

주성분회귀와 고유값회귀에 대한 감도분석의 성질에 대한 연구[†]

신재경¹ · 장덕준²

^{1,2}창원대학교 통계학과

접수 2009년 1월 20일, 수정 2009년 3월 12일, 게재확정 2009년 3월 17일

요약

회귀분석에서 설명변수들 사이에 상관이 높으면 최소제곱추정법에서 구한 회귀계수들의 정도가 떨어진다. 다중공선성이라 불리는 이 현상은 실제 자료분석에서 심각한 문제를 야기시킨다. 이 다중공선성의 문제를 극복하기 위한 여러 가지 방법이 제안되었다. 능형회귀, 축소추정량 그리고 주성분분석에 기초한 주성분회귀와 고유값회귀등이 있다. 지난 수십 년간 많은 통계학자들은 일반적인 중회귀에서 감도분석에 관해 연구하였으며, 주성분회귀, 고유값회귀와 로지스틱 주성분회귀에 대해서도 같은 주제로 연구하였다. 이 모든 방법에서 주성분분석은 중요한 역할을 하였다. 또한, 많은 통계학자들이 주성분분석과 관련된 다변량 방법에서 감도분석에 대해 연구를 하였다. 본 연구논문에서는 주성분회귀와 고유값회귀를 소개하고, 또한 주성분회귀와 고유값회귀에서 감도분석의 방법을 소개하고, 마지막으로 이들 두 방법에 대한 감도분석의 성질에 대해 논의하였다.

주요용어: 감도분석, 고유값회귀, 다중공선성, 영향함수, 주성분분석, 주성분회귀.

1. 서론

일반적으로 중회귀분석에서 독립변수들 사이에 상관이 높으면 보통의 최소제곱법을 사용하면 추정량이 불안정하게 된다는 사실은 잘 알려져 있다. 이런 현상을 다중공선성이라고 하고 이런 문제를 극복하기 위한 방법들이 다양하게 연구되어 왔다. 이런 방법들에는 능형회귀 (ridge regression : RR), 축소추정량 (shrinkage estimators), 주성분에 기초한 주성분회귀 (principal component regression : PCR)와 고유값회귀 (latent root regression : LRR)등이 있다.

지난 수십년 동안 많은 통계학자들은 일반적 중회귀에서 감도분석에 대해 연구하여 왔다. 그들 중 Radhakrishnan과 Kshirsagar (1981), Critchley (1985), Jolliffe (1986) 그리고 Pack 등 (1987)등은 주성분분석에서 감도분석에 대해 연구하였다. 또한, Shin 과 Chang (2003), Shin 과 Moon (1999), Pregibon (1981), Williams (1987), Walker 와 Birch (1988)등은 각각 principal component regression (PCR), logistic regression (LR), generalized linear model (GLM), ridge type regression과 같이 서로 다른 유형에서 같은 주제인 감도분석에 관해 연구하였다. 본 연구 논문에서는 다중공선성의 문제를 극복하는 방법 중에서 주성분회귀와 고유값회귀에서 감도분석을 하여 그들의 성질을 살펴보았다.

2장에서는 주성분회귀와 감도분석의 방법을 소개하고 3장에서는 고유값회귀와 감도분석의 방법을 소개하고 마지막으로 4장에서는 예제를 통해 각각의 방법에 대한 그들의 특성을 살펴보았다. 또한 앞으로의 연구 과제에 대해서도 논의한다.

[†] 이 논문은 2007년도 창원대학교 연구비에 의하여 연구되었음.

¹ 교신저자: (641-773) 경남 창원시 사림동 9번지, 창원대학교 통계학과, 교수.

E-mail: jkshin@changwon.ac.kr

² (641-773) 경남 창원시 사림동 9번지, 창원대학교 통계학과, 교수.

