CHAPTER 5

SIMULATION - APPLICATION

Several authors have conducted simulations in order to verify that ridge estimators are better than least squares in certain cases. The interested reader is referred to Gibbons (1981), Gunst and Mason (1977), Hoerl, Kennard, and Baldwin (1975), McDonald and Galarneau (1975), Wichern and Churchill (1978), Dempster et al. (1977), and Lawless and Wang (1976). In this section we will present a simulation based on the pattern of Wichern and Churchill. Our aim is to compare the relative performance of estimators (ridge and the LS estimator) with respect to their MSE so as to identify cases where ridge estimators provide a good alternative to the LS estimator.

5.1 Description of the Simulation

We use the five parameter model $\mathbf{Y} = \mathbf{X\beta} + \mathbf{U}$,

where **X** is a 30×5 matrix of explanatory variables, **Y** is a 30×1 response vector, $\boldsymbol{\beta}$ is a $(5+1) \times 1$ vector of parameters and **U** is a 30×1 vector of errors.

STEP 1: Thirty observations are generated for each explanatory variable. The explanatory variables are generated by:

$$X_{ij} = \left(1-\alpha^2\right)^{1/2} Z_{ij} + \alpha Z_{i6} \qquad i = 1,2,\ldots,30 \qquad j = 1,2,3$$

$$X_{ij} = \left(1-\alpha_*^{\,2}\right)^{1/2} Z_{ij} + \alpha_* Z_{i6} \qquad i = 1,2,\ldots,30 \qquad j = 4,5$$

where $Z_{i1}, Z_{i2}, \ldots, Z_{i6}$ are independent standard normal numbers and α^2, $\alpha_*^{\,2}$ are coefficients leading to the following correlation matrix:

Correlation matrix					
	X_1	X_2	X_3	X_4	X_5
X_1	1	α^2	α^2	$\alpha\alpha_*$	$\alpha\alpha_*$
X_2	α^2	1	α^2	$\alpha\alpha_*$	$\alpha\alpha_*$
X_3	α^2	α^2	1	$\alpha\alpha_*$	$\alpha\alpha_*$
X_4	$\alpha\alpha_*$	$\alpha\alpha_*$	$\alpha\alpha_*$	1	${\alpha_*}^2$
X_5	$\alpha\alpha_*$	$\alpha\alpha_*$	$\alpha\alpha_*$	${\alpha_*}^2$	1

Table 5.1: The correlation matrix

The explanatory variables are then standardized so that $\mathbf{X'X}$ is in correlation form. Three different combinations of α and α_* are investigated:

CASE 1: Both α and α_* are equal to 0.99, the condition number λ_1/λ_5 = 581.

CASE 2: α is equal to 0.99, α_* is equal to 0.10 and λ_1/λ_5 = 165.

CASE 3: α is equal to 0.70, α_* is equal to 0.30 and λ_1/λ_5 = 8.

Case 1 is a case of extreme multicollinearity while 2 represents a mixed case. Case 3 represents a moderate situation.

STEP 2: For each design matrix we use two coefficient vectors: $\boldsymbol{\beta}_L$, the normalized eigenvector corresponding to the largest eigenvalue of $\mathbf{X'X}$ and $\boldsymbol{\beta}_S$, the normalized eigenvector corresponding to the smallest eigenvalue of $\mathbf{X'X}$. This choice seems appropriate since these eigenvectors give the maximum (for $\boldsymbol{\beta}_L$) and minimum (for $\boldsymbol{\beta}_S$) MSE considering however certain constraints (McDonald and Galarneau, 1975).

STEP 3: Observations on the dependent variable are determined by:

$$Y_i = \beta_0 + \beta_1 X_{i1} + \ldots + \beta_5 X_{i5} + U_i \qquad i = 1, 2, \ldots, 30$$

where $X_{i1}, X_{i2}, \ldots, X_{i5}$ are the original unstandardized variables, $U_i \sim N(0, \sigma^2)$, and β_0 is zero. $\beta_1, \ldots, \beta_5$ are the appropriate eigenvector values. Four values of σ are investigated:

σ= 0.1, 0.5, 1.0, 5.0 or equivalently four signal-to-noise ratios $\rho = \boldsymbol{\beta}'\boldsymbol{\beta}/\sigma^2$ =100, 4, 1, 0.04. The dependent variable is standardized so that $X'y$ is a vector of correlations.

STEP 4: Additional samples of size 100 are generated; $\mathbf{X}'\mathbf{X}$, $\boldsymbol{\beta}$ remain fixed while e_i and hence the dependent variable change.

