

Shrinkage Structure of Ridge Partial Least Squares Regression¹⁾

Jong-Duk Kim²⁾

Abstract

Ridge partial least squares regression (RPLS) is a regression method which can be obtained by combining ridge regression and partial least squares regression and is intended to provide better predictive ability and less sensitive to overfitting. In this paper, explicit expressions for the shrinkage factor of RPLS are developed. The structure of the shrinkage factor is explored and compared with those of other biased regression methods, such as ridge regression, principal component regression, ridge principal component regression, and partial least squares regression using a near infrared data set.

Keywords : Partial Least Squares Regression, Principal Component Regression, Ridge Partial Least Squares Regression, Ridge Principal Component Regression, Shrinkage Factor

1. 서론

다음의 회귀모형을 고려한다.

$$y = X\beta + \epsilon \quad (1.1)$$

여기서 y 는 $n \times 1$ 반응벡터, X 는 $n \times p$ 설계행렬, β 는 $p \times 1$ 모수벡터, ϵ 은 $n \times 1$ 오차벡터로 평균이 0 , 분산이 $\sigma^2 I$ 라고 가정한다. X 의 계수는 r 로서 불완전계수(rank-deficiency)일 수 있는 일반적인 경우를 가정한다. X 와 y 는 중심화 또는 표준화된 것으로 간주한다.

자주 사용될 비정칙값 분해(singular value decomposition, SVD)의 부호는 다음과 같이 정의한다. 행렬 X 의 SDV는 $X = UDV'$ 으로 쓸 수 있으며 여기서 U 는 열들이

1) 이 논문은 2005년도 부산외국어대학교 학술연구 조성비에 의하여 연구되었음

2) 부산시 남구 우암동 산 55-1, 부산외국어대학교 데이터경영학과 교수
E-mail : jdkim@pufs.ac.kr

XX' 의 고유벡터로 구성된 $n \times n$ 행렬, V 는 열들이 $X'X$ 의 고유벡터로 구성된 $p \times p$ 행렬, D 는 r 개의 선도주대각원소가 X 의 비정칙값이 크기순으로 나열되고 나머지는 0인 $n \times p$ 행렬이다. U 와 V 는 직교행렬이다(즉 $U'U = UU' = I_n$, $V'V = VV' = I_p$). X 의 계수가 r 일 때 U, D, V 를 각각 분할하여 $U = (U_r, U_{n-r})$, $D = \begin{pmatrix} D_r & 0 \\ 0 & 0 \end{pmatrix}$, $V = (V_r, V_{p-r})$ 로 쓰면 X 의 SVD는 $X = U_r D_r V_r'$ 이며 여기서 $U_r = (u_1, \dots, u_r)$, $D_r = \text{diag}(\delta_1, \dots, \delta_r)$, $V_r = (v_1, \dots, v_r)$ 로서 $U_r' U_r = I_r$ 이고 $V_r' V_r = I_r$ 이다. 그리고 $X'X = V_r D_r^2 V_r' = \sum_{j=1}^r \lambda_j v_j v_j'$ 이고 $\lambda_j = \delta_j^2$ 은 고유값이다.

본문에서 인용될 QR-분해는 다음과 같다. 행렬 A 가 $p \times m$ 행렬이라고 하자 ($p \geq m$). 그러면 $A = QR$ 의 분해가 존재하는데 여기서 Q 는 $Q'Q = I_m$ 을 만족하는 $p \times m$ 행렬이고 R 은 $m \times m$ 상삼각행렬이다.

설명변수들에 다중공선성이 존재하거나 설명변수의 수가 관측수 보다 큰 경우는 보통최소제곱회귀(ordinary least squares regression, OLS) 보다는 능형회귀(ridge regression, RR), 주성분회귀(principal component regression, PCR), 부분최소제곱회귀(partial least squares regression, PLS) 등의 편향회귀가 유용하게 사용된다. 이들 회귀의 추정량의 유클리드 놈(norm)은 OLS의 것보다 작다는 의미에서 축소 추정량(shrinkage estimator)이라 할 수 있다. Frank와 Friedman(1993)은 이 세 추정량의 축소 성질을 주방향, 즉 X 의 SVD에 의해 주어지는 방향에서 비교하였다. Vigneau 외(1997)는 RR과 PCR의 장점을 결합한 능형주성분회귀(ridge principal component regression, RPCR)를 제안하였다. Hoy 외(2002)는 RR과 PLS를 결합한 능형부분최소제곱회귀(ridge partial least squares regression, RPLS)를 보였다. Kim(2007)은 다른 접근법을 이용한 RPCR과 새로운 RPLS를 유도하였고 이들이 과대적합에 덜 민감하고 좋은 예측력을 가질 수 있음을 보였다.

본 논문에서는 Kim(2007)이 제안한 RPCR과 RPLS 추정량의 축소구조를 살펴보고자 한다. 논문의 순서는 다음과 같다. 2절에서는 선형회귀에서 축소를 정의하고 축소 추정량의 일반식을 도입하고 이것을 이용하여 축소추정량의 평균제곱오차를 유도한다. 3절에서는 PLS의 축소인자 식의 몇 가지 다른 형태를 보인다. 4절에서는 RPCR과 RPLS의 축소인자의 식을 유도하고 그 특성을 논한다. 5절에서는 실제 데이터 셋으로 RPCR과 RPLS의 축소인자의 값이 RR, PCR, PLS의 축소인자의 값에 비해 어떻게 달라지는지를 본다.

2. 축소와 축소추정량

벡터모수 θ 의 추정량 $\hat{\theta}$ 의 질을 평가하는 한 방법은 그것의 평균제곱오차(mean squared error, MSE)를 보는 것으로 다음과 같이 정의된다.

$$\begin{aligned} \text{MSE}(\hat{\theta}) &= E[\text{tr}(\hat{\theta} - \theta)(\hat{\theta} - \theta)'] \\ &= E[(\hat{\theta} - \theta)'(\hat{\theta} - \theta)] \end{aligned}$$

이것은 다음과 같이 분해된다.

$$\begin{aligned} \text{MSE}(\hat{\theta}) &= E[(\hat{\theta} - E(\hat{\theta}))'(\hat{\theta} - E(\hat{\theta}))] + (E(\hat{\theta}) - \theta)'(E(\hat{\theta}) - \theta) \\ &= \sum \text{Var}(\hat{\theta}_j) + \sum (\text{Bias in } \hat{\theta}_j)^2 \end{aligned}$$

즉 $\hat{\theta}$ 의 MSE를 분산항과 편향제곱항으로 나눈 것으로, 전자를 $\text{Var}(\hat{\theta})$, 후자를 $(\text{Bias in } \hat{\theta})^2$ 으로 흔히 표현한다. 여기서 $\text{Var}(\hat{\theta})$ 의 실제 의미는 $\text{tr}(\text{Var}(\hat{\theta}))$ 이다.

OLS 추정량은 선형이다. 회귀모형 (1.1)에서 X 가 완전열계수이면 OLS 추정량은 유일하지만 완전열계수가 되지 못하면 $y = X\beta + \epsilon$ 은 과대모수화되어 무수히 많은 β 가 존재한다. 그렇지만 여전히 $\min_{\beta} \|y - X\beta\|$ 의 $\hat{\beta}_{ols} = (X'X)^+ X'y$ 와 $\hat{y}_{ols} = X(X'X)^+ X'y$ 를 사용할 수 있다. 첨자 $+$ 는 Moore-Penrose 일반화역행렬을 의미한다. $\hat{\beta}_{ols}$ 은 유일하지 않지만 \hat{y}_{ols} 은 유일하다. 그러나 이 추정량은 데이터의 자원을 모두 사용하기 때문에 안정적이지 못하다.