2. 주성분회귀에서 감도분석

2.1. 주성분회귀

다음과 같은 일반적인 회귀모형을 생각하자.

$$y = \mathbf{1}\beta_0 + X\beta + \epsilon, \quad \epsilon \sim N(0, \sigma^2\mathbf{I}). \quad (2.1)$$

여기서 y 는 종속변수의 $(n \times 1)$ 벡터이고, $\mathbf{1}$ 은 모든 요소가 1인 $(n \times 1)$ 벡터이고 X 는 독립변수들의 $(n \times p)$ 행렬이고, ϵ 은 오차항의 $(n \times 1)$ 벡터이다. 또한, μ 와 Φ 는 아래 첨자가 나타내는 관련된 변수들의 평균벡터와 공분산행렬을 각각 나타낸다. 즉, μ_x 는 x 의 평균벡터이고, Φ_{xx} 는 x 의 공분산 행렬이고, Φ_{xy} 는 x 와 y 의 공분산 행렬이고, Γ_{xx} 는 x 의 상관 행렬 등을 나타낸다.

주성분회귀에서 먼저 독립변수들에 주성분분석을 적용한다. 주성분분석에는 2가지 방법이 있는데 하나는 분산-공분산행렬에 기초한 주성분분석이고 다른 하나는 상관행렬에 기초한 주성분 분석이다. 일반적으로 전자보다 후자가 더 자주 사용되어 본 논문에서도 후자인 상관행렬에 기초한 주성분분석에 대해 논의 한다.

이제, 상관행렬을 스펙트럼 분해하며 다음과 같다.

$$\Gamma_{xx} = (\Phi_{xx})_D^{-1/2} \Phi_{xx} (\Phi_{xx})_D^{-1/2} = V_1 \Lambda_1 V_1^T + V_2 \Lambda_2 V_2^T. \quad (2.2)$$

여기서, D 는 “대각”을 의미하고 Λ_1 과 Λ_2 는 각각 관심있는 고유치들과 나머지 고유치들의 대각 행렬들 이고, V_1 과 V_2 는 각각의 고유치에 대응하는 고유벡터들의 행렬들 이다. q 개의 관심있는 고유치에 대한 q 주성분은 다음과 같이 나타낼 수 있다.

$$z = V_1^T (\Phi_{xx})_D^{-1/2} (\mathbf{x} - \mu_x).$$

그러면 $(1, \mathbf{z}^T)$ 에서 y 의 회귀를 생각할 수 있다. 즉, $y = \alpha_0 + \mathbf{z}^T \alpha$. 회귀계수는 다음과 같이 얻을 수 있다.

$$\begin{pmatrix} \alpha_0 \\ \alpha \end{pmatrix} = \begin{pmatrix} 1 & 0^T \\ 0 & \Lambda_1 \end{pmatrix}^{-1} \begin{pmatrix} \mu_y \\ V_1^T (\Phi_{xx})_D^{-1/2} \Phi_{xy} \end{pmatrix} = \begin{pmatrix} \mu_y \\ \Lambda_1^{-1} V_1^T (\Phi_{xx})_D^{-1/2} \Phi_{xy} \end{pmatrix}.$$

표준화된 원래 변수들인 $\mathbf{x}^* = (\Phi_{xx})_D^{-1/2} (\mathbf{x} - \mu_x)$ 에 대한 회귀계수 벡터는 다음과 같이 나타난다.

$$\hat{\beta}^* = V_1 \Lambda_1^{-1} V_1^T (\Phi_{xx})_D^{-1/2} \Phi_{xy}. \quad (2.3)$$

2.2. 주성분회귀에서 감도분석

영향곡선 혹은 영향함수에 기초한 회귀결과에서 i 번째 관측 값의 영향의 측정은 중요하다 (Hampel, 1974). 그것은 $T = T(F)$ 를 분포함수 F 의 범함수로 표현되는 모수라고 두면, \mathbf{x} 에서의 이론적인 영향함수 (theoretical influence function) $I(\mathbf{x}; T)$ 는 다음과 같이 정의된다. $I(\mathbf{x}, y; T) = T^{(1)} = [dT/dc]_{\epsilon=0}$.

