BIASED ESTIMATION TECHNIQUES FOR MULTIPLE LINEAR REGRESSION

by

PHILLIP DEAN WITTMER

B. S., Kansas State University, 1975

A MASTER'S REPORT

submitted in partial fulfillment of the

requirements for the degree

MASTER OF SCIENCE

Department of Statistics

KANSAS STATE UNIVERSITY
Manhattan, Kansas

1976

Approved by:

George A. Milliken
Major Professor

ACKNOWLEDGEMENTS

I would like to express my deepest gratitude to my Major Professor, Dr. George Milliken, for his guidance and encouragement in this endeavor. He always had time for my problems and questions. My thanks also to Dr. Krishna Akkina and Dr. Holly Fryer for their additional support and suggestions.

Finally, a special "Thank You" to my wife Gloria, not only for her help with typing and drawings, but also for her neverending patience and devotion throughout my graduate studies.

TABLE OF CONTENTS

CHAPTER 1

INTRODUCTION

Multiple linear regression is one of the most widely used of all statistical methods. Statisticians and non-statisticians alike have found it to be a useful tool for modeling the response of a dependent variable as it is influenced by independent variables.

When dealing with experimental data, one often encounters the problem of near multicollinearity of the data vectors. Usually, this is due to high correlations between two or more of the explanatory variables. When this occurs, the least squares estimators often contain values which are useless in the sense that they are extremely large or even of the wrong sign. (Wrong in the sense that they deviate from the accepted theory for the related field.) Hoerl and Kennard [6] refer to estimates of that type as unstable estimates.

This report investigates techniques for estimating the parameters of the linear model when near multicollinearity exists between the independent variables. The linear model is:

$$Y = X\beta + \varepsilon \quad (1.1)$$

where; (1) Y is an nx1 vector of observations on the dependent variable.

(2) X is an nxp matrix of observations on the independent variables such that $\rho(X)=p$ where $p \leq n$.

(3) ε is an nx1 vector of unobservable random errors such that

(a) $E(\varepsilon)=0$ and
(b) $E[\varepsilon\varepsilon']= \sigma^2 I$

Unless otherwise specified, X'X is assumed to be in the form of a pxp correlation matrix. See Appendix I for details of transforming the matrix to this form and obtaining the estimate of β.

The Ordinary Least Squares (OLS) estimator $\hat{\beta}$, of β is;

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

That estimator is unbiased for β and, by the Gauss-Markov Theorem, has minimum variance among the class of linear unbiased estimators of β. Computationally, the least squares procedure is good if X'X is well conditioned, i.e., not singular or near singular. If the matrix is ill-conditioned, the analyst will be tempted to delete variables in an attempt to remove the multicolinearities. This is hardly satisfactory when the model is correct as specified. We therefore look for more useful estimators, which are biased, but have smaller mean square error (MSE).

Before we discuss biased estimation techniques, we need to examine the characteristics of OLS estimators when the X'X matrix is ill-conditioned. The covariance matrix of the least squares estimator is

$$\mathrm{Var}(\hat{\beta}) = \sigma^2(X'X)^{-1}$$

Let the distance between the OLS estimator and the true but unknown value of β be denoted by $L_1 = ||\hat{\beta}-\beta||$. Then,

(1) $L_1^2 = (\hat{\beta}-\beta)'(\hat{\beta}-\beta)$ (1.2)

(2) $E(L_1^2) = \sigma^2 tr(X'X)^{-1}$ (1.3)

(3) $E[\hat{\beta}'\hat{\beta}] = \beta'\beta + \sigma^2 tr(X'X)^{-1}$ (1.4)

and if $\varepsilon \sim N(0,\sigma^2 I)$, then

(4) $\mathrm{Var}(L_1^2) = 2\sigma^4 tr(X'X)^{-2}$

Denote the characteristic roots of X'X by;

$$\lambda_{max} = \lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_p = \lambda_{min} > 0,$$

then (1.3) can be rewritten as

$$E(L_1^2) = \sigma^2 \sum_{i=1}^{p}(1/\lambda_i) \quad (1.5)$$

and if the errors are normal,

$$\text{Var}(L_1^{\,2}) = 2\sigma^4 \, \Sigma(1/\lambda_i)^2 \tag{1.6}$$

Hence, lower bounds for $E(L_1^{\,2})$ and $\text{Var}(L_1^{\,2})$ are σ^2/λ_{min} and $2\sigma^4/\lambda^2_{min}$ respectively.

When one or more of the λ_i are small, this indicates a linear dependence of the ith column vector on the other p - 1 vectors, and we say X'X is ill-conditioned. If this occurs, the distance from $\hat{\beta}$ to β is large, as indicated by coefficients, $\hat{\beta}_i$, large in absolute value or with reversed signs, as alluded to earlier. By definition, the least squares estimate is that value of β which minimizes

$$\Phi(\beta) = (Y - X\beta)'(Y - X\beta) \tag{1.7}$$

The λ_i measure the sensitivity of the solution to (1.7) and thus should be utilized to construct "better estimates". The criterion for determining which estimators are "better" differs with the biased technique. Most authors strive to minimize the MSE.

The biased estimation techniques outlined in the remainder of the report all utilize the small λ_i's in one way or another to aid in the estimation of β. The ways in which these near-singularities are utilized to predictive advantage are the basic differences in the techniques. They all achieve a reduction in the length of the vector of estimated coefficients ($\hat{\beta}'\hat{\beta}$) when compared with the length of the vector of OLS coefficients.

Another view of the problem that ill-conditioning creates in the OLS estimates is given by Webster, Gunst and Mason [12]. They partition X as $X = [x_j \vdots X_j^*]$ where x_j is the vector of observations for the jth independent variable, and X_j^* is the remaining p-1 observation vectors. If c_{jj} is the