17.1) 식 (17-2) 의 $E[L^2(k)] \le E[L^2]$ 을 만족시키는 $k>0$ 의 값이 항상 존재함을 증명하라.

$$Var(b_i) = \frac{\sigma^2}{1-r_{12}^2}, \quad i=1,2,\cdots \qquad (17\text{-}2)$$

변수들은 모두 표준화된 변수이다. 즉:

$\sum y_i = \sum x_{1i} = \sum x_{2i} = 0$

$\sum y_i^2 = \sum x_{1i}^2 = \sum x_{2i}^2 = 1$

이런 경우에는 절편항 $\beta_0$가 필요 없음을 유의하여 주길 바란다.

$$X'X = \begin{bmatrix} \sum x_{1i}^2 & \sum x_{1i}x_{2i} \\ \sum x_{2i}x_{1i} & \sum x_{2i}^2 \end{bmatrix} = \begin{bmatrix} 1 & r_{12} \\ r_{21} & 1 \end{bmatrix}$$

$$X'y = \begin{bmatrix} \sum x_{1i}y_i \\ \sum x_{2i}y_i \end{bmatrix} = \begin{bmatrix} r_{1y} \\ r_{2y} \end{bmatrix}$$

$r_{12}$ : 변수 $x_1$과 $x_2$ 사이의 표본상관계수

$r_{iy}$ : 변수 $x_i$와 $y$간의 표본상관계수

$E(L^2) = E[(b-\beta)'(b-\beta)]$

$b : (X'X)^{-1}X'y$ (OLS estimator)

$b(k) = (X'X + kI)^{-1}X'y$, $k>0$ (Ridge estimator)

} Start here

$$E(L^2(k)) = E[(b(k)-\beta)'(b(k)-\beta)] = var(\hat{\beta}(k)) + bias(\hat{\beta}(k))^2$$
$$= \sigma^2 tr[(X'X+kI)^{-1}X'X(X'X+kI)^{-1}] + k^2\beta'(X'X+kI)^{-2}\beta$$
$$= \sigma^2 \sum_{i=1}^{p} \lambda_i(\lambda_i+k)^{-2} + k^2\beta'(X'X+kI)^{-2}\beta$$

$$E[L^2] = E[(b-\beta)'(b-\beta)] = E[trace((b-\beta)'(b-\beta))]$$
$$= E[trace((b-\beta)(b-\beta)')]$$
$$= trace(E[(b-\beta)(b-\beta)'])$$
$$= trace(Var(b))$$
$$= \sum_{i=1}^{p} Var(b_i)$$
$$= \sigma^2 \sum_{i=1}^{p} \lambda_i^{-1}$$

* $b = (X^TX)^{-1}X^Ty$

$Var(b) = \sigma^2(X^TX)^{-1} = \sigma^2 P\Lambda^{-1}P^T$, $\Lambda = diag(\lambda_1,\cdots,\lambda_p)$

$trace(Var(b)) = \sigma^2 trace(\Lambda^{-1}P^TP)$

$= \sigma^2 trace(\Lambda^{-1}) = \sigma^2 \sum_{i=1}^{p} \lambda_i^{-1}$

$$E(L^2(k)) = \sigma^2 \sum_{i=1}^{p} \lambda_i(\lambda_i+k)^{-2} + k^2\beta'(X'X+kI)^{-2}\beta$$
$$= \sigma^2 \sum_{i=1}^{p} \lambda_i(\lambda_i+k)^{-2} + k^2\beta'(P\Lambda P^T + kPP^T)^{-2}\beta$$
$$= \sigma^2 \sum_{i=1}^{p} \lambda_i(\lambda_i+k)^{-2} + k^2\beta'P(\Lambda+kI)^{-2}P^T\beta$$
$$= \sigma^2 \sum_{i=1}^{p} \frac{\lambda_i}{(\lambda_i+k)^2} + k^2 \sum_{i=1}^{p} \frac{1}{(\lambda_i+k)^2}a_i^2, \quad P^T\beta = P^T[\beta_1\cdots\beta_p]^T = [a_1\cdots a_p]^T$$

$$\frac{\partial}{\partial k}E[L^2(k)] = \frac{\partial}{\partial k}\sigma^2 \sum_{i=1}^{p} \frac{\lambda_i}{(\lambda_i+k)^2} + k^2 \sum_{i=1}^{p} \frac{1}{(\lambda_i+k)^2}a_i^2$$
$$= -2\sigma^2 \sum_{i}^{p} \frac{\lambda_i}{(\lambda_i+k)^3} + 2k \sum_{i}^{p} \frac{1}{(\lambda_i+k)^2}a_i^2 - 2k^2 \sum_{i}^{p} \frac{1}{(\lambda_i+k)^3}a_i^2$$

$$\therefore \frac{\partial}{\partial k}E[L^2(k)]\Big|_{k=0} = -2\sigma^2 \sum_{i}^{p} \frac{1}{\lambda_i^2} < 0$$

$\sigma^2 > 0$ 이고 $\lambda_i > 0$ 이므로, $k=0$ 에서 MSE 의 기울기는 "엄격한 음수 (Strictly negative)" 입니다.