이제 $A \equiv (X'X)^+ X'$, $B \equiv X(X'X)^+ X'$ 로 두자. B 는 벡터 y 를 X 의 열들에 의해 생성되는 공간에 사영시키는 사영행렬이다. 회귀모형 (1)에서 $\hat{\theta} = Ly$ 를 하나의 선형 추정량이라 하자. 그러면 $E(\hat{\theta}) = LX\beta$ 이고 $\text{Var}(\hat{\theta}) = \sigma^2 \text{tr}(LL')$ 이다. 추정량 \hat{y}_{ols} 는 비편향(unbiased)으로 $E(\hat{y}_{ols}) = BX\beta = X\beta$ 이다. 추정량 $\hat{\beta}_{ols}$ 는 $\beta \in \text{range}(X'X)^+$ 이면 비편향으로

$$E(\hat{\beta}_{ols}) = (X'X)^+ X'X\beta = \beta$$

이다. 분산항을 보면 다음과 같다. $\hat{\beta}_{ols}$ 에서 $AA' = (X'X)^+ = V_r(D_r^2)^{-1}V_r'$ 이고 $\text{tr}(AA') = \sum (1/\delta_j^2)$ 이므로

$$\text{Var}(\hat{\beta}_{ols}) = \sigma^2 \sum_{j=1}^r \frac{1}{\delta_j^2} \quad (2.1)$$

이다. B 는 사영행렬이다. 따라서 $\text{tr}(BB') = \text{rank}(X) = r$ 이고 따라서 $\text{Var}(\hat{y}_{ols}) = \sigma^2 r$ 이다.

그러므로 추정량 $\hat{\beta}_{ols}$ 의 MSE는 $X'X$ 의 고유값 $\lambda_1, \dots, \lambda_r$ 에 의존함을 알 수 있다. 식 (2.1)에서 어떤 고유값이 작으면 $\hat{\beta}_{ols}$ 의 분산은 매우 커지고 따라서 MSE가 커짐을 알 수 있다.

OLS 추정량은 다음과 같이 고유벡터의 선형결합으로 나타낼 수 있다.

$$\hat{\beta}_{ols} = V_r D_r^{-1} U_r' y = \sum_{j=1}^r \hat{\alpha}_j v_j$$

여기서 $\hat{\alpha}_j = u_j' y / \delta_j$ 이다. MSE를 줄이기 위해 OLS 추정량에서 분산을 크게 만드는

고유방향(eigendirection)을 축소시키는 방법을 생각해볼 수 있다. 이렇게 하면 당연히 편향이 발생한다. 따라서 우리는 편향의 증가량이 분산의 감소량에 비해 작기를 기대하면서 OLS 추정량의 축소를 시도한다.

모수벡터 β 의 축소추정량의 일반식으로 다음을 고려한다(Frank와 Friedman, 1993).

$$\hat{\beta}_{shr} = \sum_{j=1}^r f_{j(\cdot)} \hat{\alpha}_j v_j$$

이것은 추정량을 고유벡터 v, \dots, v_r 의 선형결합으로 나타낸 형태이며, 여기서 $f_{j(\cdot)}$ 는 축소인자(shrinkage factor)라 부르기로 한다. 축소인자가 y 에 의존하지 않으면 $\hat{\beta}_{shr}$ 는 y 에 선형이며, $f_{j(\cdot)} \neq 1$ 인 인자는 j 번째 성분(고유방향)의 편향을 증가시킨다. j 번째 성분(고유방향)의 분산은 $f_{j(\cdot)} < 1$ 이면 감소하고 $f_{j(\cdot)} > 1$ 이면 증가한다. 따라서 분산의 감소량에 비해 편향의 증가량이 작은 축소 방법이 바람직할 것이다.

이제 추정량의 MSE가 축소인자에 의해 어떻게 영향을 받는지 보자. 축소추정량이 선형이면 즉 축소인자가 y 에 의존하지 않으면 문제는 간단한 편이다. 축소추정량을 행렬형으로 쓰면

$$\hat{\beta}_{shr} = V_r F_{(\cdot)} \hat{\alpha}$$

이며, 여기서 축소인자행렬 $F_{(\cdot)} = \text{diag}(f_{1(\cdot)}, \dots, f_{r(\cdot)})$ 이고 $\hat{\alpha} = D_r^{-1} U_r' y$ 이다. 이때의 평균반응의 추정량은

$$\hat{y}_{shr} = X \hat{\beta}_{shr} = U_r D_r F_{(\cdot)} \hat{\alpha} = U_r F_{(\cdot)} U_r' y$$

이다.

축소추정량 $\hat{\beta}_{shr}$ 의 MSE를 구하면 다음과 같다. 먼저 분산항은

$$\begin{aligned} \sigma^2 \text{tr}[(V_r F_{(\cdot)} D_r^{-1} U_r')(V_r F_{(\cdot)} D_r^{-1} U_r)'] &= \sigma^2 \text{tr}(F_{(\cdot)} D_r^{-2} F_{(\cdot)}) \\ &= \sigma^2 \sum_{j=1}^r \frac{f_{j(\cdot)}^2}{\delta_j^2} \end{aligned}$$

이다. 그리고 $E(\hat{\beta}_{shr}) = E(V_r F_{(\cdot)} D_r^{-1} U_r' y) = V_r F_{(\cdot)} V_r' \beta$ 이므로 편향제곱 부분은 다음과 같다.

$$\begin{aligned} (E(\hat{\beta}_{shr}) - \beta)'(E(\hat{\beta}_{shr}) - \beta) &= (V_r F_{(\cdot)} V_r' \beta - \beta)'(V_r F_{(\cdot)} V_r' \beta - \beta) \\ &= \beta'(V_r F_{(\cdot)} V_r' - I_p)^2 \beta \\ &= \beta'(V_r F_{(\cdot)} V_r' - V_r V_r' - V_{p-r} V_{p-r}')^2 \beta \\ &= \beta' V_r (F_{(\cdot)} - I_r)^2 V_r' \beta + \beta' V_{p-r} V_{p-r}' \beta \end{aligned}$$

$$= \sum_{j=1}^r (f_{j(\cdot)} - 1)^2 (\mathbf{v}_j' \boldsymbol{\beta})^2 + \sum_{j=r+1}^p (\mathbf{v}_j' \boldsymbol{\beta})^2$$

따라서

$$\text{MSE}(\hat{\boldsymbol{\beta}}_{shr}) = \sigma^2 \sum_{j=1}^r \frac{f_{j(\cdot)}^2}{\delta_j^2} + \sum_{j=1}^r (f_{j(\cdot)} - 1)^2 (\mathbf{v}_j' \boldsymbol{\beta})^2 + \sum_{j=r+1}^p (\mathbf{v}_j' \boldsymbol{\beta})^2$$

이다. \mathbf{X} 가 완전열계수일 때, 즉 $\text{rank}(\mathbf{X}) = p$ 일 때는

$$\text{MSE}(\hat{\boldsymbol{\beta}}_{shr}) = \sigma^2 \sum_{j=1}^p \frac{f_{j(\cdot)}^2}{\delta_j^2} + \sum_{j=1}^p (f_{j(\cdot)} - 1)^2 (\mathbf{v}_j' \boldsymbol{\beta})^2$$

으로 쓸 수 있다.

