

C.1

ADDITIVITY OF COMPONENT REGRESSION EQUATIONS
WHEN THE UNDERLYING MODEL IS LINEAR

by

SIMEON SANDARAMU CHIYENDA

B.S.F., The University of British Columbia, 1974
M.S., Iowa State University, 1979

A THESIS SUBMITTED IN PARTIAL FULFILMENT OF
THE REQUIREMENTS FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

in

THE FACULTY OF GRADUATE STUDIES
Department of Forestry

We accept this thesis as conforming
to the required standard

THE UNIVERSITY OF BRITISH COLUMBIA

March 1983

In presenting this thesis in partial fulfilment of the requirements for an advanced degree at the University of British Columbia, I agree that the Library shall make it freely available for reference and study. I further agree that permission for extensive copying of this thesis for scholarly purposes may be granted by the head of my department or by his or her representatives. It is understood that copying or publication of this thesis for financial gain shall not be allowed without my written permission.

Department of Forestry

The University of British Columbia
1956 Main Mall
Vancouver, Canada
V6T 1Y3

Date April 20, 1983

ABSTRACT

This thesis is concerned with the theory of fitting models of the form $y = X\beta + \epsilon$, where some distributional assumptions are made on ϵ . More specifically, suppose that $y_j = X\beta_j + \epsilon_j$ is a model for a component j ($j = 1, 2, \dots, k$) and that one is interested in estimation and inference theory relating to $y_T = \sum_{j=1}^k y_j = X\beta_T + \epsilon_T$.

The theory of estimation and inference relating to the fitting of y_T is considered within the general framework of general linear model theory. The consequence of independence and dependence of the y_j ($j = 1, 2, \dots, k$) for estimation and inference is investigated. It is shown that under the assumption of independence of the y_j , the parameter vector of the total equation can easily be obtained by adding corresponding components of the estimates for the parameters of the component models. Under dependence, however, this additivity property seems to break down. Inference theory under dependence is much less tractable than under independence and depends critically, of course, upon whether y_T is normal or not.

Finally, the theory of additivity is extended to classificatory models encountered in designed experiments. It is shown, however, that additivity does not hold in general in nonlinear models. The problem of additivity does not require new computing subroutines for estimation and inference in general in those cases where it works.

TABLE OF CONTENTS

ABSTRACT	ii
LIST OF TABLES	v
LIST OF FIGURES	v
ACKNOWLEDGEMENT	vi
DEDICATION	viii
Chapter	
I INTRODUCTION	1
II PRELIMINARIES, NOTATION AND PROBLEM DEFINITION	4
2.1 Preliminaries and Notation	4
2.2 Problem Definition	6
III LITERATURE REVIEW	7
IV ADDITIVITY IN THE CASE $r_j \equiv p$	13
4.1 Inferences for Total Model when $\varepsilon_j \sim N(\phi, I\sigma_j^2)$	15
4.2 Inferences for Total Model when $\varepsilon_j \sim N(\phi, V\sigma_j^2)$	19
V ADDITIVITY WHEN $r_j \leq p$ WITH $r_j < p$ FOR SOME j	23
5.1 Estimation when $\varepsilon_j \sim N(\phi, I\sigma_j^2)$	24
5.2 Inference when $\varepsilon_j \sim N(\phi, I\sigma_j^2)$	29
5.3 Estimation and Inference when $\varepsilon_j \sim N(\phi, V\sigma_j^2)$	32
VI A GENERALIZATION OF THE ADDITIVITY PROBLEM	33
VII OTHER ASPECTS OF THE ADDITIVITY PROBLEM	35
7.1 The Case y_j, y_k ($j \neq k$) Dependent	36
7.1.1 Estimation for Total Model under Dependence	39

7.1.2	Inference for Total Model when y_1, y_2, \dots, y_k are JMVN	41
7.1.3	Inference for Total Model when y_1, y_2, \dots, y_k are not JMVN	43
7.2	The Case of X Random or Measured with Error ..	44
7.2.1	The Case of X Random	44
7.2.2	The Case when X is Measured with Error	46
7.3	Other General Complements	48
VIII	SOME EXTENSIONS OF THE THEORY	53
8.1	Extension to Classificatory Models	54
8.2	Extension to Nonlinear Models	56
IX	COMPUTATIONAL CONSIDERATIONS	60
X	SOME ILLUSTRATIVE EXAMPLES	67
10.1	Assessing Multivariate Normality of y_T	67
10.2	Example 1	72
10.3	Example 2	75
10.4	Example 3	76
XI	CONCLUSIONS AND REMARKS	78
REFERENCES	80
APPENDIX I	86
APPENDIX II	90
APPENDIX III	92

7.1.2	Inference for Total Model when y_1, y_2, \dots, y_k are JMVN	41
7.1.3	Inference for Total Model when y_1, y_2, \dots, y_k are not JMVN	43
7.2	The Case of X Random or Measured with Error ..	44
7.2.1	The Case of X Random	44
7.2.2	The Case when X is Measured with Error	46
7.3	Other General Complements	48
VII	SOME EXTENSIONS OF THE THEORY	53
8.1	Extension to Classificatory Models	54
8.2	Extension to Nonlinear Models	56
IX	COMPUTATIONAL CONSIDERATIONS	60
X	SOME ILLUSTRATIVE EXAMPLES	67
10.1	Assessing Multivariate Normality of y_T	67
10.2	Example 1	72
10.3	Example 2	75
10.4	Example 3	76
XI	CONCLUSIONS AND REMARKS	78
REFERENCES	80
APPENDIX I	86
APPENDIX II	90
APPENDIX III	92

LIST OF TABLES

Table

1	Results of Koziol's (1982) test for multivariate normality on three data sets	69
2	Results of Koziol's test for bivariate normality	70
3	Comparison of predicted values and residuals of unrestricted total equation with those of total conditioned equation	94

LIST OF FIGURES

Figure

1	Scatter of Residuals from Total Unrestricted Equation and from Total Conditioned Equation	95
---	---	----

ACKNOWLEDGEMENT

I would like to express my indebtedness to Dr. Antal Kozak, my supervisor, for his advice, encouragement, and for a degree of understanding of a rare order. If it were not for his encouragement and unstinting support, this thesis would hardly have reached this stage. My most sincere gratitude is due to other members of my research committee, namely Drs. Michael Schulzer of the Department of Mathematics, J. H. G. Smith, D. D. Munro, and D. H. Williams, all of the Faculty of Forestry, for providing much-needed guidance and constructive criticism at various stages of this study. I am especially grateful to Dr. Schulzer who gave freely of his time to discuss various aspects of my research. I would like to extend my appreciation to Dr. Stanley Nash, now Professor Emeritus in the Department of Mathematics at the University of British Columbia, for kindly reviewing an earlier draft of this thesis.

My two-year association with Dr. James Zidek of the Department of Mathematics at this university, prior to his sabbatical leave at Imperial College (London), was a truly rewarding experience and I would like to express my admiration for a person so selfless and so endowed with academic vision.

I would like to record my deepest gratitude to the Faculty of Forestry for financial support during the period of my studies and to the Food and Agriculture Organization for a generous fellowship. My gratitude also goes to Professor Lewis K. Mughogho for always having confidence in me and to Professors Chimphamba and Edje for supporting me whenever I

needed support. Finally, I express much gratitude and love to my wife Jane and daughters Lucy and Margaret for putting up with a husband and daddy who was hardly home during the preparation of this thesis. I am indebted to Mrs. Nina Thurston for a typing job truly well done.

There are many others that I should mention, including some very special people. To these and others, I would like to say "I have not forgotten about you."

DEDICATION

I dedicate this thesis to the loving
memory of my mother Katherine who
passed away in February 1982 while I
studied at the University of British
Columbia

CHAPTER I

1.0 INTRODUCTION

The main objective of this thesis is to formalize and extend or generalize results obtained by Kozak (1970) concerning conditions that ensure that predicted values calculated from component regression equations add up to those obtained from a corresponding total equation. Kozak (1970) derived his results within the context of forest biomass prediction using component biomass equations and a corresponding total biomass equation. He cites examples in other areas of forestry and forest research where such a problem arises. A broader view is adopted in this thesis with regard to areas of application of the 'additivity' problem. Since biomass analysis is of interest to scientists in various other disciplines of applied biology than forestry, a formalization and generalization of the additivity problem and its related statistical theory will be of value to a large number of scientists, including those in agriculture and ecology.

To fix ideas with regard to the additivity problem as perceived in this thesis, suppose as in Kozak (1970), that for some tree species weight of bole (Y_1), weight of bark on the bole (Y_2), weight of crown (Y_3 = branches and foliage), and total weight ($Y_T = \sum_{j=1}^3 Y_j$) can each be modelled as some function (in the linear regression sense) of diameter at breast height (X). In this setting, one refers to the equations

expressing each of Y_1 , Y_2 , and Y_3 as functions of X as component equations and that giving Y_T as a function of X as the total equation.

More generally, one can envisage k components of an organism or system characterized by measurable attributes Y_1, Y_2, \dots, Y_k and their sum $Y_{k+1} = \sum_{j=1}^k Y_j$ where Y_1, Y_2, \dots, Y_{k+1} are each related to a common set of p independent variables X_1, X_2, \dots, X_p according to a multiple linear regression model. Using the case (with $k = 3$) described in the preceding paragraph, Kozak (1970) states conditions under which one need only fit the component equations, the total equation being completely determined by adding coefficients of corresponding independent variables in the component equations. Kozak's (1970) results pertain to the situation where each component equation contains all independent variables under consideration.

In the sequel, our objective is essentially four-fold. First, it is intended to demonstrate, in a rather simple way, why Kozak's (1970) conditions hold and to derive explicit expressions for statistics of interest for the total equation from those of the component equations. Secondly, it will be shown that additivity can be assured even when different terms are retained in the component equations (that is, when $r_j \leq p$ with strict inequality for at least one j , where r_j is the number of independent variables in component equation j and p is the number in the total equation). This will be achieved by appropriately correcting the total equation in order to take into account the conditioning that forces some independent variables not to appear in some component equations. Thirdly, it will be shown that the case considered by Kozak (1970), where each component equation has the same number of independent variables as the total equation (that is, $r_j \equiv p$), can be derived as a

special case of a generalization of the conditioning principle mentioned above. Estimation and inference theory will be developed for the above objectives based primarily on the assumption that Y_j and Y_ℓ are independent for each $j \neq \ell$ ($j, \ell = 1, 2, \dots, k$) and an appropriate distributional assumption on the error term corresponding to model j . Because of its relevance when considering many biological phenomena, theory relating to the case where Y_j, Y_ℓ ($j \neq \ell$) are dependent will be considered. Finally, some examples will be worked out to illustrate the application of the theory.

It is important to emphasize here that the main objective of this dissertation is to investigate the additivity problem and its related statistical theory within a fairly general framework. Therefore, the use of certain models incorporating and, probably, excluding particular independent variables in the examples and elsewhere in this thesis should not be construed as suggesting that the equations are best in a predictive sense. More specifically, although the subject of the thesis has a direct bearing upon biomass prediction problems, it is not the objective here to arrive at a best biomass equation in any particular sense. On the other hand, the view is taken that the determination of best equations for prediction purposes is best left to particular applications of the theory to be presented here.

CHAPTER II

2.0 PRELIMINARIES, NOTATION AND PROBLEM DEFINITION

We begin with some definitions and rules of convention regarding notation to be adopted in the sequel.

2.1 Preliminaries and Notation

Since the development in this thesis will be concerned with linear regression models, we first seek to identify this class of models precisely. Accordingly, we define a regression model following Gallant (1971). Let $X \subset \mathbb{R}^m$, $\Omega \subset \mathbb{R}^p$ and n and p be positive integers such that $n > p$. The elements of X and Ω will be denoted by x and β , respectively. Furthermore, we shall let $\{\epsilon_t\}_{t=1}^{\infty}$ be a sequence of random variables, $\{x_t\}_{t=1}^{\infty}$ a sequence from X , $f(x, \beta)$ a real-valued function with argument (x, β) and β_0 to be a point in Ω .

Definition 1. A regression model is defined here to be the sequence of random variables $\{y_t\}_{t=1}^n$ given by

$$y_t = f(x_t, \beta_0) + \epsilon_t.$$

We emphasize that we owe the basic idea behind this definition to Gallant (1971). Our definition is not as rigorous as Gallant's (1971) but it will suffice for our purposes. Note that f is a mapping of points from the product space $X \times \Omega$ into the real line \mathbb{R}^1 (that is, $f : X \times \Omega \rightarrow \mathbb{R}^1$).

Now, suppose we denote the set of all possible regression models generated according to definition 1 by R^* and define a set of regression

models of the form

$$y_t = \phi_0(x_t) + \sum_{j=1}^p \beta_j \phi_j(x_t) + \varepsilon_t$$

where $\phi_i : X \rightarrow R^1$, $i = 0, 1, \dots, p$. Again we owe this formulation to Gallant (1971). Note that in the above, $f(x_t, \beta_0)$ in definition 1 has been replaced by $\phi_0(x_t) + \sum_{i=1}^p \beta_i \phi_i(x_t)$. Gallant (1971) designates the class of regression models whose members are specified according to the last equation by L , which is the class of linear regression models. Thus we have a second definition.

Definition 2. A regression model r^* is called a linear regression model if $r^* \in L$, where L is as defined above.

Note, for completeness, that $L \subset R^*$. This thesis will be concerned with the theory of estimation and inference for members of L under certain conditions. As is well known, any member of L can be written in matrix form.

With regard to notation, a matrix representation will be adopted throughout most of the development here. This has an obvious aesthetic appeal but, more importantly, leads to brevity and a rather compact presentation of results which would otherwise be cumbersome using ordinary scalar arithmetic. Accordingly, let y_j denote an $n \times 1$ vector of realizations of an observable random variable corresponding to the j^{th} component, X an $n \times (p + 1)$ matrix defined so that $X = (X_0 | X_1 | \dots | X_p) = (\vartheta | X_1 | \dots | X_p)$, where $X_0 = \vartheta$ is an $n \times 1$ vector with components identically equal to unity and X_i ($i = 1, 2, \dots, p$) is an $n \times 1$ vector consisting of realizations of the independent variable X_i , β_j is a $(p + 1) \times 1$ parameter vector and ε_j is the corresponding $n \times 1$ vector of errors or disturbances.

For any matrix A , say, we shall write A^t to denote the transpose of A and A^{-1} to denote the inverse of A , provided that the inverse exists. Where A^{-1} does not exist, we shall have occasion to use A^+ , the generalized inverse of A , to obtain more general results. In any case, new notation may be introduced in the discourse as the need arises but in every case our notation will be consistent with that used in standard texts in linear algebra (e.g., Noble & Daniel, 1977; Searle, 1966; Stewart, 1973; Strang, 1976).

2.2 Problem Definition

With the above conventions and notation, we shall be concerned, in this dissertation, with models of the form

$$y_j = X\beta_j + \epsilon_j \quad (j = 1, 2, \dots, k) \quad (2.2.1)$$

and

$$y_T = \sum_{j=1}^k y_j = X\beta_T + \epsilon_T \quad (2.2.2)$$

where (2.2.1) gives us k component models and (2.2.2) the corresponding total model. We shall focus interest in the sequel on characterizing estimators of β_T and on propounding a theory of inference relating to the total model under a number of assumptions concerning the behavior of the ϵ_j ($j = 1, 2, \dots, k$). The theory to be presented here will be based on well-known general linear model theory. Note that specification of the behavior of the ϵ_j leads, in general, to specification of the behavior of ϵ_T .

In the next chapter, a review is made of some of the work in the literature relating to the additivity problem before proceeding to propose a unified theory of estimation and inference for this problem in succeeding chapters.

CHAPTER III

3.0 LITERATURE REVIEW

Theoretical and applied biologists have traditionally viewed biomass as a useful index for assessing the productivity of various flora and fauna with respect to designated environments or ecosystems (Ovington, 1962). This index has also been used for cataloguing, in the form of inventories, the quantities of biological matter available at a given time in a given environment.

A thorough reading of the literature on biomass and related studies indicates development in two main directions. The early part of the literature indicates that scientists essentially sought ways of quantitatively describing biomass production and productivity of various biological organisms. This approach was especially dominant in ecological studies for many years. Quite often, systematic sampling schemes (e.g., line transects) were used to obtain data which were subsequently summarized to give crude estimates of biomass. In many cases, these estimates were reported by component of the organism or system under consideration. -In general, little or no statistical information, such as measures of precision, accompanied the summarizations. In any case, the very nature of the sampling schemes upon which the estimates were based militated against a meaningful statistical interpretation of the results. More recent literature is suggestive of a significant shift from the purely

descriptive approach to model-based, statistically-oriented methods of describing biomass. This approach not only leads naturally to the need to address questions relating to choice of proper model for use in a given situation, but more importantly perhaps, attaches particular importance to choosing estimates that are statistically reasonable. The following review of the literature on developments that led to the additivity problem will be brief and, hopefully, informative rather than exhaustive. For a more complete review, see Chiyenda (1974) or Kurucz (1969). More recent comprehensive reviews are given by Smith (1979) and Smith and Williams (1980).

The forestry literature credits Tufts (1919) with the first reported work on tree component biomass. In that work, Tufts correlated trunk circumference of fruit trees with the weight of their tops (or crowns). Following that work, many workers in forestry engaged in biomass studies of one form or another. Accordingly, considerable work has been reported in the general area of total biomass production of various tree species (Honer, 1971; Kellogg and Keys, 1968; Young and Chase, 1965) and of forest ecosystems (Ovington, 1962). Some of this work was carried out as part of ongoing inventory programmes (e.g., Honer, 1971) while others were conducted in research connected with forest fire-hazard abatement efforts (e.g., Kiil, 1967, 1968; Loomis et al., 1966; Storey et al., 1969).

In a biomass study of 13 North American tree species, Storey et al. (1955) found that dry crown weight, branchwood weight, and foliage weight were significantly related to stem diameter at the base of live crown for all the species. On the other hand, Ovington (1956) investigated and compared the forms, weights, and productivity of tree species grown in close stands. This study was motivated by silvicultural and ecological considerations. In a study similar to that of 1955, Storey and Pong

(1957) investigated and compared crown characteristics of a number of hardwood species.

Fahnestock (1960) used data collected from nine coniferous tree species in the Northern Rocky Mountain area to fit regression equations to predict crown weight and proceeded to construct crown weight tables for the species. Among the species investigated were Douglas-fir (Pseudotsuga menziesii [Mirb.] Franco), western hemlock (Tsuga heterophylla [Raf.] Sarg.) and western red-cedar (Thuja plicata Donn).

Tadaki et al. (1961) investigated the productivity of a young stand of birch (Betula platyphylla) in southern Hokkaido, Japan, and established linear relationships on logarithmic axes between basal area and stem biomass, basal area and branch biomass, and between basal area and foliage biomass. They also reported that estimated fresh and dry foliage weights did not vary with stand density but that branch biomass decreased with stand density.

Brown (1963) investigated the relations between crown weight and diameter in some Lake States red pine (Pinus resinosa Ait.) plantations and also studied the influence of site quality and stand density on the weight of individual tree crowns. Keen (1963) analysed average green weights and centres of gravity of samples of black spruce (Picea mariana [Mill.] B.S.P.), white spruce (Picea glauca [Moench.] Voss.), and balsam fir (Abies balsamiae [L.] Mill.) and investigated their variation with species, season, and location. He also derived a tabulation of weights and centres of gravity of the trees.

Young et al. (1964) used regression equations to construct fresh and dry fibre weight tables for individual tree components, groups of components, and complete trees for seven tree species. Brown (1965)

investigated the effect of site and stand density on the crown size of individual red pine and jack pine (Pinus banksiana Lamb.) trees and studied ways of estimating crown fuel weights. The study indicated that estimated amounts of foliage and branchwood per unit area varied with the age and growing conditions of the stand.

Loomis et al. (1966) used analysis of covariance to test the effect of stand density on dry foliage and branchwood weights in shortleaf pine (Pinus echinata Mill.) and found that regressions of dry foliage and branchwood weights for different stand densities were not significantly different. Dyer (1967) prepared preliminary fresh and dry weight tables for northern white cedar (Thuja occidentalis L.) and derived linear regression equations for predicting fresh and dry wood weights of various tree components as percentages of total tree fresh and dry weights.

Kiil (1967) used regression analysis to construct fuel weight tables for white spruce and lodgepole pine (Pinus contorta Dougl.) in west-central Alberta and found that a combination of diameter at breast-height and either crown width or crown length gave the most precise estimating equation for fuel weight. In a follow-up study, Kiil (1968) studied the fuel complex of 70-year-old lodgepole pine in the same area with a view to facilitating measurement and prediction of weight and size distribution of fuel components.

Kurucz (1969) obtained predictive regression equations for total and component biomass of Douglas-fir, western hemlock, and western red-cedar grown on the University of British Columbia Research Forest near Haney, British Columbia. In a study that was probably motivated as much by Kurucz's (1969) study as by others, Kozak (1970) considers the problem of additivity of component biomass regression equations for

purposes of prediction. The real essence of Kozak's (1970) work does not lie in the uniqueness of the problem he poses but, rather, in the statistical problems that it raises and the potential practical impact that a solution to these problems might have.

Other studies conducted following Kozak's (1970), while essentially underscoring the importance of the biomass estimation problem in forestry and related disciplines, did not address the additivity aspect of the problem directly. Instead, many investigators continued to look for the best set of variables giving the most parsimonious predictive equation for total and component biomass (e.g., Crow, 1971; Honer, 1971; Johnstone, 1971; Muraro, 1971; Zavitkovski, 1971; Sando and Wick, 1972).

