)	(13.03)	(-1.58)	(6.47)	(0.98)	(11.24)
GLS	2.854	6.111	-0.8357	0.02618	4.618	-0.007125	0.03178	0.001560	0.06064
	(28.48)	(6.67)	(-11.48)	(7.30)	(4.98)	(-0.44)	(4.69)	(0.25)	(8.81)
RC1	986.25	-274.33	0.2300	0.007971	65.95	-1.955	-0.01127	-0.01922	-2.343
	(1.46)	(-1.72)	(0.64)	(0.30)	(2.24)	(-1.35)	(-0.13)	(-1.61)	(-2.08)
RC2	422.46	-90.71	0.2903	0.01716	-63.62	-0.6112	0.03953	0.02185	-0.5014
	(0.63)	(-0.68)	(2.94)	(0.75)	(-3.37)	(-0.61)	(0.62)	(2.64)	(-0.45)

Table 5 (continued)

Models	B_T	$B_{N_S^2}$	$B_{P_S^2}$	$B_{K_S^2}$	B_{pH^2}	$B_{N_A^2}$	$B_{P_A^2}$	$B_{K_A^2}$
FC	-1.773 (-1.07)	-0.1233 (-5.99)	0.02817 (15.60)	-0.00002994 (-6.89)	-1.364 (-14.45)	-0.00004457 (-1.15)	-0.00008328 (-11.88)	-0.00009140 (-4.03)
GLS	-8.744 (-9.60)	-0.1756 (-9.40)	0.02455 (17.17)	-0.00002935 (-7.83)	-0.3378 (-4.63)	-0.00007798 (-2.32)	-0.00005406 (-9.61)	0.00001955 (1.11)
RC1	19.56 (0.97)	-0.1095 (-0.34)	-0.07461 (-1.54)	-0.00001599 (-0.15)	-5.387 (-1.60)	-0.0001397 (-8.45)	-0.00008401 (-4.07)	-0.00004040 (-1.46)
RC2	29.66 (2.19)	0.7381 (2.25)	-0.04694 (-1.84)	-0.00006065 (-0.87)	4.906 (2.27)	-0.00004563 (-1.98)	-0.00003972 (-2.21)	-0.00006064 (-4.00)

Table 5 (continued)

Models	B_R2	B_T2	$B_{N_S pH}$	$B_{N_S N_A}$	$B_{N_S R}$	$B_{N_S T}$	$B_{P_S P_A}$	$B_{K_S K_A}$
FC	-0.0001307	-0.1068	-0.05081	0.001716	0.004362	0.1090	-0.001811	-0.00006871
	(-19.78)	(-1.25)	(-0.93)	(1.48)	(4.72)	(1.90)	(-6.10)	(-3.67)
GLS	-0.0001070	0.3310	0.02093	0.001200	-0.003677	-0.3282	-0.001693	-0.00000446
	(-18.93)	(6.30)	(0.50)	(1.24)	(-4.57)	(-6.66)	(-6.93)	(0.31)
RC1	0.00006721	-1.007	2.433	-0.006666	0.4893	-0.2839	0.0009490	-0.00008186
	(0.47)	(-1.63)	(0.95)	(-0.80)	(1.53)	(-0.30)	(0.68)	(-7.34)
RC2	-0.0001158	-1.523	-0.3197	-0.009145	0.1745	-0.2881	-0.001041	-0.00006318
	(-1.41)	(-2.40)	(-0.16)	(-1.47)	(0.66)	(-3.49)	(-0.78)	(-4.68)

Table 5 (continued)

Models	B_{pHP_A}	$B_{N_AP_A}$	$B_{N_AK_A}$	B_{N_AR}	B_{N_AT}	$B_{P_AK_A}$	B_{RT}
FC	0.0006820 (0.77)	0.00003485 (1.77)	0.0001795 (4.39)	0.00002216 (1.11)	0.008006 (4.16)	0.00001381 (0.62)	-0.0003985 (-0.31)
GLS	0.0008844 (1.24)	0.00007273 (4.67)	0.00003747 (1.15)	0.00001940 (1.34)	0.004556 (2.93)	-0.00000490 (-0.27)	0.005738 (6.17)
RC1	0.007510 (0.46)	0.00002821 (2.07)	0.0002932 (3.09)	0.003670 (0.69)	0.002993 (1.58)	-0.00001331 (-0.64)	-0.008053 (-0.29)
RC2	0.001598 (0.20)	0.00005953 (3.05)	0.00004227 (0.51)	0.001215 (0.65)	0.005415 (2.88)	-0.00000669 (-0.40)	-0.01284 (-0.71)

Table 5 (continued)

Models	$B_{N_A P_S}$	Number of observations	R^2	Error mean square
FC	-0.001257 (-3.28)	4,569	0.36	74.32
GLS	-0.007212 (-2.36)	4,569	0.44	-
RC1	0.004354 (4.64)	4,569	0.28	-
RC2	-0.0003663 (-1.20)	4,569	0.31	-

level,[27] rejecting the hypothesis that the coefficients between the groups are equal. It was shown in the last chapter that under these conditions the least squares estimates of the FC model have aggregation biases.

Next, the assumption of the GLS model that the coefficients of the groups are equal assuming different group error variances was tested by calculating the $H_\beta/\Lambda(n-1)$ statistic in eq. (3.59). The $H_\beta/\Lambda(n-1)$ value was found to be 224.86 and this is greater than the tabulated F value with (132, 4404) d.f. at the 0.01 level. Again, the rejection of this hypothesis leads to the conclusion that the estimators of the coefficients of the GLS model have aggregation biases.

In estimating the Δ matrix for the RC model, it was found that five of the diagonal elements (those corresponding to the variables P_S, N_A^2, N_ST, K_SK_A, N_AT) were negative. Since the diagonal elements of the $\hat{\Delta}$ matrix are functions of the variances of the coefficients, they are always expected to be positive. However, sampling variance could lead to some negative estimates. The rows and columns associated with the above five elements of the $\hat{\Delta}$ matrix were set equal to zero and the resultant $\hat{\Delta}$ matrix used to estimate the coefficient vector of the RC model. However, as mentioned in the last chapter, such a procedure would lead to biased estimates and is inconsistent with the assumptions of the RC model.

From the above results, we can conclude that none of the specifications of the aggregate model are likely to provide unbiased estimates of the average of the group models' coefficients.

---

[27] The corresponding F value using the MSE criterion from the tables provided by Wallace and Toro (1969) is 1.25.

"Cost of the Wrong Decision" Criteria

To describe the population of soils, 144 soil classes were established, based on the levels of the four soil variables N_s, P_s, K_s, and pH. The intervals for the four variables are:

N_s	(percent o.m.) :	≤ 2.0 ;	2.1 - 4.0 ;	> 4.0	
P_s	(p.p.m. of P) :	≤ 4.0 ;	4.1 - 8.0 ;	8.1 - 15.0 ;	> 15.0
K_s	(p.p.m. of K) :	≤ 60 ;	61 - 120 ;	121 - 230 ;	> 230
pH	:	≤ 5.5 ;	5.6 - 6.5 ;	> 6.5 .	

The soils from the experimental data were classified according to this scheme, and the mean value of each of the four variables for all the 144 classes was calculated, together with the relative frequency of occurrence. Out of the possible 144 soil classes for each group, 12 classes with positive frequencies were obtained for group 1, 32 for group 2, 39 for group 3, 16 for group 4, and 8 for group 5, and the remaining classes in each group had zero frequencies.[28] For these soil values, we have the information regarding the group to which they belong. Hence, we can derive fertilizer recommendations and the corresponding profits for the soil classes belonging to each of the groups, using the estimates of the coefficients of the group models from Table 3. For rainfall and temperature, the averages for each of the groups are used (see Appendix B, Table 1). The prices used throughout

[28]See Table 6 in the next section for the details of the soil classes.

this study are 2 soles/kgm for potatoes, 9.118 soles/kgm for the nitrogen fertilizer, 5.76 soles/kgm for the phosphorous fertilizer, and 5.76 soles/kgm for the potassium fertilizer.[29]

Using all the group models and the different specifications of the aggregate model, analytical solutions could not be obtained in every situation. In some cases the estimated response curve reached a maximum at fertilizer levels far in excess of the experimental application rates. This is likely to occur when experimental rates applied are not large enough to expose the region of zero or negative marginal product. While it is acceptable practice to predict beyond the range of experimental data, due to the large standard error of estimates of the optimum in such cases, such recommendations would impose very large risks on farmers. We, therefore, arbitrarily confine all recommendations within 350, 350, 350 kg/ha for applied nitrogen, phosphorus, and potassium, respectively, as these are the approximate levels of maximum treatments. Use of other values for the constraints would undoubtedly affect the results obtained here, although to what extent and in what way is not known. In this study, we will restrict our study to considering the above mentioned maximum recommendations.

