〈Research Paper〉

Journal of the Korean
Statistical Society
Vol. 16, No. 2, 1987

A New Deletion Criterion of Principal Components Regression with Orientations of the Parameters

Wonwoo Lee*

ABSTRACT

The principal components regression is one of the substitutes for least squares method when there exists multicollinearity in the multiple linear regression model. It is observed graphically that the performance of the principal components regression is strongly dependent upon the values of the parameters. Accordingly, a new deletion criterion which determines proper principal components to be deleted from the analysis is developed and its usefulness is checked by simulations.

1. Introduction

The multiple linear regression (*MLR*) model is one of the most popular methods for exploring linear relationships between a response variable and a set of regressor variables. Under the assumptions of identically, independently distributed random errors, the least squares (*LS*) estimation method provides the well-known best linear unbiased estimator. Unfortunately, in many practical situations, the regressor variables are not orthogonal and hence, in such cases, the statistical properties of the *LS* estimator can be adversely affected.

Confronted with multicollinearity in the model, some alternative estimation techniques to the *LS* method have been developed to remedy the problem resulting from severe multicollinearity. Among the alternatives the principal components (*PC*) estimation method is one of prevalent methods, which leads to improve over the *LS* estimatorin

* Dept. of Applied Statistics, King Sejong Univ., Seoul, Korea

terms of mean square error (MSE).

In Section 2 the MLR model is explained in terms of the eigenvalue decomposition and the singular value decomposition, followed by a review of principal components analysis along with Massy's deletion criterion. And the new criterion, based upon the orientation of the unknown parameter vector and the value of σ^2, will be introduced in Section 3. Section 4 is devoted to the problems related to principal components regression with respect to the orientations of the parameters by graphs. Finally, the criterion is evaluated by simulation studies in Section 5 and the conclusion is in Section 6.

2. Principal Components Analysis

2.1 Linear Regression Model

The general mutiple linear regression model is a statistical model which attempts to explain the response variable, y, by a linear combination of p explantory variables, $X_1, \cdots, X_p$, that is

$$y = X\beta + \varepsilon, \tag{1}$$

where y is an $n\times 1$ vector of observed responses, X is an $n\times p$ full (column) rank matrix of nonstochastic regressor variables, β is a $p\times 1$ vector of parameters, and ε is an $n\times 1$ random vector whose mean is 0 and variance matrix is $\sigma^2 I_{n\times n}$.

For convenience, the columns of X are assumed to be standadized so that $x_j'1=0$ and $x_j'x_j=1$, $j=1,\cdots,p$.

The MLR model, as in equation (1), can be equivalently expressed by the eigenvalue decompositon or the singular value decomposition which is often useful in the regression context.

The eigenvalue decomposition results in

$$y = Z\alpha + \varepsilon, \tag{2}$$

where $Z=XV=(z_1,\cdots,z_p)$, $\alpha=V'\beta$ and V is the matrix of the eigenvectors of $X'X$, that is, $V=(v_1,\cdots,v_p)$. Note that the column vectors z_j are called the principal components.

The singular value decomposition theorem allows X to be decomposed as

$$X = ULV',$$

where U is the matrix of the eigenvectors associated with the positive eigenvalues of XX' and L is the diagonal matrix of the so called singular values of X. Thus, another reparameterized *MLR* model can be obtained as follows;

$$y = U\gamma + \varepsilon, \tag{3}$$

where $\gamma = LV'\beta$.

It is worthwhile to observe several facts; (a) $Z'Z = \Lambda$, where $\Lambda = \text{diag}\ (\lambda_1, \lambda_2, \cdots, \lambda_p)$ and λ_j's are the eigenvalues of $X'X$ (b) $L = \Lambda^{1/2}$ and (c) $\gamma = \Lambda^{1/2}\alpha$. Without loss of generality, furthermore, the eigenvalues, $\lambda_1, \cdots, \lambda_p$ are assumed to be ordered in magnitude, that is, $\lambda_1 \geq \cdots \geq \lambda_p$.

Multicollinearity refers to the near linear dependencies that may exist among the regressors. Severe multicollinearity is said to exist if there are nonzero constant c_j's such that

$$\sum_{j=1}^{p} c_j x_j \approx 0,$$

where x_j is the jth column vector of X. Multicollinearity can have adverse effects on the *LS* solution and related procedures [Belsley, Kuh, and Welch(1980), Myers(1986), etc.].

When multicollinearity is detected from the *MLR* model, an alternative to the *LS* method is the principal components regression method which provides a biased estimator of β.

2.2 Principal Components Estimator

The principal components estimator of β can be obtained from the reparameterized model (2), $y = Z\alpha + \varepsilon$. Note that the *PC*s are ordered so that z_j is the jth *PC* associated with the jth largest eigenvalue of $X'X$.