Biomass studies and methods of effectively predicting individual tree biomass continued to interest applied quantitative biologists in the mid- and late-seventies and well into the eighties. This interest in biomass is ascribable to the applicability of individual tree biomass information in addressing a wide range of ecological and forest management problems. These include large-scale biomass inventories (Young, 1978; Ker and Van Raalte, 1980), nutrient-cycling problems (Marks and Bormann, 1972; Kimmins, 1977; Kimmins and Krumlik, 1976; Kimmins et al., 1979), as well as the determination of net productivity of forest ecosystems (Whittaker et al., 1974). Many studies, such as Jacobs and Cunia (1980), Jokela et al. (1981), Keyes and Grier (1981), Schmitt and Grigal (1981), Yandle and Wiant (1981), Zavitkovski et al. (1981), Chaturvedi and Singh (1982), Freedman et al. (1982), and Singh (1982), have used regression methods to address the biomass prediction problem.

It is worth mentioning that biomass studies of various descriptions are being conducted to date. Some of these are essentially computer-

based simulations of various aspects of the biomass problem. An example of this is the FORCYTE study being undertaken by Kimmins and his associates at the University of British Columbia (see Kimmins and Scoullar, 1979; Kimmins et al., 1980). As described by its authors, "FORCYTE is an interactive simulation model designed to examine, on a site-specific basis, the long-term effects on nutrient budgets and productivity of various intensive forest management and harvesting practices." Other studies are conducted as part of ongoing national programmes aimed at identifying useful modelling procedures for predicting or, otherwise, describing biomass. An example of this is the study, again at the University of British Columbia, by Smith (1979) and Smith and Williams (1980) originally commissioned by the Canadian Forestry Service to propose the development of a comprehensive forest biomass growth model. That proposal has since been approved and work is currently under way to develop such a model.

Surprisingly, most of the studies cited earlier do not consider the additivity problem except for passing reference to Kozak's additivity result in a few instances (e.g., Ker and Van Raalte, 1980; Singh, 1982). One might surmise that this apparent lack of interest in the additivity problem might be largely due to the fact that additivity has, since its introduction into the forestry literature by Kozak (1970), been restricted to situations in which each component equation contains the same independent variables. This precludes the use of additivity in the more common and important situations where only statistically important independent variables are used in any component equation. An extension of the additivity problem to such situations along with its corresponding statistical theory would obviously be of interest. This is what is intended to be done in succeeding chapters of the discourse. One hopes that studies such as have been

cited above will, in time, benefit from or at least find useful complementary methodology in the theory to be presented in this dissertation.

CHAPTER IV

4.0 ADDITIVITY IN THE CASE $r_j \equiv p$

We consider first the models $y_j = X\beta_j + \epsilon_j$ ($j = 1, 2, \dots, k$) and $y_T = \sum_{j=1}^k y_j = X\beta_T + \epsilon_T$ where it is understood that each of these $k + 1$ models involves the same matrix X . This is the case considered by Kozak (1970).

Consider the estimation of β_T assuming $\epsilon_j \sim (\phi, I\sigma_j^2)$ or $\epsilon_j \sim (\phi, V\sigma_j^2)$ where ϕ is an $n \times 1$ null vector, I is an identity matrix of dimension n and V is a known symmetric positive definite matrix of dimension n . Note that we have not for now specified the form of the distribution function of ϵ_j as this is not necessary to obtain estimates of desirable properties. We restrict attention in this chapter to the situation where X is of full rank.

Under the assumption that $\epsilon_j \sim (\phi, I\sigma_j^2)$, ordinary least squares (OLS) fitting of the k component models yields Gauss-Markoff estimators

$$\hat{\beta}_j = (X^t X)^{-1} X^t y_j \quad (j = 1, 2, \dots, k). \quad (4.0.1)$$

The result given by (4.0.1) is completely basic and warrants no further comment except to note that the resulting $\hat{\beta}_j$ are best linear unbiased estimators (BLUE's) in the sense of the Gauss-Markoff theorem (see Graybill, 1976, p. 219; Kempthorne, 1975, p. 32; Searle, 1971, p. 88).

When $\epsilon_j \sim (\phi, V\sigma_j^2)$, generalized least squares (GLS) fitting applied to each component model leads to Gauss-Markoff estimators

$$\hat{\beta}_j = (X^t V^{-1} X)^{-1} X^t V^{-1} y_j \quad (j = 1, 2, \dots, k) \quad (4.0.2)$$

Note that in (4.0.2) the existence of V^{-1} is guaranteed by the positive definiteness of V . It is worth pointing out that the OLS estimator of β_j is in general different from the GLS estimator, except in the special case where there exists a $(p+1) \times (p+1)$ nonsingular matrix F such that $VX = XF$. This is a very special result and is due to Zyskind (1962). See Graybill (1976), Kempthorne (1975), and Searle (1971) for references to this result.

Now consider the total model. Introducing the expression for y_j given earlier into the expression for the total model, one gets

$$y_T = \sum_{j=1}^k (X\beta_j + \epsilon_j) = X\beta_T + \epsilon_T \quad (4.0.3)$$

or

$$y_T = X \sum_{j=1}^k \beta_j + \sum_{j=1}^k \epsilon_j = X\beta_T + \epsilon_T \quad (4.0.4)$$

from where it is clear that $\beta_T = \sum_{j=1}^k \beta_j$ and $\epsilon_T = \sum_{j=1}^k \epsilon_j$. Hence, the least squares estimator of β_T is given by $\hat{\beta}_T = \sum_{j=1}^k \hat{\beta}_j$, where $\hat{\beta}_j$ is given by (4.0.1) or (4.0.2) according as ordinary least squares or generalized least squares fitting is used to obtain BLUE's. Thus the total equation is completely determined by coefficients of the component equations, as Kozak (1970) pointed out.

Having shown (for $r_j \equiv p$) that the total equation is completely determined by the parameters of the component equations, it might be of interest to investigate whether statistics derived from the component analyses can be utilized to make inferences pertaining to the total equation. It will be shown, in the sequel, that this is the case in general. The results presented below will be useful for testing hypotheses

concerning the total equation and for constructing confidence intervals. In order to simplify the derivations, we shall assume that the random variables y_j and y_ℓ are independent for each $j \neq \ell$ ($j, \ell = 1, 2, \dots, k$). As we point out later, this assumption may be quite unrealistic for the phenomena being modelled and this may considerably affect the utility of the theory to be developed on the basis of this assumption.

4.1 Inferences for Total Model when $\epsilon_j \sim N(\phi, I\sigma_j^2)$

We now address the problem of inference for the total model when the ϵ_j follow a multivariate normal distribution with expectation vector ϕ and covariance matrix $I\sigma_j^2$, where ϕ and I have been defined earlier. Note that we have explicitly specified the form of the distribution of ϵ_j here since such specification is necessary for inference.

Let us begin by supposing that it is desired to discover how well the independent variables in the total equation explain the observed variation in the components of y_T . To answer this question, one needs to determine the amount of variation in the components of y_T that can be attributed jointly to these independent variables. This is the usual sum of squares due to regression.

Denote the uncorrected sum of squares due to fitting the unrestricted total equation by $SS_R(\beta_T)$ and that due to fitting a version of this model restricted so that all components of β_T other than the intercept component are set equal to zero by $SS_R(\beta_{OT})$. It is easy to show that

$$SS_R(\beta_T) = \hat{\beta}_T^t X^t y_T = (\sum_{j=1}^k \hat{\beta}_j^t) X^t (\sum_{j=1}^k y_j) \quad (4.1.1)$$

and

$$SS_R(\beta_{OT}) = \hat{\beta}_{OT} \left(\sum_{i=1}^n y_{iT} \right) = n \bar{y}_T^2 \quad (4.1.2)$$

where y_{iT} is the i^{th} component of y_T and \bar{y}_T is the mean of the components

of y_T . It follows that the desired regression sum of squares for fitting the total equation is given by

$$SS_R(\beta_T^* | \beta_{OT}) = \left(\sum_{j=1}^k \hat{\beta}_j^t \right) X^t \left(\sum_{j=1}^k y_j \right) - n \bar{y}_T^2. \quad (4.1.3)$$

As usual, the corrected total sum of squares for the total equation is

$$SS_T = y_T^t y_T - n \bar{y}_T^2 = \sum_{i=1}^n y_{iT}^2 - n \bar{y}_T^2 \quad (4.1.4)$$

and the error sum of squares is given by

$$SS_E = y_T^t y_T - SS_R(\beta_T) \quad (4.1.5)$$

or as the result of subtracting (4.1.3) from (4.1.4), that is,

$SS_E = SS_T - SS_R(\beta_T^* | \beta_{OT})$. It can be shown easily that $SS_R(\beta_T^* | \beta_{OT})$ is associated with p degrees of freedom while SS_E is associated with $n - p - 1$ degrees of freedom. Furthermore, since SS_T is associated with $n - 1$ degrees of freedom and SS_T is the sum of $SS_R(\beta_T^* | \beta_{OT})$ and SS_E , it follows from Cochran's theorem (see Hogg and Craig, 1970, p.393; Kempthorne, 1975, p. 57; Montgomery, 1976, p. 37) that $SS_R(\beta_T^* | \beta_{OT})$ and SS_E are independent. By our distributional assumption on ϵ_j ($j = 1, 2, \dots, k$) we have that $SS_R(\beta_T^* | \beta_{OT})/\sigma_T^2$ and SS_E/σ_T^2 are, respectively, noncentral chi-square with p degrees of freedom and noncentrality parameter $\beta_T^{*t} X^t X \beta_T^* / 2\sigma_T^2$ and central chi-square with $n - p - 1$ degrees of freedom. We have already noted that they are independent. Whence it follows that a test of significance for the total equation can be obtained using data and estimates relating to the component models without actually fitting the total equation. An \hat{R}^2 associated with the total equation is similarly obtained.

For examining the hypothesis that some component of β_T is equal to zero, one is often interested in constructing confidence intervals about

the component or performing a direct t-test (see Montgomery, 1976, p. 325). In either case, one requires an estimate of the covariance matrix of $\hat{\beta}_T$. Denote the true covariance matrix of $\hat{\beta}_j$ by Φ_j and that of $\hat{\beta}_T$ by Φ_T . Then we have that

$$\Phi_j = (X^t X)^{-1} \sigma_j^2 \quad (4.1.6)$$

and by the fact that $\hat{\beta}_T = \sum_{j=1}^k \hat{\beta}_j$ and that the ϵ_j ($j = 1, 2, \dots, k$) are independent, we have

$$\Phi_T = (X^t X)^{-1} \sum_{j=1}^k \sigma_j^2. \quad (4.1.7)$$

The results given in (4.1.6) and (4.1.7) are completely basic and we shall not venture to prove them here. The estimator of Φ_T will be given by

$$\hat{\Phi}_T = (X^t X)^{-1} \sum_{j=1}^k \hat{\sigma}_j^2 \quad (4.1.8)$$

where $\hat{\sigma}_j^2$ is the mean square error associated with fitting the j^{th} component equation. Hence confidence limits on the relevant component of β_T , say $\beta_{\ell T}$, will be given by

$$\hat{\beta}_{\ell T} \pm t_{\alpha/2, n-p-1} [(\sum_{j=1}^k \hat{\sigma}_j^2) C_{\ell\ell}]^{\frac{1}{2}} \quad (4.1.9)$$

where $t_{\alpha/2, n-p-1}$ is the $(1 - \alpha/2)$ 100-th percentile of the central t-distribution with $n - p - 1$ degrees of freedom and $C_{\ell\ell}$ is the $(\ell + 1)$ st diagonal element of the matrix $(X^t X)^{-1}$. Note that ℓ takes integer values in the range 0 to p inclusive. The corresponding test based on the t-distribution is performed by computing

$$t_0 = \hat{\beta}_{\ell T} / [C_{\ell\ell} \sum_{j=1}^k \hat{\sigma}_j^2]^{\frac{1}{2}} \quad (4.1.10)$$

and rejecting the null hypothesis that $\beta_{\ell T} = 0$ if $|t_0| > t_{\alpha/2, n-p-1}$. Note also that the test in (4.1.10) can be derived as a special case of a more general approach in the context of general linear hypothesis theory, as will be shown in a later part of the discourse.

The inferences based on (4.1.9) and (4.1.10) are valid if $(X^t X)^{-1}$

is a diagonal matrix. In general, however, $(X^t X)^{-1}$ is not diagonal and so results obtained from (4.1.9) and (4.1.10) can be misleading (see Montgomery, 1976, p. 326). This is so because both (4.1.9) and (4.1.10) are based on the assumption that the elements of $\hat{\beta}_T$, such as $\hat{\beta}_{iT}$ and $\hat{\beta}_{jT}$ for $i \neq j$, are independent. When $(X^t X)^{-1}$ is not diagonal, this assumption does not hold in general. Therefore, a test for $\beta_{\ell T} = 0$ versus $\beta_{\ell T} \neq 0$ must be constructed using the 'extra sum of squares' principle used in deriving (4.1.3) or using the general linear hypothesis theory alluded to above. The extra sum of squares principle is described in Draper and Smith (1981, p. 97 cf.) and in Montgomery (1976, pp. 326-328).

Finally, for constructing a confidence interval about a true value, $y_{\ell T}$, corresponding to the x-coordinate, $X_{\ell T} = (1, x_{1T}, \dots, x_{pT})$, where $X_{\ell T}$ is a row vector, one needs the covariance matrix $\hat{\phi}_{\hat{y}_T}$ of \hat{y}_T . Since $\hat{y}_T = X\hat{\beta}_T$, we know from theory (e.g., Morrison, 1976, pp. 83-84) that

$$\hat{\phi}_{\hat{y}_T} = X(\hat{\phi}_T) X^t = X(X^t X)^{-1} X^t \sum_{j=1}^k \sigma_j^2. \quad (4.1.11)$$

Hence the estimator of $\hat{\phi}_{\hat{y}_T}$ is given by

$$\hat{\hat{\phi}}_{\hat{y}_T} = X(X^t X)^{-1} X^t \sum_{j=1}^k \hat{\sigma}_j^2 \quad (4.1.12)$$

and, in particular, $\hat{\hat{\sigma}}_{\hat{y}_{\ell T}}$, the estimated standard error of $\hat{y}_{\ell T}$ corresponding to the x-coordinate $X_{\ell T}$ is

$$\hat{\hat{\sigma}}_{\hat{y}_{\ell T}} = [X_{\ell T}(X^t X)^{-1} X_{\ell T}^t \sum_{j=1}^k \hat{\sigma}_j^2]^{\frac{1}{2}}. \quad (4.1.13)$$

Therefore, the desired confidence interval for $y_{\ell T}$ is given by

$$\hat{y}_{\ell T} \pm [X_{\ell T}(X^t X)^{-1} X_{\ell T}^t \sum_{j=1}^k \hat{\sigma}_j^2]^{\frac{1}{2}} t_{\alpha/2, n-p-1}. \quad (4.1.14)$$

Results given above show that tests concerning specific components of β_T and corresponding confidence intervals can be constructed using

information obtained from analyses relating to the component equations. Thus there is no need to fit the total equation in order to make inferences about it. These observations pertain to the situation where

$\epsilon_j \sim N(\phi, I\sigma_j^2)$. We show below that the same holds when $\epsilon_j \sim N(\phi, V\sigma_j^2)$.

4.2 Inferences for Total Model when $\epsilon_j \sim N(\phi, V\sigma_j^2)$

Consider now the situation where $\epsilon_j \sim N(\phi, V\sigma_j^2)$ with V as defined earlier. It was stated in (4.0.2) that the BLUE for β_j under this distributional assumption is $(X^t V^{-1} X)^{-1} X^t V^{-1} y_j$. One must add here that this estimator is also a maximum likelihood estimator (MLE). It is desirable to motivate the derivation of $(X^t V^{-1} X)^{-1} X^t V^{-1} y_j$, mainly to clear the way for its use in making inferences. Accordingly, consider the model specification $y_j = X\beta_j + \epsilon_j$ with $\epsilon_j \sim N(\phi, V\sigma_j^2)$. Since V is positive definite, there exists an $n \times n$ nonsingular matrix P such that

$$V = P^t P. \quad (4.2.1)$$

Hence we have that

$$(P^t)^{-1} V P^{-1} = I. \quad (4.2.2)$$

Suppose one pre-multiplies the model $y_j = X\beta_j + \epsilon_j$ by $(P^t)^{-1}$; then one has

$$(P^t)^{-1} y_j = (P^t)^{-1} X\beta_j + (P^t)^{-1} \epsilon_j \quad (4.2.3)$$

which may be given equivalently by

$$y_j^* = X^* \beta_j + \epsilon_j^*. \quad (4.2.4)$$

Note that $E y_j^* = X^* \beta_j = (P^t)^{-1} X \beta_j$, since $E \epsilon_j^* = \phi$, and $E(\epsilon_j^* \epsilon_j^{*t}) = I\sigma_j^2$ so that we now have that $\epsilon_j^* \sim N(\phi, I\sigma_j^2)$. Therefore all the theory developed in section 4.1 applies to (4.2.4). Applying OLS to (4.2.4) one obtains

$$\begin{aligned}
\hat{\beta}_j &= (X^{*t} X^*)^{-1} X^{*t} y_j^* \\
&= [X^t (P^t P)^{-1} X]^{-1} X^t (P^t P)^{-1} y_j \\
&= (X^t V^{-1} X)^{-1} X^t V^{-1} y_j,
\end{aligned} \tag{4.2.5}$$

which is the generalized least squares estimator given earlier. Note that this type of fitting (GLS) which produces (4.2.5) is also commonly referred to as weighted least squares fitting and, in more general presentations, as minimum V-norm fitting (see Kempthorne, 1975).

Results useful in making inferences when $\epsilon_j \sim N(\phi, V\sigma_j^2)$ are basically similar to those derived above for $\epsilon_j \sim N(\phi, I\sigma_j^2)$ with a few important distinctions as a result of our transformation of y_j above. In testing for significance of the total equation, for instance, the regression sum of squares is given by

$$SS_R(\beta_T^* | \beta_{OT}) = \left(\sum_{j=1}^k \hat{\beta}_j^t \right) X^{*t} \left(\sum_{j=1}^k y_j^* \right) - n \bar{y}_T^{*2} \tag{4.2.6}$$

where X^* and y_j^* are as defined above and \bar{y}_T^* is the mean of the components of $y_T^* = \sum_{j=1}^k y_j^*$. Furthermore, the corrected total sum of squares for the total equation is

$$SS_T = y_T^{*t} y_T^* - n \bar{y}_T^{*2} \tag{4.2.7}$$

and the error sum of squares is

$$SS_E = y_T^{*t} y_T^* - SS_R(\beta_T), \tag{4.2.8}$$

where $SS_R(\beta_T)$ is given by the first term on the right-handside of (4.2.6).

In these sums of squares, $\hat{\beta}_j$ is as defined in (4.2.5). Again,

$SS_R(\beta_T^* | \beta_{OT})/\sigma_T^2$ and SS_E/σ_T^2 will be distributed as noted earlier except

that the noncentrality parameter associated with the distribution of the former is now $\beta_T^{*t} X^t V^{-1} X \beta_T^* / 2\sigma_T^2$. As before, it is clear that

a test of significance for the total regression equation and the associated \hat{R}^2 are obtainable from data and ancillary quantities derived from fitting

the component equations without recourse to actually fitting the total equation.

In light of (4.2.5) the covariance matrix of $\hat{\beta}_j$ is now given by

$$\hat{\beta}_j = (X^t V^{-1} X) \sigma_j^2 \quad (4.2.9)$$

and that of $\hat{\beta}_T$ by

$$\hat{\beta}_T = (X^t V^{-1} X)^{-1} \sum_{j=1}^k \sigma_j^2 \quad (4.2.10)$$

with $\hat{\beta}_j$ and $\hat{\beta}_T$ obtained respectively by simply replacing σ_j^2 by $\hat{\sigma}_j^2$ in the expressions for β_j and β_T . In this context, a $100(1 - \alpha)\%$ confidence interval for $\beta_{\ell T}$ is given by

$$\hat{\beta}_{\ell T} \pm t_{\alpha/2, n-p-1} [C^*_{\ell\ell} \sum_{j=1}^k \hat{\sigma}_j^2]^{\frac{1}{2}} \quad (4.2.11)$$

while a corresponding t-test is obtained by calculating

$$t^*_{\ell T} = \hat{\beta}_{\ell T} / [C^*_{\ell\ell} \sum_{j=1}^k \hat{\sigma}_j^2]^{\frac{1}{2}} \quad (4.2.12)$$

and rejecting the null hypothesis that $\beta_{\ell T} = 0$ if $|t^*_{\ell T}| > t_{\alpha/2, n-p-1}$.

In (4.2.11) and (4.2.12), $C^*_{\ell\ell}$ refers to the $(\ell + 1)$ st diagonal element of $(X^t V^{-1} X)^{-1}$, with ℓ specified as before. The limitations of the results given by (4.2.11) and (4.2.12) when $(X^t V^{-1} X)^{-1}$ is not diagonal are equally valid here so that one must resort to using the extra sum of squares principle or general linear hypothesis theory to obtain valid tests.

