L09 A biased estimator: Principal component estimator

1. Model, problem and remedy

   (1) Model
   In Model $Y = X\beta + \epsilon$, $\epsilon \sim (0,\, \sigma^2 I_n)$, X has full column rank. Hence β is estimable and the BLUE of β is

   $$\widehat{\beta} = X^+ Y = (X'X)^{-1}X'Y \sim (\beta,\, \sigma^2 (X'X)^{-1})$$

   with

   $$\begin{aligned}\text{MSEM}(\widehat{\beta},\, \beta) &= \text{Cov}\left(\widehat{\beta}\right) = \sigma^2 (X'X)^{-1} \text{ and}\\ \text{MSE}(\widehat{\beta},\, \beta) &= \text{tr}\left[\text{MSEM}(\widehat{\beta},\, \beta)\right] = \sigma^2 \text{tr}\left[(X'X)^{-1}\right].\end{aligned}$$

   (2) Problem
   Let $\lambda_1 \geq \cdots \geq \lambda_p > 0$ be the eigenvalues of $X'X$. Then

   $$\text{MSE}(\widehat{\beta},\, \beta) = \sigma^2 \left(\frac{1}{\lambda_1} + \cdots + \frac{1}{\lambda_p}\right).$$

   If there is multicollinearity in X, then $|X'X| = \lambda_1 \cdots \lambda_p$ is small. Consequently $\text{MSE}(\widehat{\beta},\, \beta)$ becomes large. Hence the estimator is not stable and the risk is high.

   (3) A remedy
   In $\text{MSE}(\widehat{\beta},\, \beta) = \frac{\sigma^2}{\lambda_1} + \cdots + \frac{\sigma^2}{\lambda_p}$, $\frac{\sigma^2}{\lambda_1} \leq \cdots \leq \frac{\sigma^2}{\lambda_p}$, if we keep the firs q terms and drop the rest, the $\text{MSE}(\widehat{\beta},\, \beta)$ is reduced. The resulted estimator is called a principal component estimator since $\frac{\sigma^2}{\lambda_1}, ..., \frac{\sigma^2}{\lambda_q}$ kept in the estimator are the variances of the first q principal components of $\widehat{\beta}$.

2. Expression, parameters and risk

   (1) Expression
   By EVD

   $$X'X = P\Lambda P' = (P_I,\, P_{II}) \begin{pmatrix} \Lambda_I & 0 \\ 0 & \Lambda_{II} \end{pmatrix} (P_I,\, P_{II})' = P_I \Lambda_I P_I' + P_{II} \Lambda_{II} P_{II}' \text{ and}$$
   $$(X'X)^{-1} = P\Lambda^{-1} P' = (P_I,\, P_{II}) \begin{pmatrix} \Lambda_I^{-1} & 0 \\ 0 & \Lambda_{II}^{-1} \end{pmatrix} (P_I,\, P_{II})' = P_I \Lambda_I^{-1} P_I' + P_{II} \Lambda_{II}^{-1} P_{II}'.$$

   So $\quad \widehat{\beta} = (X'X)^{-1}X'Y = \left(P_I \Lambda_I^{-1} P_I' + P_{II} \Lambda_{II}^{-1} P_{II}'\right) X'Y.$
   Drop $P_{II} \Lambda_{II}^{-1} P_{II}'$ associated with $\lambda_{q+1}, ..., \lambda_p$. We have principal component estimator

   $$\widehat{\beta}(q) = P_I \Lambda_I^{-1} P_I' X'Y.$$

   (2) Parameters
   With $\widehat{\beta}(q) = P_I \Lambda_I^{-1} P_I' X'Y$ and $Y \sim (X\beta,\, \sigma^2 I_n)$,
   $E[\widehat{\beta}(q)] = P_I \Lambda_I^{-1} P_I' X'X\beta = P_I \Lambda_I^{-1} P_I' (P_I \Lambda_I P_I' + P_{II} \Lambda_{II} P_{II}')\beta = P_I P_I' \beta$ and

   $$\begin{aligned}\text{Cov}\left(\widehat{\beta}(q)\right) &= \left(P_I \Lambda_I^{-1} P_I' X'\right) \sigma^2 I_n \left(P_I \Lambda_I^{-1} P_I' X'\right)' \\ &= \sigma^2 \left(P_I \Lambda_I^{-1} P_I'\right) \left(P_I \Lambda_I P_I' + P_{II} \Lambda_{II} P_{II}'\right) \left(P_I \Lambda_I^{-1} P_I'\right) = \sigma^2 P_I \Lambda_I^{-1} P_I.\end{aligned}$$