In some other cases, the stationary point on the profit function was either a saddle point or a minimum. Because of these analytical difficulties, the profit maximizing fertilizer levels were determined using the steepest ascent method with constraints as described earlier in the chapter. In this study the initial point for all the solutions

---

[29]These are the average prices as of March 1971. The prices for the fertilizer include the average transport costs (see Ryan, 1972, p. 181). The exchange rate is 42.5 Peruvian soles = U. S. $1.00.

was chosen to be 120 kg/hectare for each of the applied nitrogen, phosphorous and potassium fertilizers. The step lengths were 20 kg/hectare initially and 5 kg/hectare for the reduced step lengths. The inequality constraints were

$$0 \leq N_A \leq 350 \text{ kg/ha}$$

$$0 \leq P_A \leq 350 \text{ kg/ha}$$

$$0 \leq K_A \leq 350 \text{ kg/ha.}$$

The expected loss in using each of the specifications of the aggregate model, assuming the group models to be true, can be calculated by using the formula:[30]

$$E(L_A) = \sum_{i=1}^{5} \sum_{j=1}^{s_i} [(\Pi_{ij} - \Pi_{Aj}) P_{ij}] \tag{4.1}$$

where $E(L_A)$ = expected loss in soles/hectare in using each of the specifications of the aggregate model

s_i = number of soil classes in group i

P_{ij} = probability of soil analysis for the jth soil class in the ith group

Π_{ij} = profits obtained from using the fertilizer recommendations for the jth soil class and the ith group model from Table 6

[30] In calculating the profits, only the variable costs associated with the level of fertilizers applied are considered. If other costs of production are taken into account, then the profits would be lower. The prices of potatoes and the fertilizers are also assumed to be constant, in spite of the increased production implied by this analysis.

Table 6. Optimal fertilizer recommendations, yields, and profits derived from using the group models for the different soil classes

Models	Soil analysis number	Soil code[a]	Means of soil variables (ppm) N_S	P_S	K_S	pH	Optimal amts of applied fert. (kg./ha.) N_A	P_A	K_A	Optimal yield (kg/ha) Y	Profits soles/ha.	Prob. of soil analysis
Group 1	1	2122	2.6	3.0	104.0	6.5	350	350	0	16.2	25,567	.002626
	2	2123	2.6	4.0	104.0	6.6	350	350	0	18.5	30,115	.005252
	3	2133	3.4	4.0	197.5	7.0	350	350	0	25.4	44,007	.005909
	4	2232	2.7	5.0	147.3	6.4	350	350	0	18.2	29,580	.007878
	5	2233	3.4	6.5	206.3	6.9	350	350	0	25.8	44,771	.012475
	6	2242	3.6	9.0	241.0	6.5	350	350	0	24.2	41,516	.002626
	7	2243	3.6	7.8	257.5	6.9	350	350	0	26.3	45,767	.012257
	8	2343	3.9	12.0	391.0	6.8	350	350	0	27.9	48,969	.003064
	9	3123	8.4	4.0	84.7	7.7	350	350	0	28.6	50,341	.007878
	10	3223	8.8	5.0	98.0	7.7	350	350	0	32.6	58,411	.002626
	11	3343	4.5	14.0	410.0	6.7	350	350	0	29.6	52,476	.003064
	12	3443	5.4	40.0	873.0	7.0	350	0	0	36.3	69,440	.003064
	13	Average	4.2	7.8	226.5	6.9	350	350	0	25.9	45,021	.068724
Group 2	14	1121	1.8	3.3	90.0	4.1	154	350	141	14.5	23,090	.021011
	15	1131	1.9	3.5	133.5	4.3	153	350	119	14.6	23,484	.004815
	16	1211	1.4	9.0	13.8	4.0	200	255	190	8.0	10,410	.005252
	17	1221	1.7	7.3	96.0	4.3	160	288	129	10.0	14,902	.022981
	18	1231	1.9	6.7	136.7	4.2	145	301	101	11.1	17,216	.013789
	19	1311	1.5	11.3	13.4	4.0	243	210	205	7.1	8,729	.015756
	20	1321	1.4	10.0	65.0	4.0	192	236	158	7.1	9,136	.005909
	21	1433	1.9	17.0	202.0	7.1	270	99	110	9.0	13,829	.003064
	22	1443	2.0	17.0	247.0	6.9	256	99	80	9.4	14,957	.003064
	23	2131	2.1	4.0	150.0	4.3	159	350	112	14.6	23,430	.002408
	24	2211	3.7	6.5	52.5	3.8	302	304	205	15.8	24,572	.010504
	25	2221	3.3	7.4	87.2	4.4	277	287	176	14.7	22,820	.010068
	26	2231	2.8	7.4	174.7	4.8	198	286	99	13.2	21,108	.035894

.e 6 (continued)

:ls	Soil analysis number	Soil code[a]	Means of soil variables (ppm) N_S	P_S	K_S	pH	Optimal amts. of applied fert. (kg./ha.) N_A	P_A	K_A	Optimal yield (kg/ha) Y	Profits soles/ha.	Prob. of soil analysis
	27	2321	2.3	12.0	85.0	5.1	276	197	177	9.7	13,852	.002845
	28	2331	3.1	12.5	165.7	5.0	303	186	142	11.9	18,347	.011819
	29	2333	2.1	15.0	195.0	7.5	257	138	109	9.8	15,292	.003064
	30	2431	3.0	24.4	213.5	4.9	350	0	133	16.0	28,083	.005691
	31	2432	3.9	19.0	210.0	5.7	350	106	141	16.2	27,285	.002626
	32	2433	2.1	16.0	208.0	7.2	266	119	105	9.7	15,188	.003064
	33	2441	2.7	18.0	254.0	4.7	319	80	99	10.9	17,457	.009192
	34	3121	9.2	2.7	91.7	4.5	350	350	213	31.2	54,304	.047713
	35	3131	9.5	2.9	172.7	5.0	350	350	160	32.4	57,092	.034581
	36	3141	4.2	4.0	240.0	5.3	236	350	85	21.1	35,916	.002626
	37	3211	4.1	9.0	50.0	3.9	350	280	220	15.8	24,207	.005252
	38	3221	9.3	5.0	80.3	4.2	350	350	215	28.6	49,217	.007223
	39	3222	9.5	6.0	100.0	5.7	350	310	199	31.8	56,133	.002626
	40	3231	6.9	7.5	165.0	5.1	350	290	163	24.5	41,979	.022324
	41	3241	5.0	7.5	319.5	4.9	284	285	49	19.8	33,725	.006128
	42	3331	4.2	13.2	178.4	4.9	350	202	148	15.2	24,342	.007004
	43	3332	4.3	11.0	190.0	5.6	350	244	142	17.0	27,420	.002626
	44	3341	5.0	10.0	404.0	5.0	287	236	3	19.5	34,021	.003064
	45	3441	4.1	20.0	240.0	5.5	350	80	173	17.1	29,422	.002626
	46	Average	4.7	7.5	135.5	4.7	350	306	173	20.2	33,221	.336616
ıp 3	47	1122	1.8	3.5	87.8	5.7	350	268	350	55.5	102,973	.012257
	48	1212	1.4	7.0	60.0	5.9	350	232	350	49.2	90,860	.002626
	49	1222	1.4	8.6	68.3	5.8	350	239	350	48.3	88,914	.010664
	50	1312	2.0	12.7	46.7	5.7	350	242	350	44.5	81,339	.016415
	51	1321	1.7	11.5	96.8	5.3	350	297	350	49.4	90,543	.018385
	52	1343	1.8	14.7	449.5	7.0	350	61	350	34.2	62,614	.008974
	53	1421	2.0	24.0	104.0	5.0	350	274	350	37.8	67,506	.001094