To construct a $100(1 - \alpha)\%$ confidence interval about a true value of y_T , say $y_{\ell T}$, corresponding to the x-coordinate, $X_{\ell T} = (1, x_{1T}, \dots, x_{pT})$, for $X_{\ell T}$ a row vector, one can show easily that the covariance matrix of $\hat{y}_T = P^t \hat{y}_T^*$ is estimated by

$$\hat{\beta}_{\hat{y}_T} = X(X^t V^{-1} X)^{-1} X^t \sum_{j=1}^k \hat{\sigma}_j^2. \quad (4.2.13)$$

Therefore, a confidence interval for $y_{\ell T}$ is given by

$$\hat{y}_{\ell T} \pm [X_{\ell T} (X^t V^{-1} X)^{-1} X_{\ell T}^t \sum_{j=1}^k \hat{\sigma}_j^2]^{\frac{1}{2}} t_{\alpha/2, n-p-1}. \quad (4.2.14)$$

Once again, results obtained in this section show that tests concerning specific components of β_T and associated confidence intervals

are constructible from information relating to analyses of the component equations. This result holds when $\varepsilon_j \sim N(\phi, V\sigma_j^2)$. We already showed that it holds when $\varepsilon_j \sim N(\phi, I\sigma_j^2)$. In both cases, our derivations are based on the assumption that y_j and y_ℓ are independent for each $j \neq \ell$ ($j, \ell = 1, 2, \dots, k$). This essentially completes our consideration of the problem of estimation and inference for the total model when $r_j \equiv p$.

The problem of additivity as defined here when $r_j \equiv p$ is mathematically nice and, in a sense, trivial. The problem, however, has obvious practical limitations as a result of requiring that $r_j \equiv p$ since, in practice, one would like to retain in each component equation only those of the p independent variables that are statistically important. It is, therefore, of interest to consider the consequences for estimation and inference for the total model when $r_j \leq p$ with the possibility that $r_j < p$ for all j . This is the situation not considered by Kozak (1970). We consider this case in the next chapter.

CHAPTER V

5.0 ADDITIVITY WHEN $r_j \leq p$ WITH $r_j < p$ FOR SOME j

We now relax the requirement that each component model contain all the independent variables in the total equation. Specifically, if the total equation contains p independent variables (each assumed important), we shall allow the component equations to contain only statistically important independent variables among the p variables. Thus r_j , the number of independent variables in component equation j , may be less than p and strictly so for at least one j . This admits the possibility that $r_j < p$ for each j provided that in that case each independent variable in the total equation is contained in at least one component equation. This will be consistent with additivity as defined here.

With $X = (\emptyset | X_1 | \dots | X_p)$ as defined earlier, we now consider models of the form

$$y_j = X\beta_j + \epsilon_j \quad (j = 1, 2, \dots, k) \quad (5.0.1)$$

and

$$y_T = \sum_{j=1}^k y_j = X\beta_T + \epsilon_T \quad (5.0.2)$$

where the matrix X is common to the $k + 1$ models. However, the latter models differ from those specified earlier in the following respects.

In (5.0.1), β_j is a $(p + 1) \times 1$ vector defined so that β_j has an intercept component with $r_j \leq p$ of the remaining p components nonzero and the other $p - r_j$ equal to zero. The relative positions of the zero

and nonzero components in the last p cells of β_j will be such as correspond to the presence or absence of particular X_i ($i = 1, 2, \dots, p$) in the j^{th} component equation. β_T will be as defined before with $p + 1$ nonzero components. Our definition of β_j implies that the effective X matrix, say X_j , corresponding to component equation j is necessarily different for each j except when β_j and β_ℓ have nonzero components in identical positions for $j \neq \ell$ ($j, \ell = 1, 2, \dots, k$). Assume that ϵ_j and ϵ_T are distributed as specified in chapter IV and also that y_j and y_ℓ (hence ϵ_j and ϵ_ℓ) are independent for $j \neq \ell$.

5.1 Estimation when $\epsilon_j \sim N(\phi, I\sigma_j^2)$

We first consider the problem of estimation for the total equation when $\epsilon_j \sim N(\phi, I\sigma_j^2)$. First, we state an intuitive result for the estimation problem and then demonstrate its validity. Note that model (5.0.1) is equivalent to model (2.2.1) with a condition adjoined, namely,

$$y_j = X\beta_j + \epsilon_j \quad \text{D. } b_{q_j} = \phi \quad (j = 1, 2, \dots, k) \quad (5.1.1)$$

where b_{q_j} is a vector of zeros corresponding to the vector of components of β_j in (5.0.1) which are set equal to zero in the j^{th} component equation. It is clear that estimation relating to (5.1.1) can be achieved via constrained minimization. Denote the resulting solution by $\hat{\beta}_j^*$ ($j = 1, 2, \dots, k$) and the corresponding conditioned fit by

$$\hat{y}_j = X\hat{\beta}_j^* \quad (j = 1, 2, \dots, k). \quad (5.1.2)$$

Furthermore, let y_T be as defined in chapter IV; we shall occasionally refer to y_T as the unconditioned total model. Also, let y_{TC} be termed the 'conditioned' total equation in a sense to be defined momentarily and let $\hat{y}_{TC}^* = \sum_{j=1}^k X\hat{\beta}_j^*$. Then the conditioned total predictive equation is given by

$$\hat{y}_{TC} = \hat{y}_T - (\hat{y}_T - \hat{y}_{TC}^*) \quad (5.1.3)$$

where the factor $\hat{y}_T - \hat{y}_{TC}^*$ is a correction or conditioning factor which corrects the unconditioned fitted total equation (obtained in chapter IV) for parameters that are set equal to zero in some of the component equations. Note that (5.1.3) implies that $\hat{y}_{TC} = \hat{y}_{TC}^* = X \sum_{j=1}^k \hat{\beta}_j^*$ and that

$$\hat{\beta}_{TC} = \sum_{j=1}^k \hat{\beta}_j^* = \hat{\beta}_T - (\hat{\beta}_T - \sum_{j=1}^k \hat{\beta}_j^*). \quad (5.1.4)$$

In (5.1.4), $\hat{\beta}_{TC}$ and $\hat{\beta}_T$ are the estimator of the parameter vector of the total conditioned equation and that of the unconditioned total equation, respectively. Equation (5.1.4) implies that the parameter vector of the conditioned total equation is estimated by adding the estimates of the parameter vectors corresponding to the conditioned component equations. Our immediate task is to demonstrate that this result is mathematically valid. To do so, we state our problem as one of minimization subject to constraints.

Let us solve the estimation problem associated with fitting

$$y_T = X\tilde{\beta}_T + \epsilon_T \quad (5.1.5)$$

subject to

$$\tilde{\beta}_T = \sum_{j=1}^k \beta_j^*. \quad (5.1.6)$$

Note that β_j^* in (5.1.6) refers to the parameter vector corresponding to the conditioned component equation j . This problem is solved by minimizing the Lagrangian function

$$S(\tilde{\beta}_T, \theta) = (y_T - X\tilde{\beta}_T)^t (y_T - X\tilde{\beta}_T) + 2\theta^t (\tilde{\beta}_T - \sum_{j=1}^k \beta_j^*), \quad (5.1.7)$$

where $2\theta^t$ is a vector of Lagrange multipliers. Now, differentiating (5.1.7) with respect to the elements of $\tilde{\beta}_T$ and θ , respectively, one obtains

$$\frac{\partial S}{\partial \tilde{\beta}_T} = 2X^t X \tilde{\beta}_T + 2\theta - 2X^t y_T \quad (5.1.8)$$

$$\frac{\partial S}{\partial \theta} = \tilde{\beta}_T - \sum_{j=1}^k \beta^*_j. \quad (5.1.9)$$

Equating both (5.1.8) and (5.1.9) to zero leads to

$$X^t X \hat{\beta}_T + \hat{\theta} = X^t y_T \quad (5.1.10)$$

$$\hat{\beta}_T = \sum_{j=1}^k \hat{\beta}^*_j. \quad (5.1.11)$$

Our solution vector $\hat{\beta}_T$ must satisfy both (5.1.10) and (5.1.11). From

(5.1.10) one has

$$\begin{aligned} X^t X \hat{\beta}_T = X^t y_T - \hat{\theta} &\Rightarrow \hat{\beta}_T = (X^t X)^{-1} X^t y_T - (X^t X)^{-1} \hat{\theta} \\ &= \hat{\beta}_T - (X^t X)^{-1} \hat{\theta}. \end{aligned} \quad (5.1.12)$$

Now, (5.1.12) $\Rightarrow \hat{\theta} = X^t X(\hat{\beta}_T - \hat{\beta}_T)$ and since $\hat{\beta}_T = \sum_{j=1}^k \hat{\beta}^*_j$ by (5.1.11), one has that

$$\hat{\theta} = X^t X(\hat{\beta}_T - \sum_{j=1}^k \hat{\beta}^*_j). \quad (5.1.13)$$

Hence, putting (5.1.13) into (5.1.12) one gets

$$\begin{aligned} \hat{\beta}_T &= \hat{\beta}_T - (X^t X)^{-1} X^t X(\hat{\beta}_T - \sum_{j=1}^k \hat{\beta}^*_j) \\ &= \hat{\beta}_T - (\hat{\beta}_T - \sum_{j=1}^k \hat{\beta}^*_j), \end{aligned} \quad (5.1.14)$$

a result given earlier in equation (5.1.4). This establishes the validity of that result.

Since $\tilde{\beta}_T$, the parameter vector of the total conditioned equation, is determined by β^*_j ($j = 1, 2, \dots, k$), it is important to discuss the estimation of β^*_j here. It should be emphasized that β^*_j is a $(p+1) \times 1$ vector having $p - r_j$ of its components equal to zero. We are explicitly assuming that the intercept component of β^*_j is nonzero. In estimating β^*_j , it is important to recognize that one does so conditionally on some specified components being assumed equal to zero. In the following, we cast the problem within the framework of general linear hypothesis theory.

Consider testing the general hypothesis $H : K^t \beta_j = m$, where β_j is a $(p + 1) \times 1$ parameter vector of the model (2.2.1), K^t is any matrix of s rows and $p + 1$ columns and m is a vector, of order s , of specified constants. We shall require that K^t be of full row rank, that is $r(K^t) = s$, where $r(\cdot)$ denotes the rank of the argument. One is interested here in estimating β_j under the null hypothesis $H : K^t \beta_j = m$. Designate the parameter vector under the null hypothesis by β_j^* . Using constrained least squares (see Searle, 1971, pp. 113-114), the desired estimator is given by

$$\hat{\beta}_j^* = \hat{\beta}_j - (X^t X)^{-1} K [K^t (X^t X)^{-1} K]^{-1} (K^t \hat{\beta}_j - m), \quad (5.1.15)$$

where $\hat{\beta}_j$ is the unconstrained estimator of β_j (with all independent variables included). When the hypothesis is of the form $H : b_{q_j} = \phi$ for b_{q_j} a subset of β_j of order q_j , we have

$$K^t = [I_{q_j} \ 0], \quad m = \phi, \quad s = q_j. \quad (5.1.16)$$

Now partition β_j , $\hat{\beta}_j$ and $(X^t X)^{-1}$ as follows

$$\beta_j = \begin{pmatrix} b_{q_j} \\ b_{p_j} \end{pmatrix}, \quad \hat{\beta}_j = \begin{pmatrix} \hat{b}_{q_j} \\ \hat{b}_{p_j} \end{pmatrix}, \quad (X^t X)^{-1} = \begin{pmatrix} T_{q_j q_j} & T_{q_j p_j} \\ T_{p_j q_j} & T_{p_j p_j} \end{pmatrix}$$

where $p_j + q_j = p + 1$. Then the estimator of β_j^* is

$$\hat{\beta}_j^* = \begin{pmatrix} \phi \\ \hat{b}_{p_j} - T_{p_j q_j} T_{q_j q_j}^{-1} \hat{b}_{q_j} \end{pmatrix}. \quad (5.1.17)$$

If X is partitioned as $X = (X_{q_j} | X_{p_j})$, then the estimator in (5.1.17) is equivalently given by

$$\hat{\beta}_j^* = \begin{pmatrix} \phi \\ (T_{p_j p_j} X_{p_j}^t - T_{p_j q_j} T_{q_j q_j}^{-1} T_{q_j p_j} X_{p_j}^t) y_j \end{pmatrix}. \quad (5.1.18)$$

Observe that if the columns of X_{q_j} are orthogonal to those of X_{p_j} , then $(X^t X)^{-1}$ is block-diagonal so that

$$\hat{\beta}^*_{j} = \begin{pmatrix} \phi \\ \hat{b}_{p_j} \end{pmatrix} = \begin{pmatrix} \phi \\ T_{p_j p_j} X_{p_j}^t y_j \end{pmatrix} . \quad (5.1.19)$$

Equation (5.1.19) expresses the expected fact that when the columns of X_{q_j} and those of X_{p_j} are orthogonal, then fitting only the last p_j components of X will yield the same \hat{b}_{p_j} as fitting all components of X conditionally on the first q_j having zero coefficients. The consequence for estimation of having all columns of X mutually orthogonal should be obvious from this. In general, the estimator of β for any linear model depends upon variables not included in the model, including those that are not known.

When a subset of a set of predictor variables is statistically unimportant, it is common practice to fit an equation which simply ignores the unimportant subset. Unless the latter subset is orthogonal to the important one in the original set, the resulting fit will not be conditioned in the sense defined above. Hence, the corresponding estimator of the parameter vector is different from that of a corresponding conditioned fit, again unless orthogonality holds. The estimator is also of smaller order and is given by the non-null part of $\hat{\beta}^*_{j}$ in (5.1.19). It follows from (5.1.18) and (5.1.19) that one can correct the latter estimator and fill it out appropriately to obtain the corresponding conditioned estimator. In view of this, the remainder of this chapter will be based on conditioned component equations in the sense just defined.

5.2 Inference when $\varepsilon_j \sim N(\phi, I\sigma_j^2)$

When testing the significance of the total conditioned equation, the regression sum of squares is given by

$$SS_R(\beta_{TC}^* | \beta_{OTC}) = (\sum_{j=1}^k \hat{\beta}_j^{*t}) X^t (\sum_{j=1}^k y_j) - n\bar{y}_T^2. \quad (5.2.1)$$

Other sums of squares relating to this particular testing problem are obtained in an obvious way. We omit further details which, again, are completely obvious. We consider, instead, the problem of testing specific hypotheses relating to the total conditioned equation.

It has been demonstrated above that the estimator of the parameter vector β_{TC} for the total conditioned equation is determined by summing corresponding components of conditioned component equations. Analytically, there are two possible ways in which a particular component of $\hat{\beta}_{TC}$ might turn out to be zero under additivity. First, a component of $\hat{\beta}_{TC}$ may be zero as a result of the cancellation law when adding negative and positive elements. While this is arithmetically plausible, it is not reasonable given that we have assumed a priori that each independent variable is statistically important. Secondly, if a particular independent variable has an estimated coefficient of zero in each individual component equation, then the corresponding component of $\hat{\beta}_{TC}$ will be zero. Again, this is unlikely since it is contrary to the hypothesis that each independent variable is important.

The foregoing suggests that an hypothesis which states that some component (or vector of a subset of components of β_{TC}) is zero is not a reasonable hypothesis. Instead, an hypothesis that some component (or vector of a subset of components of β_{TC}) is equal to c , where c is a known scalar (vector with all its components) different from zero, is a reasonable hypothesis. The specification of c may be based on

past experience or related analyses.

There are two ways in which c may be specified. The specification of c may be direct as described in the preceding paragraph. On the other hand, c may be specified indirectly by simply specifying c_j ($j = 1, 2, \dots, k$) in a series of subhypotheses relating to each component conditioned equation. We emphasize that the value of c is not specifically determined in the latter case, but it is determined under additivity of the conditioned component equations when they are fitted under the subhypotheses specified by c_j . We shall refer to the test of hypothesis concerning c when c is specified directly as a direct test of the hypothesis c . The corresponding test when c is specified indirectly through the c_j ($j = 1, 2, \dots, k$) will be referred to as an indirect test. We use this terminology only for its mnemonic appeal.

Let the conditioned total model specified by β_{TC} be designated as the full model. Recall that $\hat{\beta}_{TC} = \sum_{j=1}^k \hat{\beta}_j^*$ and that the sum of squares regression due to fitting this model is given by $SS_R(\beta_{TC}^* | \beta_{OTC})$ as given in (5.2.1). Now consider fitting a version of the full model restricted so that some specified component(s) of β_{TC} is(are) given by the scalar (vector) c . Let the latter model be indexed by the parameter vector β_{TC}^c . Note that $\hat{\beta}_{TC}^c$ is obtained by fitting the model indexed by β_{TC} subject to the further restriction that some specified component(s) of β_{TC} is(are) equal to c . Designate the sum of squares regression due to fitting the model indexed by β_{TC}^c by $SS_R(\beta_{TC}^{*c} | \beta_{OTC}^c)$. This sum of squares is given by

$$SS_R(\beta_{TC}^{*c} | \beta_{OTC}^c) = (\hat{\beta}_{TC}^c)^t X^t (\sum_{j=1}^k y_j) - n\bar{y}_T^2. \quad (5.2.2)$$

Since the model indexed by β_{TC}^c is a restriction on the model indexed by β_{TC} , it follows that $SS_R(\beta_{TC}^{*c} | \beta_{OTC}^c) \leq SS_R(\beta_{TC}^* | \beta_{OTC})$. Hence the

sum of squares

$$\psi(c) = SS_R(\beta_{TC}^* | \beta_{OTC}) - SS_R(\beta_{TC}^{*c} | \beta_{OTC}^c) \quad (5.2.3)$$

can be interpreted as the extra sum of squares for testing the hypothesis that a specified component(s) of β_{TC} is(are) equal to c . Thus, using $\psi(c)$ and the error sum of squares for the full model leads to a direct test for c .

Suppose, now, that c is specified indirectly by specifying c_j ($j = 1, 2, \dots, k$). As in an earlier part of the sequel, let β_j^* be the parameter vector corresponding to conditioned component model j . Suppose we fit the model indexed by β_j^* under the subhypothesis that some component of β_j^* is equal to c_j . This leads to estimators $\hat{\beta}_j^{*c}$ ($j = 1, 2, \dots, k$), where the superscript c on $\hat{\beta}_j^*$ denotes the further restriction c_j . Again if we denote the total conditioned model with the further restriction c imposed indirectly through the c_j as being indexed by β_{TC}^c , it follows by our result on additivity that $\hat{\beta}_{TC}^c = \sum_{j=1}^k \hat{\beta}_j^{*c}$. The regression sum of squares due to fitting the latter model is given by

$$SS_R(\beta_{TC}^{*c} | \beta_{OTC}^c) = (\sum_{j=1}^k (\hat{\beta}_j^{*c})^t) X^t (\sum_{j=1}^k y_j) - n\bar{y}_T^2. \quad (5.2.4)$$

Once again, one has that $SS_R(\beta_{TC}^{*c} | \beta_{OTC}^c) \leq SS_R(\beta_{TC}^* | \beta_{OTC})$, so that again the sum of squares

$$\psi^*(c) = SS_R(\beta_{TC}^* | \beta_{OTC}) - SS_R(\beta_{TC}^{*c} | \beta_{OTC}^c) \quad (5.2.5)$$

can be interpreted as the extra sum of squares for testing the hypothesis that a specified component(s) of β_{TC} is(are) equal to c via the c_j . Thus, using $\psi^*(c)$ and the sum of squares error for the full model leads to an indirect test for c .

It is worth emphasizing that the procedure for testing a hypothesis concerning c directly or indirectly is applicable whether c is a scalar

or a vector.

Tests of hypotheses concerning individual components of β_{TC} using the usual t-statistic and/or univariate confidence intervals about such components can be obtained in a manner similar to that described in chapter IV provided an estimate of the covariance matrix of $\hat{\beta}_{TC}$ is available. These tests and confidence intervals are especially likely to be misleading here, however, since the conditioning makes the components of $\hat{\beta}_{TC}$ correlated unless the columns of X are orthogonal. Therefore, tests on individual components of β_{TC} are best performed as outlined in the preceding paragraph. On the other hand, a confidence interval can be constructed about $y_{\ell T}$, say, corresponding to a given x-coordinate in an obvious way.

5.3 Estimation and Inference when $\varepsilon_j \sim N(\phi, V\sigma_j^2)$

Results presented in sections 5.1 and 5.2 relate to the distributional assumption $\varepsilon_j \sim N(\phi, I\sigma_j^2)$. These results carry over to the case $\varepsilon_j \sim N(\phi, V\sigma_j^2)$, for V as defined in chapter IV with the obvious modification that wherever X , y_j and y_T occur, in the various expressions, they are replaced by $(P^t)^{-1}X$, $(P^t)^{-1}y_j$ and $(P^t)^{-1}y_T$ in that order, where $P^tP = V$ with P as defined in that chapter. Thus, most of the results will involve V^{-1} as demonstrated before.

In the following chapter, it is demonstrated that the problem of additivity when $r_j \equiv p$ can be treated as a special case of additivity when $r_j \leq p$. Such a demonstration provides a basis for constructing a unified theory relating to the additivity problem.

CHAPTER VI

6.0 A GENERALIZATION OF THE ADDITIVITY PROBLEM

Within the framework of the conditioning principle described in the preceding chapter, the fitting of the component and corresponding total equations when $r_j \equiv p$ can be considered as a problem of fitting subject to 'null' or 'trivial' conditions. By null or trivial conditions we mean here that no further conditions are imposed on the β_j ($j = 1, 2, \dots, k$) beyond the basic additivity requirement that $\sum_{j=1}^k \beta_j = \beta_T$. The point to observe here is that there is no requirement that any component(s) of β_j be equal to zero. In terms of result (5.1.3) in the preceding chapter, this implies that $\hat{y}_{TC}^* = \hat{y}_T$ so that (5.1.3) reduces to

$$\hat{y}_{TC} = \hat{y}_T. \quad (6.0.1)$$

Note that \hat{y}_T , \hat{y}_{TC} and \hat{y}_{TC}^* were all defined in the previous chapter. Thus the correction factor or the conditioning factor is identically zero when $r_j \equiv p$.