   So $\quad \widehat{\beta}(q) \sim (P_I P_I' \beta,\; \sigma^2 P_I \Lambda_I^{-1} P_I).$

(3) Risks
With $\widehat{\beta}(q) \sim \left(P_I P_I' \beta,\, \sigma^2 P_I \Lambda_I^{-1} P_I'\right)$, $\beta - E(\widehat{\beta}(q) = (I - P_I P_I')\beta = P_{II} P_{II}' \beta$. So

$$r(\widehat{\beta}(q),\, \beta) = \text{MSEM}(\widehat{\beta},\, \beta) = \sigma^2 P_I \Lambda_I^{-1} P_I' + (P_{II} P_{II}')\beta\beta'(P_{II} P_{II}')$$

Ex: The parameters and risk of the principal component estimator have been derived based on Y. They can also be derived based on the BLUE $\widehat{\beta}$.
$\widehat{\beta}(q) = P_I \Lambda_I^{-1} P_I' X'Y = P_I \Lambda_I^{-1} P_I' (X'X)(X'X)^{-1} X'Y = P_I P_I' \widehat{\beta}$.
With $\widehat{\beta} \sim (\beta,\, \sigma^2 (X'X)^{-1})$,

$$\widehat{\beta}(q) \sim (P_I P_I' \beta,\, \sigma^2 P_I P_I' (X'X)^{-1} P_I P_I') = (P_I P_I' \beta,\, \sigma^2 P_I \Lambda_I^{-1} P_I').$$

So $$r(\widehat{\beta}(q),\, \beta) = \text{MSEM}(\widehat{\beta},\, \beta) = \sigma^2 P_I \Lambda_I^{-1} P_I' + (P_{II} P_{II}')\beta\beta'(P_{II} P_{II}').$$

3. Making $\widehat{\beta}(q)$ better than BLUE

(1) Sufficient and necessary condition for $\widehat{\beta}(q)$ to be better than BLUE
$\widehat{\beta}(q)$ dominates $\widehat{\beta} \Longleftrightarrow P_{II}' \beta\beta' P_{II} \le \sigma^2 \Lambda_{II}^{-1}$

Proof. The PC estimator $\widehat{\beta}(q)$ is better than the BLUE $\widehat{\beta}$ by the risk MSEM$(\cdot, \cdot)$

$$\begin{aligned}
&\Longleftrightarrow \quad \text{MSEM}(\widehat{\beta}(q),\, \beta) \le \text{MSEM}(\widehat{\beta},\, \beta) \\
&\Longleftrightarrow \quad \sigma^2 P_I \Lambda_I^{-1} P_I' + P_{II} P_{II}' \beta\beta' P_{II} P_{II}' \le \sigma^2 (P_I \Lambda_I^{-1} P_I' + P_{II} \Lambda_{II}^{-1} P_{II}') \\
&\Longleftrightarrow \quad P_{II} P_{II}' \beta\beta' P_{II} P_{II}' \le \sigma^2 P_{II} \Lambda_{II}^{-1} P_{II}' \\
&\overset{*}{\Longleftrightarrow} \quad P_{II}' \beta\beta' P_{II} \le \sigma^2 \Lambda_{II}^{-1} \qquad \text{since } A \le B \Longrightarrow CAC' \le CBC'
\end{aligned}$$

(2) A sufficient condition for $\widehat{\beta}(q)$ to be better than $\widehat{\beta}$.
$P_{II}' \beta\beta' P_{II} \le \frac{\sigma^2}{\lambda_{q+1}} I \Longrightarrow \widehat{\beta}(q)$ dominates $\widehat{\beta}$

Proof. Note that $\frac{\sigma^2}{\lambda_{q+1}} I \le \text{diag}\left(\frac{\sigma^2}{\lambda_{q+1}}, ..., \frac{\sigma^2}{\lambda_p}\right) = \sigma^2 \Lambda_{II}^{-1}$.
So $P_{II}' \beta\beta' P_{II} \le \frac{\sigma^2}{\lambda_{q+1}} I \Longrightarrow P_{II}' \beta\beta' P_{II} \le \sigma^2 \Lambda_{II}^{-1} \Longrightarrow \widehat{\beta}(q)$ dominates $\widehat{\beta}$.