함수가 $k=0$ 에선 연속이고 미분가능하며 기울기가 음수라는 것은, 아주 작은 양수 $k$를 증가시키는 초기구간에서는 편향(bias)이 증가하는 속도보다 분산이 감소하는 속도가 압도적으로 빠르다는 것을 의미합니다.

따라서 평균값의 정리 (Mean Value Theorem) 또는 테일러 전개 (Taylor expansion)에 의해, $E[L^2(k)] < E[L^2(0)] = E[L^2]$를 만족하는 엄격한 양수 $k>0$가 항상 존재합니다.

17.2) 능형회귀(Ridge Regression)의 총평균제곱오차 $E[L^2(k)]$가 식 (17.17)과 같이 됨을 상세히 밝히고 그의 미분이

$$\frac{dE[L^2(k)]}{dk} = -2\sigma^2\sum_{i=1}^{p}\frac{\lambda_i}{(\lambda_i+k)^3} + 2k\sum_{i=1}^{p}\frac{\lambda_i a_i^2}{(\lambda_i+k)^3}$$

이 됨을 증명하라.

* $y = X\beta+\epsilon = XPP^T\beta+\epsilon = (XP)(P^T\beta)+\epsilon$
$= Z\alpha+\epsilon$

$Z = XP$, $\alpha = P^T\beta$, $\hat{\alpha} = (Z'Z)^{-1}Z'y = (P'X'XP)^{-1}P'X'y = P'P\Lambda^{-1}P'P\cdot P'X'y$

$\hat{\alpha}(k) = (Z'Z+kI_n)^{-1}Z'y = (P'X'XP+kI_n)^{-1}P'X'y$
$= P'P(\Lambda+kI_n)P'P\,P'X'y$
$= (\Lambda+kI_n)P'X'y = (\Lambda+kI_n)\hat{Z^Ty}$

$Z'Z = (XP)'XP = P'X'XP = P'P\Lambda P'P$
$= \Lambda$

$E[L^2(k)] = E[(\hat{\alpha}(k)-\alpha)'(\hat{\alpha}(k)-\alpha)]$
$= tr\{Var[\hat{\alpha}(k)]\} + bias(\hat{\alpha})^2$
$= \sigma^2 tr[\Lambda(\Lambda+kI)^{-2}] + \hat{\alpha}'[(\Lambda+kI)^{-1}\Lambda-I]'[(\Lambda+kI)^{-1}\Lambda-I]\hat{\alpha}$

$= \sigma^2\sum_{i=1}^{p}\frac{\lambda_i}{(\lambda_i+k)^2} + \sum_{i=1}^{p}\frac{k^2a_i^2}{(\lambda_i+k)^2}$ … (17.17)

$\Rightarrow E[L^2(k)] = E[(b(k)-\beta)'(b(k)-\beta)]$

$b(k) = (X^TX+kI_p)^{-1}X^Ty$
$= (P\Lambda P^T+kPP^T)^{-1}X^Ty$
$= (P(\Lambda+kI_p)P^T)^{-1}X^Ty$
$= P(\Lambda+kI_p)^{-1}P^TX^Ty$
$= P(\Lambda+kI_p)^{-1}Z^Ty$ $(\because Z = XP)$

$E[(b(k)-\beta)'(b(k)-\beta)] = E[(Pb(k)-P\beta)'(Pb(k)-P\beta)]$
$= E[(\hat{\alpha}(k)-\alpha)'(\hat{\alpha}(k)-\alpha)]$
$= tr[(\hat{\alpha}(k)-\alpha)'(\hat{\alpha}(k)-\alpha)]$
$= tr[(\hat{\alpha}(k)-\alpha)(\hat{\alpha}(k)-\alpha)']$

$= tr\{[\hat{\alpha}(k)-E[\hat{\alpha}(k)]+E[\hat{\alpha}(k)]-\alpha][\hat{\alpha}(k)-E[\hat{\alpha}(k)]+E[\hat{\alpha}(k)]-\alpha]^T\}$
$= tr\{(\hat{\alpha}(k)-E[\hat{\alpha}(k)])(\hat{\alpha}(k)-E[\hat{\alpha}(k)])'\}$ … ①
$+ tr\{(\hat{\alpha}(k)-E[\hat{\alpha}(k)])(E[\hat{\alpha}(k)]-\alpha)'\}$ … ②
$+ tr\{(E[\hat{\alpha}(k)]-\alpha)(E[\hat{\alpha}(k)]-\alpha)\}$ … ③