In terms of a result given in (5.1.14), the above implies that

$$\hat{\beta}_T = \sum_{j=1}^k \hat{\beta}_j^* \quad (6.0.2)$$

so that (5.1.14) then reduces to

$$\hat{\hat{\beta}}_T = \hat{\beta}_T. \quad (6.0.3)$$

Equations (6.0.1) through (6.0.3) suggest that the problem of additivity when $r_j \equiv p$ can be treated as a special case of additivity when $r_j \leq p$.

In this connection, the theory of estimation and inference described in chapter V reduces to that presented in chapter IV when $r_j \equiv p$. This generalization is significant, at least theoretically, since it makes it possible to visualize the additivity problem as defined here as one very general problem which can be studied under one unified theory of estimation and inference.

CHAPTER VII

7.0 OTHER ASPECTS OF THE ADDITIVITY PROBLEM

The development of the theory has, thus far, been based upon the assumption that y_j and y_ℓ (and hence ϵ_j and ϵ_ℓ) are independent for each $j \neq \ell$ ($j, \ell = 1, 2, \dots, k$). As indicated earlier in the thesis, however, there are examples of applications where this assumption is simply not tenable. The implications of this, especially for inference, are well worth considering and will be examined in this chapter. There will be occasion also to consider other general complements of the additivity problem such as that of the matrix X not of full column rank and of V not necessarily positive definite. These latter generalizations are useful when considering certain classes of the general linear model. In particular, they permit the extension of the theory of additivity as developed here to classificatory models which are generally associated with designed experiments and are routinely analysed using analysis of variance procedures. Furthermore, it is noteworthy that it is generally assumed in most applications of regression analysis that the X matrix is fixed (that is, that the independent variables are either known or are measured without error). Yet it is quite conceivable that the independent variables may themselves be random, just as the dependent variable y , or they may be fixed but measured with error. It is reasonable to consider briefly the implications for analysis of these possibilities, at least for completeness.

7.1 The Case y_j, y_ℓ ($j \neq \ell$) Dependent

As a preamble, suppose that X_1, X_2, \dots, X_k are multivariate normal random vectors, each of dimension m , with mean vectors μ_i ($i = 1, 2, \dots, k$) and corresponding covariance matrices Σ_i ($i = 1, 2, \dots, k$). Now define $U = \sum_{i=1}^k X_i$ and suppose one is interested in the distribution of U . If, in addition, it is assumed that the vectors X_i ($i = 1, 2, \dots, k$) are independent, then by a well-known theorem in multivariate analysis (see Muirhead, 1982, p. 14), it follows that U is m -variate normally distributed with mean vector $\mu = \sum_{i=1}^k \mu_i$ and covariance matrix $\Sigma_U = \sum_{i=1}^k \Sigma_i$. On the other hand, if the vectors X_i ($i = 1, 2, \dots, k$) are dependent, we know only that

$$EU = \sum_{i=1}^k \mu_i \quad (7.1.1')$$

and

$$\Sigma_U = \sum_{i=1}^k \Sigma_i + \sum_{i \neq j} \Sigma_{ij}, \quad (7.1.2)$$

where Σ_i is the variance-covariance matrix of X_i and Σ_{ij} is the covariance matrix of X_i and X_j ($i \neq j$). Hence it follows that

$U \sim (\sum_{i=1}^k \mu_i, \sum_{i=1}^k \Sigma_i + \sum_{i \neq j} \Sigma_{ij})$. Note that we have only specified the parameters of the distribution of U and not its form. Indeed, as far as is known, without making any assumptions concerning the form of dependency of the summands in the definition of U , the exact distribution of U under dependence is largely an outstanding problem in mathematical statistics. However, as suggested above, given some knowledge of the form of dependency among the k vectors X_i , it is possible to obtain a distribution for U (Olkin, 1983, personal communication). Furthermore, one might surmise that the distribution of U under dependence of the component vectors might be derivable as a multivariate generalization

of the univariate analogue considered by Springer (1979, pp. 72-75). Even so, however, the explicit representation of such a distribution is likely to be nontrivial.

The representation of \mathbb{I}_U given in (7.1.2) derives from a simple generalization of the univariate case to the multivariate case. For the analogous univariate result, see Mood, Graybill, and Boes (1974, p. 179). Note that independence of the vectors X_i ($i = 1, 2, \dots, k$) implies that $\mathbb{I}_{ij} \equiv \emptyset$, the null matrix, in (7.1.2).

Now let $U = y_T$ and $X_j = y_j$ ($j = 1, 2, \dots, k$); then for y_j, y_ℓ dependent for each $j \neq \ell$, it follows that $y_T = \sum_{j=1}^k y_j \sim (X \sum_{j=1}^k \beta_j, \sum_{j=1}^k \mathbb{I}_j + \sum_{i \neq j} \sum \mathbb{I}_{ij})$. For simplicity in what follows, we shall mostly use \mathbb{I}_T to designate $\sum_{j=1}^k \mathbb{I}_j + \sum_{i \neq j} \sum \mathbb{I}_{ij}$. Even given that the y_j ($j = 1, 2, \dots, k$) are individually multivariate normal, under dependence it is not known what the distribution of y_T is exactly. What is known is that y_T is either multivariate normal or is not multivariate normal (see Kale, 1970). Examples are found in the literature of linear combinations of normal random variables which are themselves (the linear combinations, that is) not normal (e.g., Rosenberg, 1965; Behboodian, 1972) and of marginally normal random variables whose joint distributions are not normal (e.g., Ruymgaart, 1973). These results of course generalize to vector random variables. The overall implication of this is that lack of knowledge of the exact distribution of y_T and, in particular, its probable non-normality renders the construction of a small-sample theory of inference considerably more difficult.

Given lack of knowledge of the exact distribution of y_T , a small-sample theory of inference for the total regression model is constructible on the basis of normality of y_T if one can demonstrate that the vectors

y_j ($j = 1, 2, \dots, k$) are jointly multivariate normal. This is the case because it is well-known that if y_1, y_2, \dots, y_k are jointly multivariate normal, then every linear function of these y_j 's is multivariate normal. (Note that the y_j 's are vectors here.) This result follows from a characterization of the bivariate normal distribution which generalizes to other joint multivariate normal distributions (see Rao, 1965, pp. 437-438). Therefore, in our instance, under dependence of the y_j 's, normal theory can be used to construct inferences concerning y_T or β_T or both if it can be shown that y_1, y_2, \dots, y_k are jointly multivariate normal. This suggests the need for methods of assessing multivariate normality based upon realizations of the vectors y_1, y_2, \dots, y_k . Graphical methods of assessing multivariate normality have been proposed in the literature (e.g., Healy, 1968; Cox, 1968; Andrews, Gnanadesikan, and Warner, 1973). Other authors have proposed analytical significance tests for testing for multivariate normality (e.g., Malkovich and Afifi, 1973; Hawkins, 1981). More recently, however, Koziol (1982) introduced a test for assessing multivariate normality which is fairly easy to use and has some nice properties. If a test for joint multivariate normality such as Koziol's (1982) leads one to entertain joint multivariate normality, then one proceeds to make inferences concerning y_T or β_T based upon the usual normality assumptions. If, on the other hand, joint multivariate normality is rejected, then one can either appeal to asymptotic results to construct approximate tests or resort to nonparametric approaches. We shall discuss the latter approach only briefly in this thesis. But first, let us tackle the problem of estimation.

7.1.1 Estimation for Total Model under Dependence

As observed earlier, estimation should not, in general, be hampered by lack of knowledge of the distribution of y_T and, in particular, by its non-normality. Consider estimation for the total model when $\epsilon_j \sim N(\phi, I\sigma_j^2)$ and when $\epsilon_j \sim N(\phi, V\sigma_j^2)$. It is demonstrated in this section that when the y_j 's are dependent, the concept of additivity, as defined here, does not hold. This follows from the following reasoning.

When the ϵ_j are dependent, one has that $\epsilon_T \sim (\phi, \sum_{j=1}^k \phi_j + \sum_{i \neq j} \phi_{ij})$ in general. In the case where $\epsilon_j \sim N(\phi, I\sigma_j^2)$ ($j = 1, 2, \dots, k$), the first term in the variance-covariance matrix of ϵ_T reduces to $I\sigma_T^2$, where $\sigma_T^2 = \sum_{j=1}^k \sigma_j^2$. Furthermore, it is easy to see that $\phi_{ij} = I\rho_{ij}\sigma_i\sigma_j$ so that $\sum_{i \neq j} \phi_{ij} = I \sum_{i \neq j} \rho_{ij}\sigma_i\sigma_j$. Therefore, under dependence of the ϵ_j with $\epsilon_j \sim N(\phi, I\sigma_j^2)$, $\epsilon_T \sim (\phi, I\tilde{\sigma}^2)$ where $\tilde{\sigma}^2 = \sum_{j=1}^k \sigma_j^2 + \sum_{i \neq j} \rho_{ij}\sigma_i\sigma_j$. Thus the variance-covariance matrix of ϵ_T is diagonal. On the other hand, when $\epsilon_j \sim N(\phi, V\sigma_j^2)$, the first term in the variance-covariance matrix of ϵ_T becomes $V\sigma_T^2$, with σ_T^2 as defined above, so that $\epsilon_T \sim (\phi, V\sigma_T^2 + \sum_{i \neq j} \phi_{ij})$ where ϕ_{ij} is not diagonal.

Recall that under the assumption of independence of the y_j 's and V positive definite, the transformation matrix P , such that $P^t P = V$, was the same for each component model and the total model. As a consequence, additivity followed naturally since

$$\begin{aligned}
 \hat{\beta}_T &= (X^t V^{-1} X)^{-1} X^t V^{-1} y_T \\
 &= (X^t V^{-1} X)^{-1} X^t V^{-1} \sum_{j=1}^k y_j \\
 &= (X^t V^{-1} X)^{-1} X^t V^{-1} [y_1 + \dots + y_k] \\
 &= \sum_{j=1}^k \hat{\beta}_j.
 \end{aligned} \tag{7.1.1.1}$$

Note that the above result holds also when $V = I$. Now suppose that under dependence of the y_j ($j = 1, 2, \dots, k$), the variance-covariance matrix of ε_T may be written in the form $V_T \sigma_T^2$, for some matrix V_T . First, note that when $\varepsilon_j \sim N(\phi, V \sigma_j^2)$ with V positive definite, there is no guarantee that V_T is positive definite, though we know that it is at least positive semi-definite. Secondly, even if V_T were positive definite, it is obvious that the matrix P_T such that $V_T = P_T^t P_T$ is not necessarily equal to the matrix P which transforms each of the component equations. This holds when $\varepsilon_j \sim N(\phi, I \sigma_j^2)$ because $I \sigma_T^2 \neq I \tilde{\sigma}^2$. Consequently, when the y_j are dependent, $\hat{\beta}_T$ cannot be determined from the additivity property. This result must obviously hold for more general V and V_T . Assuming positive definiteness of V and V_T , nonadditivity is demonstrated as follows. Note that

$$\begin{aligned}
 \hat{\beta}_T &= (X^t V_T^{-1} X)^{-1} X^t V_T^{-1} y_T \\
 &= (X^t V_T^{-1} X)^{-1} X^t V_T^{-1} \sum_{j=1}^k y_j \\
 &= \sum_{j=1}^k \hat{\beta}_j * \\
 &\neq \sum_{j=1}^k \hat{\beta}_j = (X^t V^{-1} X)^{-1} X^t V^{-1} \sum_{j=1}^k y_j. \quad (7.1.1.2)
 \end{aligned}$$

Indeed, equality holds only if $V_T = V$ which implies that $\Sigma_{ij} = 0$ for all $i \neq j$ and, therefore, independence of the y_j 's. The above results suggest that when the y_j 's are dependent, the parameters of the total equation must be determined by actually fitting the corresponding total equation rather than from additivity. An exception to this would be in those cases where dependence is so weak that the first term in the covariance matrix dominates the second term (namely $\sum_{i \neq j} \dagger_{ij}$) in the sense that the entries of each \dagger_{ij} are close to zero. But this simply implies that independence largely obtains.

Since inference for the total model is essentially linked to estimation, it follows that inferences concerning parameters of the total model when the y_j are dependent can only be made after fitting the relevant total models directly. Therefore, under dependence of the y_j , one must actually fit the total model in order to estimate its parameters and make inferences about them.

Let us return to the problem of inference under dependence. Since, as observed earlier, ϵ_T may or may not be multivariate normal, one has several options. First, one can test for joint multivariate normality (JMVN) of y_1, y_2, \dots, y_k along the lines of Koziol (1982). If the test shows that joint multivariate normality is tenable, then one uses normal theory to make inferences concerning the total equation as described below. If joint multivariate normality is not tenable, one may examine $y_T = \sum_{j=1}^k y_j$ directly for normality, since lack of joint multivariate normality does not rule out the possibility that y_T is normal. If both joint multivariate normality of the y_j 's ($j = 1, 2, \dots, k$) and normality of y_T by direct examination of $\sum_{j=1}^k y_j$ are not tenable, then one may use non-parametric procedures. If certain conditions are met, one may use asymptotic results to arrive at approximate inferences (see Arnold, 1981, sections 10.1, 10.3) when y_T is not normal. This latter is not considered further here. However the other approaches are considered briefly below.

7.1.2 Inference for Total Model when y_1, y_2, \dots, y_k are JMVN

When a procedure for assessing joint multivariate normality such as Koziol's (1982) leads to the conclusion that the assumption of joint multivariate normality is reasonable for y_1, y_2, \dots, y_k , one treats

$y_T = \sum_{j=1}^k y_j$ as multivariate normal with mean $X\beta_T$ and variance-covariance matrix $\hat{\Sigma}_T = \sum_{j=1}^k \hat{\Sigma}_j + \sum_{i \neq j} \sum \hat{\Sigma}_{ij}$. The normality of y_T when y_1, y_2, \dots, y_k

are jointly multivariate normal is a standard result in mathematical statistics as suggested in section 7.1.

Recall that if $\epsilon_j \sim N(\phi, I\sigma_j^2)$ ($j = 1, 2, \dots, k$), then with the ϵ_j jointly multivariate normal, $\epsilon_T \sim N(\phi, I(\sum_{j=1}^k \sigma_j^2 + \sum_{i \neq j} \rho_{ij} \sigma_i \sigma_j))$.

With these conditions, fitting $y_T = X\beta_T + \epsilon_T$ directly by ordinary least squares yields BLUE's for β_T and $\tilde{\sigma}^2 = \sum_{j=1}^k \sigma_j^2 + \sum_{i \neq j} \rho_{ij} \sigma_i \sigma_j$ in the sense of the Gauss-Markoff theorem. These estimators, denoted by $\hat{\beta}_T$ and $\hat{\sigma}^2 = \text{MSE}$, respectively, are also maximum likelihood estimators. It is noteworthy that the components of $\tilde{\sigma}^2$ are not estimable directly from the total model. However, the usual analysis of variance tests for the total model based upon the above estimates are valid and confidence intervals may be constructed on individual components of β_T and on y_T essentially as explained in chapter IV. This approach is also applicable if a direct examination of $y_T = \sum_{j=1}^k y_j$ suggests that it is normally distributed.

When $\epsilon_j \sim N(\phi, V\sigma_j^2)$ and the ϵ_j are jointly multivariate normal, then $\epsilon_T \sim N(\phi, V\sum_{j=1}^k \sigma_j^2 + \sum_{i \neq j} \rho_{ij})$. If it is possible to write $V\sum_{j=1}^k \sigma_j^2 + \sum_{i \neq j} \rho_{ij}$ as $V_T\sigma_T^2$ for some positive definite matrix V_T , then it is clear by results given in chapter IV that there exists a matrix P_T such that $V_T = P_T^t P_T$. Therefore, generalized least squares applied to the total equation yields BLUE's which are also maximum likelihood estimators. Tests of hypotheses concerning β_T or its components can be achieved as outlined in section 4.2. If, on the other hand, the variance-covariance matrix of ϵ_T is simply positive semi-definite rather than positive definite, then an approach such as is used by Zyskind (1967) and Zyskind and Martin (1969) may be employed for estimation and inference.

7.1.3 Inference for Total Model when y_1, y_2, \dots, y_k are not JMVN

It has been noted above that when y_1, y_2, \dots, y_k are not jointly multivariate normal, it is still possible that $y_T = \sum_{j=1}^k y_j$ is multivariate normal. It has also been indicated that when y_T is multivariate normal, then estimation and inferences relating to the total equation fitted directly can be carried out as outlined in the preceding section.

When y_1, y_2, \dots, y_k are not jointly multivariate normal and $y_T = \sum_{j=1}^k y_j$ is not normal, one of the options left for inferences for the total model is via use of nonparametric procedures. It is not the objective here to pursue the subject in detail but, for purposes of completeness, to indicate what procedures are available and possible references.

Randles and Wolfe (1979) give a nonparametric approach to testing the slope in simple linear regression. Clearly, this is of limited use in our context. However, both the estimation problem and inference procedures for more general regression problems are considered in chapter 9 of Hollander and Wolfe (1973). Other methods of dealing with non-normality of y_T in estimation and inference for the total equation is to use any of a number of so-called robust regression techniques. One such technique is known as robust ridge regression proposed by Hogg (1979). For further references to some of these robust regression techniques see Montgomery and Peck (1982, section 9.3). Finally, before leaving the subject of nonparametric approaches and how they might be applied to fitting a total model under non-normality, it is worth mentioning two fairly novel nonparametric methods which are applicable to regression situations. These are the jackknife and the bootstrap. For a reference to use of the jackknife in regression see Miller (1974) and for application of bootstrap techniques in regression see Efron (1979).

7.2 The Case of X Random or Measured with Error

It is generally assumed in regression applications that the independent variables are either fixed and known or that they are measured without error. Indeed, in most of the development in this thesis, this has been assumed implicitly. In such situations, only the dependent variable y is assumed random. In many biological applications, however, it is often the case that both y and the independent variables are random. Alternatively, it may well be that the independent variables are in fact fixed but are measured with error. In the following, these two possibilities are considered in the light of their implication for estimation and inference for the total model. Attention is restricted in both cases to the situation of independent y_j 's. Without losing sight of the additivity problem it will suffice here to examine the consequence for estimation and inference on a component equation which, for simplicity, will be referred to without the j subscript as $y = X\beta + \epsilon$.

7.2.1 The Case of X Random

Sampson (1974) distinguishes between two related regression schemes. One scheme is that in which the independent variables are constant or fixed, as is often assumed. He refers to this simply as regression analysis. The other is that in which the independent variables are random variables (or realizations of random variables). This latter regression scheme is referred to as multivariate analysis of regression. We concern ourselves in the present section with the latter scheme. The objective here is not to provide a detailed analysis of the situation but to highlight the effect that randomness of X may have on estimation and inference for component model j and, hence, for the total equation. The following is largely due to Sampson (1974).

The multivariate analysis of regression scheme assumes that the vector y and the vectors consisting of the columns of the matrix X form a multivariate random variable (or a realization of a multivariate random variable). In the present case, it will be assumed that the joint distribution is multivariate normal. Denote the continuous random variable corresponding to the independent variable by X (where X is p -dimensional) and let X^* be a realization of X . The random variable corresponding to the dependent variable is denoted by y and its realization by y^* . With a sample of size n $\{(y_i, x_i), i = 1, 2, \dots, n\}$, let $z_i^t = (y_i, x_i^t)$ and the corresponding realizations $\{(y_i^*, x_i^*), i = 1, 2, \dots, n\}$ and $\{z_i^*, i = 1, 2, \dots, n\}$. In the multivariate analysis of regression, it is assumed that for $1 \leq i \leq n$, z_i are independently and identically distributed according to $N(\phi, \Phi)$. In the multivariate analysis of regression model, the parameters equivalent to β and σ^2 in the regression analysis model are $\Phi_{22}^{-1} \Phi_{21}$ and $\Phi_{11} - \Phi_{12} \Phi_{22}^{-1} \Phi_{21}$, where

$$\Phi = \begin{pmatrix} \Phi_{11} & \Phi_{12} \\ \Phi_{21} & \Phi_{22} \end{pmatrix} \quad (7.2.1.1)$$

and Φ_{21} is $p \times 1$. As stated by Sampson (1974), the justification for the appropriateness of $\Phi_{22}^{-1} \Phi_{21}$ as a parameter vector is that for $1 \leq i \leq n$, $E(y_i - x_i^t \gamma)^2$ is minimized for $\gamma = \Phi_{22}^{-1} \Phi_{21}$, so that $x_i^t \Phi_{22}^{-1} \Phi_{21}$ is the best linear predictor of y_i in the sense of minimizing squared error loss. Thus, when one speaks of regression coefficients in the multivariate analysis of regression situation, one speaks of $\Phi_{22}^{-1} \Phi_{21}$. In addition, under this scheme

$$\mathcal{L}(y|X = X^*) = N(X^* \Phi_{22}^{-1} \Phi_{21}, (\Phi_{11} - \Phi_{12} \Phi_{22}^{-1} \Phi_{21})I), \quad (7.2.1.2)$$

where I is an identity matrix of dimension n and $\mathcal{L}(\cdot)$ denotes the

probability distribution law of the argument. The relationship between regression analysis and (7.2.1.2) should be fairly obvious. Without going into further technical details, we state results concerning estimation and inference in multivariate analysis of regression and how they relate to corresponding results in regression analysis with fixed or nonrandom X .

