

C
001.642
JAH

**ANALYZING THE RIDGE REGRESSION TECHNIQUES
INTRODUCED FOR SOLVING THE PROBLEMS OF
MULTICOLLINEARITY**

A PROJECT REPORT PRESENTED BY

ABOBACKER JAHUFER
✓

to the Board of Study in Statistics and Computer Science of the
POST GRADUATE INSTITUTE OF SCIENCE

in partial fulfillment of the requirement
for the award of the degree of

MASTER OF PHILOSOPHY IN STATISTICS

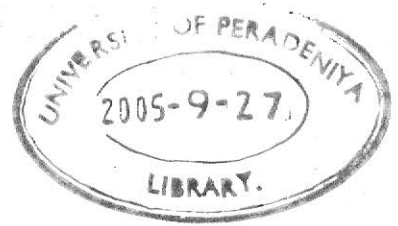
of the

UNIVERSITY OF PERADENIYA

SRI LANKA

2004

580477



ANALYZING THE RIDGE REGRESSION TECHNIQUES
INTRODUCED FOR SOLVING THE PROBLEMS OF
MULTICOLLINEARITY

A. Jahufer

Department of Mathematics and Computer Science

South Eastern University of Sri Lanka

Sammanthurai

Sri Lanka

Regression Analysis is one of the most widely used statistical techniques for analyzing multifactor data. Its broad appeal results from the conceptually simple process of using an equation to express the relationship between a set of variables. Regression analysis is also interesting theoretically because of the elegant underlying mathematics. Successful use of regression analysis requires an appreciation of both the theory and the practical problems that often arise when the technique is employed with real world data.

In the model fitting process the most frequently applied and most popular estimation procedure is the Ordinary Least Square Estimation (OLSE). The significant advantage of OLSE is that it provides minimum variance unbiased linear estimates for the parameters in the linear regression model.

In many situations both experimental and non-experimental, the independent variables tend to be correlated among themselves. Then inter-correlation or multicollinearity among the independent variables is said to exist. A variety of interrelated problems are created when multicollinearity exists. Specially, in the model building process, multicollinearity among the independent variables causes high variance (if OLSE is used) even though the estimator is still the minimum variance unbiased estimator in the class of linear unbiased estimators.

The main objective of this study is to show that the unbiased estimation does not mean good estimation when the regressors are correlated among themselves or multicollinearity exists. Instead, it is tried to motivate the use of biased estimation allowing small bias and having a low variance, which together can give a low mean square error.

In literature several biased estimation procedures were introduced for solving the problem of multicollinearity. Among them the biased regression technique namely, Ridge Regression Estimation, was first introduced by Hoerl (1964), and further developed by Hoerl and Kennard (1970a, b). Restricted Ridge Regression Estimation introduced by Sarkar (1992), Modified Ridge Regression Estimation introduced by Swindel (1976), Liu Estimation introduced by Liu Kejian (1993) and Restricted Liu Estimation introduced by S. Kaciranlar, G.P.H. Styan and H.J. Werner (1999) were frequently used biased estimation methods. These methods were rapidly developed in the recent years.

In this research work, five independent and one dependent standard normal pseudo-random (numbers) variables of multicollinearity data were generated by using *Monte Carlo Simulation Study* with the correlation ($\rho = 0.9, 0.95$ and 0.99) between the independent variables. For the analysis 100 observations for each variable are generated and fit the model by using the above **unbiased** and **biased** estimation methods. The stochastic properties of these methods were analyzed. The superiority of the biased estimation method is suggested and recommended to use and fit the model for the multicollinear real data.