① … $Var(\hat{\alpha}(k))$
② … 0
③ … $bias(\hat{\alpha}(k))^2$

$\therefore E[L^2(k)] = Var(\hat{\alpha}(k)) + bias(\hat{\alpha}(k))^2$
$= \underbrace{Var((Z'Z+kI_p)^{-1}Z^Ty)}_{①} + \underbrace{\{E[\hat{\alpha}(k)]-\alpha\}^2}_{②}$

① : $(Z'Z+kI_p)^{-1}Z^TZ(Z'Z+kI_p)^{-1}\sigma^2$
$(\Lambda+kI_p)^{-1}\Lambda(\Lambda+kI_p)^{-1}\sigma^2$

$= \sigma^2\sum_{i}^{p}\frac{\lambda_i}{(\lambda_i+k)^2}$

② $\{E[\hat{\alpha}(k)]-\alpha\}^2$

$= \{E[(Z'Z+kI_p)^{-1}Z^Ty]-\alpha\}^2$
$= \{(Z'Z+kI_p)^{-1}Z^TX\beta-\alpha\}^2$
$= \{(\Lambda+kI_p)^{-1}P^TX^TX\beta-\alpha\}^2$
$= \{(\Lambda+kI_p)^{-1}P^TP\Lambda P^T\beta-\alpha\}^2$
$= \{(\Lambda+kI_p)^{-1}\Lambda\alpha-\alpha\}^2$
$= [\{(\Lambda+kI_p)^{-1}\Lambda-I_p\}\alpha]^2$

$= \sum_{i=1}^{p}\left\{\frac{\lambda_i-\lambda_i-k}{\lambda_i+k}\right\}^2a_i^2$
$= \sum_{i=1}^{p}\frac{k^2a_i^2}{(\lambda_i+k)^2}$

$\therefore$ ① + ② $= \sigma^2\sum_{i}^{p}\frac{\lambda_i}{(\lambda_i+k)^2} + \sum_{i=1}^{p}\frac{k^2a_i^2}{(\lambda_i+k)^2}$

$\therefore E[L^2(k)] = \sigma^2\sum_{i}^{p}\frac{\lambda_i}{(\lambda_i+k)^2} + \sum_{i=1}^{p}\frac{k^2a_i^2}{(\lambda_i+k)^2}$

$\frac{\partial}{\partial k}E[L^2(k)] = \sigma^2\sum_{i}^{p}-2\frac{\lambda_i}{(\lambda_i+k)^3} + \sum_{i}^{p}\frac{2ka_i^2}{(\lambda_i+k)^2} + \sum_{i}^{p}-2\frac{k^2a_i^2}{(\lambda_i+k)^3}$
$= -2\sum_{i}^{p}\frac{\lambda_i\cdot\sigma^2}{(\lambda_i+k)^3} + \sum_{i}^{p}\frac{2ka_i^2(\lambda_i+k)-2k^2a_i^2}{(\lambda_i+k)^3}$
$= -2\sum_{i}^{p}\frac{\lambda_i\sigma^2}{(\lambda_i+k)^3} + \sum_{i}^{p}\frac{2k\lambda_ia_i^2}{(\lambda_i+k)^3}$ □

17.3) 식 (17.19)의 일반능형회귀추정량 $b(K) = (X'X+K)^{-1}X'y$ 가 사용될 때에 총평균제곱오차 $E[L^2(K)]$가 $K_i = \frac{\sigma^2}{\alpha_i^2}$ 에서 최소가 됨을 증명하라.

$L^2(K) = (b(K)-b)'(b(K)-b)$

$E[L^2(K)] = E[(b(K)-b)'(b(K)-b)]$

$= E[tr\{(b(K)-b)(b(K)-b)'\}]$

$= tr\{E[(b(K)-b)(b(K)-b)']\}$

$= tr\{Var(b(K))\} + bias(b(K))^2$

$\frac{\partial}{\partial K} E[L^2(K)] = -2\sum_1^p \frac{\lambda_i \sigma^2}{(\lambda_i+K)^3} + \sum_1^p \frac{2K\lambda_i \alpha_i^2}{(\lambda_i+K)^3}$

$= -2\sum_1^p \frac{\lambda_i}{(\lambda_i+K)^3}(\sigma^2 - K\alpha_i^2)$

if $K = \frac{\sigma^2}{\alpha_i^2}$, $\frac{\partial}{\partial K} E[L^2(K)] = 0$

$\therefore E[L^2(K)]$ is minimum at $K = \frac{\sigma^2}{\alpha_i^2}$

---

17.4) 식 (17.27)이 성립됨을 증명하라.