(3) Selecting q such that $\widehat{\beta}(q)$ is better that $\widehat{\beta}$.
Select q such that $0 < \lambda_{q+1} \le \frac{\sigma^2}{\|P_{II}'\beta\|^2}$. Then $\widehat{\beta}(q)$ dominates $\widehat{\beta}$.

Proof. $0 < \lambda_{q+1} \le \frac{\sigma^2}{\|P_{II}'\beta\|^2} \Longrightarrow 0 < \|P_{II}'\beta\|^2 \le \frac{\sigma^2}{\lambda_{q+1}} \Longrightarrow \begin{pmatrix} \|P_{II}'\beta\|^2 & 0 \\ 0 & 0 \end{pmatrix} \le \frac{\sigma^2}{\lambda_{q+1}} I_{p-q}$.
But $P_{II}' \beta\beta' P_{II} \ge 0$ with rank 1, and $\|P_{II}'\beta\|^2$ is a positive eigenvalue since
$(P_{II}' \beta\beta' P_{II})(P_{II}'\beta) = (P_{II}'\beta)(\beta' P_{II} P_{II}' \beta) = \|P_{II}'\beta\|^2 (P_{II}'\beta)$.
Thus by EVD $\quad P_{II}' \beta\beta' P_{II} = Q \begin{pmatrix} \|P_{II}'\beta\|^2 & 0 \\ 0 & 0 \end{pmatrix} Q' \in R^{(p-q)\times(p-q)}$.
But $\begin{pmatrix} \|P_{II}'\beta\|^2 & 0 \\ 0 & 0 \end{pmatrix} \le \frac{\sigma^2}{\lambda_{q+1}} I_{p-q} \Longrightarrow Q \begin{pmatrix} \|P_{II}'\beta\|^2 & 0 \\ 0 & 0 \end{pmatrix} Q' \le Q \frac{\sigma^2}{\lambda_{q+1}} I_{p-q} Q' = \frac{\sigma^2}{\lambda_{q+1}} I_{p-q}$.
Therefore $P_{II}' \beta\beta' P_{II} = Q \begin{pmatrix} \|Q_{II}'\beta\|^2 & 0 \\ 0 & 0 \end{pmatrix} Q' \le \frac{\sigma^2}{\lambda_{q+1}} I_{p-q}$
By (2), the domination holds.

Comments: The cut-off point for λ_{q+1}, $\frac{\sigma^2}{\|P_{II}'\beta\|^2}$, depends on σ^2 and β, and hence can only estimated.

L10: A mixed estimator: Mixed BLUE

1. Two models with one set of parameters

   (1) Two models with one set of parameters
   Consider two models with two sets of data but one set of parameters

$$\begin{array}{l} Y_1 = X_1\beta + \epsilon_1,\ \epsilon_1 \sim (0,\, \sigma^2\Sigma_1) \\ Y_2 = X_2\beta + \epsilon_2,\ \epsilon_2 \sim (0,\, \sigma^2\Sigma_2). \end{array}$$

   (2) Two BLUEs for β
   Assume that $X_1 \in R^{n_1\times p}$ and $X_2 \in R^{n_2\times p}$ are both of full column ranks.
   Then β has BLUEs from two models. Based on model 1,

$$\begin{array}{rcl} \widehat{\beta}_1 & = & (\Sigma_1^{-1/2}X_1)^{+}\Sigma_1^{-1/2}Y_1 \\ & = & \left[\left(\Sigma_1^{-1/2}X_1\right)'\left(\Sigma^{-1/2}X_1\right)\right]^{-1}\left(\Sigma^{-1/2}X_1\right)'\Sigma^{-1/2}Y_1 \\ & = & (X_1'\Sigma_1^{-1}X_1)^{-1}X_1'\Sigma_1^{-1}Y_1. \end{array}$$