Table 6 (continued)

Models	Soil analysis number	Soil code[a]	Means of soil variables (ppm)				Optimal amts. of applied fert. (kg./ha.)			Optimal yield (kg/ha)	Profits	Prob. of soil analysis
			N_S	P_S	K_S	pH	N_A	P_A	K_A	Y	soles/ha.	
	54	1422	1.6	21.2	107.4	5.9	350	177	350	34.7	62,447	.016633
	55	1423	1.2	18.5	91.0	6.7	350	92	350	32.5	58,880	.010504
	56	1432	1.5	18.5	164.0	6.0	350	174	350	37.1	67,167	.005252
	57	1443	1.9	19.6	423.0	7.0	350	42	350	30.3	54,943	.008974
	58	2121	2.8	3.3	99.2	5.1	350	350	350	61.3	113,713	.019479
	59	2123	3.4	2.5	110.0	6.9	350	118	350	59.0	95,209	.005252
	60	2132	3.7	4.0	172.0	6.3	350	181	350	53.1	99,099	.003060
	61	2133	3.5	3.9	150.8	6.9	350	111	350	49.8	93,149	.002845
	62	2231	3.6	7.0	130.0	4.8	350	350	350	59.4	110,061	.005909
	63	2232	3.1	6.0	162.0	6.0	350	214	350	51.6	95,776	.006128
	64	2233	4.0	5.0	150.0	6.8	350	121	350	50.5	94,478	.002626
	65	2242	3.6	7.0	404.0	5.7	350	248	350	52.4	96,961	.002626
	66	2312	2.2	14.2	51.1	5.8	350	199	350	41.7	76,139	.013351
	67	2342	3.6	13.5	557.0	6.0	350	196	350	41.8	76,283	.005252
	68	2412	2.4	16.0	60.0	5.6	350	222	350	41.6	75,609	.002626
	69	2421	2.3	25.7	102.0	4.8	350	296	350	37.4	66,492	.003940
	70	2422	2.3	30.3	102.7	5.8	350	156	350	27.0	47,154	.012694
	71	2432	2.7	29.3	199.3	5.9	125	163	131	20.4	37,208	.008536
	72	2442	3.0	30.6	340.6	6.1	0	167	0	23.4	45,158	.005472
	73	2443	2.4	22.0	860.0	6.7	350	65	350	21.4	36,947	.005691
	74	3121	4.2	4.0	72.0	5.0	350	350	350	62.1	115,305	.003502
	75	3122	7.1	3.7	105.0	6.2	350	202	350	66.9	126,492	.007878
	76	3131	4.2	2.5	175.0	5.3	350	330	350	62.6	116,568	.005252
	77	3141	4.8	2.8	310.0	5.2	350	320	350	63.7	118,820	.013131
	78	3222	8.0	5.0	120.0	6.4	350	171	350	70.2	133,475	.002626
	79	3231	5.4	7.3	167.8	5.3	350	300	350	59.3	110,267	.036332
	80	3232	5.1	7.6	164.3	6.1	350	191	350	53.9	100,693	.026483
	81	3233	4.9	6.0	150.0	6.9	350	106	350	52.3	98,355	.002626

Table 6 (continued)

Models	Soil analysis number	Soil code[a]	Means of soil variables (ppm)				Optimal amts. of applied fert. (kg/ha)			Optimal yield (kg/ha)	Profits	Prob. of soil analysis
			N_S	P_S	K_S	pH	N_A	P_A	K_A	Y	soles/ha.	
	82	3241	4.9	5.4	301.7	5.4	350	287	350	59.2	110,254	.015102
	83	3242	5.4	7.3	347.8	5.8	350	235	350	56.2	104,841	.005909
	84	3331	4.5	10.0	200.0	5.0	350	322	350	55.6	102,584	.002626
	85	3341	4.5	10.0	290.0	5.4	350	275	350	53.0	97,966	.002626
	86	Average	3.3	11.3	179.7	5.8	350	220	350	47.8	88,083	.345371
Group 4	87	1113	2.0	4.0	45.0	7.6	263	350	350	9.4	10,791	.002626
	88	1123	1.2	4.0	111.0	7.4	261	350	350	12.8	17,588	.002626
	89	1133	1.3	3.0	155.0	8.0	257	350	350	18.9	29,821	.021668
	90	1223	1.7	7.0	89.5	7.4	263	328	350	11.8	15,811	.005252
	91	1233	1.3	5.6	137.6	7.2	258	321	350	13.1	18,496	.010068
	92	2113	2.2	4.0	60.0	7.7	266	350	350	12.1	16,138	.002626
	93	2133	3.3	3.4	171.9	8.3	284	350	350	27.9	47,614	.020792
	94	2213	2.1	6.0	60.0	7.6	273	350	350	11.6	15,016	.002626
	95	2223	3.7	9.0	100.0	7.4	285	298	350	15.5	23,360	.002408
	96	2233	3.4	6.2	189.8	7.9	288	350	350	23.6	38,890	.030860
	97	2243	4.0	9.0	261.0	7.6	293	350	350	18.3	28,231	.003064
	98	2323	3.7	13.0	96.7	7.0	261	100	350	12.8	20,121	.007223
	99	2423	3.8	18.5	103.8	7.4	257	66	350	16.4	27,788	.004377
	100	3233	9.2	4.4	227.5	7.5	350	350	350	42.9	76,928	.005252
	101	3333	4.5	13.0	146.5	7.1	273	122	350	16.8	27,910	.004815
	102	3433	6.8	33.2	197.9	7.2	271	0	141	24.5	45,658	.009849
	103	Average	3.3	8.1	157.3	7.7	284	350	350	19.4	30,655	.136135

Table 6 (continued)

Model	Soil analysis number	Soil code[a]	Means of soil variables (ppm)				Optimal amts. of applied fert. (kg/ha)			Optimal yield (kg/ha)	Profits	Prob. of soil analysis
			N_S	P_S	K_S	pH	N_A	P_A	K_A	Y	soles/ha.	
Group 5	104	1143	1.9	2.0	267.0	8.1	350	350	0	48.3	89,746	.023638
	105	2133	3.2	2.4	183.9	8.5	350	350	350	40.2	71,655	.012694
	106	2143	2.9	2.9	377.3	8.4	350	350	0	41.9	77,072	.025607
	107	3123	5.6	3.8	96.0	8.1	350	350	350	29.1	49,314	.017728
	108	3133	5.9	4.0	144.1	8.1	0	350	350	32.3	59,039	.007223
	109	3223	6.1	5.0	85.0	8.0	0	350	350	34.2	62,831	.003940
	110	3232	5.1	5.7	202.3	6.3	350	350	350	19.7	30,622	.017072
	111	3242	5.3	7.0	242.0	6.2	350	350	140	17.6	27,691	.005252
	112	Average	3.9	3.6	230.7	7.9	350	350	142	39.1	70,651	.113154

[a]Each of the digits in the numbers for the soil code refers to a soil class associated with the N_S, P_S, K_S, and pH variables, respectively. For example, 2341 refers to the 2.1-4.0 interval of N_S, 8.1-15.0 interval of P_S, > 230 interval of K_S and $\leq$ 5.5 interval of pH.

Π_{Aj} = profits obtained from using the fertilizer recommendations from each of the specifications of the aggregate model for the jth soil class and the ith group model.

An application of the above formula resulted in an expected loss of 4754 soles/hectare for the FC model, 5908 soles/hectare for the GLS model, and 8281 soles/hectare for the RC model. The FC model gave the least expected loss of all the three specifications and hence was chosen as the 'best' specification for generating optimum fertilizer recommendations to farmers for those soils not known to belong to any particular group.