$\hat{\mathbf{y}}_{shr}$ 의 MSE는 다음과 같다. 분산항은

$$\sigma^2 \text{tr}[(\mathbf{U}_r \mathbf{F}_{(\cdot)} \mathbf{U}_r')(\mathbf{U}_r \mathbf{F}_{(\cdot)} \mathbf{U}_r')'] = \sigma^2 \text{tr}(\mathbf{F}_{(\cdot)}^2) = \sigma^2 \sum_{j=1}^r f_{j(\cdot)}^2$$

이다. 그리고 $E(\hat{\mathbf{y}}_{shr}) = E(\mathbf{U}_r \mathbf{F}_{(\cdot)} \mathbf{U}_r' \mathbf{y}) = \mathbf{U}_r \mathbf{F}_{(\cdot)} \mathbf{D}_r \mathbf{V}_r' \boldsymbol{\beta}$ 이므로 편향제곱 부분은

$$\begin{aligned} (\mathbf{U}_r \mathbf{F}_{(\cdot)} \mathbf{D}_r \mathbf{V}_r' \boldsymbol{\beta} - \mathbf{U}_r \mathbf{D}_r \mathbf{V}_r' \boldsymbol{\beta})' (\mathbf{U}_r \mathbf{F}_{(\cdot)} \mathbf{D}_r \mathbf{V}_r' \boldsymbol{\beta} - \mathbf{U}_r \mathbf{D}_r \mathbf{V}_r' \boldsymbol{\beta}) &= (\mathbf{V}_r' \boldsymbol{\beta})' \mathbf{D}_r^2 (\mathbf{F}_{(\cdot)} - \mathbf{I}_r)^2 \mathbf{V}_r' \boldsymbol{\beta} \\ &= \sum_{j=1}^r \delta_j^2 (f_{j(\cdot)} - 1)^2 (\mathbf{v}_j' \boldsymbol{\beta})^2 \end{aligned}$$

이다. 따라서

$$\text{MSE}(\hat{\mathbf{y}}_{shr}) = \sigma^2 \sum_{j=1}^r f_{j(\cdot)}^2 + \sum_{j=1}^r \delta_j^2 (f_{j(\cdot)} - 1)^2 (\mathbf{v}_j' \boldsymbol{\beta})^2$$

이다.

만약 축소인자가 결정적이면, 즉 \mathbf{y} 에 의존하지 않으면, 모든 $f_{j(\cdot)} \neq 1$ 에서 편향을 증가시킨다. 반면 $f_{j(\cdot)} < 1$ 은 분산을 감소시키고, $f_{j(\cdot)} > 1$ 은 분산을 증가시킨다. 따라서 $f_{j(\cdot)}$ 가 1보다 크면 분산과 편향이 둘 다 증가하여 MSE는 필연적으로 커진다. 그러나 축소인자가 확률적이면 이렇게 단정하여 말할 수 없다. PLS가 이 경우에 해당한다.

3. 부분최소제곱회귀의 축소인자

PLS는 H. Wold(1975)에 의해 개발되었으며, 다음의 과정을 반복하여 가중값벡터 w_k , 인자벡터 t_k 를 연속적으로 계산한다.

1. $w_k = X_k' y / \|X_k' y\|$
2. $t_k = X_k w_k$
3. $X_{k+1} = X_k - P_{t_k} X_k$, 단 $P_{t_k} = t_k(t_k' t_k)^{-1} t_k'$

m 개 인자의 PLS 계수벡터 추정량은 다음과 같다.

$$\hat{\beta}_{pls}^{(m)} = W_m (W_m' X' X W_m)^{-1} W_m' X' y, \text{ 단 } W_m = (w_1, \dots, w_m)$$

이와 같이 PLS의 회귀해는 반복 알고리즘으로 주어지고 인자를 구할 때 X 뿐만 아니라 y 의 정보도 이용하기 때문에 복잡한 비선형이 되어 그 특성을 이해하기 어렵다.

PLS의 비반복적 해도 구해졌다(Helland, 1988). $m \geq 1$ 에서 다음의 Krylov 공간을 고려한다.

$$K^{(m)} = \text{span}(X' y, (X' X) X' y, \dots, (X' X)^{m-1} X' y)$$

그러면 위의 PLS 알고리즘에서의 가중값벡터 집합 $\{w_1, \dots, w_m\}$ 에 의해 생성되는 공간과 Krylov 공간은 동일하다. m 개 잠재인자의 PLS 추정량은 다음의 최적화의 해이다.

$$\begin{aligned} \min_{\beta} \|y - X\beta\| \\ \text{such that } \beta \in K^{(m)} \end{aligned}$$

$p \times m$ 행렬 S_m 을 열들이 부공간 $K^{(m)}$ 을 생성하는 한 행렬이라 하면 m 개 인자의 PLS 추정량은

$$\hat{\beta}_{pls}^{(m)} = S_m (S_m' X' X S_m)^{-1} S_m' X' y$$

이다.

m 개 인자의 PLS의 비반복적 해로 다음의 형태도 가능하다(Kim과 Moon, 2005).

$$\hat{\beta}_{pls}^{(m)} = V_r D_r^{-1} G_{(m)} U_r' y$$

여기서 $G_{(m)} = Q_m (Q_m' Q_m)^{-1} Q_m'$, $Q_m = (D_r^2 U_r' y, D_r^4 U_r' y, \dots, D_r^{2m} U_r' y)$ 이고, V_r , D_r , U_r 은 비정칙값 분해 $X = U_r D_r V_r'$ 에서 나온 것이다.

대표적인 편향회귀인 능형회귀(RR)와 주성분회귀(PCR)의 축소인자는 다음과 같은 쉽게 확인된다.

$$\text{RR(능형상수 } c): f_{j(rr)} = \frac{\delta_j^2}{\delta_j^2 + c}, \quad j = 1, 2, \dots, r$$

$$\text{PCR}(m \text{ 개 인자}): f_{j(pcr)} = \begin{cases} 1, & 1 \leq j \leq m \\ 0, & m+1 < j \leq r \end{cases}$$

RR에서 c 는 조절항으로 각 고유방향에 따른 축소량을 조절하는데 $\delta_j^2 > 0$ 이고 $c > 0$ 이므로 모든 고유방향으로 항상 축소한다. PCR에서 $f_{j(\cdot)} = 1$ 인 것은 고유방향에 따른 주성분이 그대로 포함되고 $f_{j(\cdot)} = 0$ 인 것은 그것이 완전히 빠짐을 의미한다. 따라서 이들 추정량은 어떤 방향으로 어떻게 축소되는지가 쉽게 드러난다. RR에서 $c = 0$, PCR에서 모든 j 에서 $f_{j(\cdot)} = 1$ 이면 OLS가 됨은 자명하다.

PLS 추정량은 반응변수에 대해 복잡한 비선형이기 때문에 축소의 성질을 알기가 어렵다. Goutis(1996)는 PLS 회귀해도 축소추정량이라는 것과 다음이 성립함을 보였다.

$$\|\hat{\beta}_{pls}^{(1)}\| \leq \|\hat{\beta}_{pls}^{(2)}\| \leq \dots \leq \|\hat{\beta}_{pls}^{(r)}\| = \|\hat{\beta}_{ols}\|$$

여기서 괄호안의 위첨자는 사용된 PLS 인자의 개수이다.

Frank와 Friedman(1993)은 PLS 해의 축소인자가 다음과 같다고 하였다.

정리 3.1. m 개 인자의 PLS 추정량의 축소인자는 다음과 같다.

$$f_{j(pls)}^{(m)} = \sum_{k=1}^m \delta_j^{2k} b_k, \quad j = 1, 2, \dots, r$$

여기서 δ_j 는 행렬 X 의 j 번째 비정칙값이며, $b = (b_1, \dots, b_m)'$ 은 $b = S^{-1}s$ 로서 S 는

(k, l) 번째 원소가 $S_{kl} = \sum_{j=1}^r \delta_j^{2(k+l+1)} \hat{\alpha}_j^2$ 인 $m \times m$ 행렬이고 s 는 k 번째 원소가

$s_k = \sum_{j=1}^r \delta_j^{2(k+1)} \hat{\alpha}_j^2$ 인 $m \times 1$ 벡터이다.

이 정리의 유도과정은 Kim(2006)에서 볼 수 있다. Frank와 Friedman(1993)은 축소인자가 $f_{j(pls)} > 1$ 인 것도 있음을 언급하였다. 즉 어떤 고유방향으로는 팽창하는 것도 있다는 것이다.

Lingjaerde와 Christopherson(2000)는 다른 축소인자의 식을 보였으며 다음과 같다.