우리는 표준화된 회귀계수 벡터 $\hat{\beta}^*$ 에서 데이터의 아주 작은 변화에 대한 영향을 연구한다. 표준화 회귀계수에 대한 영향함수 $I(\mathbf{x}, y; \hat{\beta}^*) \equiv \hat{\beta}^{*(1)}$ 즉, empirical influence curves (EIC)를 다음과 같이 구할 수 있다.

$$\begin{aligned} \hat{\beta}^{*(1)} &= (V_1 \Lambda_1^{-1} V_1^T)^{(1)} (\Phi_{xx})_D^{-1/2} \Phi_{xy} + V_1 \Lambda_1^{-1} V_1^T \left\{ (\Phi_{xx})_D^{-1/2} \right\}^{(1)} \Phi_{xy} \\ &\quad + V_1 \Lambda_1^{-1} V_1^T (\Phi_{xx})_D^{-1/2} \Phi_{xy}^{(1)}. \end{aligned} \quad (2.4)$$

수식의 오른쪽에 있는 $(V_1\Lambda_1^{-1}V_1^T)^{(1)}$ 는 Tanaka (1989)가 제안한 고유치와 관련된 영향함수로부터 구할 수 있다.

$\hat{\beta}^{*(1)}$ 은 벡터량이므로 요약통계량으로서 간단하게 벡터 $\hat{\beta}^{*(1)}$ 의 유클리디안 노름 $\|\hat{\beta}^{*(1)}\|$ 를 계산할 수 있다.

3. 고유값회귀에서 감도분석

3.1. 고유값회귀

고유값회귀에서는 주성분분석을 모든 독립변수와 종속변수에 적용한다. 고유값회귀에서도 주성분회귀에서처럼 상관행렬에 기초한 주성분을 생각한다. 여기서 작은 고유치에 대응하는 주성분들 중에 종속변수 y 의 계수가 역시 작은 것을 non-predictive multicollinearities라 하고 이들은 y 의 예측에 별 도움을 주지 않는 것으로 생각된다.

다시 앞 절의 식 (2.1)을 가정하자. 또한, $Y_i^* = (Y_i - \bar{Y})/\eta$ 라 두자, 여기서 $\eta^2 = \sum_{i=1}^n (Y_i - \bar{Y})^2$ 이다. 행렬 $A = [Y^* : X]$ 로 정의한다. 즉, 표준화된 종속변수와 독립변수들의 $(n \times (p+1))$ 행렬이다. 그러면 $A^T A$ 는 종속변수와 독립변수들의 상관행렬이 된다. j 번째 latent 벡터의 요소를 $\gamma_j^T = (\gamma_{0j}\gamma_{1j}\cdots\gamma_{pj})$ 로 표시하고 $\gamma_j^{0T} = (\gamma_{1j}\gamma_{2j}\cdots\gamma_{pj})$ 라고 두자. 즉, γ_j^0 는 γ_j 에서 첫 번째 요소를 제외한 모든 요소들을 포함한다. 마지막으로 $\Gamma = (\gamma_0\gamma_1\cdots\gamma_p)$ 와 $\Lambda = \text{diag}(\lambda_0, \lambda_1, \cdots, \lambda_p)$ 로 두자. LRR에 대한 보다 상세한 내용은 Webster 등 (1974)을 참조하기 바란다.

회귀계수 β 의 일반적인 최소제곱 추정량을 $A^T A$ 의 고유치와 고유벡터로 나타내면 다음과 같다.

$$\hat{\beta} = -\eta \sum_{j=0}^p \alpha_j \gamma_j^0. \quad (3.1)$$

여기서,

$$\alpha_j = \gamma_{0j} \lambda_j^{-1} / \left(\sum_{r=0}^p \gamma_{0r}^2 / \lambda_r \right), \quad j = 0, \cdots, p, \quad (3.2)$$

이고 residual sum of square는 다음과 같다.