$b_g = (I - P_S P_S')b$ … (17.27)

〈주성분 회귀 (Principle Component Regression)〉

$y = X\beta + \epsilon = Z\alpha + \epsilon$ … (17.22)

$\Lambda\hat{\alpha} = Z'y$ … (17.23)  } *

$\beta$의 최소제곱추정량은 : $b = P\hat{\alpha}$ … (17.24)

$b = \hat{\beta} = PP^T\hat{\beta} = P\hat{\alpha}$

$P = (P_g, P_S)$

$\alpha' = (\alpha_g', \alpha_S')$

$b_g = P_g\hat{\alpha}_g = P_g\Lambda_g P_g' X'y$

* $y = X\beta + \epsilon = XPP^T\beta + \epsilon$

$= Z\alpha + \epsilon$ $(\because XP = Z,\ P^T\beta = \alpha)$

* $\Lambda\hat{\alpha} = \Lambda P^T\hat{\beta} = \Lambda P^T(X'X)^{-1}X'y$

$= \Lambda P^T(P\Lambda P^T)^{-1}X'y$

$= \Lambda P^T P\Lambda^{-1}P^T X'y$

$= P^T X'y = (XP)'y = Z'y$

(Solution)

$P = [P_1\ P_2]$ $\quad PP^T = [P_1\ P_2]\begin{bmatrix}P_1^T\\P_2^T\end{bmatrix} = P_1P_1^T + P_2P_2^T = I_n$

$P^TP = \begin{bmatrix}P_1^T\\P_2^T\end{bmatrix}[P_1\ P_2] = \begin{bmatrix}P_1^TP_1 & P_1^TP_2\\P_2^TP_1 & P_2^TP_2\end{bmatrix} = \begin{bmatrix}I_1 & 0\\0 & I_2\end{bmatrix}$

$\hat{\alpha} = P^Tb = \begin{bmatrix}P_1^T\\P_2^T\end{bmatrix}b = \begin{bmatrix}P_1^Tb\\P_2^Tb\end{bmatrix}$

$\hat{\alpha} = \begin{bmatrix}\hat{\alpha}_1\\\hat{\alpha}_2\end{bmatrix} = \begin{bmatrix}P_1^Tb\\P_2^Tb\end{bmatrix}$ $\therefore \hat{\alpha}_1 = P_1^Tb$ (alternatively $\hat{\alpha} = (Z^TZ)^{-1}Z^Ty = (P_X^TX_C^TX_CP_X)^{-1}P_X^TX_C^Ty$ $= P_X^TP_X^TX_C^TX_CP_XP_X^TX_C^Ty = P^Tb$

we also know that $P_1\hat{\alpha} = b_1$

$\therefore b_1 = P\hat{\alpha}_1 = PP_1^Tb = P_1P_1^Tb = (I - P_1P_1^T)b$ $\quad \therefore P\hat{\alpha} = b$

---

17.5) 식 (17.28), (17.29), (17.30)을 증명하라.

(a) $E[b_g] = \beta - \sum_{i=g+1}^{p}(P_i'\beta)P_i$ … (17.28)

(b) $Var(b_g) = \sigma^2\sum_{i=1}^{g}\lambda_i^{-1}P_iP_i'$ … (17.29)

(c) $E[(b_g-\beta)'(b_g-\beta)] = \sigma^2\sum_{i=1}^{g}\lambda_i^{-1} + \sum_{i=g+1}^{p}(P_i'\beta)^2$ … (17.30)

(Solution)

(a) $b_g = P_g\hat{\alpha}_g = P_gP_g^Tb = (I - P_SP_S^T)b$

$E[b_g] = E[(I-P_SP_S^T)b] = \beta - P_SP_S^T\beta$

$P_S = [P_{g+1}\ \cdots\ P_p]$

$\therefore P_SP_S^T = P_{g+1}P_{g+1}' + \cdots + P_pP_p'$

$\therefore E[b_g] = \beta - \sum_{i=g+1}^{p}P_iP_i'\beta$ (this is scalar)

$= \beta - \sum_{i=g+1}^{p}(P_i'\beta)P_i$

(b) $Var(b_g) = Var[(I-P_SP_S')b]$

$= (I-P_SP_S')Var(b)(I-P_SP_S')^T$

$= \sigma^2(I-P_SP_S')(X^TX)^{-1}(I-P_SP_S')^T$

$= \sigma^2(I-P_SP_S')(P\Lambda P^T)^{-1}(I-P_SP_S')^T$

$= \sigma^2(I-P_SP_S')P\Lambda^{-1}P^T(I-P_SP_S')'$

$= \sigma^2(I-P_SP_S')P\Lambda^{-1}P^T(I-P_SP_S')$

$= \sigma^2\{P\Lambda^{-1}P^T - P\Lambda^{-1}P^TP_SP_S' - P_SP_S'P\Lambda^{-1}P^T + P_SP_S'P\Lambda^{-1}P^TP_SP_S'\}$