Sensitivity of Optimum Fertilizer Recommendations to Changes in the Soil Variables

In the previous section, we used the cost of the wrong decision analysis and the distribution of experimental soils to obtain a choice of a specification of the aggregate model. Here we will consider in detail how the fertilizer recommendations change when the level of N_s, P_s, K_s, and pH change. This would enable us to obtain some idea of the necessity of soil testing before making fertilizer recommendations.

Two sets of soil data are used and fertilizer recommendations for all soils within each data set are generated. For one set of data, we have the knowledge of the groups to which all the soil values belong and for the other we have no such knowledge. For the soils used in gathering the experimental data, we have the information regarding the group to which they belong. Hence we can derive fertilizer recommendations using the estimates of the coefficients for the appropriate group models

from Table 3. For rainfall and temperature, the averages for each of the groups was used. The steepest ascent method with constraints as described earlier is used to generate optimum fertilizer levels. The detailed results of the optimum levels of N_A, P_A, and K_A and the corresponding yields for the experimental soils using each of the estimated group models are given in Table 6. Apart from the soils in the first group, the fertilizer recommendations are sensitive to the various combinations of soil test values. As expected, the optimum yields also vary substantially between the different groups and the different soils. The optimum yields are about 1 1/2 to 2 1/2 times the average yields for the experimental data for all the groups.

The second set of soils consists of a sample of 1541 Sierra soils which were submitted to La Molina for soil testing for potatoes. This set of soils was classified into the 144 soil classes described earlier. For these soils we have no information regarding the group to which they belong; hence, the FC model is used to generate fertilizer recommendations.

The results of the optimum levels of N_A, P_A, and K_A for the La Molina soils using the FC model are given in Table 7. A total of 125 classes were considered.[31] The recommendations for an average soil with average weather were 350, 310, and 325 kg/ha for N_A, P_A, and K_A, respectively, an approximate ratio of 10:9:9. Using similar

---

[31]Out of the 144 soil classes, 8 with a total probability of 0.05 had very low fertilizer recommendations. These soils had high P_S (> 30) and high K_S (> 350) or low N_S (< 1.5) which were beyond the range for the experimental data used in estimating the models under consideration. Eleven of the soil classes had zero probabilities; hence, these nineteen soil classes were omitted from subsequent analysis.

Table 7. Optimal fertilizer recommendations, yields and profits derived from using the fixed coefficients model for the different soil classes

Soil analysis no.	Soil code[a]	Means of soil variables (p.p.m.)				Optimal amts. of applied fert. (kg./ha.)			Optimal yield kg./ha.	Profits soles/ha.	Prob. of soil analysis
		N_S	P_S	K_S	pH	N_A	P_A	K_A	Y		
1	1111	1.0	4.0	55.0	4.5	350	350	350	37.2	65,534	.001298
2	1121	1.5	2.0	98.0	4.8	350	350	350	43.0	77,261	.000649
3	1131	1.8	3.0	142.0	5.2	350	350	344	42.7	76,599	.001947
4	1211	2.0	8.0	40.0	4.5	350	321	350	32.2	55,801	.000649
5	1231	1.3	5.0	170.0	5.0	350	350	334	37.7	66,719	.001298
6	1421	1.5	30.0	100.0	5.3	350	104	334	17.2	28,152	.000649
7	1431	2.0	32.0	170.0	5.4	350	81	306	18.6	31,404	.000649
8	2111	3.3	2.6	47.6	4.8	350	350	350	43.9	79,047	.004543
9	2121	3.0	2.6	93.6	4.7	350	350	350	43.4	77,983	.006489
10	2131	3.3	2.5	170.7	5.1	350	350	333	44.9	81,063	.012980
11	2141	2.5	2.0	286.3	5.1	350	350	290	45.2	81,861	.001947
12	2211	3.1	6.2	53.0	4.8	350	343	350	38.3	67,887	.006489
13	2221	3.6	6.3	101.8	5.1	350	350	350	40.0	71,131	.002596
14	2231	3.4	5.8	152.3	5.0	350	350	341	39.5	70,288	.003894
15	2241	3.6	5.0	350.0	5.2	350	350	266	41.1	73,839	.001298
16	2321	3.2	11.0	92.7	4.5	350	300	350	30.0	51,746	.001947
17	2331	2.7	9.3	197.3	5.0	350	320	314	33.5	58,779	.001947
18	2341	3.1	13.0	240.0	4.6	350	275	293	27.5	47,323	.000649
19	2411	3.3	18.0	50.0	5.4	350	228	350	24.5	41,387	.000649
20	2431	2.3	21.0	170.0	5.4	350	199	315	21.0	35,015	.000649
21	2441	3.2	33.0	234.0	5.1	178	23	122	18.1	33,573	.000649
22	3111	5.7	2.0	47.5	5.1	350	350	350	49.2	89,682	.007788
23	3121	6.5	2.1	95.2	5.0	350	350	350	49.8	90,891	.026609
24	3131	6.6	2.3	159.1	4.8	350	350	338	49.5	90,248	.026609
25	3141	5.9	2.3	393.3	5.0	350	350	250	48.3	88,362	.011033
26	3211	6.8	6.0	41.0	4.6	350	342	350	41.5	74,261	.003245
27	3221	6.9	6.5	104.1	4.9	350	341	350	41.0	73,283	.009086

Table 7 (continued)

Soil analysis no.	Soil code[a]	Means of soil variables (p.p.m.)				Optimal amts. of applied fert. (kg./ha.)			Optimal yield kg./ha. Y	Profits soles/ha.	Prob. of soil analysis
		N_S	P_S	K_S	pH	N_A	P_A	K_A			
28	3231	6.3	6.6	172.9	4.8	350	339	325	40.4	72,153	.012980
29	3241	6.0	6.2	362.5	4.9	350	343	254	41.1	74,121	.007139
30	3311	7.2	11.3	58.3	4.6	350	293	350	34.1	60,037	.001947
31	3321	6.1	12.1	88.9	4.5	350	288	350	32.0	55,737	.005192
32	3331	8.5	11.3	170.0	4.8	350	298	322	35.3	62,485	.004543
33	3341	5.2	10.5	352.2	4.8	350	292	253	32.6	57,617	.003894
34	1431	6.8	37.1	172.3	5.0	350	11	304	25.7	46,331	.013629
35	1112	1.9	2.4	42.0	6.1	350	350	350	44.9	81,055	.003245
36	1122	1.5	3.5	82.5	6.0	350	350	350	40.8	72,722	.001298
37	1132	1.7	2.4	182.4	5.9	350	350	329	44.9	81,164	.005192
38	1142	2.0	4.0	300.0	6.1	350	350	285	41.2	73,967	.000649
39	1212	1.8	6.0	50.0	6.3	350	350	350	37.4	66,013	.001298
40	1222	1.5	6.4	89.6	6.0	350	350	350	37.3	65,792	.003245
41	1232	1.7	6.4	165.9	6.3	350	350	335	37.5	66,229	.004543
42	1242	1.7	8.0	444.0	6.1	350	327	222	32.7	57,504	.000649
43	1322	1.7	11.0	88.2	6.0	350	304	350	29.9	51,529	.007139
44	1332	1.2	11.0	175.5	6.5	350	306	320	28.5	48,821	.001298
45	1342	1.4	12.5	480.0	6.2	350	276	203	24.7	42,265	.001298
46	1412	1.7	17.0	43.3	5.9	350	240	350	23.0	38,236	.001947
47	1422	1.7	22.3	107.5	5.8	350	192	337	19.6	32,185	.002596
48	1432	1.5	35.0	125.0	5.5	152	7	138	16.3	30,337	.000649
49	2112	3.2	2.8	45.6	6.0	350	350	350	45.2	81,516	.005192
50	2122	3.0	2.5	89.4	6.0	350	350	350	47.1	81,347	.014278
51	2132	2.7	2.1	187.4	6.0	350	350	327	46.5	84,409	.009735
52	2142	2.9	2.9	255.7	5.8	350	350	301	44.8	81,029	.004543
53	2212	2.8	6.0	35.0	6.5	350	350	350	38.6	68,435	.001298
54	2222	2.9	6.5	78.5	6.2	350	343	350	37.7	66,745	.002596
55	2232	2.9	6.3	152.9	6.1	350	350	340	39.6	70,504	.009735
56	2242	3.4	6.4	298.0	6.0	350	350	286	39.9	71,401	.007139