$$SSE = (Y - \hat{Y})^T (Y - \hat{Y}) = \eta^2 \left(\sum_{j=0}^p \gamma_{0j}^2 / \lambda_j \right)^{-1}.$$

q 개의 latent 벡터들인 $\gamma_0, \gamma_1, \cdots, \gamma_{q-1}$ 이 non predictive multicollinearities라고 가정하자. 위의 일반적인 최소제곱추정량은 $\alpha_0 = \alpha_1 = \cdots = \alpha_{q-1} = 0$ 으로 두면 수정할 수 있다. 수정된 추정량은 다음과 같이 표시된다.

$$\hat{\beta}_{LRR} = -\eta \sum_{j=q}^p \alpha_j \gamma_j^0. \quad (3.3)$$

여기서,

$$\alpha_j = \gamma_{0j} \lambda_j^{-1} / \left(\sum_{r=q}^p \gamma_{0r}^2 / \lambda_r \right), \quad j = q, q+1, \cdots, p. \quad (3.4)$$

이와 같은 경우의 residual sum of square는 다음과 같다.

$$SSE_{LRR} = (Y - \hat{Y})^T (Y - \hat{Y}) = \eta^2 \left(\sum_{j=q}^p \gamma_{0j}^2 / \lambda_j \right)^{-1}.$$

3.2. 고유값회귀에서 감도분석

식 (3.1)과 (3.2)에 나타나는 양 들은 y 와 x 의 결합분포함수 F 의 범함수들 이다. 고유값회귀에서 영향함수 $I(x, y; \hat{\beta}_{LRR}^{(1)}) \equiv \hat{\beta}_{LRR}^{(1)}$ 는 다음과 같이 구할 수 있다.

$$\hat{\beta}_{LRR}^{(1)} = \left(\sum_{j=q}^p f_j \gamma_j^0 \right)^{(1)} = \sum_{j=q}^p (f_j^{(1)} \gamma_j^0 + f_j \gamma_j^{0(1)}). \quad (3.5)$$

여기서, $f_j = -\eta \alpha_j$ 이고

$$\begin{aligned} f_j^{(1)} = & -(\gamma_{0j}^{(1)} \eta \lambda_j^{-1} + \gamma_{0j} \eta^{(1)} \lambda_j^{-1} + \gamma_{0j} \eta \lambda_j^{-1(1)}) \left(\sum_{j=q}^p \gamma_{0j}^2 \lambda_j^{-1} \right)^{-1} \\ & + \gamma_{0j} \eta \lambda_j^{-1} \sum_{j=q}^p (\gamma_{0j}^{(2)} \lambda_j^{-1} + \gamma_{0j}^2 \lambda_j^{-1(1)}) \left(\sum_{j=q}^p \gamma_{0j}^2 \lambda_j^{-1} \right)^{-2}. \end{aligned} \quad (3.6)$$

$\gamma_j^{0(1)}, \gamma_{0j}^{(1)}, \gamma_{0j}^{2(1)}, \lambda_j^{(1)}$ 와 $\lambda_j^{-1(1)}$ 은 Tanaka (1988)가 제안한 식으로부터 구할 수 있다. 주성분회귀에서 처럼 요약통계량으로서 간단하게 벡터 $\hat{\beta}_{LRR}^{(1)}$ 의 유클리디안 노름 $\|\hat{\beta}_{LRR}^{(1)}\|$ 를 계산할 수 있다.

4. 예제 및 토의

각 지역에 있는 학교들로부터 수집된 Equal Educational Opportunity (EEO) 데이터 (Chatterjee와 Price, 1977)를 두 분석 방법에 적용하였다. 이 데이터는 3개의 독립변수와 1개의 종속변수를 가지는 70개의 관측값(학교)들로 구성되어 있다.

먼저 주성분회귀에 대해 살펴보기로 하자. EEO 데이터의 상관행렬에 기초한 PCA의 결과는 표 4.1과 같고, 종속변수와 독립변수의 표준화한 행렬인 Γ_{AA} 에 기초한 PCA의 결과는 표 4.2와 같다.