$= \sigma^2\{P\Lambda^{-1}P^T - P_S\Lambda_S^{-1}P_S^T - P_S\Lambda_S^{-1}P_S^T + P_S\Lambda^{-1}P_S'\}$

$= \sigma^2\{\sum_1^p\frac{P_iP_i^T}{\lambda_i} - \sum_{i=g+1}^{p}\frac{P_iP_i^T}{\lambda_i}\}$

$= \sigma^2\sum_{i=1}^{g}\frac{P_iP_i^T}{\lambda_i}$ □

$Var(b_g) = Var(P_gP_g^Tb)$

$= P_gP_g^TVar(b)P_gP_g^T$

$= \sigma^2P_gP_g^T(X^TX)^{-1}P_gP_g^T$

$(\cdots) = \sigma^2P_g\Lambda_g^{-1}P_g^T$

$= \sigma^2\sum_{i=1}^{g}\lambda_i^{-1}P_iP_i'$

(c) $E[(b_g-\beta)'(b_g-\beta)] = \sigma^2\sum_{i=1}^{g}\lambda_i^{-1} + \sum_{i=g+1}^{p}(P_i'\beta)^2$

$E[(b_g - E[b_g] + E[b_g] - \beta)'(b_g - E[b_g] + E[b_g] - \beta)]$

$= E[tr\{(b_g - E[b_g] + E[b_g] - \beta)'(b_g - E[b_g] + E[b_g] - \beta)\}]$

$= tr(Var(b_g)) + bias(b_g)^2$

$= tr(\sigma^2\sum_{i=1}^{g}\frac{1}{\lambda_i}P_iP_i^T) + \{\underbrace{\beta - \sum_{i=g+1}^{p}(P_i'\beta)P_i}_{P_SP_S^T\beta} - \beta\}^2$

$= \sigma^2\sum_{i=1}^{g}\frac{1}{\lambda_i}tr(P_iP_i^T) + (P_SP_S^T\beta)'(P_SP_S^T\beta)$

$= \sigma^2\sum_{i=1}^{g}\frac{1}{\lambda_i}tr(P_i^TP_i) + \beta'P_SP_S^TP_SP_S^T\beta$

$= \sigma^2\sum_{i=1}^{g}\frac{1}{\lambda_i} + \sum_{i=g+1}^{p}(P_i^T\beta)^2$ □

17.6) 제약조건이 $C\beta = 0$ 이 아니고 $C\beta = m$ 일 때에 제한 최귀추정량 $\tilde{\beta}$ 을 구하고 식 (17.34)의 편의의 크기가 무엇인가를 보여라.

$$E[\tilde{\beta}] - \beta = (X'X)^{-1}C'[C(X'X)^{-1}C']^{-1}C'\beta \quad \cdots (17.34)$$

p. 594
"이 제한조건 아래에서 $\beta$의 최소제곱추정량은

$$\tilde{\beta} = \left\{ I - (X'X)^{-1}C'[C(X'X)^{-1}C']^{-1}C \right\} b$$

가 됨을 식 (17.29)에서 증명하였다."

$C\beta = C(X'X)^{-1}X^T X\beta = 0$

let $X_r^T = C(X'X)^{-1}X^T \Leftrightarrow X_r = X(X'X)^{-1}C^T$

$\tilde{\beta} = \underset{C\beta = 0}{\operatorname{argmin}} (y - X\beta)^T(y - X\beta)$

$\hat{z} = \underset{X_r^T z = 0}{\operatorname{argmin}} (y - z)^T(y - z)$

$\therefore z \in \ker(X_r^T) = \operatorname{im}(X_r)^{\perp}$

$\therefore \hat{z} = (\Pi_X - \Pi_{X_r})y \quad (\because \text{closest vector property})$

$X_r(X_r'X_r)^{-1}X_r^T = X(X'X)^{-1}C'[C(X'X)^{-1}C']^{-1}C(X'X)^{-1}X'y$

$\therefore \tilde{\beta} = \hat{\beta} - (X'X)^{-1}C'[C(X'X)^{-1}C']^{-1}C\hat{\beta}$

If the restriction is $C\beta = m \Leftrightarrow C\beta = C\beta' \Leftrightarrow C(\beta - \beta') = 0$
$\Leftrightarrow C\beta^* = 0 \ (\beta^* = \beta - \beta')$

$\tilde{\beta}^* = \hat{\beta}^* - (X'X)^{-1}C'[C(X'X)^{-1}C']^{-1}C\hat{\beta}^*$
$\tilde{\beta} - \beta' = \hat{\beta} - \beta' - (X'X)^{-1}C'[C(X'X)^{-1}C']^{-1}C(\hat{\beta} - \beta')$

$\tilde{\beta} = \hat{\beta} - (X'X)^{-1}C'[C(X'X)^{-1}C']^{-1}\{C\hat{\beta} - m\}$ ✓✓

