# Journal of Applied Sciences

ISSN 1812-5654

Journal of Applied Sciences 15 (3): 392-404, 2015
ISSN 1812-5654


ansinet
Asian Network for Scientific Information

REVIEW ARTICLE OPEN ACCESS

DOI: 10.3923/jas.2015.392.404

# Ridge Regression for Solving the Multicollinearity Problem: Review of Methods and Models

Hanan Duzan and Nurul Sima Binti Mohamad Shariff
Faculty of Science and Technology, Universiti Sains Islam Malaysia (USIM), Malaysia

ARTICLE INFO
*Article History:*
Received: September 29, 2014
Accepted: December 22, 2014

Corresponding Author:
Hanan Duzan,
Faculty of Science and Technology,
Universiti Sains Islam Malaysia (USIM),
Malaysia

ABSTRACT
For an estimation of the ridge parameter, relevant research on estimation methods released from 1964-2014 has been reviewed and new estimation methods are suggested in this study. The history of multicollinearity dates back to 1934 when the multicollinearity concept was formulated to refer to the condition when the variables handled are under influence of two or more relationships. To tackle this type of problem, another study proposed Ridge Regression (RR). By that time the aim was mainly to identify value of the ridge parameter, k, such that reduction in the variance term of the slope parameter is larger than the increment in its squared bias. Furthermore, non-zero values of the ridge parameter for which the Mean Squared Error (MSE) of the slope parameter is less than the variance of the Ordinary Least Squares (OLS) estimator of the very parameter has been proven using RR. Thus, various estimations of the RR parameter have been improved and related estimation methods suggested by a number of researchers working in this field of scientific research.

**Key words:** Linear models, multicollinearity, least squares method, ridge regression

## INTRODUCTION

The literature review process is essential to performing research. In spite of its importance in informing and improving practice and prompting discussions in the academic field, literature review is not the 'be-all' or 'end-all' in research.

It is normally impossible to interpret estimates of individual coefficients when the explanatory variables are mainly highly inter-correlated, no matter what the goal of multiple regression analysis is. In addition, review of the literature indicates that the multicollinearity problem has been handled in a variety of different ways. A bulk of research has also been performed on RR. The different studies estimated the shrinkage ridge parameter (k) in various ways and compared its obtained values with those drawn from the Least Squares Estimator (LSE). For example, Hoerl and Kennard (1970a), who first introduced this parameter, employed RR estimators to tackle the multicollinearity problem. A small positive number ($k \geq 0$) was added to the diagonal elements of the X'X matrix from multiple regressions and the resultant an estimator was:

$$\hat{\beta}_{RR} = (X'X + kI_p)^{-1}X'\underline{Y},\ k \geq 0 \quad (1)$$

which is referred to as RR estimator. Additionally, if C is ($n\times p$) observed matrix of regressors and b is p vector of unknown parameters, then, in comparison with LSE for a positive value of k, the RR estimator (Eq. 1) may warrant a lower MSE. Recently, estimation of value for the ridge parameter, k, has received considerable consideration and researchers have adopted widely-varying approaches to estimation of k as in Hoerl and Kennard (1970a, b), Hoerl *et al*. (1975), McDonald and Galarneau (1975), Lawless and Wang (1976), Dempster *et al*. (1977), Gibbons (1981), Khalaf and Shukur (2005), Alkhamisi *et al*. (2006), Alkhamisi and Shukur (2008) and Muniz and Kibria (2009). Based on simulation studies, performance of ridge estimators has been evaluated by many researchers (Hoerl and Kennard, 1970a, b; Hoerl *et al*., 1975; McDonald and Galarneau, 1975; Lawless and Wang, 1976; Dempster *et al*., 1977; Gibbons, 1981; Khalaf and Shukur, 2005; Alkhamisi *et al*., 2006; Alkhamisi and Shukur, 2008;

Muniz and Kibria, 2009). Additionally, with a preset number of regressors and using the MSE as the performance measure, in most of the cases data were generated from populations with non-normal or normal distributions. Moreover, in most of the studies when regression analysis was employed, observations were presumed to be equally and independently distributed (iid), despite that the iid assumption is much powerful in real-life contexts. As an example, the mean rate of an event's occurrence differs from case to another and may depend on a multitude of variables. Count data regression is more appropriate than OLS when examining the incidence rate per unit of time subject to some covariates. Examples on such circumstances encompass numbers of takeover bids, patents, accident insurance cases, criminal careers and bank failures. If the mean of counts is not large, then normal approximation and the OLS method may be suitable.

First of all, the least squares method in presence of multicollinearity is suggested as a background in order to assimilate the literature. Then, an inventory of the estimation models and methods published from 1964-2014 has been performed. This inventory was intended to help in proposing some estimators so as to eventually determine value of the ridge parameter. Furthermore, so as to determine where practitioners and researchers have laid emphasis as regards the method (σ) to use and periodicity of forecasting, the reviewed literature has been synthesized and classified. It was found that for handling the multicollinearity problem the RR technique proved to be the optimum estimation method. By and large, because of the ridge trace and as alternative, we mainly intend to identify the suitable value of a function, namely, the coefficient of determination ($R^2$) of the least squares regression, on the rest variables.

## REVIEW OF METHODS AND MODELS

The study (Pasha and shah, 2014) entailed comprehensive investigation of RR problems in multicollinear data, together with the ridge estimator's properties. By regressing the number of persons on five variables, the eigen values, variance inflation factors and standardization problem were studied through empirical comparison of OLS with RR methods. As well, some methods have been proposed for identifying the bias parameter, K. Hoerl *et al*. (1975) provided another approach to selection of the K value of 0.116. However, Hoerl (1962) argued that their estimators are better than that of Hoerl *et al*. (1975). They performed simulations supporting their estimators and emphasized that these estimators had smaller MSE values in more than 50% of the studied problems. In addition, a KHKB value of 0.58 was found to produce the smallest MSE even though Wichn and Wahba (1988) showed that the KHKB can not function well. Lawless and Wang (1976) modified their estimator by making use of eigenvalues and found that the new estimator is effective and therefore, KLW equals 0.42. The method employed in this study seems to be much logical in view of the small bias and MSE values obtained. According to this method, the KPZ value is very close to the ridge trace. Hence, the KPZ and KHKB proved to be the best and regardless of method, the RR is better than OLS regression (Pasha and Shah, 2004).

For prediction of compressive strength of concrete, Chopra *et al*. (2013) employed two regression methods: Traditional regression and RR. Values of the regression coefficients have been varied drastically such that negative coefficients have been transformed into positive and positive coefficients have been transformed into negative when regression analysis was employed and data were reduced or raised. Eventually, the traditional method did not prove to be credible for forecasting the compressive strength of concrete. In consequence, both in the case of reduction and augmentation of data, frequent minimum effect existed that has no or negligible, impact on the coefficients when performing RR (Chopra *et al*., 2013).

The researchers (Muniz *et al*., 2012) reviewed 19 existing estimators and proposed new ones for estimation of ridge parameter. Investigation was conducted based on Monte Carlo simulations. Furthermore, these researchers examined many models where sample size; the unknown coefficient vector; correlations between explanatory variables; number of variables incorporated into the model and variance of the random error were adjusted. Moreover, two thousand replications were performed in each simulation run and simulation results were presented for each model in tables and figures. Simulation results illustrated that values of the MSE depend on correlations between the independent variables; random error variance and number of correlated variables. Consequently, if sample size increases, the MSE decreases even if random error variance and correlations between independent variables are high. It is worth mentioning that in all cases considered the suggested estimators had smaller MSE values than the values produced by OLS regression and other existing estimators. Further, this study reviewed existing estimators and proposed new estimators for the ridge parameter, k, based on the studies of Kibria (2003), Khalaf and Shukur (2005) and Alkhamisi *et al*. (2006). Additionally, the unknown coefficient vector, β, sample size (n), correlations between the explanatory variables (γ) and variance of the random error (σ) had impacts on performance of estimators. Based on outcomes of simulations, some restrictions could be highlighted like variations in the experimental conditions investigated. Additionally, to evaluate performance of the estimators, the MSE measure has been employed. Augmentation of the number of interrelated variables of σ, besides augmentation of associations between the exploratory variables did negatively impact the MSE which too increased. Thereupon, the MSE decreases as sample size increases, even when correlations between σ and independent variables are large. The proposed estimators had smaller MSE values in all cases than the OLS estimators. Five of the former, namely, kKM4, kKM5, kKM8, kKM10 and kKM12, besides kK1 and kK2, performed better than the others in terms of the MSE. By and large, five of the suggested estimators; kKM4, kKM5, kKM8, kKM10 and kKM12, are much useful and can be recommended to practitioners. Moreover, since these five estimators outperform the others, the kKM8 and kKM12 estimators are particularly recommended when using models with high residual variances (Muniz *et al*., 2012).

