Least squares with linear constraints

$Y = \Theta + U = XB + U$, the usual linear model.

Introduce linear constraints on the parameter subject to a hypothesis $H_0: AB = 0$, where A is a $q \times p$ matrix, rank $A = q \leq p$.

We want to calculate the estimates when H_0 is true $\hat{\Theta}_H$ and $\hat{B}_H$ are used as symbols for these estimates

$$XB = \Theta \Rightarrow X^T X B = X^T \Theta \Rightarrow B = (X^T X)^{-1} X^T \Theta$$

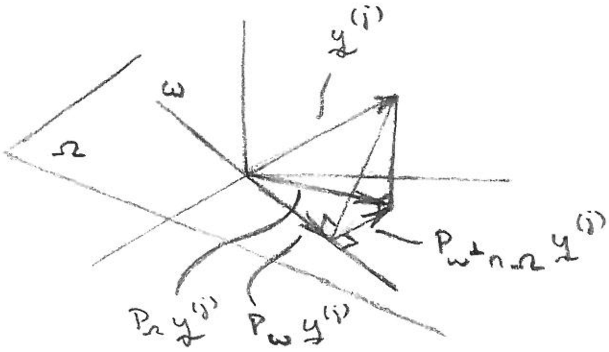
$$AB = 0 \Rightarrow A(X^T X)^{-1} X^T \Theta = 0 \Leftrightarrow A_1 \Theta = 0$$

$$\Rightarrow \underset{\sim}{\theta}^{(j)} \in \Omega \cap \mathcal{N}(A_1), \quad j = 1, \dots, d, \quad A_1 = A(X^T X)^{-1} X^T$$

Let $\omega = \Omega \cap \mathcal{N}(A_1)$ and $\hat{\Theta}_H = P_\omega Y$

$$\hat{U}_H = Y - \hat{\Theta}_H = Y - P_\omega Y = (I_n - P_\omega) Y = (I_n - P_\omega)(\Theta + U) = (I_n - P_\omega) U$$

$$E_H = (Y - \hat{\Theta}_H)^T (Y - \hat{\Theta}_H) = Y^T (I_n - P_\omega) Y = U^T (I_n - P_\omega) U$$



$$P_\omega Y = (P_\Omega - (P_\Omega - P_\omega)) Y$$
$$= (P_\Omega - P_{\omega^\perp \cap \Omega}) Y$$

cf. B 3.2

$$\omega^\perp \cap \Omega = (\Omega \cap \mathcal{N}(A_1))^\perp \cap \Omega$$
$$= (\Omega^\perp + \mathcal{R}(A_1^T)) \cap \Omega \quad \text{cf. B 2.1 and B 2.4}$$
$$= \underset{\sim}{0} + \mathcal{R}(A_1^T) \cap \Omega = \mathcal{R}(A_1^T), \text{ as}$$
$$A_1^T = X(X^T X)^{-1} A^T = XD \Rightarrow \mathcal{R}(A_1^T) \subset \Omega$$

$$P_{\omega^\perp \cap \Omega} = A_1^T (A_1 A_1^T)^{-1} A_1, \quad \text{cf. B 1.7}$$
$$= X(X^T X)^{-1} A^T (A(X^T X)^{-1} A^T)^{-1} A (X^T X)^{-1} X^T$$

$$\hat{\Theta}_H = P_\Omega Y - P_{\omega^\perp \cap \Omega} Y = \hat{\Theta} - P_{\omega^\perp \cap \Omega} Y$$

$$= X\hat{B} - X(X^TX)^{-1}A^T(A(X^TX)^{-1}A^T)^{-1}\underbrace{(X^TX)^{-1}X^TY}_{\hat{B}}$$

$$= X(I_p - (X^TX)^{-1}A^T(A(X^TX)^{-1}A^T)^{-1})\hat{B}$$

$\hat{\Theta}_H = X\hat{B}_H$ leads to

$$\hat{B}_H = (I_p - (X^TX)^{-1}A^T(A(X^TX)^{-1}A^T)^{-1})\hat{B}$$

When rank $A = s < q$ $(A(X^TX)^{-1}A^T)^{-1}$ can be interchanged with $(A(X^TX)^{-1}A^T)^-$ in the expressions for $\hat{\Theta}_H$ and $\hat{B}_H$