표 4.1 상관행렬에 기초한 PCA의 결과

	Z_1	Z_2	Z_3
X_1	0.5761	-0.6794	-0.4544
X_2	0.5754	0.7320	-0.3648
X_3	0.5805	-0.0513	0.8127
고유치	2.9520	0.0401	0.0080
기여율	0.9840	0.0134	0.0027
누적기여율	0.9840	0.9974	1.0000

표 4.2 Γ_{AA} 에 기초한 PCA의 결과

	Z_0	Z_1	Z_2	Z_3
Y	0.0141	0.0200	0.9478	0.3179
X_1	-0.4518	0.6795	-0.1906	0.5457
X_2	-0.3718	-0.7321	-0.1626	0.5471
X_3	0.8109	0.0426	-0.1972	0.5493
고유치	0.0078	0.0397	0.7532	3.1993
기여율	0.0020	0.0099	0.1883	0.7998
누적기여율	0.0020	0.0119	0.2002	1.0000

표 4.1에서 우리는 처음 2개의 주성분들만 선택하여 PCR을 하였다. 왜냐하면 나머지 고유치는 상대적으로 너무 작기 때문이다. 표 4.2에서는 latent 벡터인 γ_0 와 γ_1 이 non-predictive multicollinearities인 것을 알 수 있다. 왜냐하면 고유치 $\lambda_0(=0.0078)$ 와 $\lambda_1(=0.0397)$ 이 작고 그리고 $|\gamma_{00}|(=0.0141)$ 와 $|\gamma_{01}|(=0.0200)$ 의 값이 역시 작다(cut-off value : $\lambda_j < 0.05$ 그리고 $|\gamma_{0j}| < 0.1$). 따라서, LRR은 (3.3)식에서 “q” 대신 “2”를 바꾸어주면 구할 수 있다. PCR과 LRR을 실행한 결과 표준화 계수 벡터인 $\hat{\beta}^*$ 와 $\hat{\beta}_{LRR}^*$ 는 다음의 표 4.3과 같다.

표 4.3 PCR과 LRR의 결과

Variable	$\hat{\beta}^*$	$\hat{\beta}_{LRR}^*$
X_1	-0.1195	0.1517
X_2	0.4276	0.1227
X_3	0.1251	0.1582

다음은 2장과 3장에서 설명한 감도분석의 영향함수들로부터 개개의 관측값들에 대한 영향을 살펴보기로 하자. $\hat{\beta}^{*(1)}$ 와 $\hat{\beta}_{LRR}^{(1)}$ 에 대한 각각의 요약 통계량인 $\|\hat{\beta}^{*(1)}\|$ 와 $\|\hat{\beta}_{LRR}^{(1)}\|$ 의 인덱스 플롯은 그림 4.1과 같다.

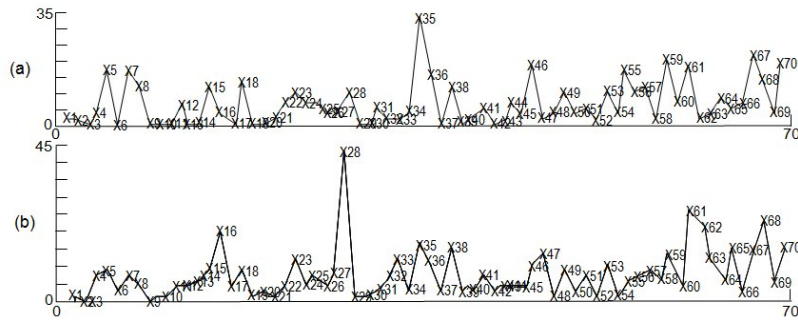


그림 4.1 $\|\hat{\beta}^{*(1)}\|$ 와 $\|\hat{\beta}_{LRR}^{(1)}\|$ 의 인덱스 플롯((A) $\|\hat{\beta}^{*(1)}\|$, (B) $\|\hat{\beta}_{LRR}^{(1)}\|$)

그림 4.1을 살펴보면 (a)에서는 35번째의 관측값이 (b)에서는 28번째 관측값이 영향이 가장 큰 것으로 나타나서 분석방법에 따라 영향력이 큰 관측값이 서로 다른 것을 알 수 있다.