The researchers (Kalatzis *et al*., 2008) used panel data and investigated the investment decisions, taking into consideration presence of financial limitations in 373 large Brazilian corporations in the period 1997-2004. As to use of RR to handle the problem of multicollinearity of model variables, these researchers employed Bayesian econometric model classified by firm's capital intensity. Priors were presumed for the parameters, classifying the model into two categories based on fixed and random effects. In addition, the study considered the normal (Student's t) distributions for model's error. A recurrent predictive density criterion was adopted for comparisons between the investment models. It is worth mentioning that presence of multicollinearity results in important modifications to the parameters and evidence of differences in model parameters' estimates with or without RR based on the findings. Based on the estimation, probably financial restraints are highly important for capital-intensive firms because of their growing agency costs, higher extents of property diversification, higher fixed costs and lower profitability indexes. Furthermore, explanation of investment decisions is highly taken into account in the investment theory literature. Hence, understanding the roles of various factors is highly important, particularly in developing economies, e.g., Brazil, since improvements in financial markets may lead to reducing transactions and information costs which can usually affect investment decisions.

Investment decisions are not as much recurrent as firm's other decisions and decision rules of thumb are difficult to develop. Moreover, roles of information asymmetries in credit markets and the sources of conflicts between capital owners and managers, like agency costs, adverse selections and moral hazards, have been investigated in the literature since they influence the costs of debts. From an empirical perspective, many studies have examined factors related to identification of investment decisions and adoption of liquidity variables in recognition of the critical effects of internal funds on investment. Moreover, because of employment of micro panel data, it is possible to overlook the representative firm and employ firm heterogeneity. Furthermore, firms are often grouped on the basis of their sizes to control for heteroskedasticity. Usually, investment is the dependent variable in the regressions in which the explanatory variables include sales, debt and cash flow, amongst other factors. These variables, mainly sales and cash flow, are associated. However, few discussions of this problem have been found in this study. Thus, the researchers investigated such considerations as the model and data employed and how multicollinearity was dealt with. As well, the variables were examined to check for presence of multicollinearity. Afterwards, the study handled the Bayesian procedure, including parameters' posteriori densities. In addition, the researchers performed model selection because of the criterion of ordinate predictive density. This research also examined the investment decisions of 373 large Brazilian firms from 1997 to 2004. In addition, the study dealt with the role of multicollinearity in predictor variables in assignment of importance to the various investment factors. A Bayesian technique has been employed and substitute specifications have been taken into account for the model, including errors' Student's t versus normal distribution and fixed versus random effects. Based on the criteria for model selection employed, the optimal model was the model of fixed effects and Student's t distribution. Hence, classification of firms by capital intensity is the 'be-all' and 'end-all' outcome of the study on the basis of its proposal. Furthermore, it is feasible to control for and separate, impacts of financial limitations from others in view of this classification. When comparing classification of the firms considered in this study by capital intensity and size, the investment rate proved to correlate more strongly with capital intensity than with size of firm where correlations of investment rate with capital intensity and associations of firm size with investment rate had coefficients of 0.0508 and 0.0069, respectively.

Based on the principal economic finding of this study, firms are subject to financial limitations, especially the capital-intensive companies. In addition, according to RR estimates, capital-intensive companies are financially restrained and more dependent on external fund sources for their investments. Besides, it has been evidenced that cash flow does not act as proxy for future profitability in view of the high cash flow coefficients and low profitability rates. Furthermore as sales influence investment in these companies negatively, the 'sales' variable draws importance of the 'cash flow' variable in the estimation without RR. The estimated coefficients vary in importance based on comparison of these estimates with those obtained from regression without consideration of multicollinearity. Therefore, this proposal considers multicollinearity to a great extent (Kalatzis *et al*., 2008).

In the study (Zaka and Akhter, 2013), Relative Least Squares Method (RLSM), a RR method and least squares (LSM) method were employed to determine parameters of the two-parameter power function distribution. Furthermore, the sampling behavior of estimates was determined based on Monte Carlo simulation. Additionally, this study employed Total Deviation (TD) and the MSE to determine the finest of the three estimators investigated. Therefore, this study determined the optimum estimation method on the basis of different sample sizes and values of parameters. In addition, this study employed the RR method, RLSM and LSM to determine the two power function distribution parameters. In this study, the RR estimator was identified by taking various values of the ridge coefficient, k. In addition, by using the two-parameter power function distribution, the study compared between these three methods to identify the most accurate method, that is, the method producing the lowest MSE. Eventually, the LSM parameters' estimates were found to be very close to the corresponding true values and the corresponding MSE and TD values were very low. The parameter estimates of the RR method and RLSM were close to the respective true values but still not as much close as the LSM estimates since the MSE and TD values produced by the latter two methods were higher than the corresponding values produced by the LSM method. Moreover, the TD and MSE values of all estimators of the scale parameter and all estimators of the shape parameter tended to increase with sample size. In consequence, this study recommended use of the LSM method for estimating parameters of the power function distribution (Zaka and Akhter, 2013).

The researchers (Cule and De Iorio, 2012) considered use of RR which is a common penalized regression method, with data having very high dimensions and more covariates than observations. The dilemma of out-of-sample prediction was the motivation and the data were high-density genotype data obtained from re-sequencing study or genome-wide association. It was formerly proven that RR is more effective for prediction than other penalized regression methods. The choice of a suitable parameter to control for extent of coefficient estimate's shrinkage is one of the problems of RR. Hence, these researchers suggested a method for choosing the ridge parameter by controlling for the predicted observations' variance in the model. In addition, these researchers proved that their own method outperforms subset selection based on univariate tests of association and another penalized regression method, i.e., Hyper Lasso regression, in terms of enhanced prediction error because of simulated data. When the outputs were binary (representing controls and cases as is the setting typically for genome-wide association studies), these researchers extended their approach to regression problems and identified the method on true data example comprising case-control and genotype data on bipolar disorder obtained from the Genetic Association Information Network and the Welcome Trust Case Control Consortium. The researchers considered the out-of-sample prediction problem in high-dimensional regression context when data included less observations that predictors. Further, prediction of an observable characteristic (i.e., phenotype) of interest was the researchers' motivation on the basis of individual's genetic information and probably, other covariates. The problem of phenotype prediction was also tackled in this study based on genetic data. Additionally, the study outlined the distinctive issue and selected ridge parameter in RR (Hoerl and Kennard, 1970a, b; Cule and De Iorio, 2012).

By combining Principal Component Regression (PCR) estimator with an ordinary RR estimator in regression model suffering from the multicollinearity problem, this study (Chandra and Sarkar, 2012) proposed new estimator, referred to the restricted r-k class estimator when linear limitations binding regression coefficients are of stochastic nature. Furthermore, performance of the proposed r-k class estimator in the mixed regression model was compared with those of the stochastic ridge regression and the mixed regression estimators in terms of the Mean Square Error Matrix (MSEM) measure. Moreover, this study examined conditions conducive to dominance of the proposed estimator over the other two estimators. In addition, these researchers performed Monte Carlo study and numerical evaluation to test performance of tests, including superiority conditions of the proposed estimator over the other two estimators. The study then suggested investigation of the r-k class estimator in presence of prior stochastic information. The researchers as well highlighted need for comparing the suggested estimator with the Mean Relative Error (MRE) and other competing estimators based on the MSEM measure. The term $\beta_m$ (r, k) was introduced and its performance was compared with SRRRE and MRE in terms of the MSEM criterion. In addition, the study performed tests to assess the associated conditions and gave simulation work and numerical illustration to evaluate effectiveness of the tests in assessing the conditions of dominance of the $\beta_m$ (r, k) estimator over the other two in terms of the MSEM measure. The empirical findings suggested that $\beta_m$ (r, k) is better as estimator than each of $\beta_m$ and $\beta_m$ (r, k), in view of the MSEM measure (Chandra and Sarkar, 2012).