정리 3.2. m 개 인자의 PLS 회귀의 축소인자는 다음과 같다.

$$f_{j(pls)}^{(m)} = 1 - \prod_{k=1}^m \left(1 - \frac{\lambda_j}{\mu_k^{(m)}} \right), \quad j = 1, 2, \dots, r$$

여기서 $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_r$ 은 $X'X$ 의 고유값이고, $\mu_1^{(m)} \geq \mu_2^{(m)} \geq \dots \geq \mu_r^{(m)}$ 은 $W_m'X'XW_m$ 의 고유값, W_m 은 PLS 알고리즘에서의 가중값벡터 $\{w_1, \dots, w_m\}$ 으로 구성된 $p \times m$ 행렬이다.

PLS 회귀의 축소인자는 다음과 같이 쓸 수도 있다(Kim, 2006).

정리 3.3. m 개 인자의 PLS 회귀의 축소인자는 다음과 같다.

$$f_{j(pls)}^{(m)} = \frac{g_j' U_r' y}{u_j' y}, \quad j = 1, 2, \dots, r$$

단 u_j 는 U_r 의 j 번째 열벡터, g_j' 은 $G_{(m)}$ 의 j 번째 행벡터, $G_{(m)} = Q_m(Q_m'Q_m)^{-1}Q_m'$, $Q_m = (D_r^2 U_r' y, D_r^4 U_r' y, \dots, D_r^{2m} U_r' y)$ 이다.

Frank와 Friedman(1993)은 m 개 인자의 PLS 회귀는 m 번째 고유값과 근처에서 계수를 확장한다고 결론을 내렸다. 그러나 Butler과 Denham(2000)은 다른 예를 통해 항상 그렇게 되는 것은 아니라고 하였다. Lingjaerde와 Christopherson(2000)은 축소인자의 특성에 관해 몇 가지 결과를 보였다.

앞서 언급한 바와 같이 PLS 회귀의 주 문제점은 $\hat{\beta}_{pls}$ 가 복잡한 비선형이기 때문에 이것의 정확한 통계적 성질을 알기가 매우 어렵다는 것이다. $\hat{\beta}_{pls}$ 와 \hat{y}_{pls} 의 MSE를 구하는 문제도 마찬가지이다. 따라서 PLS 알고리즘에서 얻어진 가중값 행렬 W_m 을 결정적(deterministic)으로 간주하는 방법이 주로 사용된다. 따라서 축소인자도 고정이라고 가정하여 근사적인 방법을 시도해볼 수 있다. 즉 2절에서 구한 $\hat{\beta}_{shr}$ 와 \hat{y}_{shr} 의 MSE 결과를 그대로 적용하면 m 개 인자의 $\hat{\beta}_{pls}$ 와 \hat{y}_{pls} 의 MSE는 각각

$$\begin{aligned} \text{MSE}(\hat{\beta}_{pls}) &= \sigma^2 \sum_{j=1}^r \frac{(f_{j(pls)}^{(m)})^2}{\delta_j^2} + \sum_{j=1}^r (f_{j(pls)}^{(m)} - 1)^2 (v_j' \beta)^2 + \sum_{j=r+1}^p (v_j' \beta)^2 \\ \text{MSE}(\hat{y}_{pls}) &= \sigma^2 \sum_{j=1}^r (f_{j(pls)}^{(m)})^2 + \sum_{j=1}^r \delta_j^2 (f_{j(pls)}^{(m)} - 1)^2 (v_j' \beta)^2 \end{aligned}$$

이며, 여기서 $f_{j(pls)}^{(m)}$ 는 PLS의 j 번째 축소인자이다.

이 MSE에 따르면 $f_{j(pls)} > 1$ 이면 \hat{y}_{pls} 의 MSE는 OLS에 비해 분산과 편향제곱이 모두 증가되어 해만 끼치게 된다. Frank와 Friedman(1993)은 축소인자에서 $f_{j(pls)} > 1$ 으로 나오는 것은 1로 대체하면 PLS 회귀의 수행이 더 향상될지 모른다고 추측하였다. 그러나 축소인자 $f_{j(pls)}$ 가 y 의 복잡한 비선형이므로 명확하지 않다.

4. 능형부분최소제곱회귀의 축소인자

능형회귀(RR)와 주성분회귀(PCR)의 결합한 능형주성분회귀(ridge principal component regression, RPCR)는 다음과 같이 쓸 수 있다(Kim, 2007).

$$\hat{\beta}_{rpcr}^{(m)} = V_m D_m^{-1} (I_m + c D_m^{-2})^{-1} U_m' y$$

여기서 V_m 은 V_r 의 첫 m 개 열들로 구성된 $r \times m$ 행렬, U_m 은 U_r 의 첫 m 개 열들로 구성된 $n \times m$ 행렬, D_m 은 D_r 의 m 차 선도주부분행렬(leading principal submatrix), c 는 능형상수이다. 이 추정량의 대수식은

$$\hat{\beta}_{rpcr}^{(m)} = \sum_{j=1}^m \left(\frac{u_j' y}{\delta_j + c/\delta_j} \right) v_j$$

이다. 이 회귀방법은 PCR에서처럼 영에 가까운 비정칙값 즉 확실히 랜덤오차인 부분을 제거하고 더불어 RR에서처럼 작은 비정칙값을 위해 적절한 값 c/δ_j 로 추가 조정 해주어 더 나은 회귀방법이 되리라 기대된다.

정리 4.1. m 개 인자의 RPCR의 축소인자는 다음과 같다.

$$f_{j(rpcr)}^{(m)} = \begin{cases} \frac{\delta_j^2}{\delta_j^2 + c}, & 1 \leq j \leq m \\ 0, & m+1 < j \leq r \end{cases}$$

여기서 δ_j 는 X 의 j 번째 비정칙값이다.

증명. RPCR의 추정량 $\hat{\beta}_{rpcr}^{(m)} = V_m D_m^{-1} (I_m + c D_m^{-2})^{-1} U_m' y$ 는

$$\hat{\beta}_{rpcr}^{(m)} = V_r D_r^{-1} (I_r + c D_r^{-2})^{-1} E U_r' y$$

로 쓸 수 있으며 여기서 $r \times r$ 행렬 E 는 m 차 선도주부분행렬(leading principal submatrix)은 $m \times m$ 의 단위행렬이고 나머지 원소는 전부 0인 행렬이다. 이 추정량은 다시

$$\hat{\beta}_{rpcr}^{(m)} = V_r (I_r + c D_r^{-2})^{-1} E D_r^{-1} U_r' y$$

로 쓸 수 있다. 여기서 $F_{(\cdot)} \equiv (I_r + c D_r^{-2})^{-1} E$ 로 두면 $\hat{\beta}_{rpcr}^{(m)} = V_r F_{(\cdot)} D_r^{-1} U_r' y$ 의 형태가 되고 $F_{(\cdot)}$ 는 m 개의 선도주대각의 원소는 $\delta_j^2 / (\delta_j^2 + c)$, $1 \leq j \leq m$ 이고 나머지는 전부

0인 대각행렬이므로 증명된다.

여기서 RPCR의 축소인자는 PCR에서 선택된 첫 m 개 성분의 고유방향을 각각 $\delta_j^2/(\delta_j^2+c)$ 의 비율로 축소함을 알 수 있다. 능형상수 c 는 작은 양의 값이므로 고유값이 큰 고유방향은 축소량이 작으나 고유값이 작은 고유방향은 축소량이 커지게 된다.

RR과 PLS의 결합한 능형부분최소제곱회귀(ridge partial least squares regression, RPLS)는 다음과 같이 쓸 수 있다(Kim, 2007).

$$\hat{\beta}_{rpls}^{(m)} = V_r D_r^{-1} (I_r + c D_r^{-2})^{-1} G_{(m)} U_r' y$$

여기서 $G_{(m)} = Q_m (Q_m' Q_m)^{-1} Q_m'$, $Q_m = (D_r^2 U_r' y, D_r^4 U_r' y, \dots, D_r^{2m} U_r' y)$ 이고, V_r , D_r , U_r 은 1절에서 정의된 대로이다. 이것의 대수식은

$$\hat{\beta}_{rpls}^{(m)} = \sum_{j=1}^r \left(\frac{g_j' U_r' y}{\delta_j + c/\delta_j} \right) v_j$$

임을 보일 수 있다. 이 추정량은 RR의 장점과 PLS의 장점을 같이 가지는 것으로 보인다. 이제 이 추정량의 축소인자의 식을 보면 다음과 같다.