Table 7 (continued)

Soil analysis no.	Soil code[a]	Means of soil variables (p.p.m.)				Optimal amts. of applied fert. (kg./ha.)			Optimal yield (kg./ha.)	Profits soles/ha.	Prob. of soil analysis
		N_S	P_S	K_S	pH	N_A	P_A	K_A	Y		
57	2312	3.1	10.0	55.0	6.0	350	307	350	33.4	58,431	.001947
58	2322	3.3	11.7	96.4	6.2	350	297	350	31.0	53,770	.007139
59	2332	3.1	11.9	181.0	6.2	350	294	317	30.7	53,399	.007788
60	2342	3.3	10.7	401.3	6.1	350	296	235	31.2	54,864	.005841
61	2412	2.3	16.0	50.0	6.0	350	250	350	24.9	42,072	.000649
62	2422	2.8	22.6	95.0	6.1	350	183	342	20.8	34,613	.003245
63	2432	3.0	28.3	178.3	5.9	350	126	307	20.0	33,764	.005841
64	2442	2.8	33.3	358.3	5.8	0	0	0	20.7	41,429	.001947
65	3112	5.5	1.4	48.6	6.1	350	350	350	51.5	94,228	.004543
66	3122	6.1	2.6	100.8	5.8	350	350	350	48.5	88,219	.012980
67	3132	5.9	2.5	177.2	5.8	350	350	331	48.4	88,057	.012980
68	3142	5.6	1.9	329.9	5.9	350	350	273	49.4	90,349	.005841
69	3212	6.5	6.5	57.5	5.9	350	337	350	41.5	74,422	.001298
70	3222	5.0	6.0	97.8	5.9	350	350	350	42.2	75,690	.005192
71	3232	6.7	6.0	175.9	6.0	350	350	331	43.5	78,311	.012980
72	3242	5.8	6.1	268.4	5.9	350	350	297	42.7	76,842	.004543
73	3322	6.3	10.4	104.0	5.9	350	315	348	37.0	65,539	.003245
74	3332	6.2	11.8	158.8	6.0	350	293	325	34.2	60,223	.011033
75	3342	6.1	11.8	345.0	6.0	350	287	255	33.3	59,070	.007788
76	3412	6.7	19.0	60.0	6.4	350	223	350	26.7	45,822	.000649
77	3432	6.7	26.7	170.4	5.9	350	138	311	23.8	41,277	.013629
78	1113	1.6	2.2	46.9	8.1	350	350	350	37.6	66,470	.011682
79	1123	1.7	1.9	82.3	8.2	350	350	350	37.4	65,893	.015576
80	1133	1.7	2.7	167.4	8.0	350	350	335	36.8	64,834	.010384
81	1143	1.3	3.5	302.8	7.9	350	350	284	34.3	60,247	.003245
82	1213	1.8	6.1	52.8	6.9	350	350	350	36.2	63,595	.005841
83	1223	1.8	5.9	104.4	8.0	350	350	350	31.6	54,408	.005192
84	1233	1.7	6.8	152.3	7.3	350	350	340	33.2	57,739	.010384

Table 7 (continued)

Soil analysis no.	Soil code[a]	Means of soil variables (p.p.m.)				Optimal amts. of applied fert. (kg./ha.)			Optimal yield (kg./ha.)	Profits	Prob. of soil analysis
		N_S	P_S	K_S	pH	N_A	P_A	K_A	Y	soles/ha.	
85	1243	1.7	5.0	348.3	7.3	350	350	267	35.8	63,299	.001947
86	1313	1.5	15.0	50.0	7.4	350	264	350	21.1	34,262	.001298
87	1323	1.7	10.3	101.0	7.5	350	322	350	27.7	46,862	.006489
88	1333	1.5	10.5	162.8	7.7	350	311	325	25.0	41,696	.007788
89	1343	1.8	13.3	280.0	8.7	350	291	279	16.1	24,441	.001947
90	1413	1.8	25.0	40.0	7.5	350	165	350	14.4	21,960	.001298
91	1423	1.0	17.0	65.0	6.7	350	245	350	20.8	33,831	.000649
92	1433	2.0	18.0	170.0	7.3	350	237	317	19.9	32,354	.000649
93	1443	1.6	29.6	403.6	7.9	0	24	0	11.3	22,308	.004543
94	2113	3.1	2.4	44.9	8.0	350	350	350	40.5	72,188	.033099
95	2123	3.1	2.8	91.7	7.7	350	350	350	40.5	72,164	.046728
96	2133	3.1	2.9	167.7	7.7	350	350	335	40.5	72,352	.024662
97	2143	3.0	2.7	315.5	7.7	350	350	279	39.8	71,261	.008437
98	2213	3.0	5.9	49.8	7.7	350	350	350	34.8	60,878	.017523
99	2223	2.8	6.3	91.7	7.8	350	350	350	34.2	59,650	.029854
100	2233	2.9	6.5	162.6	8.0	350	350	337	31.6	54,541	.021417
101	2243	3.2	6.8	288.6	7.5	350	346	282	34.2	60,042	.009086
102	2313	2.8	11.4	51.8	7.5	350	303	350	27.8	47,211	.009086
103	2323	3.2	11.1	93.9	7.6	350	312	350	27.6	46,711	.020119
104	2333	3.2	11.4	168.8	7.5	350	310	324	28.5	48,851	.014278
105	2343	2.9	10.3	297.5	8.2	350	318	276	25.2	42,399	.007788
106	2413	2.6	24.8	52.5	6.9	350	164	350	18.1	29,262	.002596
107	2423	3.2	23.9	98.0	7.3	350	177	340	18.0	29,032	.016225
108	2433	3.1	31.6	176.1	7.5	350	89	305	16.4	26,973	.009735
109	2443	2.9	32.8	459.2	7.9	0	0	0	15.3	30,605	.014278
110	3113	5.6	2.8	47.7	7.4	350	350	350	44.3	79,846	.015576
111	3123	6.2	2.8	94.6	7.4	350	350	350	44.9	81,069	.023364
112	3133	5.2	2.7	169.7	7.4	350	350	334	44.3	79,799	.017523

Table 7 (continued)

Soil analysis no.	Soil code[a]	Means of soil variables (p.p.m.)				Optimal amts. of applied fert. (kg./ha.)			Optimal yield (kg./ha.)	Profits	Prob. of soil analysis
		N_S	P_S	K_S	pH	N_A	P_A	K_A	Y	soles/ha.	
113	3143	5.3	3.5	306.9	7.2	350	350	282	42.8	77,231	.009086
114	3213	5.3	6.2	45.8	7.1	350	350	350	40.0	71,113	.003894
115	3223	5.8	6.2	96.0	7.6	350	350	350	38.4	68,039	.018172
116	3233	6.0	6.3	159.8	7.8	350	350	337	37.6	66,365	.013629
117	3243	6.3	6.2	393.7	7.4	350	350	250	38.5	68,863	.005841
118	3313	4.2	10.3	53.8	7.3	350	312	350	31.8	55,211	.002596
119	3323	5.8	10.7	90.9	7.4	350	309	350	31.6	54,819	.007139
120	3333	5.5	11.4	168.5	7.5	350	309	323	30.9	53,670	.006489
121	3343	5.2	12.6	393.9	7.6	350	281	236	26.5	45,592	.004543
122	3413	9.1	40.0	60.0	7.5	350	4	346	24.1	43,055	.000649
123	3423	5.3	24.7	94.3	7.2	350	167	342	20.3	33,799	.009735
124	3433	5.6	30.0	172.2	7.6	350	112	309	18.2	30,357	.012980
125	Average	4.2	10.8	158.0	6.7	350	310	325	31.0	53,724	-

[a]See footnote a, Table 6, for an explanation of the soil codes.

experimental data for 1960-1964 years, McCollum and Valverde S. (1968) obtained a ratio of 10:10:6 with lower levels of 272, 289 and 161 kg/ha for N_A, P_A, and K_A, respectively. One reason for the difference is that the limits for the McCollum and Valverde S. data were 160 kg/ha for all the three applied nutrients; whereas in the data used here, it was approximately 350 kg/ha.