또한 제안한 $\hat{\beta}^{*(1)}$ 및 $\hat{\beta}_{LRR}^{(1)}$ 의 타당성을 살펴보기 위해 $\hat{\beta}_i^{*(1)}$ 및 $\hat{\beta}_{LRR_i}^{(1)}$ (EIC_i)와 다음과 같이 정의되는 SIC_i (sample influence curve)와의 관계를 조사하였다.

$$SIC_i(\beta^*) = -(n - 1)(\hat{\beta}_{(i)}^* - \hat{\beta}^*), \quad i = 1, \dots, n.$$

여기서 $\hat{\beta}^*$ 와 $\hat{\beta}_{(i)}^*$ 는 각각 PCR에서 모든 표본과 i -번째 관측값을 제외한 표본에 대한 표준화 계수 β^* 의 추정치를 나타낸다.

$$SIC_i(\hat{\beta}_{LRR}) = -(n - 1)(\hat{\beta}_{LRR_i} - \hat{\beta}_{LRR}), \quad i = 1, \dots, n.$$

여기서 $\hat{\beta}_{LRR}$ 과 $\hat{\beta}_{LRR_i}$ 는 각각 LRR에서 모든 표본과 i -번째 관측값을 제외한 표본에 대한 표준화 계수 β^* 의 추정치를 나타낸다.

그림 4.2의 (a)와(b)는 각각 PCR과 LRR의 $\|\hat{\beta}^{*(1)}\|$ 와 $\|\hat{\beta}_{LRR}^{(1)}\|$ 에 대한 EIC 와 SIC 의 산점도 (scatter diagram)이다. 이 산점도를 보면 (a)와 (b) 모두 대부분의 값이 $SIC = EIC$ 인 직선상에 있는 것을 알 수 있으며, 이 산점도는 SIC_i 대신에 EIC_i 를 실용적으로 사용할 수 있다는 것을 보여준다. 그런데 (a)와 (b)를 비교해 보면 (a)가 (b)보다 훨씬 더 $SIC = EIC$ 인 직선상에 있는 것을 보여준다. 이것은 (a)인 PCR이 (b)인 LRR보다 더 타당하다는 것을 일 수 있다.

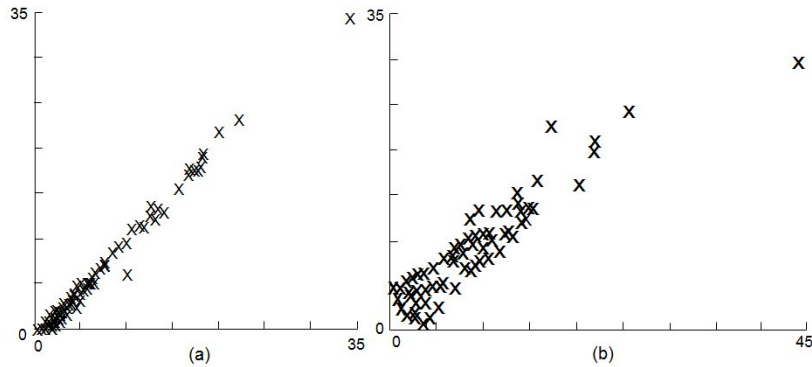


그림 4.2 $\|\hat{\beta}^{*(1)}\|$ 와 $\|\hat{\beta}_{LRR}^{(1)}\|$ 에 대한 EIC (수평축)과 SIC (수직축)의 산점도((A) $\|\hat{\beta}^{*(1)}\|$, (B) $\|\hat{\beta}_{LRR}^{(1)}\|$)

본 연구에서는 단순하게 두 분석방법에서 하나의 예를 통해 한 가지 통계량인 노름인 경우만을 살펴봤는데 앞으로 PCR에서의 주성분 수의 선택에 따른 변화와 LRR에서는 cut-off값의 변화에 따라 각 변수들의 변화와 두 분석방법이 어떻게 변화되는지를 살펴보는 것도 유익한 결과를 얻을 수 있을 것이다. 또한 두 분석방법의 적합 정도도 단순히 그림에만 의존하는 것이 아니라 앞으로는 적절한 통계량에 의한 비교도 가능할 수 있을 것이다.

