SOME ALTERNATIVES

TO

LEAST SQUARES ESTIMATION

IN

LINEAR MODELLING

A thesis presented in partial fulfilment
of the requirements for the degree of
Master of Science in Statistics
at Massey University

John Reynolds

1977

ABSTRACT

The effects of non-standard conditions on the application of the Gauss-Markov Theorem are discussed and methods proposed in the literature for dealing with these effects are reviewed. The multicollinearity problem, which is typified by imprecise least squares estimation of parameters in a multiple linear regression and which arises when the vectors of the input or predictor variables are nearly linearly dependent, is focussed upon and a class of alternative biased estimators examined. In particular several members of the class of biased linear estimators or linear transformations of the Gauss-Markov least squares estimator are reviewed. A particular generalized ridge estimator is introduced and its relation to other techniques already existing in the literature is noted. The use of this estimator and the simple ridge regression estimator is illustrated on a small data set. Further comparisons of the estimator, the ridge estimator and other generalized ridge estimators are suggested.

ACKNOWLEDGEMENTS

I wish to record my gratitude to my supervisor Dr Richard J. Brook for his encouragement and guidance in the preparation of this thesis. Grateful thanks are owed to Miss Maryanne Collins who patiently typed the tedious manuscript.

TABLE OF CONTENTS

Page

LIST OF FIGURES

1. INTRODUCTION

The solution of a system of overdetermined or overidentified linear equations requires some kind of approximation method. The most common method of arriving at a solution for B in the overdetermined system of linear equations,

$$XB = Y$$

where X is an $n \times p$ matrix of full column rank p, B is an unknown $p \times 1$ vector of parameters and y is an $n \times 1$ vector, and in which $n > p$, is the method of least squares. The least squares solution identifies the $p \times 1$ vector which minimizes the Euclidean norm of Y-XB.

The source of the overdetermination or inconsistency in the system of linear equations is usually attributed to the presence of some kind of error component in the n realizations of the p+1 variables which form X and Y. Statisticians often make very specific assumptions about the error content of the n realizations. Errors are usually assumed, in the lack of any knowledge concerning their origin, to be generated by some sort of random device which may be represented by a probability density. The realizations of the p variables which make up the matrix X and which are often controllable are usually assumed to be measurable without error whereas the vector variable Y is usually assumed to contain the randomly generated errors. Thus statisticians have concerned themselves with the linear model,

$$y = X\beta + \epsilon$$

where ϵ is an $n \times 1$ vector of stochastic errors which are independent of the measurements of the p variables which make up the matrix S, and, have used the method of least squares to extract an approximation to, or an estimate of, the unknown vector of parameters, β. Under various assumptions about X,y and ϵ, and under various restrictions on possible methods of approximation, the method of least squares has other optimal features besides the norm minimization property mentioned above. If, however, these assumptions are not met in practice the other optimal features may disappear.

The purpose of this thesis is to review some of the work which has been completed or is currently in progress, concerning the effect of the relaxation of these assumptions and restrictions on the optimality properties of least squares and to review some of the alternatives to least squares which have been developed in response to these effects. The conditions of the Gauss-Markov Theorem, which are presented in Chapter 2, form the framework for the review and it is the effect of the relaxation of these conditions on the least squares procedure which is presented in Chapter 3. In Chapter 3 it is established that multicollinearity in the matrix X is one non-standard condition which can have serious effects on least squares estimation of the parameter vector. A class of alternatives to the least squares estimator, namely biased estimators, is focussed upon in Chapter 4. These estimators were designed originally to tackle the multicollinearity problem but many variants of these biased estimation procedures have been constructed with different goals in mind. A particular member of a subclass of these biased estimators, a doubly ridged estimator, is introduced in Chapter 5. The doubly ridged estimator, which is a generalized ridge estimator, displays many of the advantages and disadvantages of the well known ridge estimator. The application of the ridge and doubly ridged estimators to a small but well known test problem - the Longley data - is undertaken in Chapter 6. A summary, Chapter 7, which also includes suggestions for further investigations in the search for alternatives to least squares, completes the thesis.