Using an FC model, Ryan's (1972) recommendations for the three nutrients for similar data were 251, 250, and 126 kg/hg, respectively. The current (1971) recommendations of the National Center of Soil Analysis at La Molina, Peru, are about 130, 40, and 105 kg/ha for N_A, P_A and K_A, respectively, for an average soil. These are quite below the above mentioned recommendations. Unlike the present study and that of Ryan (1972), the La Molina Laboratory does not have the analytical procedures to address themselves to a simultaneous consideration of all the factors in crop response. A more detailed comparison of the recommendations generated in this study with the recommendations of Ryan and the La Molina Laboratory is presented in the next section.

Nitrogen levels recommended using the fixed coefficients model generally resulted in about the maximum at 350 kgm/ha. Organic matter only slightly influenced the recommendation of applied nitrogen. However, very high levels of P_s and K_s resulted in very low recommendations of N_A. Phosphorus recommendations depend strongly on the level of soil P_s. For soils with a P_s of less than 7 p.p.m., the maximum at 350 kgm/ha is recommended. For soils with P_s greater than 7 p.p.m., the recommendations fall gradually as P_s rises so that for a soil P_s of about 25-30 p.p.m., the recommendation is about 100-120 kgm/ha. Higher pH

resulted in a slightly higher P_A and higher K_s in a slightly lower P_A recommendation. Potassium recommendations are inversely related to the level of soil K_s and to a lesser extent on the level of soil P_s and pH. With a soil K_s of about 350 p.p.m., the recommended K_A is about 250 kg/ha. Higher soil P_s results in a slightly lower recommendation for K_A.

It may be noted that we have not considered any risk aspects of fertilizer decisions. A consideration of these aspects would undoubtedly change the optimum recommendations, depending on the utility functions of the farmers.[32] However, this is beyond the scope of this study.

Value of Soil Testing and Buol's Soil Classification System

The aim in this section is threefold: to compute the value of soil testing using each of the group models and the FC model, to compute the value of using the Buol's classification system, and to compare the recommendations of this study from the FC model to those of Ryan and La Molina Laboratory. The same underlying procedure of the "cost of wrong decision" analysis as used by Havlicek and Seagraves (1962) is followed here.

The value of soil testing for each of the group models is derived as follows. First, a general recommendation for fertilizers is derived for an average soil belonging to the group under consideration. The expected profits derived from using this recommendation regardless

---

[32]See Ryan (1972) for a good review of literature on studies related to these problems.

of soil test results are compared with the expected profits from using the group model to generate recommendations based on soil tests. The formula used in calculating the value of soil testing for a group is given by

$$E(G_i) = \sum_{j=1}^{s_i} [(\Pi_{ij} - \Pi_{iA}) P_{ij}] \tag{4.2}$$

where $E(G_i)$ = value of soil testing for the ith group, $i = 1, 2, \ldots, 5$

s_i = the number of soil classes in group i

Π_{ij} = profits obtained from using the fertilizer recommendations for the jth soil class as predicted by the ith group model from Table 6

Π_{iA} = profits obtained from using the fertilizer recommendations for an average soil in the ith group as predicted by the ith group model

P_{ij} = probability of soil analysis for the jth soil class in the ith group, according to the frequency distributions of soils in the experiments.

The results of applying the above formula to the group models are given in Table 8. It can be seen that the value of soil testing is quite low at 443 soles/ha for the group 1 soils, and ranges from 1484 to 3185 soles/ha for the other group soils.[33] The value of using the soil test

[33]The soles/hectare can be converted to U. S. $/acre by dividing with a conversion factor of 105.02.

Table 8. Value of soil testing for the group models and the fixed coefficients model, using experimental data

Model	Expected profits using recommendations derived from the group models and the FC model	Expected profits using avg. recom. from group models and the FC model	Value of soil testing
	(soles/ha.)	(soles/ha.)	(soles/ha.)
Group 1	43,955	43,512	443
Group 2	30,935	28,331	2,604
Group 3	89,421	31,937	1,484
Group 4	34,229	32,382	1,847
Group 5	63,817	60,632	3,185
All groups	54,140	52,109	2,031
FC	63,155	55,708	7,447

results for all the groups was found to be 2031 soles/ha, which is about 4 percent of the expected profits obtained in using optimum fertilizer recommendations.

A similar calculation for the FC model using the experimental soils data results in a value of soil testing of 7447 soles/ha. This result leads to the conclusion that the value of soil testing in generating fertilizer recommendations is about 11.8 percent of the expected profits of 63155 soles/ha for the FC model.

Another important question that can be asked is what the gains are in knowing that a soil belongs to a particular group. The value of Buol's classification system can be found by assuming each of the group models to be true, and comparing the expected profits derived from the optimum recommendations from the group models to the expected profits from the recommendations generated by the FC model. The formula used in this calculation is given by

$$E(\beta_i) = \sum_{j=1}^{s_i} (\Pi_{ij} - \Pi_{Fj}) P_{ij} \tag{4.3}$$

where $E(\beta_i)$ = value of knowing that a soil comes from group i

s_i = number of soil classes in group i

Π_{ij} = profits obtained from using the fertilizer recommendations from the ith group model for the jth soil class and the ith group model

Π_{Fj} = profits obtained from using the fertilizer recommendations from the fixed coefficient model for the jth soil class and the ith group model

P_{ij} = probability of soil analysis for the jth soil class in the ith group, according to the frequency distribution of experimental soils.

The results of this calculation are given in Table 9. The value of knowing that a soil comes from group i varies from 1467 soles/ha for group 5 to 9915 soles/ha for group 1. Combining all the groups, the value of Buol's classification system is about 4754 soles/ha. This is about two times the value of soil testing for all the groups and about 9 percent of the expected profits of 54140 soles/ha using the group models. Thus, the value of knowing to which group a given soil belongs can be substantial.

Using the above analysis, the recommendations generated in this study are also compared to recommendations generated by Ryan (1972) and La Molina Soil Laboratory. Assuming that the FC model is the true model, the expected value of the recommendations from the FC model over the recommendations for the model used by Ryan (1972) is given by

$$E(FC) = \sum_{j=1}^{125} (\Pi_j - \Pi_{Rj}) P_j \tag{4.4}$$

where $E(FC)$ = value of FC model compared to the model used by Ryan

Π_j = profits from recommendations for the jth soil class using the FC model

Π_{Rj} = profits from Ryan's recommendations for the jth soil class using the FC model

P_j = probability of the jth soil class according to the frequency distribution of the La Molina soil samples.

Table 9. Value of Buol's soil classification system

Model	Expected profits using recommendations derived from the group models	Expected profits using recommendations derived from the FC model	Value of Buol's classification system
	(soles/ha.)	(soles/ha.)	(soles/ha.)
Group 1	43,955	34,040	9,915
Group 2	30,935	24,638	6,297
Group 3	89,421	85,461	3,960
Group 4	34,229	31,443	2,786
Group 5	63,817	62,350	1,467
All groups	54,140	49,386	4,754

The results of this calculation showed that the value of the FC model in comparison with the model used by Ryan is 18,034 soles/ha, or 27.9 percent of the expected profits of 64,621 soles/ha for the FC model.

A similar calculation used in comparing the recommendations from the FC model over the La Molina recommendations resulted in a value of 24,258 soles/ha, or 37.5 percent of the expected profits for the FC model. Thus, the cost of choosing the Ryan model or the La Molina model is considerable if the FC model is correct.

Summarizing this section, it was found that the value of soil testing for the different group models and the FC model ranged from 443 to 7447 soles/ha. The value of Buol's classification for each of the groups ranged from 1467 to 9915 soles/ha and was on the average about two times the value of soil testing. Thus, knowing the group to which a soil belongs is of more value than knowing the soil test results. Assuming the FC model to be true, the cost of using the recommendations given by Ryan and the La Molina Laboratory was substantial and ranged from 27.9 to 37.5 percent of the expected profits for the FC model.