정리 4.2. m 개 인자의 RPLS의 축소인자의 식은 다음과 같이 쓸 수 있다.

$$f_{j(rpls)}^{(m)} = \frac{g_j' U_r' y}{(1 + c/\delta_j^2) u_j' y}, \quad j = 1, 2, \dots, r$$

여기서 g_j' 은 $G_{(m)}$ 의 j 번째 행벡터, $G_{(m)} = Q_m (Q_m' Q_m)^{-1} Q_m'$, $Q_m = (D_r^2 U_r' y, D_r^4 U_r' y, \dots, D_r^{2m} U_r' y)$ 이고, u_j 는 U_r 의 j 번째 열, δ_j 는 X 의 j 번째 비정칙값, c 는 능형상수이다.

증명. 다음의 RPLS 추정량

$$\hat{\beta}_{rpls}^{(m)} = V_r D_r^{-1} (I_r + c D_r^{-2})^{-1} G_{(m)} U_r' y$$

에서 시작한다. $D_r^{-1} (I_r + c D_r^{-2})^{-1}$ 은 j 번째 대각원소가 $\delta_j/(\delta_j^2+c)$ 인 $r \times r$ 대각행렬이고, $G_{(m)}$ 을 r 개의 행벡터 g_1', \dots, g_r' 으로 이루어진 분할행렬로 쓰면 $G_{(m)} U_r' y$ 는 j 번째 원소가 $g_j' U_r' y$ 인 $r \times 1$ 의 벡터로 나타내어진다. 이 둘의 곱 $D_r^{-1} (I_r + c D_r^{-2})^{-1} G_{(m)} U_r' y$ 는 j 번째 원소가 $g_j' U_r' y/(\delta_j + c/\delta_j)$ 인 $r \times 1$ 벡터인데, 이 벡터는 다시 j 번째 대각원소가 $\frac{g_j' U_r' y}{(\delta_j + c/\delta_j) u_j' y}$ 인 $r \times r$ 대각행렬과 j 번째 원소가

$u_j'y$ 인 $r \times 1$ 벡터의 곱으로 나타낼 수 있다. 전자 대각행렬은 다시 j 번째 대각원소가 $\frac{g_j' U_r' y}{(1+c/\delta_j^2) u_j' y}$ 인 $r \times r$ 대각행렬과 j 번째 대각원소가 $1/\delta_j$ 인 $r \times r$ 대각행렬의 곱으로 쓸 수 있다. 따라서 추정량은 2절에서 보인 $V_r F_{(.)} D_r^{-1} U_r' y$ 의 형태가 되며 증명된다.

RPLS 추정량의 다른 형태는

$$\hat{\beta}_{rpls}^{(m)} = V_r (I_r + c D_r^{-2})^{-1} V_r' W_m (W_m' X' X W_m)^{-1} W_m' X' y$$

이다. 이것은 다음과 같이 바로 확인된다.

$$\begin{aligned} \hat{\beta}_{rpls}^{(m)} &= V_r D_r^{-1} (I_r + c D_r^{-2})^{-1} G_{(m)} U_r' y \\ &= V_r (I_r + c D_r^{-2})^{-1} V_r' V_r D_r^{-1} G_{(m)} U_r' y \\ &= V_r (I_r + c D_r^{-2})^{-1} V_r' W_m (W_m' X' X W_m)^{-1} W_m' X' y \end{aligned}$$

이 추정량 식에서 다음의 축소인자의 식을 얻는다.

정리 4.3. m 개 인자의 RPLS의 축소인자의 식은 다음과 같이 쓸 수 있다.

$$f_{j(rpls)}^{(m)} = \frac{1}{\delta_j^2 + c} \sum_{k=1}^m \delta_j^{2(k+1)} b_k, \quad j = 1, 2, \dots, r$$

여기서 δ_j 는 행렬 X 의 j 번째 비정칙값, c 는 능형상수, b_k 는 $m \times 1$ 벡터 $b = (K_m' X' X K_m)^{-1} K_m' X' y$ 의 k 번째 원소이며 K_m 은 Krylov 행렬 $K_m(X' X, X' y) = (X' y, (X' X) X' y, \dots, (X' X)^{m-1} X' y)$ 이다.

증명. 다음의 RPLS 추정량

$$\hat{\beta}_{rpls}^{(m)} = V_r (I_r + c D_r^{-2})^{-1} V_r' W_m (W_m' X' X W_m)^{-1} W_m' X' y$$

에서 시작한다. Krylov 행렬 $K_m = [X' y, (X' X) X' y, \dots, (X' X)^{m-1} X' y]$ 의 QR-분해는 $K_m = W_m R$ 로 쓸 수 있으며 여기서 R 은 상삼각행렬이며 정칙이다(Di Ruscio, 2000). 따라서 $W_m = K_m R^{-1}$ 로 쓸 수 있다. 이것을 이용하면 $W_m (W_m' X' X W_m)^{-1} W_m' X' y$ 는 $K_m (K_m' X' X K_m)^{-1} K_m' X' y$ 임이 바로 확인된다. 이제 $b \equiv (K_m' X' X K_m)^{-1} K_m' X' y$ 로 두면

$$\hat{\beta}_{rpls}^{(m)} = V_r (I_r + cD_r^{-2})^{-1} V_r' K_m b$$

이다. Krylov 행렬 $K_m = (V_r D_r U_r y, V_r D_r^3 U_r y, \dots, V_r D_r^{2m-1} U_r y)$ 이므로

$$V_r' K_m = (D_r U_r y, D_r^3 U_r y, \dots, D_r^{2m-1} U_r y)$$

이다. 그리고 $(I_r + cD_r^{-2})^{-1}$ 은 j 번째 대각원소가 $\delta_j^2 / (\delta_j^2 + c)$ 인 $r \times r$ 대각행렬이고 $V_r' K_m$ 은 j 번째 행이 $(\delta_j u_j' y, \delta_j^3 u_j' y, \dots, \delta_j^{2m-1} u_j' y)$ 인 $r \times m$ 행렬이다. 따라서 $(I_r + cD_r^{-2})^{-1} V_r' K_m b$ 는 j 번째 행이

$$\left(\frac{1}{\delta_j^2 + c} \delta_j^3 u_j' y, \frac{1}{\delta_j^2 + c} \delta_j^5 u_j' y, \dots, \frac{1}{\delta_j^2 + c} \delta_j^{2m+1} u_j' y \right) b$$

인 $r \times 1$ 벡터이며 이것은

$$\frac{1}{\delta_j^2 + c} (\delta_j^4, \delta_j^6, \dots, \delta_j^{2(m+1)}) b u_j' y / \delta_j$$

로 쓸 수 있다. 그러므로 $(I_r + cD_r^{-2})^{-1} V_r' K_m b$ 는 j 번째 대각원소가 $\frac{1}{\delta_j^2 + c} \sum_{k=1}^m \delta_j^{2(k+1)} b_k$ 인 $r \times r$ 대각행렬과 j 번째 원소가 $\hat{\alpha}_j = u_j' y / \delta_j$ 인 $r \times 1$ 벡터의 곱으로 쓸 수 있다. 여기서 b_k 는 벡터 b 의 k 번째 원소이다. 이제 RPLS 추정량은 $V_r F_{(j)} \hat{\alpha}$ 의 형태가 되며 따라서 증명된다.

정리 4.2의 축소인자와 정리 4.3의 축소인자는 동일함을 보일 수 있다. 이들 축소인자 식을 보면 RPCR이 PCR에 하는 것과 같은 방식의 축소가 이루어짐을 알 수 있다. 즉 RPLS의 축소인자는 PLS를 각 고유방향으로 $\delta_j^2 / (\delta_j^2 + c)$ 의 비율로 더 축소한다. 고유값이 큰 고유방향은 그 축소량이 미미하나 고유값이 작은 고유방향은 축소량이 커지게 하는 역할을 한다.