The least squares and ridge estimates are determined using the standardized variables and then the estimated coefficients are transformed back to the original model. The *k* values and the standard deviations are computed for the following rules:

	Rules (m)
1.Hoerl Kennard (HK)	1. $k_{HK} = s^2/\max(\hat{a}_i^2)$
2.Hoerl Kennard and Baldwin (HKB)	2. $k_{HKB} = ps^2/\hat{\boldsymbol{\alpha}}'\hat{\boldsymbol{\alpha}} = ps^2/\hat{\boldsymbol{\beta}}'\hat{\boldsymbol{\beta}}$
3.Lawless Wang (LW)	3. $k_{LW} = ps^2 \Big/ \sum_{i=1}^{5} \lambda_i \hat{a}_i^2 = ps^2/\hat{\boldsymbol{\alpha}}'\mathbf{X}'\mathbf{X}\hat{\boldsymbol{\alpha}} = ps^2/\hat{\boldsymbol{\beta}}'\boldsymbol{\Lambda}\hat{\boldsymbol{\beta}}$

Table 5.2: Investigated rules

The performance of the estimators is evaluated in terms of the averaged total squared errors

$$MSE\left(\hat{\beta}_i^m(k)\right) = \frac{1}{100}\sum_{j=1}^{100}\left(\hat{\beta}_{ij}^m(k) - \beta_i\right)^2 \qquad i = 0,1,..,5\ ,\quad m = 1,2,3 \quad \text{and}$$

$$TMSE(m) = \sum_{i=0}^{5} MSE\left(\hat{\beta}_i^m(k)\right) \qquad m = 1,2,3$$

where β_0 zero, $\beta_1,...,\beta_5$ the appropriate eigenvectors, and $\hat{\beta}_{ij}^m(k)$ the estimates in terms of the original model. In order to compare the ridge estimators with the least squares we also compute the ratio

$$R_m = \frac{TMSE(m)}{TMSE(LS)} \qquad m = 1,2,3.$$

5.2 *The Simulation Results*

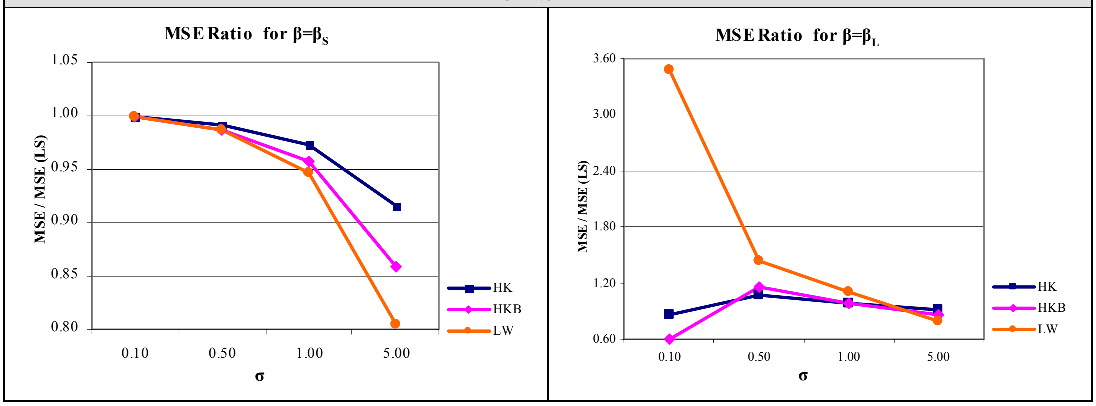
The results of the simulation are presented in the Appendix (Part 2, Table 2). The comments made below are based on these results.

5.2.1 Mean Squared Error (MSE)

The main theoretical justification for the construction of ridge estimators is that they have smaller MSE than the least squares estimator. Therefore in order to measure the improvement one can check the ratio of the estimated MSE for a particular ridge estimator to the estimated MSE for LS. The ratio for LS is obviously 1.

These ratios are plotted in the next figure. Each plot presents the ratio for each of the three estimators (HK, HKB, and LW) as a function of σ. The left (right) graph presents the results for $\boldsymbol{\beta} = \boldsymbol{\beta}_S$, $(\boldsymbol{\beta} = \boldsymbol{\beta}_L)$. Each point plotted represents the average of 100 samples.