The *PC* estimation procedure simply amounts to deleting some *PC*s from the model (2) and, then, applying the *LS* method to the restricted form as

$$y = Z_1\alpha_1 + Z_2\alpha_2 + \varepsilon,$$

where $Z_1 = (z_1, \cdots, z_r)$, $Z_2 = (z_{r+1}, \cdots, z_p)$, $\alpha_1 = (\alpha_1, \cdots, \alpha_r)'$, and $\alpha_2 = (\alpha_{r+1}, \cdots, \alpha_p)'$. Assuming that the columns of Z_2 are the deleted *PC*s, the restricted model is

$$y = Z_1\alpha_1 + \varepsilon$$

and the *LS* estimator of α_1 is $a_{1,LS} = (Z_1'Z_1)^{-1}Z_1'y$. Thus, the *PC* estimator of β can be given as

$$\begin{aligned} b_{PC} &= V_1 a_{1,LS} \\ &= V_1(Z_1'Z_1)^{-1}Z_1'y, \end{aligned}$$

where $V_1 = (v_1, \cdots, v_r)$.

However, it is important how to choose the principal components to be deleted. Massy(1965) suggested two criteria for deleting PCs: the eigenvalue criterion and the t-value criterion.

The eigenvalue criterion, which is usually used in PC analysis, leads the principal components corresponding to small eigenvalues to be deleted. The major issue is to determine which eigenvalues are small. In general, it is difficult to choose appropriate PCs to be retained in the analysis unless $\lambda_r \gg \lambda_{r+1}$ and $\lambda_{r+1} \approx 0$. While, the t-value criterion is based upon the t-test for each coefficient.

The two criteria, the eigenvalue criterion and the t-value criterion, would yield nearly identical results if the following condition is met:

$$R_{yz_1}{}^2 \geq R_{yz_2}{}^2 \geq \cdots \geq R_{yz_p}{}^2, \tag{4}$$

where $R_{yz_j}{}^2$ denotes the squared sample correlation coefficient between y and z_j. Note that $R_{yz_j}{}^2 = \lambda_j a_{j,LS}{}^2$, $j=1,\cdots,p$ and $\lambda_1 \geq \cdots \geq \lambda_p$. In such a case, the t-test statistics have the the same order as the eigenvalues so that the two criteria coincide when an appropriate significace level is chosen. However, the condition (4) may not, in general, hold since the order of $R_{yz_j}{}^2$ depends mainly upon the order of the unknown parameter γ_j, $j=1,\cdots,p$. Note also that,

$$E(\lambda_j a_{j,LS}{}^2) = \gamma_j{}^2 + \sigma^2, \quad j=1,\cdots,p. \tag{5}$$

It is important to emphasize that when the condition (4) is not satisfied, neither of the two criteria should be used as a deletion criterion. Indeed, a new deletion criterion that accounts for the orientation of γ can be developed.

3. New Deletion Criterion

Since the PC estimator resulting from applying the LS method to the reduced set of the PCs is a biased estimator it is natural to consider the MSE of b_{PC}. Furthermore, the necessary and sufficient condition for which the MSE of the PC estimator is smaller than that of the LS estimator can be developed and used to define a new deletion criterion.

3.1 *MSE:* A Performance Measure for Biased Estimator

The MSE is one of the most useful performance measures for biased estimators. The expression of the MSE of the PC estimator b_{PC}, can be decomposed into components as follows:

$$MSE(b_{PC})=\sum_{j=1}^{p} MSE(b_{j,PC})$$
$$=E(b_{PC}-Eb_{PC})'(b_{PC}-Eb_{PC})$$
$$+(Eb_{PC}-\beta)'(Eb_{PC}-\beta) \quad (6)$$

The first and second terms in (6) will be regarded as the 'variance' and the 'squared-bias' components, respectively. It is well known that using a biased estimator, a large reduction in the 'variance' component is expected while some increase in the 'squared-bias' component is accepted.

It is often of interest to consider the matrix form of MSE, denoted $MtxMSE(b_{PC})$,

$$MtxMSE(b_{PC})=E(b_{PC}-\beta)(b_{PC}-\beta)'$$
$$=\mathrm{Var}(b_{PC})+\mathrm{Bias}(b_{PC})\mathrm{Bias}(b_{PC})', \quad (7)$$

where $\mathrm{Bias}(b_{PC})=Eb_{PC}-\beta$.

In order to compare the performances of b_{LS} and b_{PC} in terms of MSE, either form in (6) or (7) can be used. That is, either $MSE(b_{LS})-MSE(b_{PC})$ or $MtxMSE(b_{LS})-MtxMSE(b_{PC})$ can be used since the former is the nonnegative if and only if the latter is positive semi-definite [see Theobald(1974)].