$\text{bias}(\tilde{\beta}) = E[\tilde{\beta}] - \beta = -(X'X)^{-1}C'[C(X'X)^{-1}C']^{-1}\{C\beta - m\}$ ✓✓

17.7) 제약조건이 $C\beta = 0$ 이 아니고 $C\beta = m$ 일 때 제한최귀추정량 $\tilde{\beta}$ 의 총평균제곱오차가 무엇인지 구하여라.

$$MSE[\tilde{\beta}] = (E[\tilde{\beta}] - \beta)^T(E[\tilde{\beta}] - \beta)$$
$$= \operatorname{tr}(\operatorname{Var}(\tilde{\beta})) + \text{Bias}(\tilde{\beta})^2$$

$\operatorname{Var}(\tilde{\beta}) = \operatorname{Var}(b - (X'X)^{-1}C'(C(X'X)^{-1}C')^{-1}(Cb - m))$
$= \operatorname{Var}((I - (X'X)^{-1}C'(C(X'X)^{-1}C')^{-1}C)b) \quad \cdots ①$

$= \left\{ I - (X'X)^{-1}C'[C(X'X)^{-1}C']^{-1}C \right\} \operatorname{Var}(b) \left\{ I - (X'X)^{-1}C'[C(X'X)^{-1}C']^{-1}C \right\}'$
$= \sigma^2 \left\{ I - (X'X)^{-1}C'[C(X'X)^{-1}C']^{-1}C \right\} (X'X)^{-1} \left\{ I - (X'X)^{-1}C'[C(X'X)^{-1}C']^{-1}C \right\}'$

$\text{Bias}(\tilde{\beta}) = (X'X)^{-1}C'(C(X'X)^{-1}C')^{-1}(m - Cb)$

$\therefore \text{Bias}(\tilde{\beta})^T\text{Bias}(\tilde{\beta}) = (m - Cb)'(C(X'X)^{-1}C')^{-1}C(X'X)^{-2}C'[C(X'X)^{-1}C']^{-1}(m - Cb) \quad \cdots ②$

$MSE(\tilde{\beta}) = \operatorname{tr}① + ②$

---

17.8) 식 (17.37)의 행렬 $M = \sigma^2 C(X'X)^{-1}C' - C\beta\beta'C'$ 이 양반정치행렬이 되기 위한 필요충분조건이 식 (17.38)의 $\beta'C'[C(X'X)^{-1}C']^{-1}C\beta \le \sigma^2$ 임을 증명하라.

$\sigma^2 C(X'X)^{-1}C' - C\beta\beta'C'$ is positive definite matrix

$C \in \mathbb{R}^{p \times (k+1)}$
$\beta \in \mathbb{R}^{(k+1)\times 1}$
$C\beta \in \mathbb{R}^p$

(Forward proof ⇒)

$\forall x \in \mathbb{R}^p,\ x^TMx > 0 \Rightarrow \sigma^2 x^TC(X'X)^{-1}C'x - x^TC\beta\beta'C'x > 0$

Let $\Sigma = C(X'X)^{-1}C'$ and $C\beta = \theta$, $\beta'C' = \theta'$
$\Sigma$ : if $(X'X)^{-1}$ is positive definite and $C$ is full rank, $\Sigma$ is also a symmetric positive definite matrix.

$x^TMx > 0 \Rightarrow \sigma^2 x^T\Sigma x - x^T\theta\cdot\theta^Tx > 0$
$\sigma^2(x^T\Sigma x) > (x^T\theta)^2$

let $x = \Sigma^{-1}\theta$
$\Rightarrow \sigma^2(\theta^T\Sigma^{-1}\Sigma\Sigma^{-1}\theta) > (\theta^T\Sigma^{-1}\theta)^2$
$\Rightarrow \sigma^2(\theta^T\Sigma^{-1}\theta) > (\theta^T\Sigma^{-1}\theta)^2$
$\Rightarrow \sigma^2 > (\theta^T\Sigma^{-1}\theta) = \beta'C'[C(X'X)^{-1}C']^{-1}C\beta$

(backward proof)
Assumption: $\theta'\Sigma^{-1}\theta \le \sigma^2$
WTS: M is positive definite. Meaning: $\forall x \in \mathbb{R}^p,\ (x'\theta)^2 \le \sigma^2 x'\Sigma x$

$x'\theta = x'I\theta = x'\Sigma^{1/2}\Sigma^{-1/2}\theta = (x'\Sigma^{1/2})(\Sigma^{-1/2}\theta)$

Let $u = \Sigma^{1/2}x$ and $v = \Sigma^{-1/2}\theta$
Then $x'\theta = u'v$.
Apply the Cauchy-Schwarz Inequality: $(u'v)^2 \le (u'u)(v'v)$

$(x'\theta)^2 \le (x'\Sigma^{1/2}\Sigma^{-1/2}x)(\theta'\Sigma^{-1/2}\Sigma^{-1/2}\theta) = (x'\Sigma x)(\theta'\Sigma^{-1}\theta)$

However, in our assumption, $\theta'\Sigma^{-1}\theta \le \sigma^2$. Thus,

$(x'\theta)^2 \le \sigma^2(x'\Sigma x)$ □

17. 식 (17.39)의 축소추정량 $\tilde{\beta}_S = cb$ 의 총평균제곱오차를 구하라.

p. 597.