The researchers (Burt *et al*., 1987) adopted a heuristic criterion to choose acceptable bias level in RR. The criterion was dependent on non-central F-test of the stochastic restraints implied in the ridge estimator. A suitable significance level for the test was based on combined employment of weak and strong MSE criteria. In addition, the procedure was conducted to determine Cobb-Douglas production function for the Central Valley of California on the basis of factor shares as priors instead of the null vector. Based on preliminary results, combined SMSE/WMSE measure with more reasonable priors chooses estimator with a lower bias than the ridge trace. A heuristic measure was employed to choose the bias level in the SPR model. Further, alternative formulation of the SPR procedure was conducted in combination with sample-based measure to assess the bias level in the estimation of the parameters of the Cobb-Douglas production function for the Central Valley of California. Moreover, factor shares were employed as the point to force the ridge estimator in its direction instead of the null vector because the factor shares warrant powerful point estimates of factor elasticity. This study started by reviewing the MSE criterion and the SPR and Ordinary Ridge Regression (ORR) estimators. Then, it introduced strategy for comparing alternative ridge estimators. Finally, the study investigated the ORR, SPR and OLS parameter estimates for the application and based on the findings it drew conclusions. Outcomes of this study, in addition to the empirical application, led to two major conclusions: (i) More bias may be justified by the MSE criterion because the procedure of ridge trace estimator selection proved to introduce a lot of bias to the estimation. Albeit such estimators may always be close to the true parameter vector because of specification error in the model and errors in the independent variables, these issues may challenge any objective analysis and (ii) To choose the bias level, the SPR estimator combined with objective sample criterion proved to be effective in empirical estimation of the Cobb-Douglas production functions, particularly when suitable prior estimates of factor elasticities are employed so as to determine the point in parameter space towards which the estimation vector is inclined (Burt *et al*., 1987).

In effect, a large deal of research on RR investigated choice of the ridge parameter, k. Further, in the statistics literature many algorithms were developed for estimation of k. In the study (Al-Hassan, 2008), seven approaches to estimation of the ridge parameter were examined. So as to assess performance of these estimators, the researchers suggested a simulation approach on the basis of the minimal MSE measure. Based on the simulation approach, two estimators; HKB and GM, were proposed and claimed to be effective under specific conditions. Moreover, many procedures were introduced in the literature to develop ridge estimators that hinge necessarily on an approach to selection of the constant k. Through Monte Carlo simulations, the strengths of correlations and the numbers of observations and variables were modified and performance of some ridge estimators was evaluated by comparing them in view of the mean squares measure. Furthermore, five thousand replications were conducted for each combination of strengths of correlations and numbers of observations and variables. In view of the simulation outcomes, it was found that the GM outperformed the other estimators, both when correlations were moderate and low in strength. However, the HKB outperformed the GM and other estimators in the case of high correlations. In addition, the majority of estimators outperformed the GM in the cases when sample sizes were large and correlations were high. Performances of the HSL, KS and HK estimators were most of the time nearly equivalent. They all performed better than the AM but poorer than the LW. Furthermore, values of the kAM estimates were very large in comparison with other ks which makes the AM highly biased estimator. In consequence, this study does not recommend use of the AM in practice (Al-Hassan, 2008).

Ridge regression which is a type of biased linear estimation method, is a more suitable technique than OLS estimation when handling highly intercorrelated predictor variables in the linear regression model:

$$\vec{Y} = X\vec{\beta} + \vec{u}$$

The study (Khalaf, 2012) evaluated two proposed RR parameters from the MSE viewpoint. Further, this study performed simulation research to examine performance of the proposed estimators relative to performance of the HK, HKB and OLS estimators. As to ridge parameters in the studied situations, the findings indicated that the proposed estimators were more effective than the OLS and the other estimators. By and large, based on the MSE measure, the proposed estimators can improve the HKB, HK and OLS estimates greatly. In addition, the proposed estimators can offer great opportunity for large reduction in the MSE when the degree of multicollinearity as measured by CN, is large. As well, RR is more than simply last resort to restore least squares linear regression in the case of near or full, collinearity of predictor variables. This linear regression technique proved to be beneficial when collinearity is a problem. Based on the MSE, classical multiple linear regressions involve multicollinearity problems and it was confirmed that when input data are multicollinear then the RR performs quite well. This proposal entailed evaluation of two methods based on the MSE via simulations to specify k. Furthermore, the OLS estimator was dominated by these estimators in all of the examined cases and improvement in the suggested estimators is remarkable from a MSE perspective based on the simulation method. However, there is still no consensus on the optimum or most general means of selecting k. However numerous strategies for selection of the optimum value of k have been suggested. Since no rule for selection of k has been evaluated to date, in terms of the MSE the respective ridge estimator proved to be uniformly better than the OLS estimator. Nonetheless, the optimal method for estimation of k is yet an unresolved problem (Khalaf, 2012).

Multicollinearity can be highly potential in studies with two or more predictor variables. The design matrix becomes almost singular in presence of multicollinearity and thus, X and the respective X'X are not of full rank. In such case, the OLS estimate can not be identified. Therefore, presence of multicollinearity in data needs careful attention. In this study (Singh, 2010), the RR mainly aimed at resolving the problem of multicollinearity. The RR technique proposed by Hoerl and Kennard (1970a, b) has become a common tool for analysis of data characterized with high multicollinearity. Addition of small positive quantities to the diagonal elements of the X'X matrix prior to inverting it has also been suggested. Stated otherwise, these researchers replaced:

$$\hat{\beta}_R = (X'X + KI)^{-1} X'Y \text{ with } \hat{\beta} = (X'X)^{-1} X'Y$$

where, $\hat{\beta}_R$ and $\hat{\beta}$ are ridge and OLS estimates of the parameter vector, β, respectively. Though it proved to be biased, the Ridge Estimate (RE) has smaller MSE than the OLS estimator. In addition to its properties, its relation with other estimators and with non-Bayesian and Bayesian interpretations, critical appraisal of choice of the biasing parameter has been carried out. According to this study, the RR plays major role in resolving the problem of multicollinearity. Moreover, Hoerl and Kennard (1970a, b) endorsed addition of small positive quantity to the diagonal elements of the X'X matrix prior to its inversion. That is, they suggested:

$$\hat{\beta}_R = (X'X + KI)^{-1} X'Y$$

where, $\hat{\beta}_R$ is a RE of the parameter vector, β. The RE has smaller MSE than the OLS estimator despite the fact that it proved to be biased. Multicollinearity proved to be of frequent occurrence, not only in theoretical problems but also in problems with particular types of data. Further, the design matrices approach singularity when multicollinearity is present in the data and thus, X and X'X are not of full rank and in such case OLS estimates cannot be provided. So, dropping some highly-correlated variables has been effective strategy that may be taken into account as simple solution to this problem. To mitigate the multicollinearity problem, dropping one or more variables from the model, particularly if the variables more important than to be removed from analysis, may leads to specification bias and in certain cases the cure may be worse than the disease itself. Collection of further data is another solution but this may frequently be expensive or impracticable in many cases.

We may be interested in extracting the greatest possible information from the data we have at our disposal. Then, testing for multicollinearity is required from the onset. Moreover, the RR proved to be a substitute estimation technique when there is extremely high degree of multicollinearity in the data (Vinod, 1978). Albeit it is in general highly advanced solution of multicollinearity, the RR reduces the MSE greatly, thus generating more reliable estimates of β. The RE is a small positive number added to the design matrix's diagonal element before its inversion. Although it was proven that this action is biased, the RE has lower MSE than the OLS estimate and it is comparable with other biased estimators. Originally, the RR was developed to tackle singularity. Anders (2001) mentioned that RR is application of Tikhonov Regularization (TR) which is an approach that has been explored in the approximation theory literature for almost as long as RR has been employed in statistics. Even though many choices for $k\in(0, 1)$ are possible, every choice was denoted with subscript k since it produces new ridge estimator. It is worth highlighting that the RR technique must be implemented cautiously in applied research. Numerous methods for selection of k have been proposed. However, no solid recommendation for an optimum k has been given so far. Obtaining a value of k for a problem will still be something of art. By employing Goldstein and Smith (1974) unbiased estimator of the MSE of biased estimators, Vinod (1978) suggested development of non-spherical confidence intervals that are centered at RE. However, the underdeveloped theory of hypothesis testing with RR reduces its utility (Schmidt, 1976). Nonetheless, these theories still have demerits that require further investigations (Singh, 2010).