3.2 Unbiased Optimal Deletion Criterion

In order for the MSE of b_{PC} to be smaller than that of b_{LS}, the difference, $MtxMSE(b_{LS})-MtxMSE(b_{PC})$ must be positive semi-definite. Since the matrix form of the MSE of b_{PC} is

$$MtxMSE(b_{PC})=\sigma^2 V_1\Lambda_1^{-1}V_1'+V_2\alpha_2\alpha_2'V_2',$$

where $\Lambda_1=\mathrm{diag}(\lambda_1,\cdots,\lambda_r)$ and $\Lambda_2=\mathrm{diag}(\lambda_{r+1},\cdots,\lambda_p)$, the difference of the two is

$$MtxMSE(b_{LS})-MtxMSE(b_{PC})$$
$$=\sigma^2V_2(\Lambda_2^{-1}-\sigma^{-2}\alpha_2\alpha_2')V_2'. \quad (8)$$

Thus, it suffices to show that $(\Lambda_2^{-1}-\sigma^{-2}\alpha_2\alpha_2')$ is positive semi-definite which is, equivalently,

$$d'(\Lambda_2^{-1}-\sigma^{-2}\alpha_2\alpha_2')d\geq 0 \quad (9)$$

for any nonzero $(p-r)\times 1$ vector d.

The relationship (9) can be rewritten as

$$\frac{d'\alpha_2\alpha_2'd}{d'\Lambda_2^{-1}d}\leq\sigma^2$$

so that, based upon the Cauchy-Schwarz inequality, the relationship is equivalent to

$$\sup_d \frac{d'\alpha_2\alpha_2'd}{d\Lambda_2^{-1}d} \leq \sigma^2.$$

Thus, the necessary and sufficient condition is obtained as

$$\alpha_2'\Lambda_2\alpha_2 \leq \sigma^2 \quad (10)$$

or, equivalently,

$$\sum_{j=r+1}^{p} \lambda_j \alpha_j^2 \leq \sigma^2 \quad (11)$$

The condition (11) depends upon the number of the deleted PCs, the length of α_2, and the magnitude of σ^2.

Note that this condition does not imply to delete the principal component associated with the smallest eigenvalue. In fact, it can be reexpressed by using the singular value decomposition as

$$\sum_{j\in D} \gamma_j^2 \leq \sigma^2, \quad (12)$$

where D is the set of the indexes whose principal components are supposed to be deleted. The set D includes the maximum number of the indexes for which γ_j^2 are small in magnitude and satisfied with the condition (12).

Since the unbiased estimators for the unknown parameters are, from (5),

$$\widehat{\gamma_j^2} = \lambda_j a_{j,LS}^2 - s^2, \quad (13)$$

the substitution of the (13) into (12) results in

$$\sum_{j\in D} (\lambda_j a_{j,LS}^2 - s^2) \leq s^2. \quad (14)$$

Thus, the inequality (14) can be used as a deletion criterion named the unbiased optimal deletion criterion. That is, based on this criterion, delete those principal components, z_j, $j\in D$ for which (14) is satisfied. The usefulness of this unbiased optimal deletion criterion will be seen by simulation studies in section 5.

Since the necessary and sufficient condition (12) is dependent on the magnitude of the γ_j^2 for $j\in D$ and σ^2, the orientation of the unknown parameter vector and the size of σ^2 are closely related to the evaluation of principal components analysis. In the following section, thus, the situation for which the regular PC estimator resulting form using either the eigenvalue criterion or the t-value criterion does not improve over the LS estimator will be introduced by considering the orientations of the unknown parameter vector.

4. Orientation of the Parameters

It has been mentioned that the order of the unknown parameter λ_j is important in determining which component should be deleted in principal components regression. Since the magnitude of γ_j is determined by α_j and $\sqrt{\lambda_j}$, the magnitudes of α_j are sometimes crucial to achieve a proper *PC* estimator. For the simple case, $p=2$, the problem related to the *PC* estimation is delineated in this section.

Consider the orientation of the unknown parameter vector in terms of α or β. Let the squared length of α be fixed as C

$$\alpha'\alpha=\beta'\beta=C \tag{15}$$

and, for illustrative purposes, consider the special case when $p=2$. Then the matrix of the correlation form of the two regressors is

$$X'X=\begin{bmatrix} 1 & r_{12} \\ r_{21} & 1 \end{bmatrix},$$

where r_{12} is the inner product of the x_1 and x_2. Without loss of generality, r_{12} is assumed to be positive. The eigenvalues of $X'X$ are $1+r_{12}$ and $1-r_{12}$ with the matrix of the eigenvectors, V, where

$$V=\begin{bmatrix} 1/\sqrt{2} & 1/\sqrt{2} \\ 1/\sqrt{2} & -1/\sqrt{2} \end{bmatrix}.$$

Assuming also that r_{12} is close to 1, there exists severe multicollinearity in this simple *MLR* model.

