DigitalCommons @WayneState

Journal of Modern Applied Statistical Methods

Volume 12 | Issue 2 Article 17

11-1-2013

A Comparison between Biased and Unbiased Estimators in Ordinary Least Squares Regression

Ghadban Khalaf
King Khalid University, Saudi Arabia

Journal of Modern Applied Statistical Methods
November 2013, Vol. 12, No. 2, 293-303.

ISSN 1538 − 9472

A Comparison between Biased and Unbiased Estimators in Ordinary Least Squares Regression

Ghadban Khalaf
King Khalid University
Saudi Arabia

During the past years, different kinds of estimators have been proposed as alternatives to the Ordinary Least Squares (*OLS*) estimator for the estimation of the regression coefficients in the presence of multicollinearity. In the general linear regression model, $\vec{Y} = X\vec{\beta} + \vec{e}$, it is known that multicollinearity makes statistical inference difficult and may even seriously distort the inference. Ridge regression, as viewed here, defines a class of estimators of $\vec{\beta}$ indexed by a scalar parameter *k*. Two methods of specifying *k* are proposed and evaluated in terms of Mean Square Error (*MSE*) by simulation techniques. A comparison is made with other ridge-type estimators evaluated elsewhere. The estimated *MSE* of the suggested estimators are lower than other estimators of the ridge parameter and the *OLS* estimator.

Keywords: *OLS* estimator, linear regression, multicollinearity, ridge regression, Monte Carlo simulation.

Introduction

Consider the multiple linear regression model

$$\vec{Y} = X\vec{\beta} + \vec{e} \tag{1}$$

where $\vec{Y}$ is an $(n \times 1)$ response vector, *X* is a fixed $(n \times p)$ matrix of independent variables of rank *p*, $\vec{\beta}$ is the unknown $(p \times 1)$ parameter vector of regression coefficients and, finally, $\vec{e}$ is an $(n \times 1)$ vector of uncorrelated errors with mean zero and common unknown variance σ^2. If $X'X$ is nonsingular, the *OLS* estimator for $\vec{\beta}$ is given by

Ghadban Khalaf is an Associate Professor in the Department of Mathematics.

$$\hat{\beta} = (X'X)^{-1} X'\vec{Y} \tag{2}$$

For orthogonal data, the *OLS* estimator in the linear regression model is strongly efficient. But in the presence of multicollinearity, the *OLS* efficiency can be reduced and hence an improvement upon it would be necessary and desirable.

The term multicollinearity is used to denote the presence of linear relationships, or near linear relationships, among explanatory variables. If the explanatory variables are perfectly linearly correlated, that is, if the correlation coefficient for these variables is equal to unity, then the parameters become indeterminate; i.e, it is impossible to obtain numerical values for each parameter separately and the method of lease squares breaks down. Conversely, if the correlation coefficient for the explanatory variables is equal to zero, then the variables are called orthogonal and there are no problems concerning the estimates of the coefficients.

When multicollinearity occurs, the least squares estimates are still unbiased and efficient but the problem is that; the estimated standard error $S_{\hat{\beta}_i}$ for the coefficient $\hat{\beta}_i$ become infinitely large; i.e, the standard error tends to be larger than it would be in the absence of multicollinearity and when $S_{\hat{\beta}_i}$ is larger than it would be, then the *t*- value for testing the significance of β_i is smaller than it should be. Thus one is likely to conclude that a variable X_i is not important in the relationship when it real1y is.

To solve the problem of multicollinearity, there is no single solution that will eliminate multicollinearity altogether. One common procedure is to select the independent variable most seriously involved in the multicollinearity and remove it from the model. This procedure often improves the standard error of the remaining coefficients and may make formerly insignificant variables significant, since the elimination of a variable reduces any multicollinearity caused by it. The difficulty with this approach is that the model now may not correctly represent the population relationship and all estimated coefficients would contain a population specification.

The procedure of increasing the sample size is sometimes recommended as another suggested procedure to solve the problem of multicollinearity. In fact this method improves the precision of an estimator and hence reduces the adverse effects of multicollinearity.