Now let rank $X = r < p$ and rank $A = q$

$P_\Omega = X(X^TX)^-X^T$ cf. B 1.7

Assume that $\operatorname{tr}(AB)$ is estimable which means that the rows of A must be linear combinations of the rows of X which again can be expressed as $A = MX$ where M $q \times n$, rank $M = q$, as rank $A = q$

$$AB = 0 \Leftrightarrow MXB = 0 \Leftrightarrow M\Theta = 0 \Leftrightarrow \underset{\sim}{\theta}^{(j)} \in \Omega \cap \mathcal{N}(M) \quad j = 1, \dots, d$$

Let $\omega = \Omega \cap \mathcal{N}(M)$ and $\hat{\Theta}_H = P_\omega Y$

$\hat{U}_H$, E_H and P_ω have the same form as before, but

$\omega^\perp \cap \Omega = \mathcal{R}(P_\Omega M^T)$ cf. B 3.3

$$P_\Omega M^T = X(X^TX)^-X^TM^T = X(X^TX)^-(MX)^T = X(X^TX)^-A^T$$

rank $P_\Omega M^T = q$ is shown indirect:

Assume $\exists \underset{\sim}{c} \neq \underset{\sim}{0}$: $P_\Omega M^T\underset{\sim}{c} = \underset{\sim}{0} \Rightarrow M^T\underset{\sim}{c} \perp \Omega$, $\Omega = \mathcal{R}(X)$

$\Rightarrow X^TM^T\underset{\sim}{c} = \underset{\sim}{0} \Rightarrow A^T\underset{\sim}{c} = \underset{\sim}{0}$ inconsistency as rank $A = q$

$$P_{\omega^\perp \cap \Omega} = P_\Omega M^T \left((P_\Omega M^T)^T P_\Omega M^T \right)^{-1} (P_\Omega M^T)^T \qquad \text{cf. B 1.8}$$

$$= X (X^TX)^- A^T \left(A (X^TX)^- X^TX (X^TX)^- A^T \right)^{-1} A (X^TX)^- X^T$$

$$= X (X^TX)^- A^T \left(A (X^TX)^- A^T \right)^{-1} A (X^TX)^- X^T$$

$$\hat{\Theta}_H = X \left(I_r - (X^TX)^- A^T \left(A (X^TX)^- A^T \right)^{-1} A \right) \hat{B}$$

$$\hat{B}_H = \left(I_r - (X^TX)^- A^T \left(A (X^TX)^- A^T \right)^{-1} A \right) \hat{B} \qquad (\text{not unique})$$

For rank $A = s < q$ we have rank $P_\Omega M^T = s$, and $(A(X^TX)^-A^T)^{-1}$ can be interchanged with $(A(X^TX)^-A^T)^-$ in the expressions for $\hat{\Theta}_H$ and $\hat{B}_H$

Now change the restriction to $H_0: AB = C$, C $q \times d$

Let $\tilde{Y} = Y - XB_0$ where B_0 is any fixed solution to $AB = C$

$$Y = XB + U \iff Y - XB_0 = XB + U - XB_0 \iff \tilde{Y} = X(B - B_0) + U$$

$$\iff \tilde{Y} = X\Lambda + U, \quad \Lambda = B - B_0$$

$$\iff \tilde{Y} = \Phi + U, \quad \Phi = X\Lambda$$

Note that $A\Lambda = A(B - B_0) = AB - AB_0 = C - C = 0$

hence $Y = XB + U$ when $H_0: AB = C$ is true is equivalent to $\tilde{Y} = X\Lambda + U$ when $H_0: A\Lambda = 0$ is true

Let $\omega = \Omega \cap \mathcal{N}(\cdot)$ and $\hat{\Phi}_H = P_\omega \tilde{Y}$

$$\hat{U}_H = \tilde{Y} - \hat{\Phi}_H = (I_n - P_\omega)\tilde{Y}; \qquad \hat{U}_H = (I_n - P_\omega) U \text{ when } H_0 \text{ is true}$$

$$E_H = \hat{U}_H^T \hat{U}_H = \tilde{Y}^T (I_n - P_\omega) \tilde{Y}; \qquad E_H = U^T (I_n - P_\omega) U \text{ when } H_0 \text{ is true}$$

$$\hat{\Lambda}_H = \left(I_r - (X^TX)^- A^T \left(A (X^TX)^- A^T \right)^- A \right) \hat{\Lambda} \qquad (\text{interchange of letters})$$