CASE 1



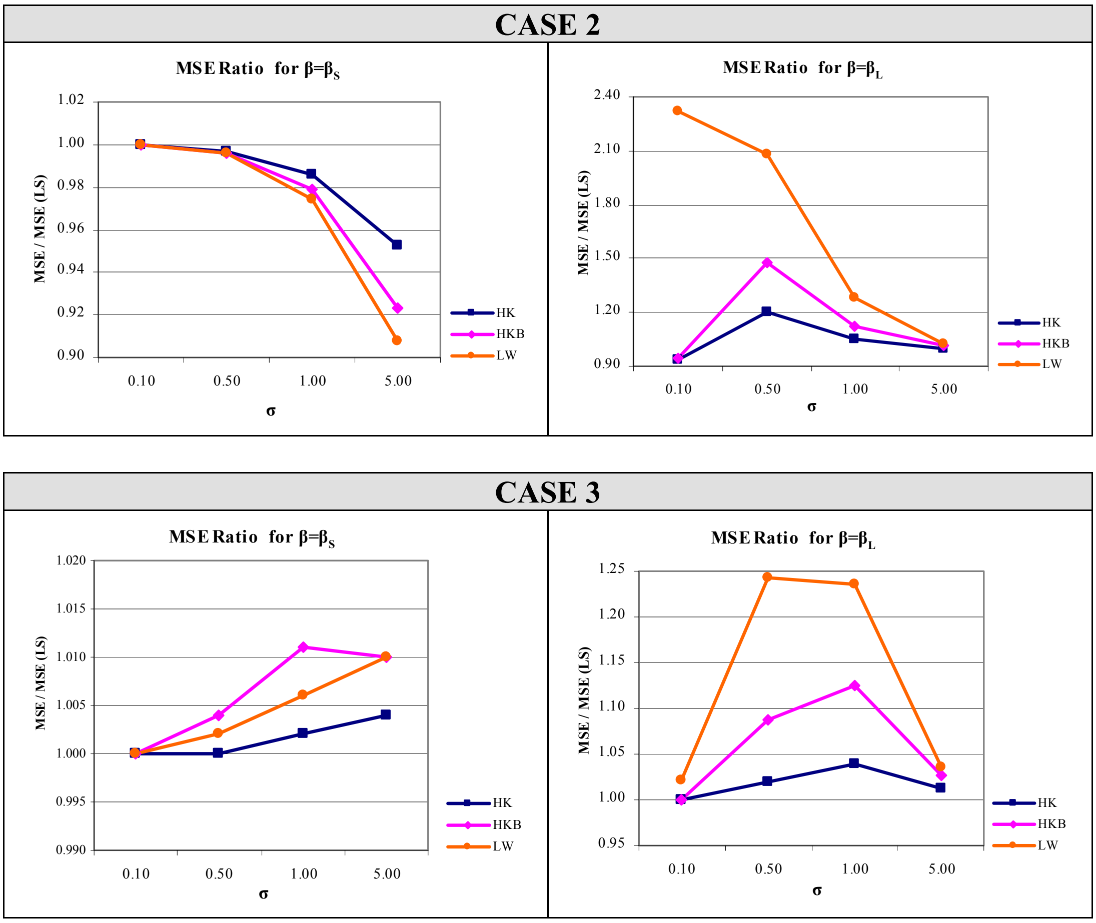


Figure 5.1: MSE Ratio

The graphs above provide the basis for these observations:

- $\boldsymbol{\beta} = \boldsymbol{\beta}_S$

All estimators are at least as good as the LS estimator for the first two cases, that is the MSE for those estimators is smaller than 1. For case 3 (where the correlation is smaller) the estimators are slightly worse than LS- the ratio ranges between 1 and 1.011.

The ratio decreases for larger values of σ again for the first two cases. In addition, in case 3 the estimators have almost the same MSE.