5. 수치 예

수치 예로서 Naes(1989)의 근적외선 분광 데이터 셋을 이용하였다. 이것은 균질 고기의 자료이며 비교적 넓은 범위의 실제 샘플을 다룬 것이다. 각 샘플(관측값)은 미리 지정된 19개의 NIR 파장을 $\log(1/\text{반사율})$ 로 변환한 것과 단백질의 비율(%)을 측정 한 것으로 구성되어 있으며 샘플수는 25이다. 고기에서의 다른 주요 성분인 지방과 물은 각각 (2%–25%)와 (60%–80%)의 구간에 있다. $\log(1/\text{반사율})$ 은 근적외선 분석에서 주로 사용되는 것인데, 가정된 회귀모형의 선형성이 잘 맞는 것이 확인되었다.

모든 변수는 사전 처리로서 단위길이표준화(상관변환)를 하였다. 이 변환된 X 의 비정칙값은 다음과 같다.

(4.3404995, 0.3505433, 0.1816692, 0.0557137, 0.0254445, 0.0157632, 0.0094138, 0.0054879, 0.0041220, 0.0038493, 0.0031614, 0.0025514, 0.0022940, 0.0017561, 0.0012941, 0.0010143, 0.0007900, 0.0007207, 0.0004674)

이것의 상태수를 구해보면 $\phi = \delta_{\max} / \delta_{\min} = 4.3404995 / 0.0004674 = 9286$ 으로 매우 높은 다중공선성이 존재함을 알 수 있다.

RR, PCR, RPCR, PLS, RPLS 추정량의 축소인자 값을 인자수가 5, 8, 11, 14, 17인 다섯 경우를 <그림 1>~<그림 5>에 각각 보였다(RR은 물론 다섯 경우에 동일하다). 여기서 RR의 편향상수로는 $c = r\hat{\sigma}^2 / \hat{\beta}_{ols}'\hat{\beta}_{ols}$ 에서 얻은 $c = 0.00000753$ 을 사용하였다. 또한 각 추정량의 놈의 OLS 추정량의 놈에 대한 비를 $\rho_{(\cdot)}$ 라 하면

$$\rho_{(\cdot)} = \|\hat{\beta}_{(\cdot)}\| / \|\hat{\beta}_{ols}\| = \sqrt{\hat{\alpha}' F_{(\cdot)} \hat{\alpha}} / \sqrt{\hat{\alpha}' \hat{\alpha}}$$

이며, 여기서 $F_{(\cdot)}$ 와 $\hat{\alpha}$ 은 2절에서 정의된 $F_{(\cdot)} = \text{diag}(f_{1(\cdot)}, \dots, f_{r(\cdot)})'$ 과 $\hat{\alpha} = (\hat{\alpha}_1, \dots, \hat{\alpha}_r)'$ 이다. <그림 6>에는 각 추정량과 OLS 추정량의 놈의 비 $\rho_{(\cdot)}$ 를 인자수에 따라 보았다.

Kim(2007)의 수치 예에서 같은 데이터 셋으로써 위의 회귀방법들의 예측력을 PRESS 기준으로 비교하였는데 이것을 요약하면 다음과 같다. 인자수가 5와 6인 PLS와 RPLS, 그리고 인자수가 6, 8, 9인 PCR과 RPCR이 가장 좋은 예측력을 보였다. 그러나 인자수가 커짐에 따라 PLS와 PCR의 예측력은 많이 떨어지는 반면 RPLS와 RPCR은 여전히 괜찮은 예측력을 보였다. 따라서 RPLS와 RPCR은 작은 수의 인자에서 뿐만 아니라 과대적합 시에도 안정적인 좋은 회귀방법으로 판단되었다. RR의 예측력은 PCR, RPCR, PLS, RPLS가 최상일 때보다는 약간 떨어지나 그 차이는 크지 않았다. 그러나 OLS는 매우 떨어지는 예측력을 보였다.

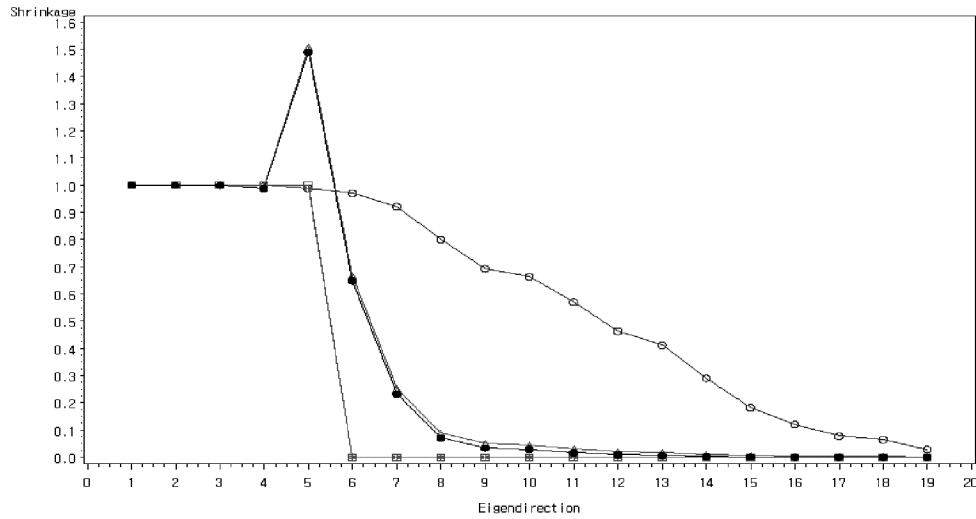
이제 <그림 1>~<그림 5>에서 먼저 PCR과 RPCR의 축소인자 값을 비교하면 다음과 같다. (참고로, 그림에서 선의 색깔의 구분은 blue는 RR, green은 PCR, red는 RPCR, brown은 PLS, black은 RPLS이다.) PCR은 고유값이 큰 소수의 고유방향은 그대로 두고 나머지는 100% 축소하는 방식이라 할 수 있다. RPCR은 정리 4.1에서 본 바와 같이 PCR의 각 고유방향으로 $\delta_j^2 / (\delta_j^2 + c)$ 의 비율로 추가 축소한다. <그림 1>의 인자수가 5인 경우는 RPCR은 PCR과 거의 같다. 이때의 인자는 고유값 δ_j^2 이 크므로 작은 값의 c 는 영향이 거의 없기 때문이다. 그러나 <그림 2>에서 <그림 5>로 가면서(인자수가 커지면서) RPCR은 PCR이 선택한 고유방향을 점점 더 많이 축소하는데 고유값이 작을수록 그 방향의 축소량은 커지게 된다. PCR의 모든 인자가 사용되는 경우는 OLS이며 이때의 RPCR은 바로 RR이다.

PLS와 RPLS의 축소인자 값을 보면 다음과 같다. PLS는 선택된 인자수 근처 앞뒤에서 일부 팽창이 일어남을 알 수 있으며 그 이후 급속히 축소하는 형태를 보인다. RPLS는 RPCR이 PCR을 축소하는 방식과 마찬가지로 각 고유방향으로 PLS를 $\delta_j^2 / (\delta_j^2 + c)$ 의 비율로 추가 축소한다. 따라서 인자수가 커질수록 RPLS의 PLS에 대한 추가 축소가 커지게 된다. 인자수가 5인 <그림 1>에서는 추가 축소는 미미하나 인자

수가 8 이상인 <그림 2>부터는 추가 축소량이 점점 커짐을 알 수 있다. 따라서 RPLS는 PLS의 결점으로 간주되는 일부 팽창하는($f_{j(pls)} > 1$) 문제를 완화하는 역할도 하는데 <그림 2>~<그림 5>에서 보면 그 완화역할이 실제 큼을 알 수 있다. 인자가 모두 다 사용되면 PLS는 OLS이며 RPLS는 RR이다.

