

C
510.73
ATA

i

(u)

COMPARISONS OF INDICATORS AVAILABLE FOR DETECTING MULTICOLLINEARITY

A PROJECT REPORT PRESENTED BY

A.A.K.U. ATAPATTU

To the Board of Study in Statistics and Computer Science of the
POSTGRADUATE INSTITUTE OF SCIENCE

*in partial fulfilment of the requirement
for the award of the degree of*

MASTER OF SCIENCE IN APPLIED STATISTICS

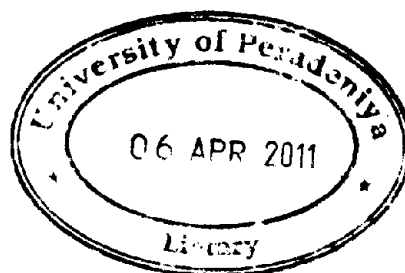
of the

UNIVERSITY OF PERADENIYA

SRL LANKA

2010

645700



COMPARISONS OF INDICATORS AVAILABLE FOR DETECTING MULTICOLLINEARITY

A.A.K.U. Atapattu

Postgraduate Institute of Science
University of Peradeniya
Peradeniya

One of the factors that affect the standard error of a partial regression coefficient is the degree to which that independent variable is correlated with the other independent variables in the regression equation. An independent variable that is very highly correlated with one or more other independent variables will have a relatively large standard error. This implies that the partial regression coefficient is unstable and will vary greatly from one sample to the next. This is the situation known as multicollinearity. Multicollinearity exists whenever an independent variable is highly correlated with one or more of the other independent variables in a multiple regression equation. Multicollinearity is a problem because it undermines the statistical significance of an independent variable. Other things being equal, the larger the standard error of a regression coefficient, the less likely it is that this coefficient will be statistically significant. It is for this reason that we need to fully understand the impact of multicollinearity on our regression analyses. Consequently, it is important that researchers be trained in detecting its existence.

Already statisticians have defined several indicators to detect multicollinearity. Many recently they have defined RED indicator, DEF and ICE factors. But when we consider these indicators we may face several difficulties. The Red indicator is always same as the indicator g introduced earlier. DEF indicator is an overall measure but ICE is not an overall measure. According to the research paper these two indicators should be lie in between 0 and 1. But according to my findings they were not lie in between 0 and 1. When multicollinearity level is increased they deviate from zero and take large negative values. This problem was informed to the reasercher, who published this paper and asked his ideas. Then he identified this situation

and asked me to improve this DEF factor and ICE values. The proposed new indicators are named as DEFm values and ICEm values and they are lie in between 0 and 1.

When we consider the other multicollinearity indicators most of the indicators are depending on sample size, the number of variables in the data set and the sign of the correlation among the explanatory variables.

To select the best indicators to detect multicollinearity some computational applications and simulation studies were used. A simulation study was done for the data set with different sample sizes and different number of independent variables.