An important result concerning estimation in multivariate analysis of regression is that although the maximum likelihood (ML) estimators are necessarily different from those in regression analysis (mainly because they are defined on different sample spaces), the corresponding ML estimates under the two models are exactly the same. Thus estimation under the two model formulations is the same. However, Sampson (1974) shows that for testing hypotheses in the two situations, the power functions are different. This is a significant result in that it stresses the importance of using a correct model in order to obtain tests with the correct power. This result is of considerable relevance in the present and other biological applications where X may in fact be random rather than fixed as is often assumed in regression situations. The implication for additivity is that testing is obviously affected by randomness of X but not estimation.

7.2.2 The Case when X is Measured with Error

In regression situations where the independent variables may reasonably be considered fixed, it is conceivable that an error may be introduced when measuring X at its fixed value. It is noteworthy that this problem is not necessarily the same as that of random X unless further assumptions are made about both X and y . The main objective here is to demonstrate the effect upon estimation and inference when X is measured

with error. For simplicity, restrict attention to the simple linear regression model

$$y = \beta_0 + \beta_1 X + \epsilon, \quad (7.2.2.1)$$

where it is assumed that $\epsilon \sim N(0, \sigma_\epsilon^2)$ and $\text{cov}(\epsilon_i, \epsilon_j) = 0$ for $i \neq j$. Now if X is measured with error, one does not observe X directly but rather observes

$$X' = X + \delta \quad (7.2.2.2)$$

where X is the true value of X and δ is a measurement error. Suppose that $\delta \sim N(0, \sigma_\delta^2)$, $X \sim N(\mu_X, \sigma_X^2)$ and that ϵ , δ , and X are independent. Then Y and X' follow a bivariate normal distribution (Snedecor and Cochran, 1973) and the regression of Y on X' is linear with regression coefficient

$$\beta'_1 = \beta_1 / (1 + \lambda), \quad (7.2.2.3)$$

where $\lambda = \sigma_\delta^2 / \sigma_X^2$. Thus it is the case that when X is measured with error, our least squares estimate of the regression of Y on X is biased in that it underestimates the true regression coefficient from fitting Y on X . When X is not normal, the above result holds in large samples and holds approximately in small samples if λ is small (see Snedecor and Cochran, 1973). Inferences concerning y or the regression coefficient are valid if X is measured with error provided that ϵ , δ , and the true X are approximately normal. However, predictions of y are less precise because of the increase in residuals as a result of errors in X .

The results given above have some relevance in the additivity problem. More importantly, they point to the need for a proper regression approach if proper estimates and inferences are to be made. For other aspects of this problem see Wald (1940), Berkson (1950), and Madansky (1959).

7.3 Other General Complements

In more general applications of the general linear model, it is not uncommon that the matrix X is not of full column rank. Suppose, first, that $\epsilon_j \sim N(\phi, I\sigma_j^2)$. Observe that X^tX is singular and, therefore, a unique solution does not exist for the least squares problem. An optimal solution is obtainable, however, by using the well-known concept of a generalized inverse. Let us begin by considering a particular generalized inverse, one commonly referred to as the Moore-Penrose inverse (Moore, 1920; Penrose, 1955) but also often called the pseudo-inverse or, simply, p -inverse. Attention is restricted to real X throughout; but first some definitions are in order, given here as theorems.

Theorem 1. Suppose X is a real $n \times (p+1)$ matrix with $\text{rank}(X) = r < p+1$. Then the $(p+1) \times (p+1)$ matrix X^tX has exactly r positive eigenvalues: $\lambda_1^2 \geq \lambda_2^2 \geq \dots \geq \lambda_r^2 > 0$ plus the zero eigenvalue with multiplicity $p+1-r$.

The next theorem is based on a well-known theorem in matrix algebra called the Singular-Value Decomposition theorem.

Theorem 2. With X satisfying theorem 1, one can always find an $n \times n$ orthogonal matrix U and a $(p+1) \times (p+1)$ orthogonal matrix G such that $\Lambda = U^tXG$ and $X = U\Lambda G^t$ with Λ the $n \times (p+1)$ matrix

$$\Lambda = \begin{pmatrix} D & \emptyset \\ \emptyset & \emptyset \end{pmatrix},$$

where D is an $r \times r$ diagonal matrix with i^{th} diagonal element $d_{ii} = \lambda_i > 0$ for $1 \leq i \leq r$. The expression of X in the form $X = U\Lambda G^t$ is termed the singular-value decomposition of X .

One must remark that U and G in the above theorem are not necessarily unique. The real importance of theorem 2 is that a decomposition

of the matrix X exists, a result which leads to definition of the Moore-Penrose inverse as follows.

Theorem 3. If, from the $n \times (p + 1)$ matrix Λ of theorem 2, one defines Λ^+ as the $(p + 1) \times n$ matrix

$$\Lambda^+ = \begin{pmatrix} D^{-1} & \emptyset \\ \emptyset & \emptyset \end{pmatrix},$$

then the Moore-Penrose inverse (pseudo-inverse) of the matrix X is given by

$$X^+ = G\Lambda^+U^t,$$

where G and U are as specified in theorem 2.

With X^+ defined as above, an optimal least squares solution for a model of the form $y = X\beta + \epsilon$ is given by

$$\hat{\beta} = X^+y. \quad (7.2.1)$$

Pertaining to the additivity problem of the discourse, result (7.2.1) implies that

$$\hat{\beta}_j = X^+y_j \quad (7.2.2)$$

for component equation j ($j = 1, 2, \dots, k$) with

$$\hat{\beta}_T = \sum_{j=1}^k \hat{\beta}_j \quad (7.2.3)$$

for the total model. Note that inference theory relating to the total model as discussed elsewhere in the thesis now incorporates X^+ in an obvious way. Note, for instance, that when $\epsilon_j \sim N(\phi, I\sigma_j^2)$, it is the case that the covariance matrix of $\hat{\beta}_T$ is given by

$$\hat{\beta}_T = X^+X^{+t} \sum_{j=1}^k \sigma_j^2, \quad (7.2.4)$$

a result which can be derived easily from (7.2.2) and (7.2.3).

When $y_j \sim N(\phi, V\sigma_j^2)$, using our usual transformation, $\hat{\beta}_j$ is given by

$$\hat{\beta}_j = X^{*+} y_j^* \quad (7.2.5)$$

so that

$$\hat{\beta}_T = X^{*+} \sum_{j=1}^k y_j^*. \quad (7.2.6)$$

In (7.2.5) and (7.2.6), X^{*+} is the Moore-Penrose inverse of $X^* = (P^t)^{-1}X$ and $y_j^* = (P^t)^{-1}y_j$. Again, the incorporation of these results in inference theory is a straightforward exercise and is omitted here. However, some remarks are in order with respect to X^+ (or X^{*+}). First, it is noteworthy that although U and G are not necessarily unique in the decomposition of X given by theorem 2, the Moore-Penrose inverse X^+ (or X^{*+}) is unique. Therefore, different U and G will lead to the same X^+ and, hence, the same optimal solution $\hat{\beta}$. It was mentioned earlier that when X is not of full column rank, there is no unique solution to the least squares problem of fitting $y = X\beta + \epsilon$ or its corresponding transform. While this is so, it is remarkable that the solution (7.2.1) or its transformed version is optimal in the sense that it is the only solution giving least 2-norm; that is, it is the best solution to the least squares problem. When X is square and nonsingular, then $X^+ = X^{-1}$, the unique inverse of X . Finally, it is important to mention that the real practical usefulness of X^+ hinges upon the ease with which it can be determined in any one problem. It turns out that X^+ is relatively easy to compute when X has a few columns. However, the task of computing X^+ becomes increasingly more difficult with an increasing number of columns in X . Since many practical problems tend to involve an X matrix with a fairly large number of columns (especially in classificatory models), the use of X^+ often presents a computational barrier. Largely because of this, a more general (weaker) generalized inverse

is used in many singular situations. The following is only an introduction to this type of inverse. For more complete treatments, see Searle (1971) and Rao and Mitra (1966).

The Moore-Penrose inverse described above satisfies the following conditions:

$$\begin{aligned}
 \text{(i)} \quad & XX^+X = X \\
 \text{(ii)} \quad & X^+XX^+ = X^+ \\
 \text{(iii)} \quad & (X^+X)^t = X^+X \\
 \text{(iv)} \quad & (XX^+)^t = XX^+.
 \end{aligned}
 \tag{7.2.7}$$

If one defines a matrix X^+ satisfying only condition (i) in (7.2.7), that is satisfying $XX^+X = X$, then X^+ is termed a generalized inverse of X (see Searle, 1971). Unlike X^+ , the Moore-Penrose inverse, X^+ is not unique. However, X^+ is considerably easier to compute than X^+ . Furthermore, any X^+ has the property that it generates all possible solutions relating to any given estimation problem and these solutions are invariant under affine transformations. The latter property is of value with regard to estimation and inference for linear functions of the parameters in a given problem. It should be pointed out that X^+ enters into inference theory in much the same way that the Moore-Penrose inverse does. Further details relating to the use of X^+ are omitted here as they can be found elsewhere (e.g., Searle, 1971; Rao and Mitra, 1971).

Most of the results presented so far are based upon the assumption that whenever $\varepsilon_j \sim N(\phi, V\sigma_j^2)$, then V is positive definite and known. While this is commonly true and lends itself to fairly straightforward mathematical manipulations, there are instances in which V is not necessarily positive definite. In addition, the elements of V may be unknown. The general approach is indicated here for the case where V is

nonnegative definite and known. The case of V unknown is considerably more difficult.

As was the case when X was not of full column rank, the concept of the generalized inverse is employed when dealing with models where V is not positive definite. In the general case, a solution for model j ($j = 1, 2, \dots, k$) would be given by

$$\hat{\beta}_j = (X^t V^+ X)^+ X^t V^+ y_j \quad (7.2.8)$$

where V^+ is any generalized inverse of V . Corresponding results for additivity and the associated inference problem generally correspond to those presented earlier with the obvious modification that V^+ is used in place of V^{-1} . Other details are given in Searle (1971, section 5.8) while another fairly instructive approach is given by Zyskind (1967) and Zyskind and Martin (1969).

CHAPTER VIII

8.0 SOME EXTENSIONS OF THE THEORY

It is well-known that data sets generated under experimental conditions according to a predetermined design model can be analysed using the general regression approach. Indeed, although the conventional analysis of variance approach is used in analysing most such data sets, the general regression approach often represents the most efficient and, at times, the only exact method of analysis, especially for unbalanced situations. In view of this link between regression analysis and conventional analysis of variance, it is reasonable to ask whether the problem of additivity, as defined here, cannot be envisaged within the context of classificatory models. It is shown, in this chapter, that an extension of the additivity problem to classificatory models is not only theoretically plausible but also makes sense in some practical situations. Secondly, in view of the growing interest in the use of nonlinear models in many branches of applied biology, it is of interest to investigate the extent to which the concept of additivity, as understood here, can be expected to hold in nonlinear situations. In a forestry context, such an investigation has an important bearing upon the determination of total volume or weight biomass of individual trees from corresponding component biomass using any of the well-known nonlinear models, such as the Chapman-Richards function. The main thrust of the development in this chapter will,

then, be directed towards establishing the extent to which the theory of additivity, as developed in chapters IV, V, and VI, applies to classificatory and nonlinear models.

8.1 Extension to Classificatory Models

A theoretical justification for the extension of the theory of additivity to classificatory models is based upon the fact that any classificatory model can be equivalently expressed in linear regression form. A special feature to note about such a model is that the incidence matrix, otherwise known as the design matrix, is, in general, not of full column rank. Therefore, no unique solution exists for the estimation problem using least squares. Hence, one either uses the unique Moore-Penrose inverse to obtain an optimal solution or uses a generalized inverse to arrive at a solution. As indicated elsewhere in the discourse, the decision to use the Moore-Penrose inverse or a generalized inverse will depend upon considerations of computational efficiency. Furthermore, the incorporation of results from the estimation problem into inference involves the simple substitution of expressions involving the appropriate generalized inverse into statistics derived in earlier chapters. To indicate the practicality of the additivity problem in the context of a classificatory model, we describe below how such a problem might arise in practice. We draw our example from the field of agriculture.

For simplicity, consider a controlled field crop experiment involving a treatments, each replicated n times. It is a simple matter to recognize the design here as a completely randomized design. We shall suppose that the leaf component of the biomass of the crop under investigation is used for human consumption as a vegetable. Further, suppose that the floral component of the crop is used as a different type of

vegetable food. Next, suppose that the seed component is used as another type of food while the remaining unusable above-ground part of the plant is burned as a fuel. If one is interested in the effect of treatment upon the accumulation of biomass (on a green weight basis) in component j ($j = 1, 2, 3, 4$) and the corresponding effect on total biomass accumulation (resulting from adding the four components), then one has the following problem. Designate the observed biomass of the j^{th} component corresponding to the r^{th} replicate by y_{ijr} ($i = 1, 2, \dots, a$; $j = 1, 2, 3, 4$; $r = 1, 2, \dots, n$). Then one would be interested in models of the form

$$y_{ijr} = \mu_j + \tau_{ij} + \epsilon_{ijr} \quad (8.1.1)$$

and

$$y_{iTr} = \sum_{j=1}^4 y_{ijr} = \mu_T + \tau_{iT} + \epsilon_{iTr}, \quad (8.1.2)$$

with appropriate assumptions on the errors. Note that by using an appropriate incidence matrix X , one may write (8.1.1) and (8.1.2), respectively, in matrix form as

$$y_j = X\theta_j + \epsilon_j \quad (j = 1, 2, 3, 4) \quad (8.1.3)$$

and

$$y_T = \sum_{j=1}^4 y_j = X\theta_T + \epsilon_T. \quad (8.1.4)$$

We remark that $\theta_j = (\mu_j, \tau_{1j}, \tau_{2j}, \dots, \tau_{aj})^t$ while

$\theta_T = (\mu_T, \tau_{1T}, \tau_{2T}, \dots, \tau_{aT})^t$. Clearly, the estimation and inference

theory presented elsewhere in this thesis can be applied to (8.1.3) and

(8.1.4) subject only to the proviso that a generalized inverse or the

Moore-Penrose inverse is used in place of $(X^tX)^{-1}$ or $(X^tV^{-1}X)^{-1}$. The

regression formulation (8.1.3) and (8.1.4) of the analysis of variance

models (8.1.1) and (8.1.2) makes it especially easy to estimate θ_j sub-

ject to the condition that certain of its components are equal to zero.

Furthermore, there is no reason why the additivity concept cannot be

applied to more complicated classificatory models such as the randomized complete block design, Latin square design, and other designs since, in each case, one can write the corresponding models in linear regression form. This demonstrates that the extension of the notion of additivity of component regression equations to classificatory models is not only theoretically plausible but also appears to make sense in practice.

8.2 Extension to Nonlinear Models

A definition was given in chapter II for a regression model in general and for a linear regression model in particular. From those definitions, it follows that any regression model $r^* \in R^*$ satisfying $R^* \notin L$ is a nonlinear regression model. Recall that L was defined in chapter II as the set of all linear regression models. Conventionally, nonlinear regression models are divided into two groups, namely the class of nonlinear regression models that can be made linear by applying an appropriate transformation to the nonlinear model and the class of nonlinear regression models for which there exists no known linearizing transformation. The two types of nonlinear regression models are generally referred to in the literature as intrinsically linear and intrinsically nonlinear, respectively (see Draper and Smith, 1981). The main objective in this section is to investigate whether the notion of additivity does make sense for these two types of nonlinear models.

We consider models of the form

$$y_j = f(X, \beta_j) + \epsilon_j \quad (j = 1, 2, \dots, k) \quad (8.2.1)$$

and

$$y_T = \sum_{j=1}^k y_j = f(X, \beta_T) + \epsilon_T. \quad (8.2.2)$$

Attention is directed here toward discovering the extent to which model (8.2.2) is arithmetically determined by the models specified in (8.2.1).

For purposes of simplicity, we restrict detailed analysis to two types of nonlinear models, namely

$$y_j = \beta_{0j} e^{\beta_{1j}X} \epsilon_j \quad (j = 1, 2, \dots, k) \quad (8.2.3)$$

and

$$y_j = \beta_{0j} e^{\beta_{1j}X} + \epsilon_j \quad (j = 1, 2, \dots, k). \quad (8.2.4)$$

Note that models specified by (8.2.3) are intrinsically linear, so that for purposes of estimation, one may transform them to linear form using a logarithmic transformation. This leads to

$$\ln y_j = \ln(\beta_{0j}) + \beta_{1j}X + \ln \epsilon_j \quad (j = 1, 2, \dots, k) \quad (8.2.5)$$

or simply

$$y_j^* = \beta_{0j}^* + \beta_{1j}X + \epsilon_j^* \quad (j = 1, 2, \dots, k). \quad (8.2.6)$$

On the other hand, models specified by (8.2.4) cannot be so transformed technically although the first member of the expression on the right of this model is linearizable. Indeed, (8.2.4) specifies the more simple forms of an intrinsically nonlinear model.

Now consider fitting the linearized form of (8.2.3) and suppose one is interested in this linearized form for prediction purposes. If, in addition, one is interested in the predictive equation for the sum of the transformed form of the components, then the parameters of the latter model are determined by additivity from the component models. This is the case since we have

$$\begin{aligned} y_T^* &= \sum_{j=1}^k y_j^* = \sum_{j=1}^k (\beta_{0j}^* + \beta_{1j}X + \epsilon_j^*) \\ &= \sum_{j=1}^k \beta_{0j}^* + (\sum_{j=1}^k \beta_{1j})X + \sum_{j=1}^k \epsilon_j^* \\ &= \beta_{0T}^* + \beta_{1T}X + \epsilon_T^*. \end{aligned} \quad (8.2.7)$$

Thus, as in the ordinary model, additivity holds here as long as the variable of interest is the transformed version of the dependent variable y_j .

Kozak (1970) made this point in his paper and, therefore, it is not new. Suppose, on the other hand, that one is really interested in the original nonlinear form as a predictive equation. In this case, the linearization is only an intermediate step aimed largely at simplifying the process of estimation. Note that in this case some of the parameters estimated from the linearized equation would need to be further transformed before being inserted in a nonlinear predictive equation. Now, if the nonlinear fitted analogue of (8.2.7) were desired, note that using additivity one would need to back-transform the expression

$$\hat{y}_T^* = \sum_{j=1}^k \widehat{\ln y_j} = \sum_{j=1}^k \widehat{\ln \beta_{0j}} + (\sum_{j=1}^k \hat{\beta}_{1j})X \quad (8.2.8)$$

to obtain

$$\begin{aligned} \hat{y}_T &= \prod_{j=1}^k \hat{y}_j = (\prod_{j=1}^k \hat{\beta}_{0j}) e^{(\sum_{j=1}^k \hat{\beta}_{1j})X} \\ &= \prod_{j=1}^k [\hat{\beta}_{0j} e^{\hat{\beta}_{1j}X}]. \end{aligned} \quad (8.2.9)$$

The result in (8.2.9) warrants some comment. Perhaps the most important of such comments is the following. If one is interested in predicting total biomass, say, as a sum of the components y_j using a model of the form (8.2.3), then one must not do so by invoking additivity of the transformed version of (8.2.3) and then re-transform (that is, back-transform). If one does so, then one gets an equation which predicts the product of the components rather than their sum. In a nutshell, one gets the wrong predictive equation. Herein lies the real virtue of a proper analysis of a modelling situation. Thus for models of the form (8.2.3), a total predictive equation of the same form cannot be determined from additivity of the parameters of the transformed component equations. It may be determined at least arithmetically, however, as a simple sum of the corresponding fitted nonlinear component equations, though the predictive

merits of such an equation may be debatable.

With respect to models of the form (8.2.4), it has been indicated that such models do not admit transformation to linear form. Therefore, the parameters would be estimated using any of a number of known iterative search techniques. As with models of the form (8.2.3), however, a predictive total equation cannot be obtained here by appealing to the additivity property since the parameters of the total predictive equation cannot be determined by adding corresponding parameters in the component equations. However, if simple prediction was the objective, then a predictive total equation may be obtained by simply adding up the predictive component equations. Once again, the predictive usefulness of such a model is largely an open question.

The foregoing discussion indicates that the additivity property, which holds almost universally for linear models, does not carry over, in general, to the class of nonlinear models. This precludes, for instance, the use of the notion of additivity in inventory and/or biomass studies if nonlinear models are used for prediction.

CHAPTER IX

9.0 COMPUTATIONAL CONSIDERATIONS

We have attempted to present, in preceding chapters, a theory of estimation and inference for the additivity problem and to indicate generalizations and extensions to other types of models. The essential objective of the discourse has been to present the additivity problem as perceived here within the general framework of linear model theory. One hopes that this objective has been achieved to a large extent. However, our derivation of expressions for estimators and associated statistics, particularly in chapter V, leaves one important question largely unanswered. This question is: Does the conditioning principle introduced to handle the additivity problem in general call for new computing subroutines or algorithms in order to obtain estimates and other statistics? We show, in this chapter, that no such subroutines or algorithms are required. All estimates and associated statistics can be computed using existing system-based software such as is provided by the various statistical packages. Examples of such packages are MIDAS (The Michigan Interactive Data Analysis System, The University of Michigan, Ann Arbor, Michigan), BMDP (Biomedical computer programmes P-series, University of California Press, Los Angeles, California), SAS (Statistical Analysis System, SAS Institute, Raleigh, North Carolina), and SPSS (Statistical Package for the Social Sciences, McGraw-Hill Inc., New York). At computing

installations where SAS is not available, BMDP is perhaps the most commendable package to use mainly because it has options for generating diagnostic plots and other information valuable in model choice and validation.