The orientations of the unknown parameter vectors in terms of α and β, which satisfy (15), are described as the circle in Figure 1. The specific α depicted in the figure may cause the regular *PC* estimator, *PC*(2nd), to be inappropriate in that $MSE[PC(2nd)]>MSE[PC(1st)]$ since the absolute value of α_2 is relatively large compared to that of α_1. Note that *PC*(1st) and *PC*(2nd) refer to the principal components estimators resulted from deleting the first and the second principal component, repectively. However, in Figure 1, the exact range of the orientations for which the regular *PC* estimator is improper, cannot be found.

In order to investigate the performances of the *PC* estimator, the corresponding orientation in terms of γ can be obtained by applying the singular value decomposition. Since $\gamma=\Lambda^{1/2}\alpha$, the orientation of γ corresponding to α lies on the ellipse $\gamma'\Lambda^{-1}\gamma=C$ in Figure

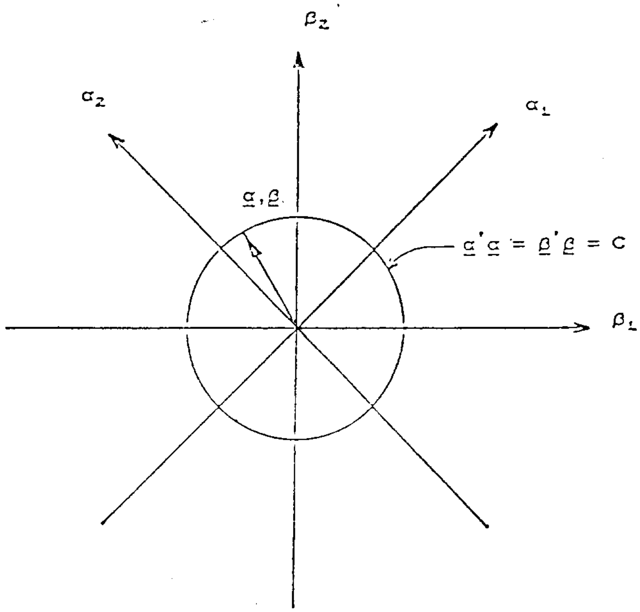

Figure 1. Specific Orientations, α and β

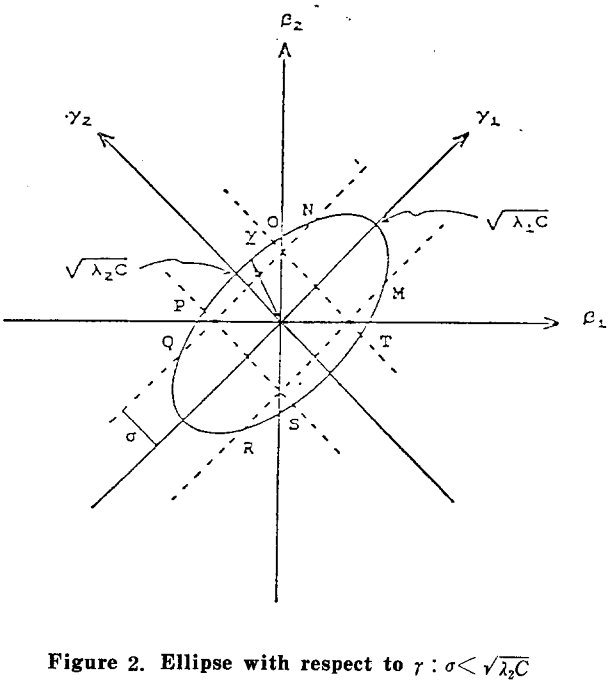

Figure 2. Ellipse with respect to γ : $\sigma < \sqrt{\lambda_2 C}$

2. Note that the ellipse is associated with the circle in Figure 1 and the dotted lines represent $\gamma_j=\pm\sigma$, $j=1,2$.

The dotted lines in the figure are based upon the assumption that σ is fixed such that

$$\sigma<\sqrt{\lambda_2 C}, \tag{16}$$

where $\pm\sqrt{\lambda_2 C}$ are the points of the ellipse at $\gamma_1=0$. Then, by the condition (12), the orientations in terms of γ can be catagorized into three regions summarized in Table 1. First, when

$$\gamma_2^2<\sigma^2<\gamma_1^2 \tag{a}$$

which is described by the chord segments MN and QR in Figure 2, the deletion of z_2 improves estimation of α in that $MSE[a_{PC(2nd)}]<MSE[a_{LS}]$. In other words, from(12), the usual PC estimator obtained through the deletion of the component associated with the smallest eigenvalue will outperform the LS estimator in terms of MSE. Note that $\gamma_j^2=\lambda_j\alpha_j^2$, $j=1,2$.