- $\boldsymbol{\beta} = \boldsymbol{\beta}_L$

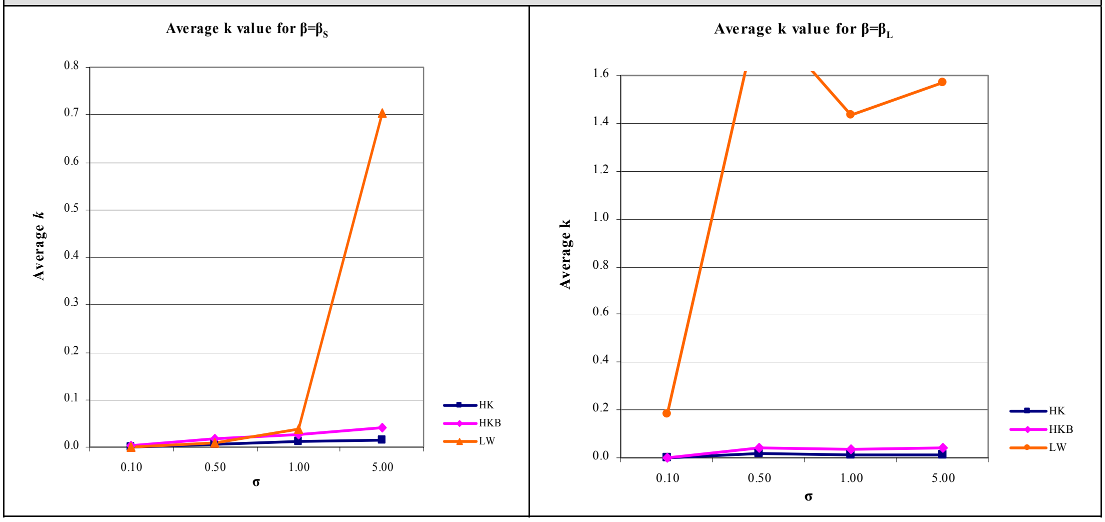
None of the estimators is constantly better than the LS estimator. In all 3 cases the ratio increases and then decreases for larger values of σ, namely it does not appear to be a monotone function of σ. Moreover, HK and HKB rules appear to perform better than the rule of LW.

As it was expected the ridge estimators perform better for higher degree of multicollinearity, that is in case 1.

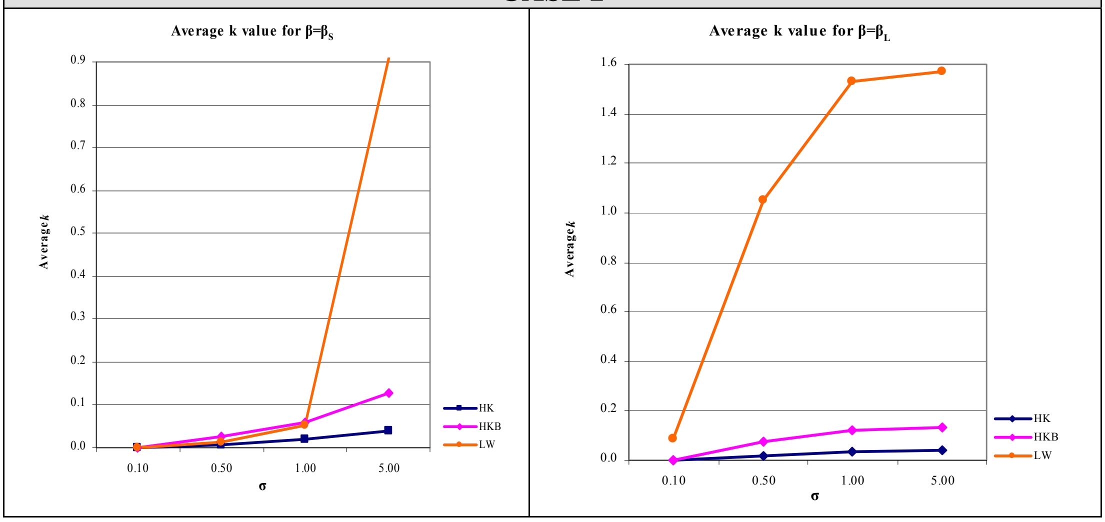
5.2.2 Average *k*

The average k and standard deviation of the k values observed in 100 samples are recorded for the three rules described above (HK-HKB-LW). The next figure presents the results for the three cases.