To motivate the derivation of the main result of this chapter, consider the estimation problem associated with fitting component equation j ($j = 1, 2, \dots, k$) as presented in chapter V. More specifically, recall that in fitting equation j , where equation j contains only statistically important independent variables, the estimator for β_j^* is given by

$$\hat{\beta}_j^* = \begin{pmatrix} \phi \\ \hat{b}_{p_j} - T_{p_j q_j} T_{q_j q_j}^{-1} \hat{b}_{q_j} \end{pmatrix}, \quad (9.0.1)$$

where \hat{b}_{q_j} and \hat{b}_{p_j} are given by the partition

$$\hat{\beta}_j = \begin{pmatrix} \hat{b}_{q_j} \\ \hat{b}_{p_j} \end{pmatrix} \quad (9.0.2)$$

of $\hat{\beta}_j$ obtained from fitting the full model (with all independent variables) while $T_{p_j q_j}$ and $T_{q_j q_j}$ are obtained from a corresponding partitioning of $(X^t X)^{-1}$ in the form

$$(X^t X)^{-1} = \begin{pmatrix} T_{q_j q_j} & T_{q_j p_j} \\ T_{p_j q_j} & T_{p_j p_j} \end{pmatrix}. \quad (9.0.3)$$

With a corresponding partitioning of the X matrix in the form

$X = \begin{pmatrix} X_{q_j} & X_{p_j} \end{pmatrix}$, the result in (9.0.1) is given equivalently by (see equation (5.1.18) in chapter V)

$$\hat{\beta}_j^* = \begin{pmatrix} \phi \\ (T_{p_j p_j} - T_{p_j q_j} T_{q_j q_j}^{-1} T_{q_j p_j}) X_{p_j}^t y_j \end{pmatrix}. \quad (9.0.4)$$

Consider now fitting y_j on the set X_{p_j} only (i.e., so that the coefficients of the components of X_{q_j} are set identically equal to zero). It is shown below that the estimator $\hat{\beta}_{p_j}^*$, say, from the latter fit is precisely equal to the non-null part of $\hat{\beta}_j^*$ in (9.0.4). Hence $\hat{\beta}_j^*$ is completely specified by simply fitting y_j on the set X_{p_j} . The consequence of this result is that no special algorithm is necessary to obtain $\hat{\beta}_j^*$ beyond those already available on standard statistical packages. We demonstrate this result formally by stating and proving the following theorem.

Theorem: Let $\chi = \{X_1, X_2, \dots, X_{m-1}\}$ be a set of predictor variables and consider fitting the linear model (with intercept)

$$y_j = X\beta_j + \epsilon_j$$

subject to a subset b_{q_j} of β_j of order q_j being equal to the zero vector. Here $X = (X_0 | X_1 | \dots | X_{m-1})$ is an $n \times m$ ($n > m$) matrix of full column rank and we assume, as usual, that $\epsilon_j \sim N(\phi, I\sigma_j^2)$. Denote the solution from fitting this constrained model by $\hat{\beta}_j^*$. Now partition X so that $X = (X_{q_j} | X_{p_j})$ for $q_j + p_j = m$, $0 < q_j < m$ in accordance with the constraining of β_j so that

$$(X^t X)^{-1} = \begin{pmatrix} X_{q_j}^t X_{q_j} & X_{q_j}^t X_{p_j} \\ X_{p_j}^t X_{q_j} & X_{p_j}^t X_{p_j} \end{pmatrix}^{-1} = \begin{pmatrix} T_{q_j q_j} & T_{q_j p_j} \\ T_{p_j q_j} & T_{p_j p_j} \end{pmatrix}.$$

Then $\hat{\beta}_j^*$ is given by

$$\hat{\beta}_j^* = \begin{pmatrix} \phi \\ (T_{p_j p_j}^{-1} - T_{p_j q_j} T_{q_j q_j}^{-1} T_{q_j p_j}) X_{p_j}^t y_j \end{pmatrix}.$$

Furthermore, if $\hat{b}_{p_j}^*$ is the solution from fitting y_j on the set X_{p_j} (that is, ignoring X_{q_j}), then

$$\hat{b}_{p_j}^* = (T_{p_j p_j} - T_{p_j q_j} T_{q_j q_j}^{-1} T_{q_j p_j}) X_{p_j}^t y_j.$$

We prove the above theorem by utilizing a well-known theorem in linear algebra concerning the inverse of a partitioned matrix. This theorem is stated here as a lemma, without proof, and we refer the reader to Graybill (1976, p. 19) or any standard text in linear algebra for a proof.

Lemma: Let W be an $n \times n$ nonsingular matrix that is partitioned as follows:

$$W = \begin{pmatrix} W_{11} & W_{12} \\ W_{21} & W_{22} \end{pmatrix}$$

where W_{ij} has size $n_i \times n_j$ for $i, j = 1, 2$ ($n_1 + n_2 = n$, $0 < n_1 < n$).

If $|W_{11}| \neq 0$ and $|W_{22}| \neq 0$, then W^{-1} is given by

$$W^{-1} = \begin{pmatrix} [W_{11} - W_{12}W_{22}^{-1}W_{21}]^{-1} & -W_{11}^{-1}W_{12}[W_{22} - W_{21}W_{11}^{-1}W_{12}]^{-1} \\ -W_{22}^{-1}W_{21}[W_{11} - W_{12}W_{22}^{-1}W_{21}]^{-1} & [W_{22} - W_{21}W_{11}^{-1}W_{12}]^{-1} \end{pmatrix}.$$

Proof of Theorem: The expression for $\hat{\beta}_j^*$ given in the theorem follows from Searle (1971, pp. 113-114), as demonstrated earlier in the thesis.

It remains to show that the second part of the theorem holds.

With $(X^t X)$ partitioned as in the theorem and using the above lemma, it follows that

$$(X^t X)^{-1} = \begin{pmatrix} T_{q_j q_j} & T_{q_j p_j} \\ T_{p_j q_j} & T_{p_j p_j} \end{pmatrix}$$

where

$$T_{q_j q_j} = [X_{q_j q_j}^t - X_{q_j p_j}^t (X_{p_j p_j}^t)^{-1} X_{p_j q_j}^t]^{-1}$$

$$T_{q_j p_j} = -(X_{q_j q_j}^t)^{-1} X_{q_j p_j}^t [X_{p_j p_j}^t - X_{p_j q_j}^t (X_{q_j q_j}^t)^{-1} X_{q_j p_j}^t]^{-1}$$

$$T_{p_j q_j} = -(X_{p_j p_j}^t)^{-1} X_{p_j q_j}^t [X_{q_j q_j}^t - X_{q_j p_j}^t (X_{p_j p_j}^t)^{-1} X_{p_j q_j}^t]^{-1}$$

$$T_{p_j p_j} = [X_{p_j p_j}^t - X_{p_j q_j}^t (X_{q_j q_j}^t)^{-1} X_{q_j p_j}^t]^{-1}.$$

Note that the existence of $(X_{q_j q_j}^t)^{-1}$ and $(X_{p_j p_j}^t)^{-1}$ follows from

the fact that X is of full column rank. Note also that the solution

$\hat{b}_{p_j}^*$ obtained from fitting y_j on the set X_{p_j} must also be given by

$$\hat{b}_{p_j}^* = (X_{p_j p_j}^t)^{-1} X_{p_j j}^t y_j.$$

Therefore, to complete the proof of the theorem, we need only show that

$$(T_{p_j p_j} - T_{p_j q_j} T_{q_j q_j}^{-1} T_{q_j p_j}) = (X_{p_j p_j}^t)^{-1}. \quad \text{To this end, note that}$$

$$\begin{aligned} T_{p_j p_j} - T_{p_j q_j} T_{q_j q_j}^{-1} T_{q_j p_j} &= [X_{p_j p_j}^t - X_{p_j q_j}^t (X_{q_j q_j}^t)^{-1} X_{q_j p_j}^t]^{-1} \\ &\quad - \{ (X_{p_j p_j}^t)^{-1} X_{p_j q_j}^t \\ &\quad [X_{q_j q_j}^t - X_{q_j p_j}^t (X_{p_j p_j}^t)^{-1} X_{p_j q_j}^t]^{-1} \\ &\quad [X_{q_j q_j}^t - X_{q_j p_j}^t (X_{p_j p_j}^t)^{-1} X_{p_j q_j}^t] (X_{q_j q_j}^t)^{-1} X_{q_j p_j}^t \\ &\quad [X_{p_j p_j}^t - X_{p_j q_j}^t (X_{q_j q_j}^t)^{-1} X_{q_j p_j}^t]^{-1} \} \end{aligned}$$

$$\begin{aligned}
&= [X_{P_j}^t X_{P_j} - X_{P_j}^t X_{q_j} (X_{q_j}^t X_{q_j})^{-1} X_{q_j}^t X_{P_j}]^{-1} \\
&\quad - \{ (X_{P_j}^t X_{P_j})^{-1} X_{P_j}^t X_{q_j} (X_{q_j}^t X_{q_j})^{-1} X_{q_j}^t X_{P_j} \\
&\quad [X_{P_j}^t X_{P_j} - X_{P_j}^t X_{q_j} (X_{q_j}^t X_{q_j})^{-1} X_{q_j}^t X_{P_j}]^{-1} \} \\
&= [I - (X_{P_j}^t X_{P_j})^{-1} X_{P_j}^t X_{q_j} (X_{q_j}^t X_{q_j})^{-1} X_{q_j}^t X_{P_j}] \\
&\quad [X_{P_j}^t X_{P_j} - X_{P_j}^t X_{q_j} (X_{q_j}^t X_{q_j})^{-1} X_{q_j}^t X_{P_j}]^{-1} \\
&= (X_{P_j}^t X_{P_j})^{-1} [X_{P_j}^t X_{P_j} - X_{P_j}^t X_{q_j} (X_{q_j}^t X_{q_j})^{-1} X_{q_j}^t X_{P_j}] \\
&\quad [X_{P_j}^t X_{P_j} - X_{P_j}^t X_{q_j} (X_{q_j}^t X_{q_j})^{-1} X_{q_j}^t X_{P_j}]^{-1} \\
&= (X_{P_j}^t X_{P_j})^{-1},
\end{aligned}$$

which is what we set out to show. Q.E.D.

Thus, we have established by the above theorem that the estimation problem associated with the generalized additivity problem as developed in chapter V does not require that new computing algorithms be developed to obtain estimates and related statistics. One simply fits a model containing what are construed to be statistically important independent variables. The parameter estimate corresponding to such a fit can then be augmented to the corresponding estimate for a full model constrained so that the unimportant independent variables have coefficients of zero. The virtue of the estimator of the parameter vector for component equation j ($j = 1, 2, \dots, k$) given in chapter V is that it is of appropriate size for additivity. However, whether $\hat{\beta}_j^*$ is obtained directly as in chapter V or indirectly by fitting a subset X_{P_j} and then augmenting the resulting estimator, its components may need to be permuted before invoking additivity to obtain the estimator for the corresponding total equation.

Such permuting ensures that appropriate components of $\hat{\beta}_j^*$ ($j = 1, 2, \dots, k$) are added to obtain $\hat{\beta}_{TC}$.

Finally, the theorem proved above is based upon the distributional assumption $\varepsilon_j \sim N(\phi, I\sigma_j^2)$. Clearly, obvious modifications in the theorem would make it hold for the case $\varepsilon_j \sim N(\phi, V\sigma_j^2)$ for V positive definite. Other generalizations are also possible.

CHAPTER X

10.0 SOME ILLUSTRATIVE EXAMPLES

In this chapter, some examples are given that illustrate the application of the theory presented in the discourse. First, however, a somewhat detailed analysis is given aimed at assessing the tenacity of the assumption of multivariate normality of $y_T = \sum_{j=1}^k y_j$ for each of three data sets. As was stated in chapter VII, the usual inferences for the total model depend critically upon the assumption that $y_T = \sum_{j=1}^k y_j$ is multivariate normal. Since there is no prior knowledge that the assumption holds, it is necessary to assess for multivariate normality in order to more appropriately qualify any inference statements in the examples.

10.1 Assessing Multivariate Normality of y_T

Koziol's (1982) method for assessing joint multivariate normality of the components y_1, y_2, \dots, y_k was used on three data sets. The first of these data sets is that used by Kozak (1970) to illustrate the additivity result presented in his paper. The second data set is British Columbia coastal western hemlock data used by Kurucz (1969). The third data set is western hemlock data from various parts of British Columbia and was obtained from the ENFOR project (Williams, 1983, personal communication). Note that ENFOR is an acronym for Energy from Forests.

Koziol's (1982) method for assessing multivariate normality is based upon a Cramér-von Mises type statistic J_n , which is computed as follows:

1. Given X_1, X_2, \dots, X_n are random k -dimensional vectors, calculate $\bar{X} = (\bar{X}_1, \bar{X}_2, \dots, \bar{X}_k)^t$ and S . Here S is the sample variance-covariance matrix of the n vectors and is $k \times k$.
2. Calculate the sample Mahalanobis squared distances Y_1, Y_2, \dots, Y_n defined by

$$Y_i = (X_i - \bar{X})^t S^{-1} (X_i - \bar{X}).$$

3. Put $Z_i = F_{(k)}(Y_i)$, $i = 1, 2, \dots, n$ and order the Z_i in ascending order so that $Z_{(1)} \leq Z_{(2)} \leq \dots \leq Z_{(n)}$. $F_{(k)}(Y_i)$ here denotes the area under the chi-square density function with k degrees of freedom between the limits of zero and Y_i (i.e., $F_{(k)}(Y_i) = \Pr[Y \leq Y_i]$).
4. Calculate J_n using

$$J_n = \sum_{i=1}^n [Z_{(i)} - (i - \frac{1}{2})/n]^2 + (12n)^{-1}.$$

Note that with three components, k is equal to three in our case. The three data sets are reproduced in Appendix I (a,b,c). The first two of the data sets in the appendix are reported in imperial units, while the third data set is given in metric units. However all analyses reported in this dissertation were carried out in metric units.

Before presenting details of the test for joint multivariate normality of the y_j 's, it is worth pointing out some technical considerations which simplify considerably the computation of the Koziol statistic J_n . In particular, this simplifies the computational formula for the sample Mahalanobis squared distances Y_i ($i = 1, 2, \dots, n$). Observe that the joint distribution of the y_j 's in this case is conditional upon the independent variables (the X 's). As a result of this and based upon the notion that the regression of y_j on the X is important (significant), the sample Mahalanobis squared distances are given by

$$Y_i = (y_i - \hat{y}_i)^t S^{-1} (y_i - \hat{y}_i)$$

or by

$$Y_i = \hat{\epsilon}_i^t S^{-1} \hat{\epsilon}_i$$

where $y_i - \hat{y}_i = \hat{\epsilon}_i = (y_{i1} - \hat{y}_{i1}, y_{i2} - \hat{y}_{i2}, \dots, y_{ik} - \hat{y}_{ik})^t$ and $S = (n - 1)^{-1} \hat{\epsilon} \hat{\epsilon}^t$ (note that $\hat{\epsilon} = (\hat{\epsilon}_1, \hat{\epsilon}_2, \hat{\epsilon}_3)^t$). The distinction between Y_i and y_i must be borne in mind here. The residual vectors used in computing Y_i above are those obtained from fitting component equations using only statistically significant independent variables for each data set. These equations are those used to obtain conditioned total predictive equations in examples 1, 2, and 3 that follow. It should be emphasized that a test for multivariate normality of $\hat{\epsilon}_1, \hat{\epsilon}_2, \dots, \hat{\epsilon}_k$ is equivalent here to a test for multivariate normality of y_1, y_2, \dots, y_T . Thus if $\hat{\epsilon}_1, \hat{\epsilon}_2, \dots, \hat{\epsilon}_k$ are jointly multivariate normal, one can speak of the multivariate normality of y_1, y_2, \dots, y_k and hence of ϵ_T and y_T .

In computing the statistic J_n for each data set, an APL programme was used to calculate the Y_i as specified above using an IBM 5100 Portable Computer. This computer is located in the Mathematics Annex at the University of British Columbia. APL is an extremely efficient language when one is dealing with matrix computations. The computation of the chi-square probabilities in step 3 was achieved by calling the IMSL (International Mathematical

and Statistical Libraries) subroutine MDCH which computes cumulative chi-square probabilities. A short fortran programme was written to call MDCH (see Appendix Id, PROGRAMME 1). Note that although DF in programme 1 is specified as 2.0, $DF = 3.0$ for the first part of this assessment problem. Finally, the ordered chi-square probabilities from step 3 were used in another fortran programme to calculate J_n (see PROGRAMME 2 in Appendix Id). The results of Koziol's (1982) test on the three data sets are summarized in Table 1 below.

Table 1. Results of Koziol's (1982) test for multivariate normality on three data sets

Data Set	Sample Size	DF	Computed Koziol Statistic (J_n)	p-value
Kozak (1970)	10	3.0	0.05799	> 0.15
Kurucz (1969)	18	3.0	0.86078	< 0.01
ENFOR	48	3.0	4.27682	<<0.01

The p-values in Table 1 are obtained by comparison with Koziol's Table 1 (Koziol, 1982). It is to be emphasized that due to small sample sizes associated with Kozak's (1970) and Kurucz's (1969) data, our p-values may be somewhat off. However, on the basis of these results, the assumption of joint multivariate normality will be entertained for Kozak's data but not for the other two data sets. Note that this conclusion is quite reasonable for the ENFOR data because of the moderate ($n = 48$) sample size.

In view of the above results (ignoring the small sample sizes in the first two data sets) it is reasonable to expect that y_T is multivariate normal for Kozak's data since it is reasonable that y_1, y_2, y_3 are jointly

multivariate normal in this data set. On the other hand, the above results suggest only that for the two other data sets y_T may or may not be normal, since it is possible for y_T to be multivariate normal even when y_1, y_2, y_3 are not jointly multivariate normal.

For both the Kurucz (1969) and ENFOR data, it was considered of some interest to check for joint bivariate normality of the y_j 's. Accordingly, Koziol's test for multivariate normality was applied to pair-wise y_j 's; thus three tests were performed on each data set. The results are summarized in Table 2 below.

Table 2. Results of Koziol's test for bivariate normality

	DF	Sample Size	Koziol's Computed Statistic (J_n)	p-value
Kurucz's (1969) data				
$(\hat{\epsilon}_1, \hat{\epsilon}_2)$	2.0	18	0.8484	<0.01
$(\hat{\epsilon}_1, \hat{\epsilon}_3)$	2.0	18	0.5023	<0.01
$(\hat{\epsilon}_2, \hat{\epsilon}_3)$	2.0	18	0.6873	<0.01
ENFOR data				
$(\hat{\epsilon}_1, \hat{\epsilon}_2)$	2.0	48	5.4432	<<0.01
$(\hat{\epsilon}_1, \hat{\epsilon}_3)$	2.0	48	3.7356	<<0.01
$(\hat{\epsilon}_2, \hat{\epsilon}_3)$	2.0	48	3.1189	<<0.01

The results in Table 2 indicate that the assumption of joint bivariate normality is rejected essentially in every case. This result is not unexpected since, having rejected trivariate normality, one expects that bivariate normality should fail to obtain in at least one of the three cases. It is also probably adequate to check bivariate normality and reject trivariate normality the first time bivariate normality fails to

hold.

It has been shown so far that joint multivariate normality of y_1, y_2, y_3 does not appear to hold for the Kurucz (1969) and ENFOR data while multivariate normality will be entertained for the Kozak (1970) data. It should be emphasized again that, in general, one should be more cautious in accepting multivariate normality for the Kozak data because of the very small sample size. However, for purposes of the examples to follow, multivariate normality will be entertained. Once again, it is reasonable then to assume y_T multivariate normal for the Kozak data. However, one is unable to decide whether or not y_T is multivariate normal for the Kurucz and ENFOR data. A direct examination of the behaviour of $y_T = \sum_{j=1}^k y_j$ is necessary to make a judgement concerning its normality or non-normality.

One way in which information can be obtained concerning the multivariate normality or lack of it for $y_T = \sum_{j=1}^k y_j$ is to fit the component models and investigate the behaviour of the empirical distribution of $\hat{e}_T = \sum_{j=1}^k \hat{e}_j$. This can be achieved, in part, by plotting a histogram of \hat{e}_T or a normal probability plot of \hat{e}_T . Unfortunately, these procedures require large enough sample sizes in order for the plots to be reasonably interpretable. Largely because of this, it was possible to examine such plots in this study only for the ENFOR data because of its moderate sample size ($n = 48$). The Kurucz data were obviously too small to be examined by this procedure.

Three component models were fitted using the ENFOR data. Bole biomass was regressed on D^2H and DCL, where D denotes diameter at breast-height, H denotes height, and CL crown length. Bark biomass was regressed on D^2H and HCL and crown biomass was regressed on DCL and HCL. The

residuals from these fitted component equations were added up and a histogram and normal probability plot constructed using the BMDP P:5D subroutine. If the histogram of $\hat{\epsilon}_T$ looks sufficiently bell-shaped, it is reasonable to conclude that y_T is normal. Similarly, normality of y_T would be suggested by a sufficiently linear normal probability plot. The plots are given in Appendix II(a,b) and both suggest that ϵ_T and hence y_T is normal for the ENFOR data.

Based upon results of this section, we can proceed as though y_T were multivariate normal for the Kozak and ENFOR data but are unable to say whether y_T is multivariate normal or not for the Kurucz data.