   Similarly based on Model 2, $\widehat{\beta}_2 = \left(X_2'\Sigma_2^{-1}X_2\right)^{-1}X_2'\Sigma_2^{-1}Y_2.$

   (3) Parameters and risks of the BLUEs

$$\begin{array}{rcl} \widehat{\beta}_1 & = & \left(X_1'\Sigma_1^{-1}X_1\right)^{-1}X_1'\Sigma_1^{-1}Y_1 \\ & \sim & \left(\beta,\, (X_1'\Sigma_1^{-1}X_1)^{-1}X_1'\Sigma_1^{-1}(\sigma^2\Sigma_1)\Sigma_1^{-1}X_1(X_1'\Sigma_1^{-1}X_1)^{-1}\right) \\ & = & \left(0,\, \sigma^2(X_1'\Sigma_1^{-1}X_1)^{-1}\right). \\ \text{Similarly, } \widehat{\beta}_2 & \sim & \left(0,\, \sigma^2(X_2'\Sigma_2^{-1}X_2)^{-1}\right) \end{array}$$

   Therefore $\mathrm{MSEM}(\widehat{\beta}_1,\, \beta) = \sigma^2(X_1'\Sigma_1^{-1}X_1)^{-1}$
   and $\mathrm{MSEM}(\widehat{\beta}_2,\, \beta) = \sigma^2(X_2'\Sigma_2^{-1}X_2)^{-1}.$

2. Combined model and mixed estimator

   (1) Combined model
   Under the assumption of the independence of Y_1 and Y_2, combining the two models, one has

$$\begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix} = \begin{pmatrix} X_1 \\ X_2 \end{pmatrix}\beta + \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \end{pmatrix},\ \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \end{pmatrix} \sim \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix},\, \sigma^2\begin{pmatrix} \Sigma_1 & 0 \\ 0 & \Sigma_2 \end{pmatrix}\right).$$

   β in this combined model has BLUE, $\widehat{\beta}_m$, called the mixed BLUE for β.

$$\begin{array}{rcl} \widehat{\beta}_m & = & \left[\begin{pmatrix} X_1 \\ X_2 \end{pmatrix}'\begin{pmatrix} \Sigma_1 & 0 \\ 0 & \Sigma_2 \end{pmatrix}^{-1}\begin{pmatrix} X_1 \\ X_2 \end{pmatrix}\right]^{-1}\begin{pmatrix} X_1 \\ X_2 \end{pmatrix}'\begin{pmatrix} \Sigma_1 & 0 \\ 0 & \Sigma_2 \end{pmatrix}^{-1}\begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix} \\ & = & \left(X_1'\Sigma_1^{-1}X_1 + X_2'\Sigma_2^{-1}X_2\right)^{-1}\left(X_1'\Sigma_1^{-1}Y_1 + X_2'\Sigma_2^{-1}Y_2\right) \\ & = & \left(X_1'\Sigma_1^{-1}X_1 + X_2'\Sigma_2^{-1}X_2\right)^{-1}\left((X_1'\Sigma_1^{-1}X_1)\widehat{\beta}_1 + (X_2'\Sigma_2^{-1}X_2)\widehat{\beta}_2\right) \end{array}$$

   (2) Weighted average of vectors or matrices with matrix-weights
   Suppose $W_1 > 0$ and $W_2 > 0$ are $p\times p$ matrices. Then $W = W_1 + W_2$ and W^{-1} are $p\times p$ positive definite matrices. Also, $W^{-1}W_1 + W^{-1}W_1 = I_p$.
   For $V_1 \in R^{p\times t}$ and $V_2 \in R^{p\times t}$,

$$W^{-1}W_1V_1 + W^{-1}W_2V_2 = W^{-1}(W_1V_1 + W_2V_2)$$

   is called the weighted average of V_1 and V_2 with matrix-weights W_1 and W_2.