Hoerl and Kennard (1970) suggested a new technique to overcome the problem of multicollinearity. This technique is called ridge regression. Ridge regression is a variant of ordinary multiple linear regression whose goal is to circumvent the predictors collinearity. It gives up the least squares as a method for estimating the parameters of the model and focuses instead of the $X'X$ matrix. This matrix will be artificially modified so as to make its determinant appreciably different from zero. This is accomplished by adding a small positive quantity, say *k* (*k* > 0), to each of the diagonal elements of the matrix $X'X$ before inverting it for least squares estimation. The resulting estimator is given by

$$\hat{\vec{\beta}}(k) = (X'X + kI_p)^{-1} X'\vec{Y}, \quad k > 0 \tag{3}$$

which coincides with the *OLS* estimator, defined by (2), when $k = 0$. The resulting estimator will be biased, but have smaller variances than $\hat{\vec{\beta}}$. This is precisely what the ridge regression estimator we study can accomplish.

The plan of this paper is as follows: Section 2 presents the proposed estimators included in the study; a novel feature is our proposed ridge estimator which, as we shall see presently, has lower *MSE*. Section 3 is described the simulation technique that we have adopted in our study to evaluate the performance of the new values of the ridge parameter we suggest. The results of the simulation study, which appear in the tables, are presented in Section 4. Finally, Section 5 contains summary and conclusions.

The Proposed Estimators

With the ridge estimator method, there arises the problem of determining an optimal value of *k*. With a good choice of *k*, one might hope to improve on the *OLS* estimator for every coefficient.

Hoerl, Kennard and Baldwin (1975) showed, through simulation, that the use of ridge estimator with the following biasing parameter

$$\hat{k} = \frac{p\hat{\sigma}^2}{\sum_{i=1}^{p} \hat{\beta}_i^2} \tag{4}$$

implies that $MSE(\hat{\vec{\beta}}(k)) < MSE(\hat{\vec{\beta}})$, where p denotes the number of parameters (excluding the intercept) and $\hat{\sigma}^2$ is the usual unbiased estimate of σ^2, defined by;

$$\hat{\sigma}^2 = \sum_{i=1}^{n} e_i^2 / (n - p - 1).$$

They showed that the probability of a smaller *MSE* using (4) increases with the number of parameters *p*. We will use the acronym *HKB* for the estimator (4).

Khalaf and Shukur (2005) suggested a modification of Hoerl and Kennard (1970) given by;

$$\hat{k} = \frac{t_{\max}\,\hat{\sigma}^2}{(n-p)\hat{\sigma}^2 + t_{\max}\hat{\beta}_{\max}^2} \tag{5}$$

which guaranteed lower *MSE*, where $t_{\max}$ is the maximum eigenvalue of $X'X$ matrix. For this estimator we will use the acronym *KS*.

From the estimators (4) and (5), we suggest as a modification of *HKB* and *KS* by multiplying them by the amount;

$$\frac{\frac{1}{2}(t_{\max} + t_{\min})}{\sum_{i=1}^{p}\left|\hat{\beta}_i\right|} = \frac{t_{\max} + t_{\min}}{2\sum_{i=1}^{p}\left|\hat{\beta}_i\right|},$$

where $t_{\min}$ is the minimum eigenvalue of the matrix $X'X$. This leads to the following estimators;

$$\hat{k}_1 = \frac{(t_{\max} + t_{\min})}{2\sum_{i=1}^{p}\left|\hat{\beta}_i\right|} \cdot \frac{p\hat{\sigma}^2}{\sum_{i=1}^{p}\hat{\beta}_i^2} \tag{6}$$

$$\hat{k}_2 = \frac{(t_{\max} + t_{\min})\, t_{\max}\,\hat{\sigma}^2}{2\sum_{i=1}^{p}\left|\hat{\beta}_i\right|((n-p)\hat{\sigma}^2 + t_{\max}\hat{\beta}_{\max}^2)}. \tag{7}$$

For our two suggested estimators, defined by (6) and (7), we use the acronym K_1 and K_2, respectively.