It has been proven that parameter's estimation on the basis of minimal residual sum of squares is non-satisfactory in presence of multicollinearity. Hoerl and Kennard (1970a, b) introduced the RR estimator as substitute method. The biasing constant or ridge parameter, plays a major role in parameter estimation in RR. So, bulky volume of research has been conducted to assess the ridge parameter. In the study (Dorugade and Kashid, 2010), a new approach was adopted so as to choose the ridge parameter. Performance of the suggested method was assessed and analyzed via simulation studies in terms of the MSE. The technique presented in this study seems to be highly acceptable because of its small MSE values. The new approach was adopted to evaluate the ridge parameter in RR. The suggested ridge estimator (D k) was founded on the number of data points (n) and strength of multicollinearity in the dataset. Performance of the suggested ridge parameter was evaluated via simulation studies. In order to compare the ratio of the average MSE, the ridge parameter proposed by Hoerl *et al.* (1975) and Khalaf and Shukur (2005) and others was endorsed. The simulation outcomes indicated that performance of the proposed ridge parameter is better than those of other ridge parameters employed in RR (Dorugade and Kashid, 2010).

Parameter estimates based on minimal residual sum of squares are potentially non-satisfactory in multiple regressions, if not incorrect, chiefly when the prediction vectors are non-orthogonal. What ids suggested by this study is estimation procedure based on addition of small positive quantities to the diagonal of X'X. The study also introduced ridge trace which is method for displaying the impacts of non-orthogonally in two dimensions. Then, it was shown how to reinforce X'X to get biased estimates with smaller MSE values. Moreover, estimation on the basis of the matrix $(X'X+KI_p)$, $k>0$, rather than on X'X, possibly can help in counteracting many of the drawbacks concomitant with the normal least squares estimates. In specific, the procedure may be employed to depict sensitivity of estimates to the specific dataset used. It can be employed to get point estimate with small MSE values. It has been demonstrated that when X'X has non-uniform eigenvalue spectrum, estimates of β in $Y = X\beta+\varepsilon$ depending on the minimum residual sum of squares criterion may have high potential for being removed far from β. This non-satisfactory condition has been introduced in estimates that are very large in magnitudes and even that may have wrong sign. The system $(X'X+KI\ \beta^* = X'Y)$ then acts more like orthogonal system by addition of small positive quantity to each diagonal element. When $K = kI$ and all solutions in the interval $0<k<1$ are obtained and because of intercorrelations between predictors, there is potential for getting two-dimensional characterization of the system and depiction of the types of potential difficulties. Furthermore, so as to improve the MSE of estimation, an investigation of the properties of the estimator $\beta^*$ has been presented and degree of improvement is augmented with rise in spread of the eigenvalue spectrum. An estimate based on $\beta^*$ is biased and use of biased estimator reveals prior limit on the regression vector $\beta^*$. However, data in any specific problem contain information which may reveal the class of reasonable generators. Then, explicit illustration of this information is the supreme objectives of the ridge trace, which, hence, can help users, get better estimates (Hoerl and Kennard, 2000).

In General Ridge Regression (GRR), p ridge parameters need to be determined. However, simple RR requires determination of one parameter only. In his textbook, Jurgen Gross described linear regression as substantial complication. According to this proposal, however, identification of the p parameters can be easily and fairly achieved. Furthermore, generalization of the GR estimator that has been derived by Hemmele; Tee Kens and de Boer has been endorsed as well. In view of weighted quadratic loss measure, this estimator which is more conservative, outperforms the estimator of Hoerl and Kennard (1970a, b). The aim of this notice is twofold. Firstly, the study supports that determination of the p ridge parameters can be achieved easily and fairly. Secondly, the study derives characteristics of the GR estimator's MSE and demonstrates that estimators may be built and that they perform in weighted MSE sense better than the OLS estimator. The study is then only concerned with GR estimators whose shrinkage intensity, i, depends on component of transformed data vector. Other estimators like that of Strawderman (1978) are dependent on all components. However, as underlined by Lawless (1981), the other estimators provide little efficiency gains only. In addition, albeit the estimator:

$$\alpha_i^*(\tau,\gamma) = \left\{ \frac{1}{2}\hat{\alpha}_i \left(1 + \gamma\hat{\alpha}_i\sqrt{1 - 4\frac{\sigma_i^2}{\hat{\alpha}_i^2}}\right) \quad \begin{matrix} |\hat{\alpha}_i| / \sigma_i \prec \tau \\ |\hat{\alpha}_i| / \sigma_i \geq \tau \end{matrix} \right.$$

performs better than Hoerl and Kennard (1970a, b) estimator on the basis of the weighted quadratic loss measure, selection of (τ, γ) needs to be made in accordance with prior information. Efficiency gains will then be small but the estimator is best if information is diffuse (De Boer and Hafner, 2005).

The monotone or separation likelihood phenomenon appears in the process of fitting a logistic model if the likelihood converges while not less than a parameter estimate diverges to ±∞. Separation mainly takes place in small samples with many unbalanced and largely predictive risk factors. Additionally, a method developed originally by Firth to remove bias of the maximum likelihood estimates was introduced which may warrant ideal solution to separation. Finite parameter estimates were also generated because of penalized maximum likelihood estimation means. Penalized likelihood ratio tests and profile penalized likelihood confidence intervals proved often to be preferable. However respective Wald tests and confidence intervals are available. Moreover, the evident benefit of the method over previous analysis options was impressively identified as a result of statistical analysis of two cancer studies. The separation problem is by no means negligible and can occur even if parameters of the underlying model have low absolute values. Furthermore, probability of separation is dependent upon number of dichotomous risk factors, sample size, the degree of balance in their distribution and magnitude of the odds ratios concomitant with them. Mehta and Patel (1995) David Firth's medication of the score function was adopted; its great practical relevance may not have been recognized fully. In addition, it was proven that separation is non-negligible problem in logistic regression and that the medication derived originally to remove bias of the maximum likelihood estimates assures typical solution to the problem. Quality of the findings presented has been improved substantially because of extensive empirical research and analyses of two varying clinical datasets. Moreover, application of Firth's methodology proved to allow for unconditional analysis. Lastly, for large parameter estimates where Firth's modification seems to be most useful, inference on the basis of profile penalized likelihood has particularly proved to be preferred over Wald-type tests and confidence intervals. In the present study, the proposed procedure allows for Monte Carlo studies of small samples which formerly led to recurrent separation cases. Therefore, this has resulted in interpretable findings. Likewise, separation took place in some resampling steps when the bootstrap was conducted, even when no separation was incorporated in analysis of the original sample. Thus, bootstrap applicability to logistic regression with small samples has been supported. Additionally, application of Firth-type inference procedures and estimation was facilitated by a Fuzzy Logic (FL) program (Heinze and Schemper, 2002).

Many different estimators of the Ridge parameter, k, have been introduced in the study (Mansson *et al*., 2010) following the studies of Alkhamisi *et al*. (2006), Khalaf and Shukur (2005) and Muniz *et al*. (2012). However, this study differs from the aforementioned ones in three respects: (i) The prediction sum of square (PRESS), MSE and maximum MSE were considered as the performance criteria; (ii) Various error variances were employed (σ is between 0.5 and 5) and (iii) The number of regressors taken into account ranged from 4-12 rather than only 2-4 which is the normal practice. Then, in order to compare performances of estimators, simulation studies were performed owing to that theoretical comparison is quite impossible. Based on results of the simulation studies, it was confirmed that augmenting correlations between independent variables leads to negative effects on the PRESS and MSE. However, raising the number of regressors has positive effects on both the PRESS and MSE. The MSE decreases as sample size is increased, even if associations between independent variables are high. In addition, it was proven that the predominant pictures of the estimators stay the same, both under the PRESS and MSE criteria. However, the distribution's error variance can influence performance of estimators. Based on the estimated PRESS and MSE, the suggested estimators $K_2$, $K_3$, $K_4$, $K_9$ and $K_{12}$ outperform the other estimators. Moreover, it was $K_{12}$ only which always had low average value of k under all simulation scenarios. Therefore, $K_{12}$ produces low, stable values of k which implies that it only produces small bias and may therefore be recommended to practitioners (Mansson *et al*., 2010).