(3) Mixed BLUE
Let $W_1 = X_1\Sigma_1^{-1}X_1$, $W_2 = X_2\Sigma_2^{-1}X_2$ and $W = W_1 + W_2$. By the formula in (1) and definition in (2),

$$\widehat{\beta} = W^{-1}W_1\widehat{\beta}_1 + W^{-1}W_2\widehat{\beta}_2 = W^{-1}(W_1\widehat{\beta}_1 + W_2\widehat{\beta}_2).$$

So the mixed BLUE is the weighted average of BLUE $\widehat{\beta}_1$ from Model 1, and BLUE $\widehat{\beta}_2$ from Model 2.

3. Improved estimator

(1) Parameters of mixed BLUE
With W_1, W_2 and W in (3) of 2, by (3) of 1,

$$\widehat{\beta}_1 \sim \left(\beta,\, \sigma^2 W_1^{-1}\right) \text{ and } \widehat{\beta}_2 \sim \left(\beta,\, \sigma^2 W_2^{-1}\right).$$

Now $E(\widehat{\beta}_m) = E[W^{-1}(W_1\widehat{\beta}_1 + W_2\widehat{\beta}_2)] = W^{-1}(W_1\beta + W_2\beta) = \beta$ and

$$\begin{aligned}
\mathrm{Cov}(\widehat{\beta}_m) &= \mathrm{Cov}[W^{-1}(W_1\widehat{\beta}_1 + W_2\widehat{\beta}_2)] = W^{-1}\mathrm{Cov}(W_1\widehat{\beta}_1 + W_2\widehat{\beta}_2)W^{-1} \\
&= W^{-1}\left(W_1\sigma^2 W_1^{-1}W_1 + W_2\sigma^2 W_2^{-1}W_2\right)W^{-1} \\
&= \sigma^2 W^{-1}(W_1 + W_2)W^{-1} = \sigma^2 W^{-1}.
\end{aligned}$$

So $\widehat{\beta}_m \sim (\beta,\, \sigma^2 W^{-1})$.

(2) Risks
For model 1, $\quad \mathrm{MSEM}(\widehat{\beta}_1,\, \beta) = \sigma^2 W_1^{-1}$
For model 2, $\quad \mathrm{MSEM}(\widehat{\beta}_2,\, \beta) = \sigma^2 W_2^{-1}$
For the combined model $\quad \mathrm{MSEM}(\widehat{\beta}_m,\, \beta) = \sigma^2 W^{-1}$

(3) Improved BLUE
Note that $W^{-1} = (W_1 + W_2)^{-1} = W_1^{-1} - W_1^{-1}(W_1^{-1} + W_2^{-1})^{-1}W_1^{-1}$ since

$$(W_1 + W_2)[W_1^{-1} - W_1^{-1}(W_1^{-1} + W_2^{-1})^{-1}W_1^{-1}] = I.$$

Hence $\quad \sigma^2 W^{-1} = \sigma^2 W_1^{-1} - \sigma^2 W_1^{-1}(W_1^{-1} + W_2^{-1})^{-1}W_1^{-1}$, i.e.,

$$\mathrm{MSEM}(\widehat{\beta}_m) = \mathrm{MSEM}(\widehat{\beta}_1) - \sigma^2 W_1^{-1}(W_1^{-1} + W_2^{-1})^{-1}W_1^{-1}.$$

But $\sigma^2 W_1^{-1}(W_1^{-1} + W_2^{-1})^{-1}W_1^{-1} \geq 0$. Hence $\mathrm{MSEM}(\widehat{\beta}_m,\, \beta) \leq \mathrm{MSEM}(\widehat{\beta}_1,\, \beta)$.
Similarly $\mathrm{MSEM}(\widehat{\beta}_m,\, \beta) \leq \mathrm{MSEM}(\widehat{\beta}_2,\, \beta)$ can be derived from

$$W^{-1} = (W_1 + W_2)^{-1} = W_2^{-1} - W_2^{-1}(W_1^{-1} + W_2^{-1})^{-1}W_2^{-1}.$$

Comments: Suppose β was estimated by a BLUE. Now with new data, one can get another BLUE. But it is better to combined them to get a mixed BLUE which dominates the BLUEs from the two models.