The Simulation Study

A simulation study was conducted in order to draw conclusions about the performance of our suggested estimators relative to *HKB*, *KS* and the *OLS* estimator. To achieve different degree of collinearity, following Kibria (2003), the explanatory variables are generated by using the following equation;

$$x_{ij} = (1-\rho^2)^{1/2} z_{ij} + \rho\, z_{ip}, \qquad i = 1, 2, \ldots, n. \qquad j = 1, 2, \ldots, p$$

where z_{ij} are independent standard normal distribution, p is the number of the explanatory variables and ρ is specified so that the correlation between any two explanatory variables is given by ρ^2 . Three different sets of correlation are considered according to the value of ρ = 0.85, 0.95 and 0.99. The n observations for the dependent variable are determined by the following equation;

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \ldots + \beta_p x_{ip} + e_i, \qquad i = 1, 2, \ldots, n$$

where e_i are *i.i.d* pseudo-random numbers. In this study, β_0 is taken to be zero and the term e_i is generated from each of the following distributions: $N(0, 1)$, $T(3)$, $T(7)$ and $F(3, 11)$. The parameters values are chosen so that $\sum_{i=1}^{p} \beta_i^2 = 1$, which is a common restriction in simulation studies (see Muniz and Kibria (2009)).

The other factors we chose to vary is the sample size and the number of regressions. We generate models consisting of 25, 50, 100 and 150 observations and with 2 and 4 explanatory variables. It is noted from the results of the previous simulation studies (see Khalaf and Shukur (2005), Alkhamisi and Shukur (2008) and Khalaf (2011)) that increasing the number of regressor and using non-normal pseudo random numbers to generate e_i leads to a higher estimated *MSE*, while increasing the sample size leads to a lower estimated *MSE*.

The criterion proposed here for measuring the goodness of an estimator is the *MSE* using the following formula;

$$MSE(\hat{\vec{\beta}}_r) = \frac{1}{5000} \sum (\hat{\vec{\beta}}_r - \vec{\beta})'(\hat{\vec{\beta}}_r - \vec{\beta}), \tag{8}$$

where $\hat{\vec{\beta}}_r$ is the estimator of $\vec{\beta}$ obtained from the *OLS* estimator or from the ridge estimator for different estimated values of *k* considered for comparison reasons and, finally, 5000 is the number of replicates used in the Monte Carlo simulation.

The Simulation Results

Tables 1 – 6 below, present the output from the Monte Carlo experiment concerning properties of the different methods that used to choose the ridge parameter *k*. The results showed that the estimated *MSE* is affected by all factors we choose to vary in the design of experiment. It is also noted that the higher the degree of correlation the higher estimated *MSE*, but this increase is much greater for the *OLS* than the ridge regression estimator. The distribution of the error term and the number of explanatory variables having a different impact on the estimators.

Table 1. Estimated *MSE* when $\rho = 0.85$ and $p = 2$.

	OLS	*HKB*	*KS*	K_1	K_2
N(0, 1)					
25	0.238	0.181	0.190	0.243	0.181
50	0.111	0.093	0.097	0.255	0.176
100	0.057	0.051	0.053	0.266	0.178
150	0.034	0.032	0.032	0.282	0.184
T(3)					
25	2.259	0.957	1.325	0.506	0.497
50	1.219	0.602	0.896	0.521	0.364
100	0.531	0.312	0.445	0.588	0.329
150	0.473	0.261	0.414	0.632	0.351
T(7)					
25	0.350	0.248	0.266	0.304	0.218
50	0.169	0.135	0.143	0.324	0.212
100	0.076	0.067	0.069	0.352	0.220
150	0.053	0.048	0.049	0.367	0.224
F(3, 11)					
25	0.853	0.502	0.584	0.383	0.289
50	0.391	0.261	0.309	0.429	0.236
100	0.178	0.139	0.157	0.481	0.276
150	0.126	0.104	0.115	0.520	0.290

Table 2. Estimated *MSE* when $\rho = 0.85$ and $p = 4$.