... 편의추정 이외에도 Stein (17.20)이 고안한 축소추정량 (Shrunken estimator)이란 것도 있는데, 이는 최소제곱추정량 $b$를 일괄적으로 축소시킨 추정량

$$\tilde{\beta}_S = cb, \quad 0 < c < 1 \qquad (17.39)$$

이다.

$$\begin{aligned}
MSE(\tilde{\beta}_S) &= E[(\tilde{\beta}_S - \beta)^T(\tilde{\beta}_S - \beta)] \\
&= E[(\tilde{\beta}_S - E[\tilde{\beta}_S] + E[\tilde{\beta}_S] - \beta)^T(\tilde{\beta}_S - E[\tilde{\beta}_S] + E[\tilde{\beta}_S] - \beta)] \\
&= E[(\tilde{\beta}_S - E[\tilde{\beta}_S])^T(\tilde{\beta}_S - E[\tilde{\beta}_S])] + E[(E[\tilde{\beta}_S] - \beta)^T(E[\tilde{\beta}_S] - \beta)] \\
&= tr(Var(\tilde{\beta}_S)) + bias(\tilde{\beta}_S)^2 \\
&= tr(c^2 Var(b)) + (c-1)^2 \beta'\beta \\
&= \sigma^2 c^2 tr((X^TX)^{-1}) + (c-1)^2 \sum_{i=1}^{p} \beta_i^2 \\
&= \sigma^2 c^2 tr(P\Lambda^{-1}P^T) + (c-1)^2 \sum_{i=1}^{p} \beta_i^2 \\
&= \sigma^2 c^2 \sum_{i=1}^{p} \lambda_i^{-1} + (c-1)^2 \sum_{i=1}^{p} \beta_i^2
\end{aligned}$$

---

17.10) 식 (17.13)의 선형모형, $y = Z\alpha + \epsilon$ 에 대하여 능형추정량

$$\hat{\alpha}(k) = (Z'Z + kI)^{-1}Z'y$$

가 사용될 때, 어떤 임의의 점 $z' = x'P$ 에서의 반응의 추정량

$$\hat{y}_k(z) = z'\hat{\alpha}(k)$$

을 생각하여 보자. 다음을 구하라.

(1) $Var[\hat{y}_k(z)]$ (2) $E[\hat{y}_k(z)] - z'\alpha$ (3) $MSE[\hat{y}_k(z)]$

(1) $$\begin{aligned}
Var[\hat{y}_k(z)] &= Var(z'\hat{\alpha}(k)) = z' Var(\hat{\alpha}(k)) z'^T \\
&= z' Var((Z'Z + kI)^{-1}Z'y) z'^T \\
&= z'(Z'Z + kI)^{-1}Z' Var(y) Z (Z'Z + kI)^{-1} z'^T
\end{aligned}$$

$y = X\beta + \epsilon = XPP^T\beta + \epsilon \quad \therefore Z = XP, \ \alpha = P^T\beta$

$Z'Z = P'X'XP = P'P\Lambda P'P = \Lambda$

$$\begin{aligned}
Var(\hat{\alpha}(k)) &= Var((Z'Z + kI)^{-1}Z'y) \\
&= (Z'Z + kI)^{-1}Z' Var(y) Z (Z'Z + kI)^{-1} \\
&= \sigma^2 (\Lambda + kI)^{-1}\Lambda (\Lambda + kI)^{-1} \\
&= \sigma^2 \Lambda (\Lambda + kI)^{-2}
\end{aligned}$$

$\therefore z' Var(\hat{\alpha}(k)) z'^T = \sigma^2 x'P\Lambda(\Lambda + kI)^{-2}P'x$ ✓✓

(2) $E[\hat{y}_k(z)] - z'\alpha$

$\hat{y}_k(z) = z'\hat{\alpha}(k)$ , $E[\hat{y}_k(z)] = z' E[\hat{\alpha}(k)]$

$\hat{\alpha}(k) = (Z'Z + kI)^{-1}Z'y$ , $E[\hat{\alpha}(k)] = (Z'Z + kI)^{-1}Z'X\beta$