The study (El-Dereny and Rashwan, 2011) employed many RR methods including GRR, Directed Ridge Regression (DRR) and ORR to solve the multicollinearity problem. As well, properties of the RR estimators and approaches to selection of biased RR parameters were investigated. Further, these researchers performed data simulation so as to compare RR methods with OLS methods. Results of this study showed that all RR methods outperform the OLS method in presence of multicollinearity. This study also studied the multicollinearity problem, methods for detecting this problem and multicollinearity effects on outcomes of multiple regression models. In addition, the study developed a number of RR models on the basis of standard deviation, MSE and $R^2$ for each model's estimators to tackle this problem. The RR models were then compared with OLS methods based on simulations of data (2000 iterations). When multicollinearity exists and the finest model is the GRR model, all RR models proved to outperform the OLS model owing to the smaller MSE values of estimators, smaller standard deviations for most estimators and greater $R^2$ values of the former than the latter models (El-Dereny and Rashwan, 2011).

This study (Fu, 1998) took into consideration the P-Bridge regression which is special family of penalized regression of the penalty function:

$$\sum |\beta_j|^\gamma \text{ with } \gamma \geq 1$$

Further, a general method was enhanced in such a way as to handle the bridge estimator. It is worth mentioning that new algorithm for the lasso ($\gamma = 1$) was developed by studying structure of the bridge estimator. The tuning parameter, $\lambda$ and the shrinkage parameter, $\gamma$, were selected by Generalized Cross-Validation (GCV). Comparisons between the bridge model ($\gamma \geq 1$) with many other shrinkage models, e.g., the OLS regression ($\gamma = 0$), the lasso ($\gamma = 1$) and ridge ($\gamma = 2$) regression methods, were held by means of simulation studies. In addition, it was confirmed that bridge regression outperforms ridge and lasso regressions. These approaches were adopted based on analysis of prostate cancer data. A number of computational limitations and advantages were also investigated. To solve bridge regression for $\gamma \geq 1$, structure of the bridge estimator was studied and a general methodology was adopted by this study. Moreover, a simple algorithm was in particular employed for the lasso regression; the shooting method. This proposal was then organized in the same way how structure of bridge estimator was dealt with. Algorithms for lasso and bridge regressions were studied and variance of the bridge estimator was discussed. By means of GCV, this study identified the tuning parameter $\lambda$ and the shrinkage parameter $\gamma$ for bridge regression. The researchers also examined special case of the orthonormal matrix, X and handled bridge penalty as Bayesian prior. Lastly, through the GCV approach, these researchers discussed advantages of the shooting method for lasso regression in addition to limitations of model selection procedure (Fu, 1998).

In the study (Marquardt and Snee, 1975), use of biased estimation in model building and data analysis was discussed. Review of the RR theory and its relation with generalized inverse regression was presented in addition to results of simulation experiment and three examples on use of RR in practice. Comments on computation procedures for ridge and generalized inverse regression, model validation and variable selection procedures were also included. The RR generates coefficients that predict and extrapolate better than the least squares and is safe procedure for selection of variables when the predictors are highly correlated according to the examples studied. Theoretical distillation outcomes and other computer programming needed for this study were produced from a model developed by Gorman and Toman (1966) and Marquardt and Snee (1975).

Assuming that the ridge estimate $\beta$ ($\lambda$) for $\beta$ in the model:

$$\underline{Y} = X\underline{\beta} + \underline{\varepsilon},\ \underline{\varepsilon} \sim N(\underline{O}, \sigma^2 I)$$

$\sigma^2$ is unknown:

$$\hat{\beta}(\lambda) = (X^T X + n\lambda I)^{-1} X^T Y$$

The study (Golub *et al*., 1979) the GCV method for finding good value $\hat{\lambda}$ for $\lambda$ from data. The estimate $\hat{\lambda}$ is the minimizer of v ($\lambda$) calculated from:

$$V(\lambda) = \frac{1}{n} \frac{\|(I - A(\lambda))Y\|^2}{[\frac{1}{2}\text{Trace}(I - A(\lambda))]^2}$$

Where:

$$A(\lambda) = X(X^T X + n\lambda I)^{-1} X^T$$

This estimate is rotation-invariant version of ordinary cross-validation or Allen's PRESS. It does not demand estimate of $\sigma^2$, despite the fact that it behaves as risk-improvement estimator. So, it may be employed when (n-p) is small or in specific cases if $p \geq n$. As well, the GCV method may be employed with singular value truncation and subset selection methods for regression and even to select from combinations of these methods. The GCV method was adopted in RR to estimate k. This estimate can be employed if the number of degrees of freedom for estimating $\sigma^2$ is small or if the true model in some cases involves greater than n parameters as it does not need estimate of $\sigma^2$. This method can also be employed for subset selection or principal components selection in place of RR or even to select from different combinations of principal components, subset selection or ridge methods. A brief numerical example indicative of behavior of this method was presented. It illustrated what experimenter may wish to do in order to investigate properties of the method as regards her/his design matrix (Golub *et al*., 1979).

With the rest independent variables in regression model, tolerance and the Variance Inflation Factor (VIF) are widely used measures of degree of multicollinearity of the ith predictor variable. Many practitioners considered numerous rules of thumb-most commonly the rule of 10 for the VIF as indicator of serious or severe multicollinearity. This rule appears both in advanced statistical textbooks and scholarly articles. When the VIF meets these threshold values the researcher often tries to lower collinearity by deleting one variable or more from analysis; analyzing data using RR or by combining two or more predictor variables into one index. These methods of treating problems concomitant with multicollinearity may create more serious problems than the ones they intend to solve. Hence, these rules of thumb were examined and threshold values for tolerance and the VIF were evaluated, taking into account many other factors that affect variance of the regression coefficients. The VIF values of 10, 20, 40 or greater, do not by themselves affect the outcomes of regression analyses or demand deletion of predictor variable or more from analysis or suggest use of RR or necessitate combining predictor variables into one index. Moreover, some interrelated problems may appear upon adoption of rules of thumb related to certain VIF and/or levels of tolerance. This particularly holds true (i) When assessing accuracy of regression coefficients and stability of outcomes of regression analysis, there is exclusive focus on multicollinearity; (ii) When cases in which the null hypothesis is rejected (or when accuracy of the estimated regression coefficient is adequate) are handled similar to cases when the null hypothesis is not rejected (or when the confidence interval for the regression coefficient is very wide to allow for meaningful conclusions) and (iii) When researchers only investigate remedies of impacts of multicollinearity that involve reducing multicollinearity (O'Brien, 2007).

The study (Farrar and Glauber, 1967) made attempt to define multicollinearity in terms of departures from hypothesized statistical condition and to fashion series of hierarchical measures at three levels of detail for presence, severity and location of multicollinearity within a dataset. In terms of generalized, multivariate normal and linear models, tests were developed. Instead of terms of levels of classical probability, pragmatic interpretation of resulting statistics, like unitless measures of correspondence between hypothesized and sample properties, was advocated. Numerical example and summary complete the paper. Tests of intercorrelation are approached as consecutive tests of dependence of every independent variable on other variables in the dataset. In consequence, this study performed computational and conceptual regression analysis in order to warrant simple and quick, yet serviceable, alternative for factor analysis of independent variable set. Ultimately, the problem was solved and no extra 'revisits' are needed. Such feeling can be much misleading. Though this does not assure cure, diagnosis is an essential first step. Miraculous 'instant orthogonalization' cannot be offered but we optimally end up. The diagnostics described here provide econometrician place to begin in. Albeit econometric problems are tractable, multicollinearity can return from the world of impossible to that of difficult in conjunction with spirit of selectivity in achieving and performing further information (Farrar and Glauber, 1967).