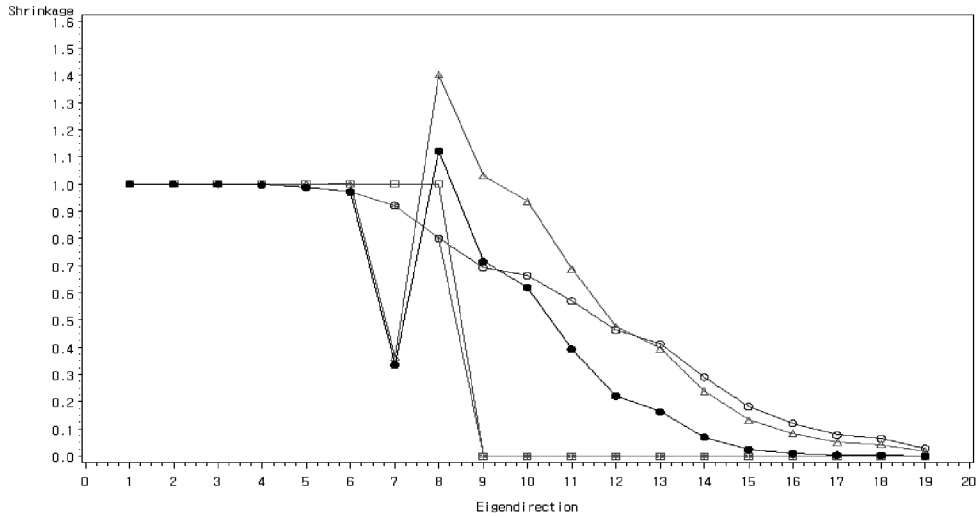
각 회귀 추정량 높의 OLS 추정량의 높에 대한 비 $\rho_{(.)}$ 는 <그림 6>에 있다. 여기서 RPCR과 RPLS의 축소 특성이 잘 나타난다. PCR과 PLS의 높 비는 인자수가 각각 11과 7인 경우에 RR과 비슷해지며 그 후 계속 증가하여 1에 접근한다. 그러나 RPCR과 RPLS의 높의 비는 인자수가 각각 13과 9에서 RR에 근접해지며 그 이후는 사실상 RR과 같아진다.

Shrinkage factor values for RR, PCR(5), RPCR(5), PLS(5), RPLS(5)



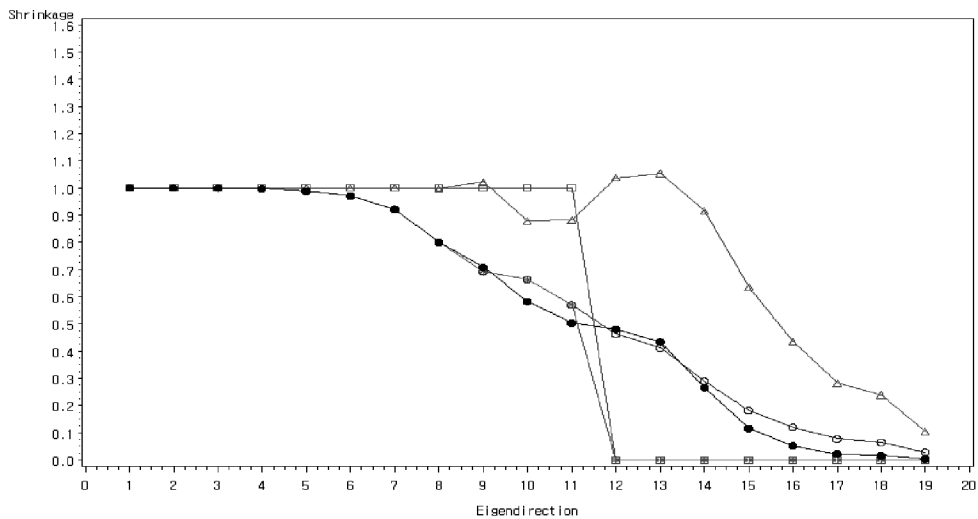
<그림 1> 인자수가 5일 때의 RR, PCR, RPCR, PLS, RPLS 추정량의 고유방향에 따른 축소인자 값. 원 ○은 RR, 정사각형 □은 PCR, 별표 *는 RPCR, 삼각형 △은 PLS, 점 ●은 RPLS의 값을 나타낸다.

Shrinkage factor values for RR, PCR(8), RPCR(8), PLS(8), RPLS(8)



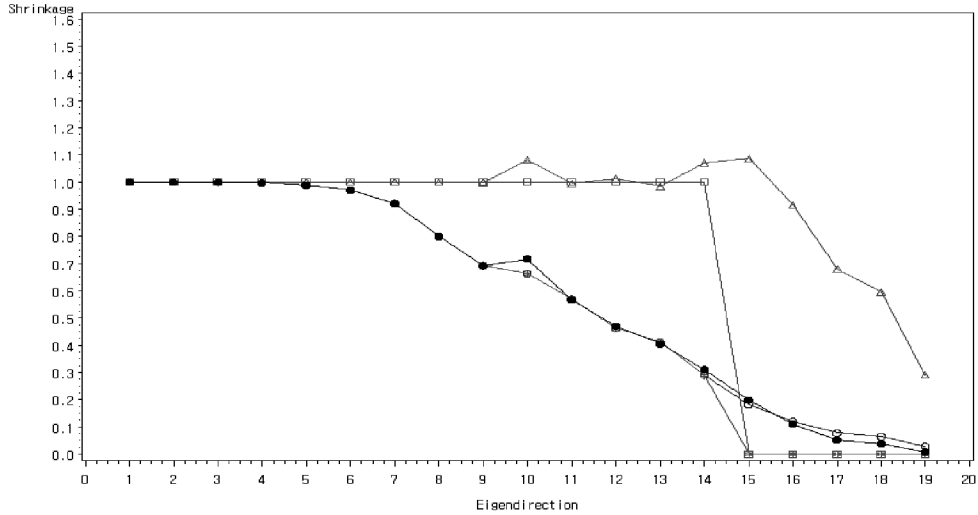
<그림 2> 인자수가 8일 때의 RR, PCR, RPCR, PLS, RPLS 추정량의 고유방향에 따른 축소인자 값. 원 ○은 RR, 정사각형 □은 PCR, 별표 *는 RPCR, 삼각형 △은 PLS, 점 •은 RPLS의 값을 나타낸다.

Shrinkage factor values for RR, PCR(11), RPCR(11), PLS(11), RPLS(11)



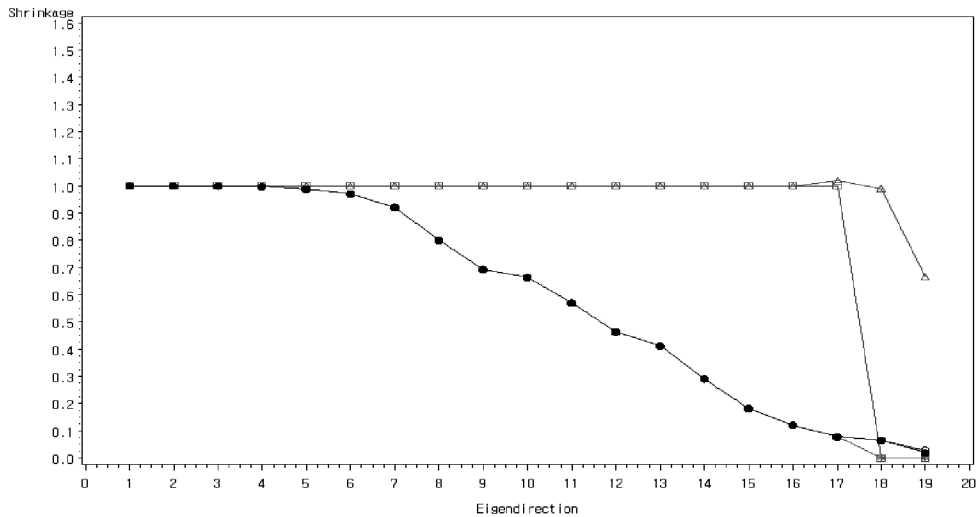
<그림 3> 인자수가 11일 때의 RR, PCR, RPCR, PLS, RPLS 추정량의 고유방향에 따른 축소인자 값. 원 ○은 RR, 정사각형 □은 PCR, 별표 *는 RPCR, 삼각형 △은 PLS, 점 •은 RPLS의 값을 나타낸다.