$Z$ 에서 실수 $= (\Lambda + kI)^{-1}PX'X\beta$

$= (\Lambda + kI)^{-1}P\Lambda\beta$

$\therefore z'E[\hat{\alpha}(k)] = z'(\Lambda + kI)^{-1}P\Lambda\beta$

$E[\hat{y}_k(z)] - z'\alpha = x'P(\Lambda + kI)P\Lambda\beta - x'\beta$

$= x'\{P(\Lambda + kI)P\Lambda - I\}\beta$ ✓✓

(3) $MSE[\hat{y}_k(z)]$

$= E[\{\hat{y}_k(z) - y\}\{\hat{y}_k(z) - y\}^T]$

$= Var[\hat{y}_k(z)] + bias[\hat{y}_k] \cdot bias[\hat{y}_k]^T$

(1) (2)²

$\therefore MSE[\hat{y}_k(z)] = \sigma^2 x'P\Lambda(\Lambda + kI)^{-2}P'x' + x'\{P(\Lambda + kI)P\Lambda - I\}\beta\beta^T\{P(\Lambda + kI)P\Lambda - I\}^T x'^T$

$z, x'$ 주의

17.11) 식 (17.26)의 주성분 회귀추정량

$$b_g = P_g \hat{a}_g$$

가 사용될 때 어떤 임의의 점 $x$에서의 반응의 추정량

$$\hat{y}_g(x) = x' b_g$$

를 고려하여 보자. 다음을 구하라.

(1) $\mathrm{Var}[\hat{y}_g(x)]$ (2) $E[\hat{y}_g(x)] - x'\beta$ (3) $MSE[\hat{y}_g(x)]$

(1) $X^TX = P\Lambda P^T$ $\quad P = [P_g \; P_s]$, $\Lambda = \begin{bmatrix} \Lambda_1 & 0 \\ 0 & \Lambda_2 \end{bmatrix}$

(Approach #1)

$$b = (X^TX)^{-1}X^Ty = P\Lambda^{-1}P^T \cdot X^Ty = [P_g\Lambda_1^{-1} + P_s\Lambda_2^{-1}]Z^Ty = P_g\Lambda_1^{-1}Z^Ty + P_s\Lambda_2^{-1}Z^Ty$$
$$= P_g\Lambda_1^{-1}P_g^TX^Ty + P_s\Lambda_2^{-1}P_s^TX^Ty$$

(Approach #2)

$y = Z_1a_1 + Z_1a_2 + \epsilon$

$$\hat{a}_1 = (Z_1^TZ_1)^{-1}Z_1^Ty = (P_g^TX^TXP_g)^{-1}P_g^TX^Ty$$
$$= (P_g^TP\Lambda P^TP_g)^{-1}P_g^TX^Ty$$
$$= \Lambda^{-1}P_g^TX^Ty$$

$$\hat{a}_1 = P_g^Tb_g = \Lambda_1^{-1}P_g^TX^Ty$$
$$\therefore b_g = P_g\Lambda_1^{-1}P_g^TX^Ty$$

$$\mathrm{Var}(b_g) = P_g\Lambda_1^{-1}P_g^TX^T\sigma^2I_n \times P_g\Lambda_1^{-1}P_g^T$$
$$= \sigma^2P_g\Lambda_1^{-1}P_g^T$$

$$\therefore \mathrm{Var}(x'b_g) = x'\mathrm{Var}(b_g)x'^T$$
$$= \sigma^2x'P_g\Lambda_1^{-1}P_g^Tx$$

$$= \sigma^2\sum_{i=1}^{g}\frac{(x'p_i)^2}{\lambda_i}$$

(2) $E[\hat{y}_k(x)] - x'\beta$

$$= E[x'b_g] = x'E[b_g] = x'P_g\Lambda_1^{-1}P_g^TX^TX\beta$$
$$= x'P_g\Lambda_1^{-1}P_g^TP_g\Lambda_1P_g^T\beta$$
$$= x'P_gP_g^T\beta$$

$$\therefore E[\hat{y}_k(x)] - x'\beta = x'[P_gP_g^T - I]\beta$$

(3) $MSE[\hat{y}_g(x)] = \mathrm{Var}(\hat{y}_g(x)) + \mathrm{bias}(\hat{y}_g(x))\mathrm{bias}(\hat{y}_g(x))^T$

$$= \sigma^2\sum_{i=1}^{g}\frac{(x'p_i)^2}{\lambda_i} + \left(x'P_gP_g^T\beta\right)^2$$

$$PP^T = [P_g \; P_s]\begin{bmatrix} P_g^T \\ P_s^T \end{bmatrix} = P_gP_g^T + P_sP_s^T = I_n$$
$$\therefore P_gP_g^T = I_n - P_sP_s^T$$

$$MSE[\hat{y}_g(x)] = \sigma^2\sum_{i=1}^{g}\frac{(x'p_i)^2}{\lambda_i} - \left[\sum_{i=g+1}^{p}(x'p_i)(p_i'\beta)\right]^2$$