CASE 1



CASE 2



CASE 3

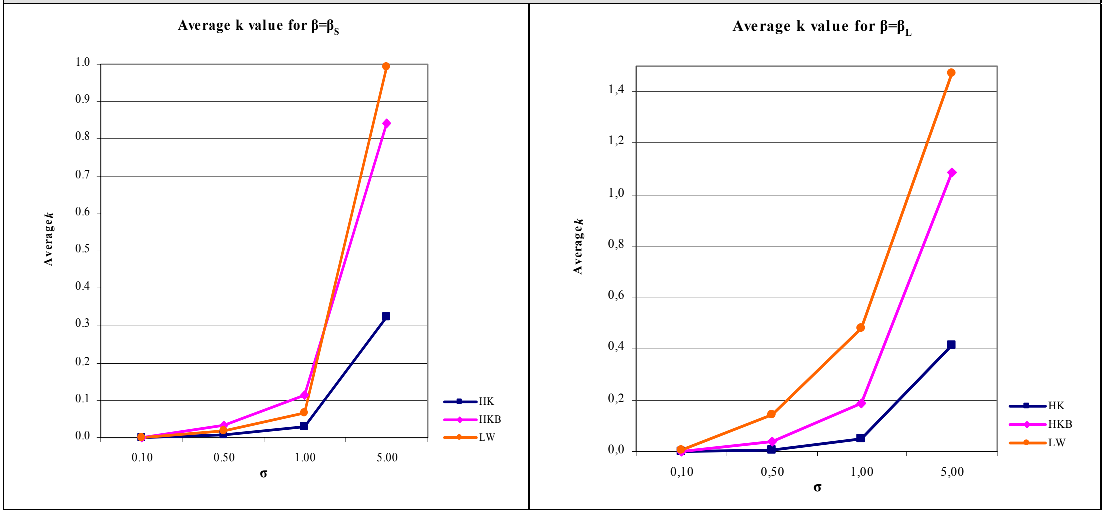


Figure 5.2: Average *k*

Note that the values for the ISRM estimator are 0.3, 0.15, and 0.7 for each of the three cases and are not presented in the figure. This rule tends to give much larger values than the other rules. The following observations can be made for each case:

Case 1: For this extreme case of multicollinearity we notice that HK and HKB rule give values for k between 0.0004 to 0.044 irrespective of the value of σ and $\boldsymbol{\beta}$. The LW k increases rapidly to 0.7 for $\boldsymbol{\beta} = \boldsymbol{\beta}_S$ and σ >1. For $\boldsymbol{\beta} = \boldsymbol{\beta}_L$ the same rule gives values of *k* larger to one except for σ =0.1 (k = 0.18).

Case 2: As in case 1 HK and HKB rule give small values for *k*, specifically, between 0.0003 and 0.1332. The values of *k* for LW rule exceed 1 for $\boldsymbol{\beta} = \boldsymbol{\beta}_L$ and σ >0.1

Case 3: All rules give larger values for *k* than for cases 1 and 2. These values increase as σ increases and for σ >1 become close or larger to 1.

In general:

- The average *k* is smaller for all estimators when $\boldsymbol{\beta} = \boldsymbol{\beta}_S$.
- Tke *k* values associated with estimators LW are not restricted to values less than one. This estimator assumes relatively large average *k* values as σ approaches one.

5.2.3 Conclusions

In our simulation we used three methods for determining the value of k, methods which are often met in other simulation studies and are easy to handle. The comparison of the three estimators for estimating ridge parameter k to the least squares estimators was made using the MSE criterion. As a general remark we could say that ridge regression estimators yielded similar or slightly better results compared to the least squares estimators.

Specifically, the performance of the ridge estimators depends on the variance of the random error, the correlations among the explanatory variables and the choice ($\boldsymbol{\beta}_L$ *or* $\boldsymbol{\beta}_S$) of the unknown coefficient vector. In this simulation one can observe the performance of the estimators when one of these factors is changed while the remaining two are fixed. As one could expect, no ridge estimator is shown to be better than the LS in all cases. While in some cases the HK or the HKB rule achieves a reduction in MSE in other cases the MSE increases. However, in case of high multicollinearity the HK rule appears to be a good alternative to the classical LS estimator.

Overall, in this study ridge regression reduces the mean squared error of the estimated coefficients under conditions of multicollinearity, low signal to noise ratios and as long as $\boldsymbol{\beta}$ is equal to the eigenvector of the smallest eigenvalue. Yet, when $\boldsymbol{\beta}$ is equal to the eigenvector of the largest eigenvalue, ridge in general performs poorly. However, our conclusions must be viewed with reservation since the size of the regression problem presented was fixed (a five parameter model with n=30) and the number of replication samples taken 100. Moreover, the size of the ratio of the number of predictors to the sample (n) is relatively small (5/30). It has been found that when the ratio is too small, then no difference between the least squares and ridge regression really exist. However, when the ratio is large, i.e., many predictors with a small sample, ridge regression has been demonstrated to be more accurate than least squares. (Dempster et al., 1977). In order to check this conclusion we also simulated data for a 5 parameter model with n=15 and thus a large ratio (1/3). For high multicollinearity of regressors (case 1 in section 5.1) the MSE of all three ridge estimators were smaller than least squares (for both $\boldsymbol{\beta}_L$ and

$\boldsymbol{\beta}_S$), especially for low signal to noise ratios. So it would be wiser to use ridge regression in respective cases.

In practice, careful investigation is needed when a researcher considers a particular regression problem so as to decide whether ridge regression is the appropriate alternative to least squares and how to choose the best k value.