Table 1. Condition (12) **with respect to** γ

Condition	Region	Proper Decision	Result
(a) $\gamma_2<\sigma<\gamma_1$	MN and QR	delete z_2	$MSE[PC(2nd)]<MSE(LS)$
(b) $\gamma_1<\sigma<\gamma_2$	OP and ST	delete z_1	$MSE[PC(1st)]<MSE(LS)$
(c) $\gamma_1>\sigma$ and $\gamma_2>\sigma$	NO, PQ, RS and TM	none	smaller $MSE(LS)$

Secondly, if

$$\gamma_1^2<\sigma^2<\gamma_2^2 \tag{b}$$

then the PC(2nd) estimator cannot have MSE as small as the LS estimator. The condition(12) requires the deletion of the first PC even though $\lambda_1>\lambda_2$. Thus any orientation of γ along the chord segments OP and ST leads to the improvement of the PC(1st) estimator over the LS estimator.

Finally, at the other chord segments (NO, PQ, RS, and TM), the LS estimator has smaller MSE than any PC estimator since

$$\gamma_1^2>\sigma^2 \text{ and } \gamma_2^2>\sigma^2 \tag{c}$$

so that deletion of any principal component cannot guarantee smaller MSE. The result can be expected by the fact that the effect of multicollinearity may be eliminated by a relatively small σ^2. Note that as σ^2 is decreased, the range of the orientations where PC estimation is preferred over LS estimation is also decreased.

In fact, only when the true parameter vector, in terms of γ, lies on the limited chord segment, MN and QR, can the regular PC estimator, PC(2nd), be used to combat the multicollinearity problem. Therefore, it is essential to examine the orientation of the parameter vector along the ellipse.

Furthermore, the magnitude of σ^2 also affects the performance of the PC estimator. As long as σ is in the interval,

$$\sqrt{\lambda_2 C} < \sigma < \sqrt{\lambda_1 C} \tag{17}$$

there is no need to suspect the capability of the PC(2nd) estimation since any orientation γ satisfies the condition $\gamma_2^2 < \sigma^2$ (see Figure 3). In addition, if $\gamma_1^2 + \gamma_2^2 < \sigma^2$ then the trivial result that the PC estimator is 0 from (12) can be obtained. This case is depicted, in Figure 3, by the chords, KL and MN. Therefore, for only the orientations along LM and NK, the PC(2nd) is recommended and for the other orientations along the ellipse the PC(2) should be used, where the PC(2) refers to the PC estimator resulted from deleting the last two principal components. It goes without saying that if

$$\sqrt{\lambda_1 C} < \sigma$$

then the PC estimator should always be 0.

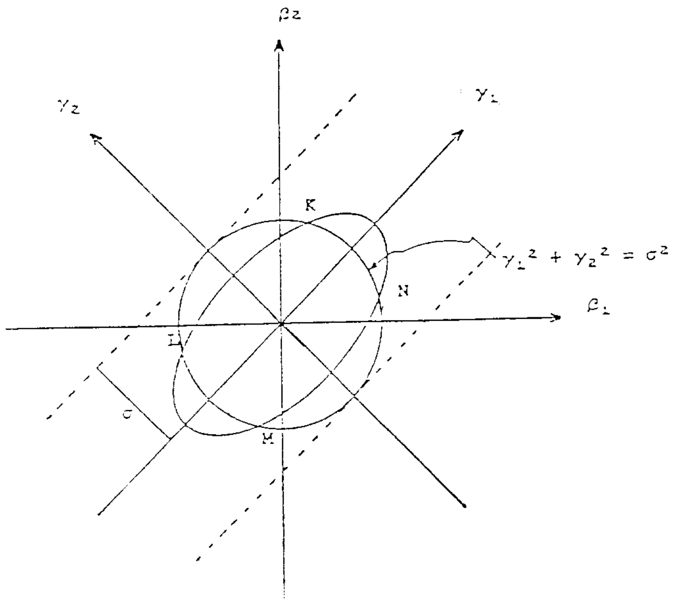

Figure 3. Ellipse with respect to γ : $\sqrt{\lambda_2 C} < \sigma < \sqrt{\lambda_1 C}$

Thus, the performances of the PC estimators depend on not only the orientation of the unknown parameter vector but also the magnitude of σ^2.

In the next section, the effects of the orientation of the parameter vector and value of σ^2 on the performance of the PC estimators are examined and the usefulness of the unbiased optimal deletion criterion will be seen by simulation studies in terms of the empirical mean square error.