CHAPTER 5
SUMMARY AND CONCLUSIONS

This study has been devoted to generating optimal fertilizer recommendations to individual potato farmers by estimating generalized crop-fertilizer production functions using potato response data from the Sierra region of Peru. Models at two levels of aggregation were considered, the group models and the aggregate models. The group models were based on pooling data from the locations into five groups on the basis of Buol's soil management classification system. The aggregate models were based on pooling data from all the experimental observations. The group models were assumed to represent the true structure. The basic assumptions underlying the structural relationship were that each group has a different coefficient vector and a different error variance. The functional form for all the models was assumed to be a generalized quadratic response relation, similar to the one used by Ryan (1972). Given the knowledge regarding the group to which any soil belongs, the group models are used to generate optimum fertilizer recommendations for that soil. On the other hand, if we do not know the group to which any given soil belongs, then an aggregate model has to be used to generate fertilizer recommendations for that soil.

In this work, three specifications of the aggregated models, the fixed coefficients (FC) model, the generalized least squares model (GLS), and the random coefficients (RC) model were considered. The assumptions and the estimation procedures for the coefficients of each of the specifications of the aggregate model were considered in detail. The relationship between the different specifications of the aggregate

model and the group models were also described. The criterion used in describing this relationship is that of the aggregate model as estimating the average of the group models. In aggregation literature (Theil, 1954), an average of the group models is the best single relation that represents the true group models for prediction purposes. A number of statistical tests were conducted to test some of the assumptions of the different specifications of the aggregate model. Bartlett's test for homogeneity of error variances between the groups resulted in a rejection of the hypothesis that they are equal. If the error variances are not equal, it was noted that the least squares estimators of the coefficients of the FC model are inefficient. An F test resulted in a rejection of the hypothesis that the coefficients of the different groups are equal. If these coefficients are not equal, it was shown that the estimators of the coefficients of the FC model are biased estimators of the average of the coefficients of the group models. Zellner's F test also led to a rejection of the hypothesis that the coefficients of the different group models are equal assuming different error variances. Thus, we accepted the hypothesis that the estimators of the coefficients of the GLS model are also biased estimators of the average of the coefficients of the group models. The RC model was shown to provide a consistent and efficient estimate of the coefficients of the average function. However, five of the diagonal elements of the estimated matrix of the coefficients of the RC model were negative, and as explained in the third chapter, this result is inconsistent with the assumptions of the RC model. From these results, we can conclude that none of the specifications of the aggregate model

are likely to provide unbiased estimates of the average of the group models' coefficients.

F tests of the equality of the coefficient vectors between all the groups in combination of two groups at a time also resulted in a rejection of the hypothesis that the coefficients between any two groups are the same. This leads to the conclusion that Buol's criteria does classify soils into groups with significantly different response coefficients.

The "cost of the wrong decision" criterion was used as a basis for the choice of a specification of the aggregate model. This criterion involves calculating the reduction in net revenues obtainable from the different specifications of the aggregate model recommendations, assuming the group models are in fact the true response curves.

Analytical solutions for the profit maximizing fertilizer levels could not be obtained in every situation. In some cases the point estimate of the optimum fertilizer levels was considerably in excess of the maximum experimental application. In other cases, the stationary point was a minimum or a saddlepoint. To overcome these problems, the profit maximizing levels were obtained using a steepest ascent search technique with lower and upper constraints of 0 and 350 kgs/hectare, respectively.

Fertilizer recommendations for different soil classes using all the models were generated. Assuming the group models to be true, the expected losses from using the different specifications of the aggregated model were calculated. For the FC model, the expected loss was 4754 soles/ha (U. S. $45.28/acre), for the GLS model it was

5908 soles/ha (U. S. $56.26/acre), and for the RC model it was 8281 soles/ha (U. S. $78.85/acre).

The fixed coefficients model was found to have the least expected loss of the three specifications and was thus chosen as the 'best' model for generating fertilizer recommendations for those soils for which we have no knowledge of the groups to which they belong.

Fertilizer recommendations and profits for the different soil classes were obtained for two cases. (1) For a soil known to belong to any given group, the experimental soils data and the group models were used to generate fertilizer recommendations. (2) For a soil not known to belong to any given group, the La Molina soils data and the FC model were used to generate fertilizer recommendations. The results show that it is important to take into account simultaneously all the factors in the response relation. The value of soil testing for the different group models and the FC model ranged from 443 soles/ha (U. S. $4.22/acre) to 7447 soles/ha (U. S. $70.91/acre). The value of Buol's classification for each of the groups ranged from 1467 to 9915 soles/ha (U. S. $13.97 to $94.43/acre) and was more than two times the value of soil testing. Assuming the FC model to be true, the cost of using the recommendations given by Ryan and the La Molina Laboratory was substantial and ranged from 27.9 to 37.5 percent of the expected profits of 64,621 soles/ha (U. S. $615.32/acre) for the FC model.

LIST OF REFERENCES

Allen, R. G. D. 1956. Mathematical Economics. Macmillan and Company, London, England.

Anderson, R. L. 1956. A comparison of discrete and continuous models in agricultural production analysis, pp. 39-61. In E. L. Baum (ed.), Economic Analysis of Fertilizer Use Data. Iowa State College Press, Ames.

Anderson, R. L., D. M. Allen and F. B. Cady. 1970. Selection of predictor variables in linear multiple regression. Technical Report No. 5, Department of Statistics, University of Kentucky, Lexington.

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

Boot, J. C. H. and A. M. de Wit. 1960. Investment demand: An empirical contribution to the aggregation problem. International Economic Review 1:1-30.

Box, M. J., D. Davies and W. H. Swann. 1969. Non-linear Optimization Techniques. Monograph No. 5. Oliver and Boyd, London, England.

Brown, W. S. and M. M. Oveson. 1958. Production functions from data over a series of years. Journal of Farm Economics 40:451-457.

Buol, S. W. 1972. Fertility capability soil classification system, pp. 15-18. In Annual Report, Agronomic-Economic Research on Tropical Soils, North Carolina State University, Raleigh.

Dillon, J. L. 1968. The Analysis of Response in Crop and Livestock Production. Permagon Press, London, England.

Draper, N. R. and H. Smith. 1966. Applied Regression Analysis. John Wiley and Sons, Inc., New York.

Goldberger, A. S. 1964. Econometric Theory. John Wiley and Sons, Inc., New York.

Griliches, Z. 1957. Specification bias in the estimates of production functions. Journal of Farm Economics 39:8-20.

Havlicek, J., Jr. and J. A. Seagraves. 1962. The "cost of wrong decision" as a guide in production research. Journal of Farm Economics 44:157-168.

Heady, E. O. and J. L. Dillon. 1961. Agricultural Production Functions. Iowa State University Press, Ames.

LIST OF REFERENCES (continued)

Heady, E. O. and J. T. Pesek. 1957. Some methodological considerations in the Iowa-TVA research project on economics of fertilizer use, pp. 144-167. In E. L. Baum (ed.), Fertilizer Innovations and Resource Use. Iowa State College Press, Ames.

Hildreth, C. H. 1957. Possible models for agronomic-economic research, pp. 176-186. In E. L. Baum (ed.), Fertilizer Innovations and Resource Use. Iowa State College Press, Ames.

Hildreth, C. H. and J. Houck. 1968. Some estimators for a linear model with random coefficients. Journal of American Statistical Association 63:584-595.

Hurst, D. C. and D. D. Mason. 1957. Some statistical aspects of the TVA-North Carolina cooperative project on determination of yield responses for corn, pp. 207-216. In E. L. Baum (ed.), Fertilizer Innovations and Resource Use. Iowa State College Press, Ames.

Jensen, D. and J. Pesek. 1959. Generalization of yield equations in two or more variables. Agronomy Journal 51:255-263.

Johnsson, A. B. 1963. Input/output relationship in fertilizers and crops. OECD. Doc. Agric. Food 65, Paris, France.