10.2 Example 1

In this section, the data given by Kozak (1970) are used to apply additivity theory as presented in chapters IV and V of this thesis. As indicated earlier, though the data are reproduced in the appendix in imperial units, all calculations here are in metric units. It is further assumed throughout that the biomass components are independent. Admittedly, this may be a tenuous assumption; however, we use it largely for purposes of demonstrating the application of the theory. For the effect of dependence on estimation and inference see the discussion in chapter VII. Let us assume further that $\epsilon_j \sim N(\phi, I\sigma_j^2)$.

First, consider fitting the component equations using both diameter and the square of diameter as independent variables. Recall that this is the case considered by Kozak (1970). Then the fitted component equations are given by

$$\hat{y}_1 = 131.39 - 19.037X + 0.95195X^2, \quad \hat{R}^2 = 0.9923$$

$$\hat{y}_2 = -1.12 + 0.205X + 0.02980X^2, \quad \hat{R}^2 = 0.9965$$

$$\hat{y}_3 = -13.08 + 1.136X + 0.08361X^2, \quad \hat{R}^2 = 0.9605.$$

The corresponding total fitted equation is

$$\hat{y}_T = 117.19 - 17.696X + 1.06540X^2, \quad \hat{R}^2 = 0.9948.$$

One can check easily that the coefficients of the total equation are obtained by adding corresponding coefficients of the fitted component equations, as Kozak (1970) demonstrated. The regression sum of squares for the total equation is 289060, to five-digit accuracy, and given that $X^t y_T = (2300.6, 56183, 1448300)^t$, one can check easily that this is the result one obtains using additivity and equation (4.2.6) of the thesis.

Next, Kozak (1970) reports that when only statistically significant (important) independent variables are used in fitting the component equations, the first equation (\hat{y}_1) involves both X and X^2 , the second (\hat{y}_2) only X^2 , and the last (\hat{y}_3) only X^2 also. The metric analogues of Kozak's specification of these fitted equations are

$$\hat{y}_1 = 131.39 - 19.037X + 0.95195X^2, \quad \hat{R}^2 = 0.9923$$

$$\hat{y}_2 = 0.822 + 0.03471X^2, \quad \hat{R}^2 = 0.9962$$

$$\hat{y}_3 = -2.342 + 0.11079X^2, \quad \hat{R}^2 = 0.9588.$$

In accordance with the extension of the concept of additivity as developed in chapter V, there is a total equation determined by additivity of the coefficients of the preceding equations. In fact, this equation is given by

$$\hat{y}_{TC} = 129.87 - 19.037X + 1.09745X^2.$$

The sum of squares regression for this conditioned total equation is given by

$$\begin{aligned} \hat{\beta}_{TC}^t X^t y_T - n \bar{y}_T^2 &= 818660 - 531860.4564 \\ &= 286799.5436. \end{aligned}$$

Since the total corrected sum of squares is 290580, it follows that \hat{R}^2 corresponding to the total conditioned equation is given by

$$\hat{R}^2 = 0.9870.$$

It is worth remarking that in terms of \hat{R}^2 , this model fits the data almost as well as the unrestricted total model, with an \hat{R}^2 of 0.9948. Other aspects of this problem, including computational details, are provided in a more detailed example in Appendix III.

In connection with this problem and related problems concerning additivity, the question naturally arises whether the variables in the conditioned total equation remain statistically significant after being incorporated into the total conditioned equation. The answer appears to be that they would be statistically significant if the variables in the conditioned component equations are not very highly correlated. However, this may not be the case if the variables are highly correlated. It should be pointed out that this has not been checked thoroughly and, thus, should be viewed here as largely a conjecture. For the Kozak data, however, the contribution to the total conditioned equation of each variable was checked by computing the increase in residual sum of squares when a particular variable is omitted from the total conditioned equation. The following partial F-values were calculated:

$$F_X = 96.27, \quad F_{X^2} = 15.64$$

The degrees of freedom for these partial F values are 1 and 7, respectively. It is clear from these results that both diameter at breast height and its square are statistically important in the total conditioned equation.

10.3 Example 2

In this example, use is made of the western hemlock data from coastal British Columbia to go through the basic computational results as in the previous example. These data were used by Kurucz (1969) and are given in Appendix I(b).

First it should be noted that, as discussed in section 10.1, it has not been possible to determine whether y_T is multivariate normal for these data or not. Therefore, inferential results given in this section relating to these data must not be viewed as strictly valid. The essence of this example is mainly to demonstrate use of the concept of additivity computationally. One would need to check that y_T is reasonably multivariate normal for inference statements to carry full weight. The calculations here are carried out in metric units and the assumption is made that the components y_j ($j = 1, 2, 3$) are independent.

An all-combinations (all subsets) procedure provided by the BMDP package (P:9R) was used to find the best variable subsets for predicting component biomass. Three components were recognized for purposes of this analysis, namely bole (y_1), bark (y_2), and crown (branches + foliage = y_3). The crown component was obtained by simply adding branch and fine branch components for individual trees. Using \hat{R}^2 as a selection criterion, the best equations were found to be

$$\hat{y}_1 = -75.708 + 0.01330X_2, \quad \hat{R}^2 = 0.9907$$

$$\hat{y}_2 = -25.782 + 0.00203X_2, \quad \hat{R}^2 = 0.9309$$

$$\hat{y}_3 = -24.765 + 0.095895X_1, \quad \hat{R}^2 = 0.8122$$

where $X_1 = (\text{height})^2$ and $X_2 = (\text{height})(\text{diameter})^2$. A corresponding unrestricted total fitted equation, using X_1 and X_2 , is given by

$$\hat{y}_T = 33.542 + 0.26819X_1 + 0.01993X_2, \quad \hat{R}^2 = 0.9655.$$

The total corrected sum of squares corresponding to the latter fit is 182559547. As in the preceding example, a conditioned total fitted equation is obtained by additivity as

$$\hat{y}_{TC} = -126.255 + 0.95895X_1 + 0.01533X_2.$$

Since $X^t y_T = (53205, 113670000, 15112000000)^t$, it follows that the sum of squares regression associated with the latter conditioned equation is

$$\begin{aligned}\hat{\beta}_{TC}^t X^t y_T - n\bar{y}_T^2 &= 332291089.2 - 18(2955.8)^2 \\ &= 175029523.7.\end{aligned}$$

Therefore, the \hat{R}^2 associated with this total conditioned equation is 0.9588. Again, if the assumption of normality of y_T held, one would conclude from this that the conditioned total equation performs well when compared with the unrestricted total equation. It may be noted that because of the great variability in the size of the trees in this data set, one needs to be careful about the predictive goodness of these models. Indeed, as mentioned elsewhere in this thesis, we are much less concerned here with using the best equations in a particular sense than with demonstrating certain aspects of additivity.

10.4 Example 3

The next example is based upon the ENFOR data for western hemlock (see Appendix I(c)). We restrict details to the level of previous examples. Using an all-combinations procedure as in the previous example, the following equations were found to be the best for predicting component biomass:

$$\begin{aligned}\hat{y}_1 &= 6.49538 + 0.01541D^2H - 0.12258DCL, \quad \hat{R}^2 = 0.9836 \\ \hat{y}_2 &= 0.93179 + 0.00247D^2H - 0.03112HCL, \quad \hat{R}^2 = 0.9561 \\ \hat{y}_3 &= -4.82066 + 0.31477DCL - 0.23344HCL, \quad \hat{R}^2 = 0.8424.\end{aligned}$$

The unrestricted total biomass equation is given by

$$\hat{y}_T = 2.05138 - 0.000074D^2H + 0.09319DCL - 0.0667HCL, \hat{R}^2 = 0.8674.$$

The corresponding total conditioned equation is given by

$$\hat{y}_{TC} = 2.60651 + 0.01788D^2H + 0.19219DCL - 0.26456HCL.$$

The sum of squares regression due to fitting the latter equation is given

by $\hat{\beta}_{TC}^t x_T^t y_T - n\bar{y}_T^2 = 2876.506$. Hence the \hat{R}^2 corresponding to this model

is 0.8215 which compares favourably with that of the unrestricted total equation. Note that even when crown variables are used for predicting crown biomass, the \hat{R}^2 is still in 0.80-0.90 range for that component.

In Appendix III, the computational details relating to the additivity problem are given using Kozak's (1970) data again. The objective there is to show how the various statistics are computed, especially the variances of the parameters in the total conditioned equation.

CHAPTER XI

CONCLUSIONS AND REMARKS

In the discourse we have generalized the additivity problem as originally posed in the context of forestry by Kozak (1970). It has been shown that the statistical theory of estimation and inference for the generalized additivity problem as defined here is constructible within the general framework of general linear model theory. It is important to recognize that both estimation and inference theory is, in general, dependent upon distributional assumptions for the ϵ_j ($j = 1, \dots, k$) and upon whether the ϵ_j are dependent or not. When the ϵ_j are dependent, it has been shown that additivity as defined here, does not hold. Furthermore, inference theory relating to the total model is complicated by the fact that although the components may follow normal distributions, it does not follow automatically that their sum is also normal. This suggests a need to investigate, or otherwise, justify the normality of y_T before inference can be drawn about it when dependence obtains among the components. In particular, it would be useful if future studies in this area could address the problem relating to the distribution of y_T directly using large enough data sets along the lines indicated in section 10.1. Large data sets that might become available through projects such as the ENFOR project might make such studies possible and worthwhile. Other directions of further investigation might be the determination of the form of the dependence among components. This might simplify the problem of determining the

distributional behaviour of y_T .

The problem of additivity has also been seen to lead to interesting but, as yet, unsolved problems in multivariate distribution theory. This is obviously a fruitful line of further research for those who are theoretically inclined.

One of the interesting results obtained here is that the additivity problem is naturally extendible to the class of linear models known as classificatory models generally encountered in designed experiments. This extension must not be construed to be accidental since any classificatory model can, in general, be expressed in regression form. The additivity problem does not, however, extend to the class of intrinsically nonlinear models. Hence the usefulness of theory relating to the additivity problem in mensurational studies involving nonlinear functions would, at best, be minimal. However, the theory should find wide applicability among ecologists and quantitative scientists interested in the assessment of biomass.

The additivity problem does not require the construction of new computing subroutines as clearly demonstrated in chapter IX. This should make it especially easy to use the theory of additivity as developed here. Finally, the examples given in the preceding chapter show that the concept of additivity is quite practical and realistic and statistically appealing.

REFERENCES

- Andrews, D. F., Gnanadesikan, R., and Warner, J. L. 1973. Methods of assessing multivariate normality. In Multivariate Analysis 3, Krishnaiah, P. R. (ed.), pp. 95-116, Academic Press, New York.
- Arnold, S. F. 1981. The theory of linear models and multivariate analysis. John Wiley and Sons, Inc., New York.
- Behboodian, J. 1972. A simple example of some properties of normal random variables. Amer. Math. Monthly, 79: 632-634.
- Berkson, J. 1950. Are there two regressions? J. Amer. Stat. Assoc., 45: 163-180.
- Brown, J. K. 1963. Crown weights in red pine plantations. U.S.D.A. Forest Serv., Lake States Forest Exp. Sta., Res. No. LS-19.
- Brown, J. K. 1965. Estimating crown fuel weights of red pine and jack pine. U.S.D.A. Forest Serv., Lake States Forest Exp. Sta., Res. Pap. LS-20.
- Chaturvedi, O. P., and Singh, J. S. 1982. Total biomass and biomass production of Pinus roxburghii trees growing in all-aged natural forests. Can. J. For. Res. 12(3): 632-640.
- Chiyenda, S. S. 1974. Prediction and estimation of slash based on total and component biomass of standing trees and on measurement of post-harvest residues. Unpublished B.S.F. thesis, The University of British Columbia, Vancouver, Canada.
- Cox, D. R. 1968. Notes on some aspects of regression analysis (with discussion). J. Roy. Stat. Soc., A131: 265-279.
- Crow, T. R. 1971. Estimation of biomass in an even-aged stand-- regression and 'mean tree' techniques. In Forest Biomass Studies. IUFRO Sect. 25: Yield and Growth Working Group on Forest Biomass Studies, University of Main Press, Oronto, pp. 35-47.
- Draper, N. R., and Smith, H. 1981. Applied regression analysis. John Wiley and Sons, Inc., New York (second edition).
- Dyer, R. F. 1967. Fresh and dry weight, nutrient elements and pulping characteristics of northern white cedar, Thuja occidentalis. Maine Agric. Exp. Sta., Techn. Bull. 27.
- Efron, B. 1979. Bootstrap methods: Another look at the jackknife. Ann. Stat. 7: 1-26.
- Freedman, B., Duinker, P. N., Morash, R., and Prager, U. 1982. A comparison of measurements of standing crops of biomass and nutrients in a conifer stand in Nova Scotia. Can. J. For. Res. 12(3): 494-502.

- Gallant, A. R. 1971. Statistical inference for nonlinear regression models. Unpublished Ph.D. dissertation, Iowa State University, Ames, Iowa.
- Graybill, F. A. 1976. Theory and application of the linear model. Duxbury Press, New York.
- Hawkins, D. M. 1981. A new test for multivariate normality and homoscedasticity. *Technometrics*, 23: 105-110.
- Healy, M. J. R. 1968. Multivariate normal plotting. *Appl. Stat.* 17: 157-161.
- Hogg, R. V. 1979. An introduction to robust estimation. In *Robustness in Statistics*, Launer, R. L., and Wilkinson, G. N. (Eds.), Academic Press, New York.
- Hogg, R. V., and Craig, A. T. 1970. Introduction to mathematical statistics. Macmillan Publ. Co., Inc., New York (third edition).
- Hollander, M., and Wolfe, D. A. 1973. Nonparametric statistical methods. John Wiley and Sons, Inc., New York.
- Honer, T. G. 1971. Weight relationships in open- and forest-grown balsam fir trees. In *Forest Biomass Studies*, IUFRO Sect. 25: Yield and Growth Working Group on Forest Biomass Studies, University of Maine Press, Orono, pp. 65-78.
- Jacobs, M. W., and Cunia, T. 1980. Use of dummy variables to harmonize tree biomass tables. *Can. J. For. Res.* 10(4): 483-490.
- Johnstone, W. D. 1971. Total standing crop and tree component distributions in three stands of 100-year-old lodgepole pine. In *Forest Biomass Studies*, IUFRO Sect. 25: Yield and Growth Working Group on Forest Biomass Studies, University of Maine Press, Orono, pp. 81-89.
- Jokela, E. J., Shannon, C. A., and White, E. H. 1981. Biomass and nutrient equations for mature Betula papyrifera Marsh. *Can. J. For. Res.* 11(2): 298-304.
- Kale, B. K. 1970. Normality of linear combinations of non-normal random variables. *Amer. Math. Monthly*, 77: 992-995.
- Keen, R. E. 1963. Weights and centres of gravity involved in handling pulpwood trees. *Pulp Pap. Res. Inst. Can., Techn. Rep.* 340.
- Kellogg, R. M., and Keays, J. L. 1968. Weight distributions in western hemlock trees. *Can. Dept. Fish. Forest., Ottawa, Bi-monthly Res. Notes* 24(4): 32-33.
- Kempthorne, O. 1975. Design and analysis of experiments. Krieger Publ. Co., Inc., New York.

- Ker, M. F., and Van Raalte, G. C. 1980. Tree biomass equations for Abies balsamiae and Picea glauca in northwestern New Brunswick. Can. J. For. Res. 11(1): 13-17.
- Keyes, M. R., and Grier, C. C. 1981. Above- and below-ground net production in 40-year-old Douglas-fir stands on low and high productivity sites. Can. J. For. Res. 11(3): 599-605.
- Kiil, A. D. 1967. Fuel weight tables for white spruce and lodgepole pine crowns in Alberta. Can. Dept. Forest. Rural Develop. Publ. 1196.
- Kiil, A. D. 1968. Weight of the fuel complex in 70-year-old lodgepole pine stands of different densities. Can. Dept. Forest. Rural Develop. Publ. 1228.
- Kimmins, J. P. 1977. Evaluation of the consequences for future tree productivity of the loss of nutrients in whole-tree harvesting. For. Ecol. Mgt. 1: 169-183.
- Kimmins, J. P., and Krumlik, G. J. 1976. On the question of nutrient losses accompanying whole-tree harvesting. IUFRO Oslo Biomass Studies, University of Maine Press, Orono, Maine, pp. 41-53.
- Kimmins, J. P., and Scoullar, K. 1979. FORCYTE: A computer simulation approach to evaluating the effect of whole-tree harvesting on the nutrient budget in Northwest forests. Univ. of B.C., Fac. of Forestry.
- Kimmins, J. P., Scoullar, K., and Feller, M. C. 1981. FORCYTE--An ecologically-based computer simulation model to evaluate the effect of intensive forest management on the productivity, economics and energy balance of forest biomass production. Univ. B.C., Fac. of Forestry.
- Kozak, A. 1970. Methods for ensuring additivity of biomass components by regression analysis. For. Chron. 46(5): 402-404.
- Koziol, J. A. 1982. A class of invariant procedures for assessing multivariate normality. Biometrika, 69: 423-427.
- Kurucz, J. 1969. Component weights of Douglas-fir, western hemlock and western red cedar biomass for simulation of amount and distribution of forest fuels. Unpublished M.F. Thesis, The University of British Columbia, Vancouver, Canada.
- Loomis, R. M., Phares, R. E., and Crosby, J. S. 1966. Estimating foliage and branchwood quantities in short-leaf pine. For. Sci. 12(1): 30-39.
- Madansky, A. 1959. The fitting of straight lines when both variables are subject to error. J. Amer. Stat. Assoc., 54: 173-205.

- Malkovich, J. F., and Afifi, A. A. 1973. On tests for multivariate normality. *J. Amer. Stat. Assoc.*, 68: 176-179.
- Marks, P. L., and Bormann, F. H. 1972. Revegetation following forest cutting: Mechanisms for return to steady-state nutrient cycling. *Science*, 176: 914-915.
- Miller, R. G. 1974. The jackknife--a review. *Biometrika*, 61: 1-15.
- Montgomery, D. C. 1976. Design and analysis of experiments. John Wiley and Sons, Inc., New York.
- Montgomery, D. C., and Peck, E. A. 1982. Introduction to linear regression analysis. John Wiley and Sons, Inc., New York.
- Mood, A. M., Graybill, F. A., and Boes, D. C. 1974. Introduction to the theory of statistics. McGraw-Hill, New York (third edition).
- Moore, E. H. 1920. On the reciprocal of the general algebraic matrix. *Bull. Amer. Math. Soc.* 26: 394-395.
- Morrison, D. F. 1976. Multivariate statistical methods. McGraw-Hill, New York (second edition).
- Muirhead, R. J. 1982. Aspects of multivariate statistical theory. John Wiley and Sons, Inc., New York.
- Noble, B., and Daniel, J. W. 1977. Applied linear algebra. Prentice-Hall Inc., Englewood Cliffs, N.J. (second edition).
- Ovington, J. D. 1956. The form, weights and productivity of tree species grown in close stands. *New Phytol.*, 55(3): 289-304.
- Ovington, J. D. 1962. Quantitative ecology and the woodland ecosystem concept. *Adv. Ecol. Res.* 1: 103-192.
- Penrose, R. A. 1955. A generalized inverse for matrices. *Proc. Cambridge Philos. Soc.* 51: 406-413.
- Randles, R. H., and Wolfe, D. A. 1979. Introduction to the theory of nonparametric statistics. John Wiley and Sons, Inc., New York.
- Rao, C. R. 1965. Linear statistical inference and its applications. John Wiley and Sons, Inc., New York.
- Rao, C. R., and Mitra, S. K. 1971. Generalized inverse of matrices and its applications. John Wiley and Sons, New York.
- Rosenberg, L. 1965. Nonnormality of linear combinations of normally distributed random variables. *Amer. Math. Monthly*, 72: 888-890.

- Ruymgaart, F. H. 1973. Non-normal bivariate densities with normal marginals and linear regression functions. *Statistica Neerlandica*, 27: 11-17.
- Sampson, A. R. 1974. A tale of two regressions. *J. Amer. Stat. Assoc.*, 69: 682-689.
- Sando, R. W., and Wick, C. H. 1972. A method of evaluating crown fuels in forest stands. U.S.D.A. Forest Serv., North Central Forest Exp. Sta., Res. Pap. NC-84.
- Schmitt, M. D. C., and Grigal, D. F. 1981. Generalized biomass estimation equations for Betula papyrifera Marsh. *Can. J. For. Res.* 11(4): 837-839.
- Searle, S. R. 1966. Matrix algebra for the biological sciences. John Wiley and Sons, New York.
- Searle, S. R. 1971. Linear models. John Wiley and Sons, New York.
- Singh, T. 1982. Biomass equations for ten major tree species of the prairie provinces. *Can. For. Serv., Northern Forest Res. Centre, Information Rept. NOR-X-242.*
- Smith, J. H. G. 1979. A review of the state of the art in complete tree utilization, short rotation forestry, biomass and growth modelling, and simulation. Univ. B.C., Fac. of Forestry.
- Smith, J. H. G., and Williams, D. H. 1980. A proposal to develop a comprehensive forest biomass growth model. Final report for the Canadian Forestry Service, Univ. B.C., Fac. of Forestry.
- Snedecor, G. W., and Cochran, W. G. 1973. Statistical methods. Iowa State Univ. Press, Ames, Iowa (6th edition).
- Springer, M. D. 1979. The algebra of random variables. John Wiley and Sons, New York.
- Stewart, G. W. 1973. Introduction to matrix computations. Academic Press, New York.
- Storey, T. G. 1969. The weights and fuel size distribution of pinyon pine and Utah juniper. Project Flambeau--an investigation of mass fire 1964-67. Pac. S. W. Forest Range Exp. Sta., Final Rep. III.
- Storey, T. G., Fons, W. L., and Sauer, F. M. 1955. Crown characteristics of several coniferous tree species. U.S.D.A. Forest Serv., Interim Tech. Rep. AFSWP-416.
- Storey, T. G., and Pong, W. Y. 1957. Crown characteristics of several hardwood tree species. U.S.D.A. Forest Serv., Interim Tech. Rep. AFSWP-968.