In the statistics literature, too many studies have acknowledged the development of RR estimator by Hoerl and Kennard (1970a, b) to tackle the multicollinearity problem in regression. The RR method received little attention in the field of econometrics as Hoerl and Kennard justified their approach on pragmatic grounds without presenting interpretation which resulted in lack of interest in RR within the econometricians' community. Difficulty in choosing appropriate value of the shrinking factor which is critical for securing dominance over OLS and the prohibitive nature of the MSE criterion, on which rests claim of this dominance, are likely to be among other reasons for reluctance to reception of RR by econometricians and which are examined in this study (Lin and Kmenta, 1982). Though the ORR estimator with given k is linear estimator that is biased, it has smaller MSE than the OLS estimator for values of k in specific interval. The advantage of ORR of this kind over OLS is, for practical purposes, illusory because ORR's interval of dominance over OLS is dependent on the real values of regression parameters. These researchers offer different interpretations of the ORR estimator which offers simple and convenient means of including such knowledge in the estimation and of decreasing size of the MSE, if we possess certain previous knowledge about the parameter space off and if this knowledge is adequately sharp (Lin and Kmenta, 1982).

The RR, perturbing the design moment matrix through k, remains in studies of ill-conditioned systems. Ridge traces, manifesting solutions as functions of k, are intended to express stability as k evolves contrary to transient instabilities in OLS. In the study (Jensen and Ramirez, 2012), the researchers examined derivative traces as analytical tools with respect to stability and developed rational representations for them. Out of ridge solutions' variances, two extra stability gauges are derivatives and variances of derivative traces, both approaching zero as k increases. Contrary to ridge traces and their derivatives, none of the latter is dependent on observed responses and both support deterministic evaluations. The broad use of RR continues, prominently now for calibration in chemical engineering and related fields e.g., Frank and Friedman (1993), Geladi (2002), Kalivas (2005) and Sundberg (1999). In the current study, these basics were complemented by consolidated approach to tracking stability of forthcoming solutions in certain applications. When evaluating stability of ridge solutions, rates of change of ridge traces were examined. Both ridge traces and their derivatives are subject to random disturbances in the observed Y. In other respects as ill conditioning exclusively resides in the matrix X, it is normal to suppose that critical characteristics, like stability, may trace back to X alone, independent of Y. A confirmative answer rests on two extra metrics, i.e., derivatives of variances of ridge estimators and variances of derivative traces. Both approach zero as the ridge parameter increases and both reflect stabilization of those distributions in deterministic way. Case studies in the highly ill-conditioned Hospital Manpower Data illustrate the fundamental outcomes. In these studies, quantities are standardized to liberate diagnostics from dependence on observational variance $\sigma^2$ which is ideally unknown. Users are warned that selections for k on the basis of other aspirations do not necessarily need to manifest the stability taken as the preliminary of Hoerl and Kennard (1970a, b). In spite of the fact as reported, that ridge traces were occasionally misconstrued, careful evaluation of derivative and associated traces can serve to prevent any false and misleading claims depending on ridge traces (Jensen and Ramirez, 2012).

By certain OLS variants, sufficient attention needs to be paid to presence of multicollinearity and its solution. The classical solution which in many situations can be impractical, is to manage further observations or delete one variable or more. Thus, an attempt is made to extract maximal information from the data one has at disposal. First, determine presence of multicollinearity. Then, apply remedial measures to relieve them. Moreover, it is proposed that to face multicollinearity one can use RR, PCR or Generalized Inverse Regression (GIR). This study (Singh, 2012) looks axiomatically only into RR to resolve the multicollinearity problem. Tychonoff (1943) proposed regularization which is known as Tikhonov Regularization and most commonly used for ill-posed problems. In statistics, this is known as RR. Hoerl and Kennard (H-K) proposed the RR technique which grew a common tool for analysis of data with high levels of multicollinearity. Hoerl and Kennard (1970a, b) suggested addition of minute positive quantities to the diagonal elements of the design matrix, X'X, prior to its inversion. It is interesting to note that this approach has frequently been employed since 1943. It is not strange then that Tychonoff (1943) published it in a Russian journal named Doklady Akademii Nauk SSSR. However, it was time when further and further qualities of RR came to light that controversy arose about who should receive credit; Tychonoff (1943) who was employing it and published it in Russian in 1943 or H-K who published it in English in 1970 (Singh, 2012).

Batah and Gore (2009) investigated a Modified Unbiased Ridge (MUR) estimator which is some new shrinkage estimator. This estimator is obtained from the UR, much in the same way how the ORR is obtained from OLS regression. The properties of the MUR were identified and results of its matrix MSE (MMSE) were obtained. The MUR was compared with the URR and ORR based on MMSE. The approach was explained using example based on data produced using the Hoerl and Kennard approach and the results highlighted that under conditions of multicollinearity the MUR is highly effective. This study investigated features of this new estimator and conditions conducive to smaller MMSE by this estimator than by the URR and ORR were derived. Much like in the URR and ORR, a value must be specified for k in the MUR. Because of the simulated data, three different means of determining k were contrasted and the optimum ridge parameters were identified. In addition, some ridge parameter estimators were presented. Results of the study were demonstrated using the Hoerl and Kennard data (Batah and Gore, 2009).

Robustness of the results of an econometric application is to a great extent ascribed to quality of the sample information. In spatial context where usually data have many irregularities, this statement is general rule that becomes particularly suitable. The study (Lauridsen and Mur, 2006) mainly aimed at investigating this question closely, paying particular attention to effect of multicollinearity. It is well known that when linear relations between regressors grow more acute, reliabilities of least squares or maximum-likelihood estimators become worse. It is worth to mention that the authors resolved the discussion in spatial context, closely looking into the shown behavior, under a number of non-favorable conditions, by the most prominent misspecification tests when collinear variables are included in regression. To this end, these researchers planned and solved a Monte Carlo simulation. Moreover, on the basis of findings, these statistics react in various ways to the posed problems. This study therefore mainly aimed at investigating effects of multicollinearity relations on the specification tests that are employed more often in the context of spatial econometric modeling. It was illustrated that the extra impacts on tests of adding extra variable in general disappear for growing multicollinearity. However, this feature interacts with spatial dependence in unpredictable way. The conducted simulation was introduced so as to support some hypotheses. For removed spatial effects, multicollinearity does not influence size of test. This applies regardless of sample size and of features of the DGP, with small deviations only for the KR test. Further, multicollinearity for any DGP and sample size, with little exceptions for the KR test, do not influence power of unadjusted tests largely under residual dependence (i.e., the KR, LM-ERR and Moran's I tests). However, for the powerful one, i.e., the LM-EL test, power grows with multicollinearity. Then, in the case where there is substantial dependence in the equation, multicollinearity increased the power functions of all tests for lag omitted endogenous variables. In effect, this proposal is nothing more than initial approximation of multicollinearity problem in cross-sectional econometric models. In addition, by analyzing limited number of combinations, this study could draw few conclusions. Nonetheless, the cases remaining to be examined, including more complex collinear patterns and/or outliers, appear to be even more appealing (Lauridsen and Mur, 2006).

As a means of overcoming deficits in least squares in multicollinearity problems, Hoerl and Kennard proposed GRR more than 40 years ago. The GRR may be considered reasonably for such problems because high-dimensional regression naturally involves correlated predictors; on the one hand because of the nature of data and on the other hand because of the dimensionality artifact. The GRR has been studied when the number of predictors, p, surpasses the sample size n. In the sense of uniquely-defined least squares estimator, new geometric interpretation for GRR is investigated and insight into its features is gained. In correlated settings, the GRR proved to have beneficial shrinkage properties that can demonstrate excellent performance in scattered high-dimensional contexts but no similar affirmations are confirmed in non-sparse contexts. The authors describe computationally-efficient representation of GRR demanding a linear number of operations in p only, therefore making GRR computationally applicable to high dimensions. The p>n condition represents ill-conditioned scenario where false associations between variables may make estimation and prediction difficult. This study examined how GRR estimators perform under this condition given their stabilizing impacts seen in n>p cases. Using a geometric logic, it can be shown that GRR characteristics when p>n shared comparable features to solution in the classic n>p setting but differed in many important ways. Similar to the classic context, shrinkage plays a role which may result both in enhanced estimation and prediction over the MLS. However, one important difference in high dimensions is that the GRR solution is restricted to subspace enclosing the MLS estimator of dimension n, at most as opposed to subspace of dimension p in the traditional context. Accordingly, the real parameter vector should be sparse in that $p_0 \leq n$ for correct estimation. As well, we may confirm correct estimation in non-sparse situations. With high interest in lasso and lasso-type regularization, high-dimensional sparse settings attracted great deals of investigations and research. In view of findings of the current study, the GRR can have outstanding performance in such settings too if the ridge matrix is appropriately selected. Thus, we should recommend a Bayesian approach to proceed which lends them naturally to ridge estimation (Ishwaran and Rao, 2014).