5. Simulation Studies

It has been observed that different orientations of the parameter vector and different values of σ^2 affect the performance of the PC estimation. This fact, demonstrated in section 4, will be thoroughly illustrated in this section by using simulated data for $p=4$. In the following simulation results, two experimental factors will be controlled: (1) orientation of β^* will change which $\beta^{*\prime}\beta^*$ will be held constant and (2) the value of σ will vary. Note that β^* is the unknown parameter vector for the standardized linear model.

5.1 Generated Data Sets: $p=4$

The explanatory variables X_1, X_2, X_3, and X_4 are established in the following way. Twenty values for X_1 and X_3 were independently preselected. And, by using these preselected values, X_2 and X_4 are generated as follows:

$$X_{i2}=4.5+6.1X_{i1}+\varepsilon_{i2},\quad i=1,\cdots,20,$$

$$X_{i4}=.75-3.75X_{i2}+5.25X_{i3}+\varepsilon_{i4},\quad i=1,\cdots,20,$$

where ε_{i2} and ε_{i4} are independent normal $(0,\ \sigma_2{}^2)$ and $(0,\ \sigma_4{}^2)$ random numbers, respectively. Note that the values of Xs are generated in terms of the original units.

The values of σ_2 and σ_4 are chosen as 2.1 and 1.3 so that the resulting X matrix along with the standardized X matrix, X^*, is given in Table 2. The eigenvalues and the condition indexes of $X^{*\prime}X^*$ is supplied in Table 3.

Instead of choosing the orientations of the unknown parameters in terms of β, it will be convenient to choose the orientations in terms of γ, where $\gamma=\Lambda^{1/2}\alpha=\Lambda^{1/2}V'\beta^*$. Thus two different orientations of the parameter vectors in terms of γ, α, β^*, and β are given in Table 4. Note that the magnitudes of γ_1 and γ_2 are so large that the first two principal components are regarded as important regressors. The first orientation of γ, denoted by (A), has a relatively large value of γ_3 and a relatively small value of γ_4,

Table 2. X and X^* Matrices

X				X^*			
10.00000	63.52738	25.70000	−104.00751	−0.29003	−0.29396	−0.09645	0.27104
12.50000	80.15732	34.10000	−119.08264	−0.09564	−0.09572	0.34031	0.22098
14.00000	89.17675	29.50000	−179.45786	0.02099	0.01179	0.10113	0.02050
16.50000	103.54194	33.20000	−212.33966	0.21538	0.18303	0.29351	−0.08869
18.00000	115.85583	31.10000	−271.52204	0.33201	0.32982	0.18432	−0.28522
12.40000	78.74884	20.50000	−189.51461	−0.10341	−0.11251	−0.36683	−0.01290
17.20000	111.64889	23.00000	−296.90512	0.26981	0.27967	−0.23684	−0.36951
15.90000	105.32143	25.50000	−260.82443	0.16873	0.20425	−0.10685	−0.24973
17.70000	117.53938	28.00000	−293.10904	0.30869	0.34989	0.02314	−0.35690
16.60000	106.79316	30.50000	−239.88858	0.22316	0.22179	0.15313	−0.18018
11.00000	72.85071	34.70000	− 90.76078	−0.21227	−0.18282	0.37151	0.31503
12.50000	85.03747	28.80000	−165.89151	−0.01788	−0.03755	0.06473	0.06555
15.00000	101.04635	29.50000	−223.86227	0.09875	0.15328	0.10113	−0.12696
7.50000	49.31779	19.90000	− 81.02018	−0.48442	−0.46334	−0.39802	0.34738
10.30000	64.45971	24.60000	−111.72797	−0.26670	−0.28284	−0.15365	0.24541
13.50000	83.79756	21.50000	−199.77653	−0.01788	−0.05233	−0.31483	−0.04698
16.20000	99.29425	24.00000	−245.03484	0.19206	0.13240	−0.18484	−0.19727
9.80000	64.08022	26.50000	−104.26253	−0.30558	−0.28737	−0.05486	0.27020
13.30000	86.78339	29.00000	−171.19733	−0.03343	−0.01674	0.07513	0.04793
13.70000	84.76920	31.50000	−152.40512	−0.00233	−0.04075	0.20512	0.11033

Table 3. The Eigenvalues and Condition Indexes

Eigenvalues	Condition Indexes
2.96075	1.0000
1.02801	1.6971
.01112	16.3173
.00012	157.0762

while the elements of the second orientation (B) are ordered as $|\gamma_1|>\gamma_2|>|\gamma_4|>|\gamma_3|$. Thus, the order of $|\gamma_3|$ and $|\gamma_4|$ may cause problems for principal components regression.