	OLS	*HKB*	*KS*	K_1	K_2
N(0, 1)					
25	0.796	0.549	0.572	0.187	0.159
50	0.334	0.255	0.268	0.216	0.162
100	0.156	0.131	0.137	0.257	0.182
150	0.103	0.090	0.093	0.284	0.196
T(3)					
25	7.387	3.781	4.330	1.049	1.205
50	6.222	2.961	4.179	1.073	1.622
100	1.685	0.969	1.318	0.509	0.409
150	1.240	0.754	1.018	0.551	0.332
T(7)					
25	1.159	0.730	0.776	0.212	0.178
50	0.504	0.362	0.389	0.255	0.184
100	0.235	0.188	0.200	0.323	0.218
150	0.152	0.127	0.135	0.353	0.231
F(3, 11)					
25	2.667	1.446	1.601	0.338	0.328
50	1.130	0.699	0.805	0.316	0.226
100	0.578	0.402	0.468	0.415	0.263
150	0.362	0.271	0.311	0.467	0.282

Table 3. Estimated MSE when $\rho = 0.95$ and p = 2.

	OLS	*HKB*	*KS*	K_1	K_2
N(0, 1)					
25	0.705	0.427	0.450	0.193	0.147
50	0.353	0.250	0.265	0.193	0.134
100	0.168	0.133	0.140	0.220	0.145
150	0.114	0.095	0.010	0.231	0.149
T(3)					
25	7.899	3.010	3.580	1.501	1.725
50	5.575	2.137	2.969	0.988	1.256
100	1.703	0.789	1.152	0.486	0.275
150	1.283	0.655	0.959	0.541	0.299
T(7)					
25	1.174	0.670	0.718	0.241	0.186
50	0.528	0.340	0.371	0.249	0.164
100	0.250	0.185	0.200	0.287	0.177
150	0.161	0.127	0.137	0.311	0.187
F(3, 11)					
25	2.556	1.223	1.372	0.401	0.343
50	1.167	0.623	0.738	0.336	0.215
100	0.566	0.346	0.419	0.397	0.224
150	0.378	0.251	0.300	0.444	0.244

Table 4. Estimated *MSE* when ρ = 0.95 and $p = 4$.

	OLS	HKB	KS	K_1	K_2
N(0, 1)					
25	2.356	1.308	1.351	0.154	0.145
50	1.057	0.673	0.705	0.125	0.099
100	0.359	0.251	0.266	0.120	0.088
150	0.323	0.245	0.258	0.194	0.138
T(3)					
25	18.886	8.484	9.026	2.145	2.299
50	14.573	7.274	8.218	1.939	2.315
100	4.122	2.079	2.527	0.284	0.194
150	3.390	1.815	2.266	0.422	0.306
T(7)					
25	3.716	1.996	2.068	0.228	0.221
50	1.502	0.892	0.948	0.146	0.114
100	0.745	0.495	0.534	0.199	0.139
150	0.478	0.341	0.368	0.239	0.162
F(3, 11)					
25	8.220	4.148	4.334	0.776	0.796
50	3.578	1.882	2.064	0.206	0.171
100	1.755	1.034	1.170	0.248	0.164
150	1.180	0.741	0.849	0.309	0.195

Table 5. Estimated *MSE* when ρ = 0.99 and $p = 2$.

	OLS	HKB	KS	K_1	K_2
N(0, 1)					
25	4.050	1.850	1.905	0.349	0.331
50	1.776	0.884	0.931	0.133	0.099
100	0.913	0.533	0.568	0.138	0.091
150	0.572	0.358	0.385	0.155	0.099
T(3)					
25	43.786	15.618	16.407	12.046	12.512
50	21.736	7.673	8.510	3.155	3.377
100	8.794	3.602	4.217	0.481	0.399
150	7.046	2.461	3.274	0.362	0.231
T(7)					
25	6.108	2.657	2.745	0.561	0.544
50	2.623	1.192	1.274	0.171	0.124
100	1.370	0.732	0.797	0.178	0.111
150	0.865	0.502	0.551	0.204	0.123
F(3, 11)					
25	12.863	4.822	5.037	1.421	1.438
50	6.402	2.550	2.779	0.343	0.296
100	3.329	1.508	1.715	0.246	0.152
150	1.901	0.899	1.060	0.279	0.151

Table 6. Estimated *MSE* when ρ = 0.99 and $p = 4$.