Dual version of the RR procedure was investigated in the study (Saunders *et al*., 1998). We can perform non-linear regression by adoption of linear regression function in high dimensional feature space. Representation of feature space can result in large, increasing number of variables used by the algorithm. So as to combat this curse of dimensionality, the algorithm allows for use of kernel functions as employed in support vector methods. This study also discusses robust family of kernel functions constructed by means of the ANOVA decomposition method from kernel corresponding to splines with infinite number of nodes. In addition, this study examined a regression estimation algorithm that combines two elements: Dual version of RR is applied to the ANOVA improvement of infinite node splines. Experimental outcomes are identified then based on the Boston Housing dataset which indicates performance of this algorithm relative to those of other algorithms. To prove that the least squares and RR algorithms outperform others, experiments were conducted on the well-known Boston housing dataset. The ANOVA kernels take into consideration a subset only of input parameters. However, they can improve the outcomes obtained on the very kernel function without implementation of the ANOVA method based on findings. Then, the dual form of the least squares and RR has been introduced. Actually, combining the dual version of RR with the ideas of Gammerman *et al*. (1998) is appealing direction of developing outcomes in this research since this helps in attaining measure of confidence for predictions output by the proposed algorithms. It is expected that plain closed-form formulas can be obtained in this case (Saunders *et al*., 1998).

The study (Yanagihara, 2013) took into consideration optimization of ridge parameters in GRR by minimizing model selection measure. As solution to the minimization problem for one model selection measure, namely, Mallows' Cp measure, can be explicitly achieved with GRR, the latter has substantial advantage over RR. However, such solution for any model selection measures (for example, GCV or Cross-Validation (CV) or the Cp measure) cannot be explicitly obtained with RR. On the other hand, since a good error variance estimate is needed in order for the Cp measure to work well; this measure compared with the GCV and CV criteria, is at disadvantage. Moreover, ridge parameters optimized by reducing the GCV measure may too be achieved by closed forms in GRR according to this study. Further, because of the GCV measure for ridge parameter optimization, we can overcome one disadvantage of GRR (Yanagihara, 2013).

A simulation study is introduced to help in examining robustness of six estimators on MLR model with combined problems of non-normal errors and multicollinearity. Then, the study (Midi and Zahari, 2012) tended to compare performances of the six estimators: MM, robust RR estimator based on MM estimator (RMM), Weighted Ridge (WRID), Ridge Least Absolute Value (RLAV), RR and OLS. By integrating the RM with robust RR, the RMM is modification to the RR. Among these estimators for numerous disturbance distribution combinations and degrees of multicollinearity, the RMM empirically proved to be the optimum. When disturbances are normal and correlation is strong, the RR with large sample size performed better marginally than the RMM, otherwise, the RMM was superior. When disturbances were normal and non-normal with no multicollinearity, the LS and MM performed better than the RMM. However, the RMM is more effective when the extent of multicollinearity is high. For several combinations of error distribution types and multicollinearity degrees, the RMM outperformed the other two estimators in view of comparisons between the WRID, robust ridge estimators, RLAV and RMM. So, based on the simulation studies when outliers and multicollinearity are both present, the RMM estimator warrants the most feasible choice (Midi and Zahari, 2012).

In face recognition applications, the problem of non-controllable illumination is substantial challenge. In the study (Huang and Yang, 2012), the researchers adopted new face recognition framework, i.e., Improved Principal Component Regression Classification (IPCRC) algorithm to overcome the multicollinearity problem in linear regression. The IPCRC method performs Principal Component Analysis (PCA) mainly so as to project face images onto face space. The first principal components are dropped intentionally to enhance robustness against illumination changes. Thereafter, Linear Regression Classification (LRC) is applied on projected data and identity is determined via the minimal reconstruction error. Based on experiments conducted on FERET and Yale B facial datasets, the suggested framework outperforms the state-of-the-art approaches and exhibits promising capabilities against variations in severe illumination. Basically for the face recognition technology, the PCA is broadly adopted to reduce dimensionality in computer vision areas. The eigenfaces based on the PCA were shown in face recognition. In PCA, data are presented as linear combination of orthonormal vectors that boost data scatter (covariance) across all images. Regarding PCA transform, to face recognition, numerous variants were introduced. In linear regression, the traditional PCA and PCA with zero average (PCAZ) will be performed according to this research (Huang and Yang, 2012).

Accurate assessment of groundwater's vulnerability to pollution is badly needed for groundwater conservation and management. The study (Ahn *et al*., 2012) stressed effective selection of the key hydrogeological parameters affecting vulnerability of groundwater to pollution in aquifer according to Genetic Algorithm (GA) and integrated model using RR (GA-RR). According to the GA-RR method, net recharge, depth to water influence of vadose zone media and topography were the hydro-geological variables that affected vulnerability of the Korean aquifer to trichloroethene pollution.

For numerous Artificial Intelligence (AI) and non-linear statistical methods like the decision trees, multinomial logistic regression, case-based reasoning and artificial neural networks, employment of these selected hydro-geological variables proved to be correct. Thus, the GA-RR approach enhances performance of AI techniques in evaluating vulnerability of groundwater to pollution and they proved to be quite effective in identifying determinant hydro-geological variables in aquifers' vulnerability to pollution based on these findings (Ahn *et al*., 2012).

## CONCLUSION

This study corresponded to review of the literature from 1964-2014 on the estimators that were intended to evaluate the ridge parameter, k, so as to provide some guidance for practicing professionals who are attempting to employ models and methods that are relevant for handling the least squares method in presence of multicollinearity. By and large, based on the investigated literature in this study:

- We overviewed the major variables that may influence multicollinearity
- Most of the models employed in RR outperform the OLS regression approach when the multicollinearity problem prevails and GRR is the optimum model as it has lower estimates' MSE and standard deviation values as well as larger $R^2$ values
- A number of studies were concerned about the multicollinearity problem, approaches to detection of this problem and its effects on results of multiple regression models

## REFERENCES


Ahn, J.J., Y.M. Kim, K. Yoo, J. Park and K.J. Oh, 2012. Using GA-Ridge regression to select hydro-geological parameters influencing groundwater pollution vulnerability. Environ. Monit. Assess., 184: 6637-6645.

Al-Hassan, Y.M., 2008. A Monte Carlo evaluation of some ridge-type estimators. Jordan J. Applied Sci., 10: 101-110.

Alkhamisi, M., G. Khalaf and G. Shukur, 2006. Some modifications for choosing ridge parameters. Commun. Stat. Theory Methods, 35: 2005-2020.

Alkhamisi, M.A. and G. Shukur, 2008. Developing ridge parameters for SUR model. Commun. Stat. Theory Methods, 37: 544-564.

Anders, B.J., 2001. Ridge Regression and Inverse Problems. Stocks University, Sweden.

Batah, F.S.M. and S.D. Gore, 2009. Ridge regression estimator: Combining unbiased and ordinary ridge regression methods of estimation. Surv. Math. Applic., 4: 99-109.

Burt, O.R., M.D. Frank and B.R. Beattie, 1987. Prior information and heuristic ridge regression for production function estimation. Western J. Agric. Econ., 12: 135-143.

Chandra, S. and N. Sarkar, 2012. A restricted r-k class estimator in the mixed regression model with autocorrelated disturbances. Indian Statistical Institute, India, July 2011-June 2012. http://www.isical.ac.in/~eru/erudp/2013-02.pdf.

Chopra, P., R.K. Sharma and M. Kumar, 2013. Ridge regression for the prediction of compressive strength of concrete. Int. J. Innov. Eng. Technol., 2: 106-111.