The observations of the response variable, for each case, are determined by

$$y_i=34.8+\beta_1X_{i1}+\beta_2X_{i2}+\beta_3X_{i3}+\beta_4X_{i4}+\varepsilon_i,\quad i=1,\cdots,20, \tag{18}$$

where ε_i are independent normal $(0,\sigma^2)$ random numbers. The values of σ are chosen as 5 and 10 in order to vary the signal-to-noise ratio. Note that the generating equation of y, (18), is equivalent to

Table 4. Two Orientations of Parameter Vectors

	(A)	(B)
γ	662.656	−88.5118
	93.2293	−88.0519
	35.1299	−2.97339
	−2.11488	5.94551
α	385.112	−51.4399
	91.9509	− 86.844
	333.203	−28.2022
	−192.875	542.225
β^*	−43.0837	−24.6925
	209.724	351.367
	56.5434	−49.7242
	−505.918	422.433
β	34.8	34.8
	−3.34998	−1.91997
	2.5	4.18844
	2.93999	−2.58542
	−1.68	1.40277

$$y_i=(34.8+\beta_1\bar{X}_1+\cdots+\beta_4\bar{X}_4)+\beta_1{}^*X_{i1}{}^*\cdots+\beta_4{}^*X_{i4}{}^*+\varepsilon_i,\quad i=1,\cdots,20,$$

since $\beta'=(\beta_0,\ X_{SD}{}^{-1}\beta^*)$, where $\beta_0=34.8$ and $X_{SD}=\mathrm{diag}(S_{X_1},\cdots,S_{X_p})$, $S_{X_j}=[\sum_i(X_{ij}-\bar{X}_j)^2]^{1/2}$.

Throughout this section, the treatment combinations will be labeled as (A5), (A10), (B5), and (B10). For example, (B5) indicates the data set whose orientation of γ and the value of σ are based on (B) and 5, respectively.

5.2 Comparisons of *PC* Estimators with *LS* Estimator

As a measure of a biased estimator $\tilde{b}$ of β, in simulation studies, the empirical mean square error (*EMSE*) can be used. The *EMSE* of $\tilde{b}$ is defined as

$$EMSE(\tilde{b})=\sum_{j=1}^{4}\sum_{l=1}^{50}\frac{(b_{jl}-\beta_j)^2}{50},\tag{19}$$

where $\tilde{b}_{jl}$ is the jth estimated coefficient of $\tilde{b}$ in the lth Monte Carlo repetition. Note that the number of Monte Carlo repetitions is 50 throughout this study.

Furthermore, it is more useful to standardize the $EMSE(\tilde{b})$ with respect to the true variances of the $b_{j,LS}$, since the resulting standardized *EMSE*s of the *LS* estimator are all equal for the four cases [see LEE(1986)]. The standardized *EMSE*(*SEMSE*) of $\tilde{b}$ is carried out as

$$SEMSE(\tilde{b})=\sum_{i}\sum_{l}\frac{(\tilde{b}_{jl}-\beta_j)^2}{50\sigma_{j,LS}^2}, \tag{20}$$

where $\sigma_{j,LS}^2$ is the variance of the jth element of b_{LS}.

Consider, now, the performances of the various *PC* estimators through the simulations. For the usual *PC* estimators, that is, the *PC*(4th), *PC*(2), and PC_t obtained by the t-value criterion with the significance level $\phi=0.01$, the results in terms of *SEMSE* are shown in Table 5.

Table 5. *SEMSE* of *PC* Estimators

	PC(4th)	*PC*(2)	PC_t
(A5)	1.6108	44.8161	1.6108
(A10)	1.1140	11.1510	7.9674
(B5)	6.3654	5.2476	5.2476
(B10)	2.3233	1.3120	1.3120

In order to utilize the unbiased optimal deletion criterion, furthermore, the unbiased estimates of γ_3^2, γ_4^2, and σ^2 are obtained in Table 6. Note that the first two principal components are out of question and hence the unbiased estimates of γ_1^2 and γ_2^2 are not considered here. Rewriting (14)

$$\sum_{j\in D}\widehat{\gamma_j^2}\leq s^2$$

the suggested *PC* estimators for each case are given in the last column of Table 6.

Table 6. Unbiased Estimates of γ_j^2, $j=3,4$ and σ^2 and the Suggested *PC*(·)

	$\widehat{\gamma_3^2}$	$\widehat{\gamma_4^2}$	$\widehat{\gamma_3^2}+\widehat{\gamma_4^2}$	s^2	Suggested *PC*(·)
(A5)	1189.27 (1234.11)	5.92 (4.47)	1195.19 (1238.58)	24.64	*PC*(4th)
(A10)	1146.57	5.22	1151.79	98.55	*PC*(4th)
(B5)	13.80 (8.84)	27.20 (35.35)	41.00 (44.19)	24.64	*PC*(3rd)
(B10)	20.91	16.90	37.81	98.55	*PC*(2)

* The true values of γ_j^2, $j=3,4$ are in the parentheses.