- Strang, G. 1976. Linear algebra and its applications. Academic Press, New York.
- Tadaki, Y., Shidei, T., Sakasegawa, T., and Ogino, K. 1961. Studies on productive structure of forest (II): Estimation of standing crop and some analyses on productivity of young birch stand (Betula platyphylla). J. Jap. For. Soc. 43(1): 19-26.
- Tufts, W. P. 1919. Pruning of deciduous fruit trees. Calif. Agric. Exp. Sta. Bull. 313.
- Wald, A. 1940. The fitting of straight lines if both variables are subject to error. Ann. Math. Stat. 11: 284-300.
- Whittacker, R. H., Bormann, F. H., Likens, G. E., and Siccama, T. G. 1974. The Hubbard Brook ecosystem study: Forest biomass and production. Ecol. Monogr. 44: 233-254.
- Yandle, D. O., and Wiant, H. V. 1981. Estimation of plant biomass based on the allometric equation. Can. J. For. Res. 11(4): 833.
- Young, H. E. 1978. Forest biomass inventory. Proc. Symposium on Complete Tree Utilization of Southern Pine, April 1978, New Orleans, Louisiana.
- Young, H. E., and Chase, A. J. 1965. Fibre weight and pulping characteristics of the logging residue of seven tree species in Maine. Maine Agric. Exp. Sta., Tech. Bull. 17.
- Young, H. E., Strand, L., and Altenberger, R. 1964. Preliminary fresh and dry weight tables for seven tree species in Maine. Maine Agric. Exp. Sta., Tech. Bull. 12.
- Zavitkovski, J. 1971. Dry weight and leaf area of Aspen (Populus tremuloides/Populus grandidentata) trees in northern Wisconsin. In Forest Biomass Studies. IUFRO Working Group on Forest Biomass Studies, Sect 25: Growth and Yield, University of Florida, Gainesville, Florida.
- Zavitkovski, J., Jeffers, R. M., Nienstaedt, H., and Strong, T. F. 1981. Biomass production of several jack pine provenances at three Lake States locations. Can. J. For. Res. 11(2): 441-447.
- Zyskind, G. 1962. On conditions for equality of best and simple least squares estimators. (Abstract) Ann. Math. Stat. 33: 1502-1503.
- Zyskind, G. 1967. On canonical forms, non-negative covariance matrices and best and simple linear estimators in linear models. Ann. Math. Stat. 38: 1092-1109.
- Zyskind, G., and Martin, F. G. 1969. On best linear estimation and a general Gauss-Markoff theorem in linear models with arbitrary non-negative covariance structure. SIAM J. Appl. Math. 17: 1190-1202.

APPENDIX I(a)

Kozak's (1970) Biomass Data

DBH (Inches) X	Bole (lbs.) Y ₁	Bark (lbs.) Y ₂	Branches (lbs.) Y ₃	Total Y _T =Y ₁ +Y ₂ +Y ₃ (lbs.)
7.2	254	29	73	356
11.0	749	60	192	1001
9.8	519	49	151	719
7.5	217	31	115	363
12.2	1025	76	222	1323
6.7	242	26	76	344
5.9	136	18	37	191
5.5	127	18	39	184
3.5	62	6	11	79
8.7	375	39	98	512

Conversion factors: 1 in = 2.54 cm, 1 lb. = 0.4535924 kg

APPENDIX I(b)

Western Hemlock Biomass Data (Kurucz, 1969)

Height (Feet)	DBH (Inches)	Branch (lbs.)	Fine Branch + Foliage (lbs.)	Bole (lbs.)	Bark (lbs.)	Total (lbs.)
20.0	3.5	25.23	36.26	15.56	3.57	80.62
14.0	1.6	3.43	7.27	3.32	0.82	14.84
34.0	6.1	55.62	71.95	60.18	16.21	203.96
53.0	11.8	314.77	309.58	318.24	52.21	994.80
41.0	8.8	130.36	168.71	169.58	35.45	504.10
55.0	10.5	231.26	260.52	278.94	44.52	815.24
55.0	11.7	289.72	275.71	346.04	56.06	967.53
93.0	21.9	1210.37	859.97	2095.34	220.94	4395.23
73.0	15.9	701.20	430.48	916.22	120.91	2168.81
92.0	17.0	529.87	423.46	1329.83	188.24	2471.40
117.0	24.5	1201.44	468.42	3806.64	487.50	5964.00
167.0	30.5	3188.59	993.78	7971.30	1221.49	13375.16
123.0	23.7	2424.59	831.81	3332.09	369.61	6958.10
131.0	26.0	1658.45	522.55	4323.78	357.79	6862.57
176.0	34.1	3876.44	2013.48	11187.62	1447.37	18524.91
175.0	36.4	4027.03	1535.28	13874.81	2337.72	21774.84
148.0	31.1	2870.00	1204.30	8920.92	1196.20	14191.62
151.0	29.0	5926.09	2149.10	7537.04	1415.22	17027.45

Conversion factors: 1 in = 2.54 cm, 1 foot = 0.3048 m, 1 lb. = .4535924 kg

DBH (cm)	Height (m)	Crown Length (m)	Crown Width (m)	Bole (kg)	Bark (kg)	Crown (kg)	Total Biomass (kg)
19.20	22.40	14.50	4.30	94.43	13.96	9.29	117.69
28.00	25.20	20.20	6.80	253.21	27.19	57.67	338.08
22.70	26.70	16.70	4.20	184.65	21.67	7.21	213.52
23.90	28.30	15.40	4.10	224.57	27.48	11.15	263.21
28.90	28.20	11.40	5.60	324.82	34.56	45.79	405.17
23.90	22.40	12.10	4.00	155.99	20.84	15.00	191.83
14.60	12.80	12.20	3.30	28.85	4.58	18.80	52.22
26.80	26.60	23.60	4.80	234.19	32.09	28.43	294.72
11.20	9.80	7.80	2.60	19.77	2.97	5.84	28.58
3.10	3.50	3.10	1.30	0.64	0.12	0.71	1.48
5.70	4.90	4.50	1.20	2.79	0.37	2.56	5.73
14.00	11.20	9.70	4.40	24.04	3.85	15.75	43.64
12.10	10.20	9.30	3.10	17.34	2.82	9.73	29.89
7.00	6.10	5.60	2.40	5.43	0.74	2.09	8.26
16.40	13.40	11.90	5.20	43.07	6.66	24.29	74.02
4.50	3.90	3.80	1.40	1.30	0.32	2.16	3.79
9.50	7.10	7.00	2.30	7.89	1.41	4.23	13.53
10.50	8.80	8.80	2.70	13.80	2.08	6.77	22.65
17.60	10.70	10.60	14.00	36.52	5.51	17.54	59.58
18.20	12.40	10.80	4.40	46.94	7.57	24.17	78.68
15.40	8.70	7.00	3.50	25.01	4.09	10.86	39.96
14.50	8.40	7.20	4.00	19.94	3.20	12.19	35.32
32.70	20.60	16.30	3.80	232.63	43.90	89.67	366.20
31.30	21.30	15.30	6.70	326.34	50.43	86.41	463.18
16.10	13.40	10.80	4.80	39.00	7.68	26.77	73.45
11.20	10.00	8.30	3.70	16.15	2.52	11.29	29.95
42.40	22.40	19.70	6.50	523.92	96.28	129.66	749.86
17.40	11.60	10.60	2.80	46.42	6.46	28.42	81.30
16.40	12.60	11.30	3.80	41.12	6.07	8.65	55.84
29.90	17.00	16.20	5.20	131.27	18.81	142.89	292.96
11.30	8.30	6.20	2.40	13.23	3.65	4.81	21.70
21.50	13.50	11.60	3.60	82.87	13.57	16.55	112.99
12.20	6.70	5.50	3.10	14.35	3.30	5.14	22.79
14.20	9.70	9.10	4.30	22.44	3.31	16.59	42.34
7.80	5.90	5.90	2.60	3.93	0.53	2.04	6.50
9.40	7.30	7.30	2.00	6.07	1.00	5.04	12.11
11.00	7.60	6.60	2.60	10.94	1.16	4.42	16.51
19.80	9.70	7.60	2.60	38.36	4.72	13.77	56.85
13.60	6.90	5.20	1.20	15.57	2.18	8.54	26.29
11.20	6.10	4.40	2.30	9.91	1.50	5.22	16.62
9.80	6.40	5.50	2.10	7.17	1.57	4.56	13.30
9.70	8.10	8.10	2.70	8.40	0.82	1.98	11.19
12.90	9.20	9.20	4.40	18.47	1.93	6.25	26.65
17.70	10.70	10.70	4.70	31.71	3.53	14.63	49.87
11.40	9.70	8.80	4.30	14.91	1.60	6.88	23.39
14.00	9.70	9.70	2.50	21.87	3.64	6.61	32.12
18.70	9.80	9.80	3.50	35.83	4.62	19.28	59.73
11.10	6.90	6.30	2.30	9.01	1.40	5.28	15.69

APPENDIX I(d)

PROGRAMME 1: Fortran programme calls IMSL subroutine MDCH to compute chi-square probabilities

```

      INTEGER IER
      REAL X(18)
      READ(5,10) (X(J),J=1,18)
10  FORMAT(8X,18F8.5)
      DO 15 I=1,18
      DF=2.0
      CALL MDCH(X(I),DF,P,IER)
      WRITE(6,20) X(I),P
20  FORMAT(' X(I)=',F8.5,5X,'P=',F10.5)
15  CONTINUE
      STOP
      END

```

PROGRAMME 2: Fortran programme computes and prints Koziol statistics J_n

```

      REAL X(18),JN
      READ(5,10) (X(J),J=1,18)
10  FORMAT(18F8.5)
      SUM=0.0
      DO 20 I=1,18
      SUM=SUM+(X(I)-((I-0.5)/18.0))**2
20  CONTINUE
      JN=SUM+ (1.0/216.0)
      WRITE(6,30) JN
30  FORMAT(' JN=',F10.6)
      STOP
      END

```

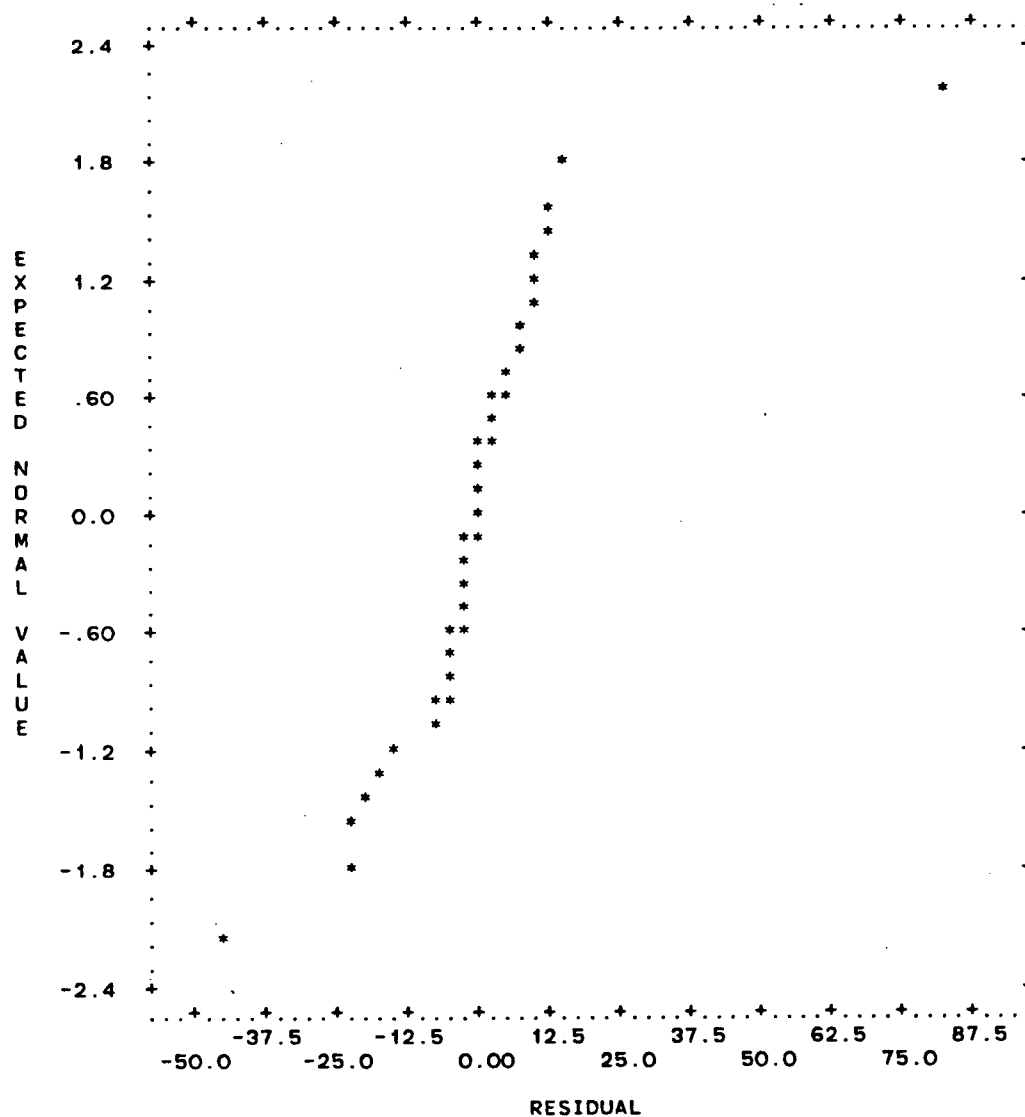
APPENDIX II(a)

Histogram of $\hat{\epsilon}_T = \sum_{j=1}^3 \hat{\epsilon}_j$ for the ENFOR data

INTERVAL NAME												FREQUENCY		PERCENTAGE		
	5	10	15	20	25	30	35	40	45	50	55	60	INT.	CUM.	INT.	CUM.
	+-----+											1	1	2.1	2.1	
*-42.000 +X													0	1	0.0	2.1
*-38.500 +													0	1	0.0	2.1
*-35.000 +													0	1	0.0	2.1
*-31.500 +													0	1	0.0	2.1
*-28.000 +													0	1	0.0	2.1
*-24.500 +													2	3	4.2	6.3
*-21.000 +XX													1	4	2.1	8.3
*-17.500 +X													1	5	2.1	10.4
*-14.000 +X													1	6	2.1	12.5
*-10.500 +X													2	8	4.2	16.7
*-7.0000 +XX													6	14	12.5	29.2
*-3.5000 +XXXXXX													13	27	27.1	56.3
*0.00000 +XXXXXXXXXXXX													8	35	16.7	72.9
*3.50000 +XXXXXXXXXX													3	38	6.3	79.2
*7.00000 +XXX													5	43	10.4	89.6
*10.5000 +XXXXXX													3	46	6.3	95.8
*14.0000 +XXX													1	47	2.1	97.9
*17.5000 +X													0	47	0.0	97.9
*21.0000 +													0	47	0.0	97.9
*24.5000 +													0	47	0.0	97.9
*28.0000 +													0	47	0.0	97.9
*31.5000 +													0	47	0.0	97.9
*35.0000 +													0	47	0.0	97.9
*38.5000 +													0	47	0.0	97.9
*42.0000 +													0	47	0.0	97.9
*45.5000 +													0	47	0.0	97.9
*49.0000 +													0	47	0.0	97.9
*52.5000 +													0	47	0.0	97.9
*56.0000 +													0	47	0.0	97.9
*59.5000 +													0	47	0.0	97.9
*63.0000 +													0	47	0.0	97.9
*66.5000 +													0	47	0.0	97.9
*70.0000 +													0	47	0.0	97.9
*73.5000 +													0	47	0.0	97.9
*77.0000 +													0	47	0.0	97.9
*80.5000 +													1	48	2.1	100.0
*84.0000 +X													0	48	0.0	100.0
*87.5000 +																
	+-----+															
	5	10	15	20	25	30	35	40	45	50	55	60				

APPENDIX II(b)

Normal probability plot of residuals ($\hat{\varepsilon}_T = \sum_{j=1}^3 \hat{\varepsilon}_j$) for the ENFOR data



APPENDIX III

A SOMEWHAT DETAILED COMPUTATIONAL EXAMPLE

In this part of the thesis, it is intended to use Kozak's (1970) data to show some of the computational details relating to the generalized additivity problem. These data are chosen partly because of their small size, making the computational exercise fairly straightforward yet making possible a demonstration of the computations involved. The computations for larger data sets (with more independent variables) are performed as typically shown here. Particular attention is given in the following to aspects of the computational details not given in section 10.2.

In the following, X_1 = diameter, X_2 = (diameter)², and the matrix X refers to the 10×3 matrix $X = (X_0 | X_1 | X_2)$ where X_0 is a column vector of 1's. Also let $X^* = (X_0 | X_2)$. The following matrices will be of use in this discussion.

$$(X^t X)^{-1} = \begin{pmatrix} 8.2031 & -0.82845 & 0.019172 \\ -0.82845 & 0.087614 & -0.002093 \\ 0.019172 & -0.002093 & 0.0000514 \end{pmatrix}$$

$$(X^{*t} X^*)^{-1} = \begin{pmatrix} 0.36948 & -0.0006217 \\ -0.0006217 & 0.00000143 \end{pmatrix}, \quad X^t y_T = \begin{pmatrix} 2300.6 \\ 56060 \\ 1443700 \end{pmatrix}$$

The unrestricted total fitted equation is given by

$$\hat{y}_T = 117.19 - 17.696X_1 + 1.06540X_2.$$

The estimated variance-covariance matrix of $\hat{\beta}_T$ for this model is $(X^t X)^{-1} \hat{\sigma}_T^2 = (X^t X)^{-1} (217.64)$, where 217.64 is the MSE associated with fitting y_T . Thus the standard errors for the parameters of this model can be obtained from the estimated variance-covariance matrix. In this case they are 42.25, 4.37, 0.1058 for $\hat{\beta}_{0T}$, $\hat{\beta}_{1T}$, and $\hat{\beta}_{2T}$ respectively.

The component equations containing only important independent variables are given in section 10.2. The corresponding total conditioned equation as determined by additivity is

$$\hat{y}_{TC} = 129.87 - 19.037X_1 + 1.09745X_2.$$

The covariance matrix of $\hat{\beta}_{TC}$, for the above equation, is obtained, under the assumption of independence of the y_j 's, as the sum of the covariance matrices of the estimated parameters of the component equations. In the present case, the estimated covariance matrix is given by

$$\hat{\Sigma}_{\hat{\beta}_{TC}} = \begin{pmatrix} 8.94206 & -0.82845 & 0.01793 \\ -0.082845 & 0.087614 & -0.002093 \\ 0.01793 & -0.002093 & 0.0000543 \end{pmatrix} \Sigma_{j=1}^3 \hat{\sigma}_j^2,$$

where $\hat{\sigma}_j^2$ is the mean square associated with fitting component model j .

In the present case, $\hat{\sigma}_1^2 = 194.04$, $\hat{\sigma}_2^2 = 0.55375$, and $\hat{\sigma}_3^2 = 42.198$. Therefore $\Sigma_{j=1}^3 \hat{\sigma}_j^2 = 236.792$. Hence the estimated standard errors for the parameter estimates in \hat{y}_{TC} are, in order, 46.02, 4.55, 0.1134, which compare favourably with those given above for the unrestricted total model. It should be pointed out here that before adding the covariance matrices, they are filled up with zeros to bring them to the full size corresponding to all variables in the conditioned total equation and the elements are permuted to correspond to the same parameters prior to addition. This part has not been exhibited in the above derivations.

Finally, the predicted values generated by the unrestricted total equation \hat{y}_T and its residuals are compared with those obtained using the total conditioned equation \hat{y}_{TC} . The results are given in Table 3(a,b) and in Figure 1. For these data, at least, the total conditioned equation performs relatively well. Of course this may partly be ascribable to the very small sample and narrow sample range.

Table. 3 Comparison of predicted values and residuals of unrestricted total equation with those of total conditioned equation

a. Unrestricted Total Equation

OBSERVED TOTAL BIOMASS	PREDICTED TOTAL BIOMASS	RESIDUAL
161.48	166.53	-5.05
454.05	454.27	-0.23
326.13	336.65	-10.51
164.65	166.53	-1.87
600.10	591.69	8.41
156.04	124.40	31.63
86.64	91.08	-4.44
83.46	77.71	5.75
35.83	43.88	-8.05
232.24	246.21	-13.97

b. Total Conditioned Equation

OBSERVED TOTAL BIOMASS	PREDICTED TOTAL BIOMASS	RESIDUAL
161.48	165.48	-4.00
454.05	454.69	-0.65
326.13	335.99	-9.86
164.65	165.48	-0.83
600.10	593.78	6.32
156.04	123.73	32.30
86.64	91.05	-4.41
83.46	78.10	5.36
35.83	47.36	-11.53
232.24	245.10	-12.86

Figure 1. Scatter of Residuals from Total Unrestricted Equation and from Total Conditioned Equation