$$= \left(I_r - (X^TX)^- A^T \left(A (X^TX)^- A^T \right)^- A \right) (\hat{B} - B_0)$$

$$\hat{B}_H = \hat{\Lambda}_H + B_0$$

$$= \hat{B} - B_0 - (X^TX)^- A^T \left(A (X^TX)^- A^T \right)^- (A\hat{B} - AB_0) + B_0$$

$$\hat{B}_H = \hat{B} - (X^T X)^- A^T (A (X^T X)^- X^T)^- (A\hat{B} - C)$$

$$\hat{\Theta}_H = X \hat{B}_H = X(\hat{\Lambda}_H + B_0) = \hat{\Phi}_H + X B_0 \quad \text{unique}$$

Note that

- $Y - \hat{\Theta}_H = Y - XB_0 - (X\hat{B}_H - XB_0) = \tilde{Y} - X(\hat{B}_H - B_0)$
  $= \tilde{Y} - X\hat{\Lambda}_H = \tilde{Y} - \hat{\Phi}_H = \hat{U}_H$

- $\hat{\Theta}_H - \Theta = X\hat{B}_H - XB = X\hat{B}_H - XB_0 - (XB - XB_0)$
  $= X(\hat{B}_H - B_0) - X(B - B_0) = X\hat{\Lambda}_H - X\Lambda$
  $= \hat{\Phi}_H - \Phi$

- $(Y - \hat{\Theta}_H)^T(\hat{\Theta}_H - \Theta) = (\tilde{Y} - \hat{\Phi}_H)^T(\hat{\Phi}_H - \Phi)$
  $= (\tilde{Y} - P_\omega \tilde{Y})^T (P_\omega \tilde{Y} - P_\omega \Phi)$
  $= \tilde{Y}^T (I_n - P_\omega) P_\omega (\tilde{Y} - \Phi)$
  $= 0$

$\hat{\Theta}_H$ is the least squares estimate as

$$(Y - \Theta)^T (Y - \Theta) = (Y - \hat{\Theta}_H + \hat{\Theta}_H - \Theta)^T (Y - \hat{\Theta}_H + \hat{\Theta}_H - \Theta)$$
$$= (Y - \hat{\Theta}_H)^T (Y - \hat{\Theta}_H) + 0 + 0 + (\hat{\Theta}_H - \Theta)^T (\hat{\Theta}_H - \Theta)$$

$(\hat{\Theta}_H - \Theta)^T (\hat{\Theta}_H - \Theta) \geq 0$ cf. A 4.4, i.e.

$(Y - \Theta)^T (Y - \Theta) \geq (Y - \hat{\Theta}_H)^T (Y - \hat{\Theta}_H)$ for all Θ

Equality for $\Theta = \hat{\Theta}_H$

Distribution theory

Model : $Y = \Theta + U$, $\underset{\sim}{\theta}^{(j)} \in \Omega$, $j = 1, \dots, d$, $\dim \Omega = r$, $n - r \geq d$

$\underset{\sim}{u}_i \sim N_d(\underset{\sim}{0}, \Sigma)$, $i = 1, \dots, n$, independent

Theorem 8.3 : $E \sim W_d(n-r, \Sigma)$ indep. of $\hat{\Theta}$

Proof : $$\left.\begin{aligned} \hat{\Theta} - \Theta = P_\Omega (Y - \Theta) &= P_\Omega U I_d \\ (I_n - P_\Omega) Y &= (I_n - P_\Omega) U I_d \end{aligned}\right\} \text{ are indep.}$$

as $P_\Omega (I_n - P_\Omega) = O$, cf. ex. 2.18

$\Rightarrow$ $\hat{\Theta}$ and $E = Y^T (I_n - P_\Omega) Y$ are independent

$E = U^T (I_n - P_\Omega) U \Rightarrow E \sim W_d(n-r, \Sigma)$, as

$\operatorname{rank}(I_n - P_\Omega) = \operatorname{tr}(I_n - P_\Omega) = n - r$, corollary 1 to theorem 2.4

Corollary : If $\Theta = XB$, X $n \times p$, $\operatorname{rank} X = p$

then $\hat{B}$ and E are independent

Proof : $\hat{\Theta} = X\hat{B} \Rightarrow \hat{B} = (X^T X)^{-1} X^T \hat{\Theta}$, hence $\hat{B}$ is a function of $\hat{\Theta}$

Theorem 8.4 : $\hat{\Theta} = P_\Omega Y$ and $\hat{\Sigma} = \frac{E}{n}$, $n - r \geq d$, is the maximum likelihood estimates for respectively Θ and Σ.