Cule, E. and M. De Iorio, 2012. A semi-automatic method to guide the choice of ridge parameter in ridge regression. arXiv: 1205.0686, May 4, 2012. http://arxiv.org/pdf/1205.0686.pdf.

De Boer, P.M. and C.M. Hafner, 2005. Ridge regression revisited. Stat. Neerlandica, 59: 498-505.

Dempster, A.P., N.M. Laird and D.B. Rubin, 1977. Maximum likelihood from incomplete data via the EM algorithm. J. R. Stat. Soc. Ser. B (Methodol.), 39: 1-38.

Dorugade, A.V. and D.N. Kashid, 2010. Alternative method for choosing ridge parameter for regression. Applied Math. Sci., 4: 447-456.

El-Dereny, M. and N.I. Rashwan, 2011. Solving multicollinearity problem using ridge regression models. Int. J. Contemporary Math. Sci., 6: 585-600.

Farrar, D.E. and R.R. Glauber, 1967. Multicollinearity in regression analysis: The problem revisited. Rev. Econ. Stat., 49: 92-107.

Frank, I.E. and J.H. Friedman, 1993. A statistical view of some chemometrics regression tools. Technometrics, 35: 109-135.

Fu, W.J., 1998. Penalized regressions: The bridge versus the lasso. J. Comput. Graphical Stat., 7: 397-416.

Gammerman, A., V. Vapnik and V. Vovk, 1998. Learning by Transduction. In: Uncertainty in Artificial Intelligence, Cooper, G. and S. Moral (Eds.)., Morgan Kaufmann, San Francisco, CA., pp: 148-155.

Geladi, P., 2002. Some recent trends in the calibration literature. Chemometr. Intell. Lab., 211: 224-10.1016/S0169-7439(01)00197-6.

Gibbons, I.R., 1981. Cilia and flagella of eukaryotes. J. Cell Biol., 91: s107-s124.

Goldstein, M. and A.F.M. Smith, 1974. Ridge-type estimators for regression analysis. J. Royal Stat. Soc. Series B (Methodological), 36: 284-291.

Golub, G.H., M. Heath and G. Wahba, 1979. Generalized cross-validation as a method for choosing a good ridge parameter. Technometrics, 21: 215-223.

Gorman, J.W. and R.J. Toman, 1966. Selection of variables for fitting equations to data. Technometrics, 8: 27-51.

Heinze, G. and M. Schemper, 2002. A solution to the problem of separation in logistic regression. Stat. Med., 21: 2409-2419.

Hoerl, A.E., 1962. Application of ridge analysis to regression problems. Chem. Eng. Prog., 58: 54-59.

Hoerl, A.E. and R.W. Kennard, 1970a. Ridge regression: Applications to nonorthogonal problems. Technometrics, 12: 69-82.

Hoerl, A.E. and R.W. Kennard, 1970b. Ridge regression: Biased estimation for nonorthogonal problems. Technometrics, 12: 55-67.

Hoerl, A.E., R.W. Kannard and K.F. Baldwin, 1975. Ridge regression: Some simulations. Commun. Stat., 4: 105-123.

Hoerl, A.E. and R.W. Kennard, 2000. Ridge regression: Biased estimation for nonorthogonal problems. Technometrics, 42: 80-86.

Huang, S.M. and J.F. Yang, 2012. Improved principal component regression for face recognition under illumination variations. IEEE Signal Process. Lett., 19: 179-182.

Ishwaran, H. and J.S. Rao, 2014. Geometry and properties of generalized ridge regression in high dimensions. Contemp. Math., 622: 81-93.

Jensen, D.R. and D.E. Ramirez, 2012. Variations on ridge traces in regression. Commun. Stat. Simul. Comput., 41: 265-278.

Kalatzis, A.E.G., C.F. Bassetto and C.R. Azzoni, 2008. Multicollinearity and financial constraint in investment decisions: A bayesian ridge regression. Proceedings of the International Conference on Applied Economics, May 15-17, 2008, Kastoria, Greece, pp: 437-447.

Kalivas, J.H., 2005. Multivariate calibration: An overview. Anal. Lett., 38: 2259-2279.

Khalaf, G. and G. Shukur, 2005. Choosing ridge parameter for regression problem. Commun. Stat. Theory Methods, 34: 1177-1182.

Khalaf, G., 2012. A proposed ridge parameter to improve the least square estimator. J. Mod. Applied Stat. Methods, Vol. 11.

Kibria, B.M.G., 2003. Performance of some new ridge regression estimators. Commun. Stat. Simul. Comput., 32: 419-435.

Lauridsen, J. and J. Mur, 2006. Multicollinearity in cross-sectional regressions. J. Geogr. Syst., 8: 317-333.

Lawless, J.F. and P. Wang, 1976. A simulation study of ridge and other regression estimators. Commun. Stat. Theory Methods, 5: 307-323.

Lawless, J.F., 1981. Mean squared error properties of generalized ridge estimators. J. Am. Stat. Assoc., 76: 462-466.

Lin, K. and J. Kmenta, 1982. Ridge regression under alternative loss criteria. Rev. Econ. Stat., 64: 488-494.

Mansson, K., G. Shukur and B.M. Golam Kibria, 2010. A simulation study of some ridge regression estimators under different distributional assumptions. Commun. Stat. Simulation Comput., 39: 1639-1670.

Marquardt, D.W. and R.D. Snee, 1975. Ridge regression in practice. Am. Stat., 29: 3-20.

McDonald, G.C. and D.I. Galarneau, 1975. A Monte Carlo evaluation of some ridge-type estimators. J. Am. Stat. Assoc., 70: 407-416.

Mehta, C.R. and N.R. Patel, 1995. Exact logistic regression: Theory and examples. Stat. Med., 14: 2143-2160.

Midi, H. and M. Zahari, 2012. A simulation study on ridge regression estimators in the presence of outliers and multicollinearity. Jurnal Teknologi, 47: 59-74.

Muniz, G. and B.G. Kibria, 2009. On some ridge regression estimators: An empirical comparisons. Commun. Stat. Simul. Comput., 38: 621-630.

Muniz, G., B.G. Kibria, K. Mansson and G. Shukur, 2012. On developing ridge regression parameters: A graphical investigation. Sort: Stat. Operations Res. Trans., 36: 115-138.

O'Brien, R.M., 2007. A caution regarding rules of thumb for variance inflation factors. Qual. Quantity, 41: 673-690.

Pasha, G.R. and M.A. Shah, 2004. Application of ridge regression to multicollinear data. J. Res. (Sci.), 15: 97-106.

Saunders, C., A.Gammerman and V. Vovk, 1998. Ridge regression learning algorithm in dual variable. Proceedings of the 15 th International Conference on machine learning, June 13-15, 2007, Morgan Kaufmann Publishers Inc. San Francisco, CA, USA, pp: 515-521.

Schmidt, P., 1976. Econometrics. Marcel Dekker, New York.

Singh, R., 2010. A survey of ridge regression for improvement over ordinary least squares. IUP J. Comput. Math., 3: 54-74.

Singh, R., 2012. Solution of multicollinearity by ridge regression. Int. J. Res. Comput. Applic. Manage., 2: 130-136.

Strawderman, W.E., 1978. Minimax adaptive generalized ridge regression estimators. J. Am. Stat. Assoc., 73: 623-627.

Sundberg, R., 1999. Multivariate calibration-direct and indirect regression methodology. Scand. J. Stat., 26: 161-207.

Tychonoff, A.N., 1943. [On the stability of inverse problems]. Doklady Akademii Nauk SSSR, 39: 195-198, (In Russian).

Vinod, H.D., 1978. Equivariance of ridge estimators through standardization: A note. Commun. Stat. Part A: Theory Meth., 7: 1157-1161.

Wichn, D. and Wahba, 1988. Applied Regression Analysis: A Research Tool. Wordsworth and Brooks, Pacific Gross, California.

Yanagihara, H., 2013. Explicit solution to the minimization problem of generalized cross-validation criterion for selecting ridge parameters in generalized ridge regression. TR-No. 13-07, May 2013, Hiroshima Statistical Research Group, Japan, pp: 1-17.þ

Zaka, A. and A.S. Akhter, 2013. Methods for estimating the parameters of the power function distribution. Pak. J. Stat. Oper. Res., 9: 213-224.