참고문헌

- Chatterjee, S. and Price, B. (1977). *Regression analysis by example*, John Wiley & Sons, New York.
- Critchley, F. (1985). Influence in principal component analysis. *Biometrika*, **72**, 627-636.
- Hampel, F. R. (1974). The influence curve and its role in robust estimation. *Journal of the American Statistical Associations*, **69**, 383-393.
- Jolliffe, I. T. (1986). *Principal components analysis*, Springer-Verlag, New York.
- Pack, P., Jolliffe, I. T. and Morgan, B. J. T. (1987). Influential observations in principal component analysis: A case study. *Journal of Applied Statistics*, **15**, 37-50.
- Pregibon, D. (1981). Logistic regression diagnostics. *The Annals of Statistics*, **9**, 705-724.
- Radhakrishnan, R. and Kshirsagar, A. M. (1981). Influence functions for certain parameters in multivariate analysis. *Communication in Statistics: Theory and Methods*, **10**, 515-529.
- Shin, J. K. and Chang, D. J. (2003). Sensitivity analysis in principal component regression with quadratic approximation. *Journal of Korean Data & Information Science Society*, **14**, 623-630.
- Shin, J. K. and Moon, S. H. (1999). Numerical investigations in choosing the number of principal component in principal component regression-case II. *Journal of Korean Data & Information Science Society*, **10**, 163-172.
- Tanaka, Y. (1988). Sensitivity analysis in principal component analysis : Influence on the subspace spanned by principal components. *Communication in Statistics: Theory and Methods*, **17**, 3157-3175. (Corrections, A 18(1989), 4305).

- Tanaka, Y. (1989). Influence functions related to eigenvalue problems which appear in multivariate methods. *Communication in Statistics: Theory and Methods*, **18**, 3991-4010.
- Walker, E. and Birch, J. B. (1988). Influence measure in ridge regression. *Technometrics*, **30**, 221-227.
- Webster, J. T., Gunst, R. F. and Mason, R. L. (1974). Latent root regression analysis. *Technometrics*, **16**, 513-522.
- Williams, D. A. (1987). Generalized linear model diagnostics using the deviance and single case deletions. *Applied Statistics*, **36**, 181-191.

A study on the properties of sensitivity analysis in principal component regression and latent root regression[†]

Jae-Kyoung Shin¹ · Duk-Joon Chang²

^{1,2}Department of Statistics, Changwon National University

Received 20 January 2009, revised 12 March 2009, accepted 17 March 2009

Abstract

In regression analysis, the ordinary least squares estimates of regression coefficients become poor, when the correlations among predictor variables are high. This phenomenon, which is called multicollinearity, causes serious problems in actual data analysis. To overcome this multicollinearity, many methods have been proposed. Ridge regression, shrinkage estimators and methods based on principal component analysis (PCA) such as principal component regression (PCR) and latent root regression (LRR). In the last decade, many statisticians discussed sensitivity analysis (SA) in ordinary multiple regression and same topic in PCR, LRR and logistic principal component regression (LPCR). In those methods PCA plays important role. Many statisticians discussed SA in PCA and related multivariate methods. We introduce the method of PCR and LRR. We also introduce the methods of SA in PCR and LRR, and discuss the properties of SA in PCR and LRR.

Keywords: Influence function, latent root regression, multicollinearity, principal component analysis, principal component regression, sensitivity analysis.

[†] This research is financially supported by Changwon National University in 2007.

¹ Corresponding author: Professor, Department of Statistics, Changwon National University, Changwon, Gyeongnam 641-773, Korea. E-mail: jkshin@changwon.ac.kr

² Professor, Department of Statistics, Changwon National University, Changwon, Gyeongnam 641-773, Korea.