Shrinkage factor values for RR, PCR(14), RPCR(14), PLS(14), RPLS(14)

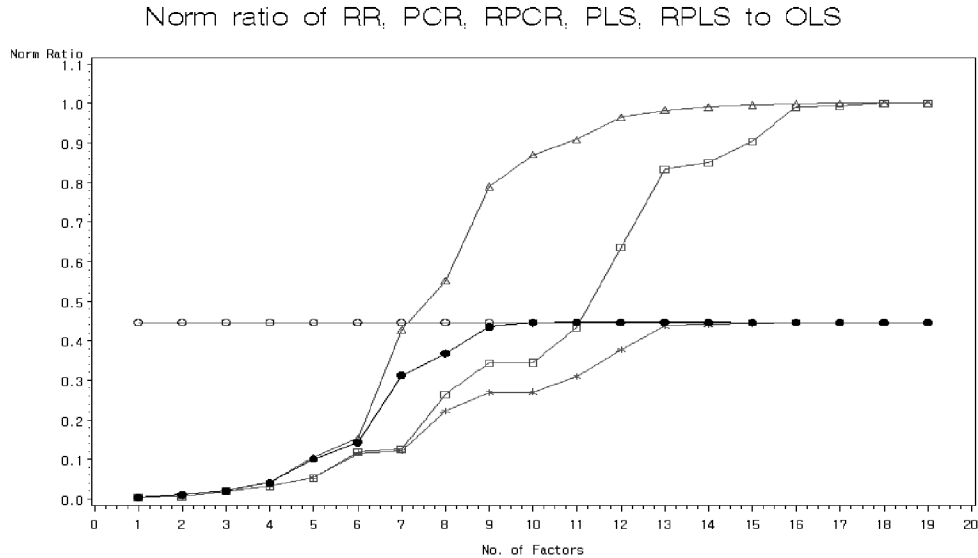


<그림 4> 인자수가 14일 때의 RR, PCR, RPCR, PLS, RPLS 추정량의 고유방향에 따른 축소인자 값. 원 ○은 RR, 정사각형 □은 PCR, 별표 *는 RPCR, 삼각형 △은 PLS, 점 ●은 RPLS의 값을 나타낸다.

Shrinkage factor values for RR, PCR(17), RPCR(17), PLS(17), RPLS(17)



<그림 5> 인자수가 17일 때의 RR, PCR, RPCR, PLS, RPLS 추정량의 고유방향에 따른 축소인자 값. 원 ○은 RR, 정사각형 □은 PCR, 별표 *는 RPCR, 삼각형 △은 PLS, 점 ●은 RPLS의 값을 나타낸다.



<그림 6> RR, PCR, RPCR, PLS, RPLS 추정량과 OLS 추정량의 놈의 비. 원 ○은 RR, 정사각형 □은 PCR, 별표 *는 RPCR, 삼각형 △은 PLS, 점 ●은 RPLS를 나타낸다.

5. 요약

다중공선성의 데이터에 사용되는 대표적인 편향회귀방법은 능형회귀(RR), 주성분회귀(PCR), 부분최소제곱회귀(PLS) 등이다. 이 회귀방법들은 계수벡터 추정량의 놈(norm)이 모두 보통 최소제곱회귀(OLS)의 추정량의 놈보다 작아진다는 의미에서 축소회귀라 부른다. 새로운 회귀방법으로 RR과 PCR을 결합한 능형주성분회귀(RPCR)가 있고 RR과 PLS를 결합한 능형부분최소제곱회귀(RPLS)가 있으며 이들도 또한 축소회귀이다. 이들 추정량은 $X'X$ 의 고유벡터들의 선형결합으로 나타낼 수 있고 따라서 각 고유방향에서 OLS에 비해 얼마나 축소되는지를 연구할 수 있다. 본 논문에서는 먼저 이들 추정량을 일반적인 축소인자의 식으로 나타내고 이를 이용하여 MSE의 일반식을 구하였으며 PLS 추정량의 MSE 식도 구하였다. 그리고 RPLS의 축소인자 식을 두 가지 다른 형태로 유도하였다. RPLS의 경우도 이 축소인자 식을 MSE의 일반식에 대입하면 MSE 식이 바로 얻어진다. 그러나 PLS나 RPLS의 축소인자는 y 의 복잡한 비선형이 되어 결정적이 아니므로 이들 추정량의 MSE는 근사적인 식이라 할 수 있다. 따라서 PLS나 RPLS를 평가하기 위해 이 MSE를 사용하는 것은 제한적이며, 경험적인 방법으로 이들 회귀의 수행성을 평가하는 것이 필요하다. 다중공선성의 대표적인 데이터인 근적외선 분광 데이터를 이용하여 이 유도된 회귀의 축소인자 값이 인자수에 따라 어떻게 변화하는지와 전체적인 축소 비율도 살펴보았다. 이들의 축소 형태를 잘 이해하면 회귀방법들의 예측력과 안정성을 파악하는데 많은 도움이 되리라 판단된다.

참고 문헌

1. Butler, N. and Denham, M. (2000). The Peculiar Shrinkage Properties of Partial Least Squares Regression, *Journal of Royal Statistical Society*, B, 62, Part 3, 585-593.
2. Di Ruscio, D. (2000). A Weighted View on the Partial Least Squares Algorithm, *Automatica*, 36, 831-850.
3. Frank, I. E. & Friedman, J. H. (1993). A Statistical View of Some Chemometrics Regression Tools (with Discussion), *Technometrics*, 35, 109-148.
4. Goutis, C. (1996). Partial Least Squares Algorithm Yields Shrinkage Estimators, *Annals of Statistics*, 24, 816-824.
5. Helland, I. S. (1988). On the structure of partial least squares regression, *Communications in Statistics - Simulation and Computation*, 17, 581-607.
6. Hoy, M., Westad, F., and Martens, H. (2002). Combining bilinear modelling and ridge regression, *Journal of Chemometrics*, 16, 313-318.
7. Kim, J. D. and Moon, S. (2005). Connecting Partial Least Squares and Generalized Ridge Regression, *Journal of the Korean Data Analysis Society*, 7, 61-71.
8. Kim, J. D. (2006). Expressions for Shrinkage Factors of PLS Estimator, *Journal of Korean Data & Information Science Society*, 17, 1169-1180.
9. Kim, J. D. (2007). Combining Ridge Regression and Latent Variable Regression, *Journal of Korean Data & Information Science Society*, 18, 51-61.
10. Lingjaerde, O. and Christopherson, N. (2000). Shrinkage Structure of Partial Least Squares, *The Scandinavian Journal of Statistics*, 27, 459-473.
11. Naes, T. (1989). Leverage and Influence Measures for Principal Component Regression. *Chemometrics and Intelligent Laboratory Systems*, 5, 155-168.
12. Vigneau, E., Devaux, M. F., Qannari, E. M., and Robert, P. (1997). Principal Component Regression, Ridge Regression and Ridge Principal Component Regression in Spectroscopy Calibration, *Journal of Chemometrics*, 11, 239-249.
13. Wold, H. (1975). Soft Modelling by Latent Variables; the Nonlinear Iterative Partial Least Squares Approach, in *Perspectives in Probability and Statistics*. Papers in Honour of M. S. Bartlett (ed. J. Gani). Academic Press, New York.

[2007년 4월 접수, 2007년 5월 채택]