Klein, L. R. 1953. A Textbook of Econometrics. Row and Peterson, New York.

Laird, R. J. and F. B. Cady. 1969. Combined analysis of yield data from fertilizer experiments. Agronomy Journal 61(6):829-834.

McCollum, R. E. and I. Carlos Valverde S. 1968. The fertilization of potatoes in Peru. I. A summary and interpretation of data from field experiments completed from 1959 through 1964 in the Sierra. North Carolina Agricultural Experiment Station Technical Bulletin No. 185, Raleigh.

National Academy of Science. 1961. Status and methods of research in economic and agronomic aspects of fertilizer response and use. Nat. Res. Council, Nat. Academic. Ser. Pub. 918, Washington, D. C.

Pesek, J. T. 1956. Agronomic problems in securing fertilizer response data desirable for economic analysis, pp. 101-112. In E. L. Baum (ed.), Economic Analysis of Fertilizer Use Data. Iowa State College Press, Ames.

Rao, C. R. 1965a. Linear Statistical Inference and Its Applications. John Wiley and Sons, Inc., New York.

LIST OF REFERENCES (continued)

Rao, C. R. 1965b. The theory of least squares when the parameters are stochastic and its application to the analysis of growth curves. Biometrika 52:447-458.

Roberts, S. M. and H. I. Lyvers. 1961. The gradient method in process control. Ind. Eng. Chem. 53:877-882.

Rust, R. H. and R. T. Odell. 1957. Methods used in evaluating the productivity of some Illinois soils. Soil Science Amer. Proc. 21:171-175.

Ryan, J. G. 1972. A generalized crop-fertilizer production function. Unpublished Ph.D. thesis, Department of Economics, North Carolina State University, Raleigh. University Microfilms, Ann Arbor, Michigan.

Scheffe, H. 1959. The Analysis of Variance. John Wiley and Sons, Inc., New York.

Swamy, P. A. V. B. 1970. Efficient inference in a random coefficient regression model. Econometrica 38(2):311-323.

Theil, H. 1954. Linear aggregation of economic relations. North-Holland Publishing Co , Amsterdam, Netherlands.

Theil, H. 1957. Specification errors and estimation of economic relationships. Review of Intl. Inst. of Stat. 25:41-51.

Thompson, W. A., Jr. 1962. The problem of negative estimates of variance components. Ann. Math. Stat. 33:273-289.

Toro-Vizcarrondo, C. and T. D. Wallace. 1968. A test of the MSE criterion for restrictions in linear regressions. Journal of Amer. Stat. Assoc. 63:558-572.

Walker, W. M. and O. H. Long. 1966. Effect of selected soil fertility parameters on soybean yields. Agronomy Journal 58:403-405.

Wallace, T. D. 1972. Weaker criteria and tests for linear restrictions in regression. Econometrica 40:689-698.

Wallace, T. D. and V. G. Ashar. 1972. Sequential methods in model construction. Review of Econ. and Stat. 54(2):172-178.

Wallace, T. D. and C. Toro-Vizcarrondo. 1969. Tables for the mean square errors test for exact linear restrictions in regression. Jour. Amer. Stat. Assoc. 64:1649-1663.

LIST OF REFERENCES (continued)

Zellner, A. 1962a. An efficient method of estimating seemingly unrelated regressions and tests for aggregation bias. Jour. Amer. Stat. Assoc. 57:348-368.

Zellner, A. 1962b. Estimation of cross-section relations: An analysis of a common specification error. Metroeconomica 14:111-117.

Zellner, A. 1969. On the aggregation problem: A new approach to a troublesome problem, pp. 265-276. In K. Fox (ed.), Economic Models, Estimation and Risk Programming--Essays in Honor of Gerhard Tintner. Springer Verlag, Berlin, Germany.

A P P E N D I C E S

Appendix A. Buol's Soil Management Classification System

Appendix A Table 1. Buol's criteria for the fertility capability soil classification system[a]

Type[b]

L = loamy soils: < 35% clay but not loamy sand or sand
C = clayey soils: > 35% clay
S = sandy soils: loamy sand or sand (USDA)
O = organic soil: > 30% o.m. to a depth of 50 cm or more

Substrata type[c]

L = loamy soils: texture as in type
C = clayey soils: texture as in type
S = sandy soils: texture as in type
R = rock or other hard root restricting layer

Condition modifier[d]

e = (exchange): CEC by Σ bases + unbuffered Al < 3; Σ cations @ pH 7 < 7 meq/100 mg soil; Σ cations + Al + H < 10
v = (vertic): very sticky, plastic clay; COLE > 0.09; > 35% clay and > 50 2:1 expanding clay
x = (x-amorphous clay): pH > 10 in 1N NaF; or positive to field NaF test
a = (Al toxic): > 60% of CEC is Al within 50 cm of surface
h = (acid): 10-60% of CEC is Al within 50 cm of surface
i = (iron): free Fe_2O_3/clay > 0.20
Ca = ($CaCO_3$): free $CaCO_3$ within 50 cm
k = (potassium): < 10% weatherable minerals in 20-2000u fraction within 50 cm
Na = (sodium): > 15% of CEC is Na within 50 cm
s = (salt): > 4 mmhos/cm^2 @ 25° within IM
g = (gley): mottles $\leq$ 2 chroma within 60 cm of surface
d = (dry): ustic or xeric environment; dry > 60 days per year
c = (cat clay): pH in I:I H_2O < 3.5 after drying; Jarosite mottles with hues 2.5 Y or yellower and chromas 6 or more within 50 cm; H_2S odor in field often sufficient

Appendix A Table 1 (continued)

---

[a]Source: Buol (1972).

[b]Texture is average of plow layer or 20 cm depth, whichever is shallower.

[c]Substrata type modifier is used if a textural change or hard root restricting layer such as rock is encountered within 50 cm.

[d]Condition is present in plow layer or 20 cm, whichever is shallower unless otherwise specified.

Appendix B. Simple Statistics of the Variables in the Experimental Data

Appendix B Table 1. Means and standard deviations of the variables from the experimental data for the group and the overall data

Variable	Units	Group 1		Group 2		Group 3	
		mean	S.D.	mean	S.D.	mean	S.D.
N_S	percent o.m.	4.21	1.95	4.73	3.51	3.30	1.81
P_S	p.p.m. P	7.84	7.45	7.54	4.93	11.27	8.16
K_S	p.p.m. K	226.46	164.94	135.48	71.20	179.7	154.32
pH	standard units	6.91	4.08	4.67	0.71	5.77	5.88
N_A	kg./ha.	127.77	52.95	140.27	51.44	144.40	53.44
P_A	kg./ha.	132.48	69.21	149.32	88.41	132.33	86.87
K_A	kg./ha.	153.12	71.13	117.69	60.48	120.94	62.57
R	m.m.	540.89	2.50	562.03	191.15	538.93	64.07
T	C	7.16	0.72	6.02	2.05	6.51	1.13
Y	metric tons/ha.	14.14	6.03	14.57	8.88	23.6	12.04
	No. of observations	314		1538		1578	

Appendix B Table 1 (continued)

Variable	Units	Group 4		Group 5		All	
		mean	S.D.	mean	S.D.	mean	S.D.
N_S	percent o.m.	3.26	1.99	3.89	1.53	3.91	2.60
P_S	p.p.m. P	8.11	8.42	3.55	1.49	8.48	7.12
K_S	p.p.m. K	157.30	50.50	230.73	108.50	170.76	120.67
pH	standard units	7.71	0.51	7.85	7.99	5.98	1.36
N_A	kg./ha.	138.84	55.09	134.89	54.41	140.04	53.25
P_A	kg./ha.	144.92	120.36	208.94	151.23	148.44	103.15
K_A	kg./ha.	121.16	60.34	116.30	74.33	121.56	64.22
R	m.m.	508.46	164.40	677.93	89.38	558.42	142.83
T	C	6.30	0.93	5.51	1.27	6.25	1.53
Y	metric tons/ha.	14.55	7.89	17.24	9.14	17.96	10.70
	No. of observations	622		517		4569	