

Title: Comparison of Predictive Abilities of Ridge-Based Remedial Measures for Multicollinearity

Author: Chona P. Guatlo

Degree: Master of Statistics

Date: May 2007

Abstract:

This study compared the predictive abilities of ridge-based remedial measures for multicollinearity using empirical data on cotton production of Negros Oriental during the crop year June 2005 to July 2006. Nineteen (19) non-climatic and climatic variables were considered in cotton yield estimation. The non-climatic variable was area planted while the climatic variables were rainfall, temperature and relative humidity of the different stages of cotton from planting to maturity. The least squares estimation generated the smallest MSEp, which means it had a relatively good fit, but had the largest MSPR among other methods considered in this study. The large difference between MSEp and MSPR indicated that there may be a more appropriate model for the data. The MSEp of RBRR ($q=10$), HEBRR, and DEBRR were closed to the least squares procedure. On the other hand, principal component regression had the largest MSEp, indicating it had the worst data fit relative to the other methods under consideration. The principal component regression, however, had the smallest difference between MSEp and MSPR supporting its appropriateness for the data. The least squares procedure had the largest prediction error based on every measure used. Among the Kubokawa and Srivastava's empirical Bayes ridge estimators, the EBRR with $q=0$ had the lowest prediction error. The methods with the lowest measures of prediction error were the principal component regression, ridge regression, and ridge trace variable selection. The ridge regression, EBRR ($q=0$), and principal component regression were able to correct some signs of least squares coefficients under multicollinearity.

Keywords: Ridge regression, empirical Bayes ridge regression, ridge trace, variable selection, principal component analysis