	OLS	*HKB*	*KS*	K_1	K_2
N(0, 1)					
25	13.319	6.484	6.547	0.971	0.981
50	6.095	3.078	3.141	0.109	0.107
100	2.708	1.466	1.520	0.061	0.050
150	1.720	0.990	1.036	0.076	0.057
T(3)					
25	169.385	72.238	73.397	58.482	59.442
50	65.170	33.982	34.732	15.466	15.685
100	30.913	15.077	15.913	2.328	2.448
150	19.922	8.885	9.738	0.505	0.556
T(7)					
25	19.789	9.337	9.473	1.739	1.756
50	8.782	4.342	4.442	0.230	0.229
100	4.068	2.152	2.240	0.077	0.063
150	2.550	1.390	1.467	0.086	0.062
F(3, 11)					
25	44.422	20.834	21.062	7.010	7.089
50	21.347	9.485	9.785	1.073	1.131
100	9.172	4.498	4.730	0.145	0.131
150	6.178	3.077	3.303	0.114	0.085

For non-normal error term in combination with $\rho = 0.95$ and $\rho = 0.99$ leads to a larger estimated *MSE* for the *OLS* estimator and the ridge parameter, especially when *n* is small, but when the sample size increases the estimated *MSE* of the suggested ridge parameters, namely K_1 and K_2 decreases substantially.

The performance of K_1 and K_2 is well for all cases when the error term is distributed as a normal and, when *n* is greater than 25 and the error term in non-normal .

When *n* is greater than 25, the modified ridge parameter performance, defined by (6) and (7), is much better than the estimators *HKB*, *KS* and the *OLS*, where K_2 has a low estimated *MSE* when the number of regressor equals 4.

Summary and Conclusions

In multiple linear regression, the effect of non-orthogonality of the explanatory variables is to pull the least squares estimates of the regression coefficients away from the true coefficients, $\vec{\beta}$, that one is trying to estimate. The coefficients can be both too large in absolute value and incorrect with respect to sign. Furthermore, the variance and the covariance of the *OLS* tend to become too large.

A slight movement away from this point can give completely different estimates of the coefficients. This is accomplished by adding a small positive quantity, *k*, to each of the diagonal elements of the matrix $X'X$. The resulting estimator is called the ridge estimator, suggested by Hoerl and Kennard (1970) and given by (3).

Several procedure for constructing ridge estimators have been proposed in the literature. These procedures were aiming at a rule (or algorithm) for selecting the constant *k* in equation (3). In fact, the best method of estimating *k* is an unsolved problem and there is no constant value of *k* that is certain to yield an estimator which is uniformly better (in terms of *MSE*) than the *OLS* in all cases.

By means of Monte Carlo simulations two suggested ridge parameters were evaluated and the result were compared with ridge parameters evaluated by Hoerl et. al (1975) and Khalaf and Shukur (2005). The estimator *HKB* performed well in this study. It appears to outperform *KS* when ρ is small and the sample size is greater than 25. The suggested estimators K_1 and K_2 performs well in our simulation. They appeared to offer an opportunity for large reduction in *MSE* when *p = 2* and the error term in normally distributed. For non-normal error term the versions of the ridge parameter has a lower estimated *MSE* when the sample size is greater than 25. K_2 is always minimizes the estimated *MSE* when the error term in normally distributed.

References

Alkhamisi, M., & Shukur, G. (2008). Developing Ridge Parameters for SUR Model. *Communication in Statistics- Theory and Methods*, *37*, 544-564.

Hoerl, A. E., & Kennard, R. W. (1970). Ridge Regression: Biased Estimation for non-orthogonal Problems. *Technometrics, Vol. 12,* 55 – 67.

Hoerl, A. E., Kennard, R. W., & Baldwin, K. F. (1975). Ridge Regression: some Simulation. *Communications in Statistics- Theory and Methods, 4,* 105 – 124.

Khalaf, G. (2011). Suggested Ridge Regression Estimators Under Multicollinearity. *Journal of Natural and Applied Sciences, University of Aden, Vol. 15,* 170 – 193.

Khalaf, G., & Shukur, G. (2005). Choosing Ridge Parameters for Regression Problems. *Communication in Statistics – Theory and Methods, 34,* 1177 – 1182.

Kibria, B. M. G. (2003). Performance of some New Ridge Regression Estimators. *Communication in Statistics- Theory and Methods, 32*, 419-435.

Muniz, G., & Kibria, B. M. G. (2009). On some Ridge Regression Estimators: An Empirical Comparisons. *Communications in Statistics-Simulation and Computation*, *38*, 621 – 630.