The *PC* estimators based on the unbiased optimal deletion criterion for cases (A5) and (A10) are the usual *PC* estimators, *PC*(4th). That for the case (B10) is the *PC* estimator resulted from deleting the last two *PCs*, *PC*(2). For the case (B5), interestingly, the *SEMSE* of the suggested *PC* estimator, *PC*(3rd), is 4.1252. It is a substantial improvement over the *PC* estimators based on the usual deletion criteria, compared with the values in Table 5.

Thus, the unbiased optimal deletion criterion determines 'how many' and 'which' principal components to be deleted. For the particular (B5) case, in addition, the new criterion should be taken in order to use principal components regression properly.

6. Concluding Remarks and Suggestions

In order to use principal components regression for combatting the multicollinearity problem, it is important to compare the magnitudes of the γ_j^2s and σ^2 to achieve a proper principal components estimator which indeed improves the least squares estimator in *MSE*. When σ^2 is in a certain range, there are limited orientations of the parameter vector for which the regular principal components estimator, which is based on the eigenvalue criterion, would remedy multicollinearity. The limited values of the parameter vector along with the size of σ are delineated in Section 4.

Thus, in this paper, a new deletion criterion that complements the weaknesses of the usual deletion criterion, named the unbiased optimal deletion criterion, has been developed. It has also been observed by simulated data that the new deletion criterion leads to delete proper principal components for all cases of the simulation. In particular, for the specific case where the usual criterion never performs well, that is, (B5), the unbiased optimal deletion criterion rather than the usual one should be recommended. Furthermore, it also gives the exact number of principal components to be deleted for the cases (A5), (A10), and (B10). It is noted that, for computing convenience only, the number of Monte Carlo repetitions is limited to 50.

The various extensions of the simulation studies may show the range of the signal-to-noise ratio for which the least squares estimator outperforms the principal components estimator even though there exists multicollinearity in the *MLR* model.

References

(1) Belsley, D. A., E. Kuh, and R. E. Welch(1980), *Regression Diagnostics: Identifying Influential Data and Sources of Collinearity*, New York: John Wiley and Sons, Inc.

(2) Farrar, D. E. and R. R. Glauber(1967), Multicollinearity in Regression Analysis: The Problem Revisited, *Review of Economics and Statistics* 49:92~107.

(3) Fomby, T. B. and R. C. Hill(1978), Deletion Criterion for Principal Components Regression Analysis, *American Journal of Agricultural Economics* 60:524~527.

(4) Greenberg, E. (1975), Minimnm Variance Properties of Principal Component Regression, *Journal of the American Statistical Association*, 70:194~197.

(5) Gunst, R. F. and R. L. Mason(1980), *Regression Analysis and Its Application: A Data-Oriented Approach*, New York and Basel: Marcel Dekker, Inc.

(6) Hocking, R. R., F. M. Speed, and M. J. Lynn(1976), A Class of Biased Estimators in Linear Regression, *Technometrics* 18:425~437.

(7) Johnson, S. R., S. C. Reimer, and T. P. Rothrock(1973), Principal Components and the Problem of Multicollinearity, *Metroeconomica* 25:306~317.

(8) Lee, Wonwoo(1986), Fractional Principal Components Regression: A General Approach to Biased Estimators, Unpublished Ph. D. dissertation, Dept. of Statistics, Virginia Tech.

(9) Mandel, J. (1982), Use of the Singular Value Decomposition in Regression Analysis, *American Statistician* 36:15~24.

(10) Massy, W. F. (1965), Principal Components Regression in Exploratory Statistical Research, *Journal of the American Statistical Association* 60:234~256.

(11) Mittelhammer, R. C. and J. L. Baritelle(1977), On Two Strategies for Choosing Principal Components in Regression Analysis, *American Journal of Agricultural Economics* 59:336~343.

(12) Montgomery, D. C. and E. A. Peck(1982), *Introduction to Linear Regression Analysis*, New York: John Wiley & Sons, Inc.

(13) Myers, R. H. (1986), *Classical and Modern Regression with Applications*, Boston: Duxbury Press.

(14) Stuart, M. (1982), A Geometric Approach to Principal Components Analysis, *American Statistician* 36:365~367.

(15) Sujan, I. (1980), Methods for Solving the Problem of Multicollinearity in Econometric Models: A Survey, *Multicollinearity and Biased Estimation*, Gottingen: Vandenhoeck & Ruprecht.

(16) Theobald, C. M. (1974), Generalizations of Mean Square Error Applied to Ridge Regression, *Journal of the Royal Statical Society* B, 36:103~106.

(17) Vinod, H. D. and A. Ullah(1981), *Recent Advances in Regression Method*, New York and Basel: Marcel Dekker, Inc.