Proof :

The simultaneous density function for $(\underset{\sim}{y}_1, \ldots, \underset{\sim}{y}_n)$:

$$f(y_{11}, \ldots, y_{nd}) = \prod_{i=1}^{n} (2\pi)^{-\frac{d}{2}} (\det \Sigma)^{-\frac{1}{2}} \exp\left(-\tfrac{1}{2}(\underset{\sim}{y}_i - \underset{\sim}{\theta}_i)^T \Sigma^{-1} (\underset{\sim}{y}_i - \underset{\sim}{\theta}_i)\right)$$

$$= (2\pi)^{-\frac{nd}{2}} (\det \Sigma)^{-\frac{n}{2}} \exp\left(-\tfrac{1}{2} \sum_{i=1}^{n} (\underset{\sim}{y}_i - \underset{\sim}{\theta}_i)^T \Sigma^{-1} (\underset{\sim}{y}_i - \underset{\sim}{\theta}_i)\right)$$

$$= (2\pi)^{-\frac{nd}{2}} (\det \Sigma)^{-\frac{n}{2}} \operatorname{etr}\left(-\tfrac{1}{2} \Sigma^{-1} \sum_{i=1}^{n} (\underset{\sim}{y}_i - \underset{\sim}{\theta}_i)(\underset{\sim}{y}_i - \underset{\sim}{\theta}_i)^T\right)$$

The likelihood function for Θ and Σ then becomes

$$L(\Theta, \Sigma) = (2\pi)^{-\frac{nd}{2}} (\det \Sigma)^{-\frac{n}{2}} \operatorname{etr}\left(-\tfrac{1}{2} \Sigma^{-1} (Y - \Theta)^T (Y - \Theta)\right)$$

$$= (2\pi)^{-\frac{nd}{2}} (\det \Sigma)^{-\frac{n}{2}} \operatorname{etr}\Big(-\tfrac{1}{2} \Sigma^{-1} \big((Y - \hat{\Theta})^T (Y - \hat{\Theta}) + 0 + 0$$
$$+ (\hat{\Theta} - \Theta)^T (\hat{\Theta} - \Theta)\big)\Big)$$

$$\ln L(\Theta, \Sigma) = -\tfrac{nd}{2} \ln(2\pi) - \tfrac{n}{2} \ln \det \Sigma - \tfrac{1}{2} \operatorname{tr}(\Sigma^{-1} E)$$
$$- \tfrac{1}{2} \operatorname{tr}\left((\hat{\Theta} - \Theta) \Sigma^{-1} (\hat{\Theta} - \Theta)^T\right)$$

$$\le -\tfrac{nd}{2} \ln(2\pi) - \tfrac{n}{2} \ln \det \Sigma - \tfrac{1}{2} \operatorname{tr}(\Sigma^{-1} E)$$

equality for $\Theta = \hat{\Theta} = P_\Omega Y$

$$\ln L(\hat{\Theta}, \Sigma) = -\tfrac{nd}{2} \ln(2\pi) - \tfrac{n}{2} \ln \det \Sigma - \tfrac{1}{2} \operatorname{tr}(\Sigma^{-1} E)$$

$$\le -\tfrac{nd}{2} \ln(2\pi) - \tfrac{n}{2} \ln \det \tfrac{E}{n} - \tfrac{n}{2} \operatorname{tr}\left(\left(\tfrac{E}{n}\right)^{-1} \tfrac{E}{n}\right)$$

with equality for $\Sigma = \hat{\Sigma} = \frac{E}{n}$, cf. A7.1

Hence

$$\ln L(\hat{\Theta}, \hat{\Sigma}) = -\tfrac{nd}{2} \ln(2\pi) - \tfrac{n}{2} \ln \det \hat{\Sigma} - \tfrac{n}{2} \operatorname{tr} I_d$$

$$L(\hat{\Theta}, \hat{\Sigma}) = (2\pi e)^{-\frac{nd}{2}} (\det \hat{\Sigma})^{-\frac{n}{2}}$$

Note that $\hat{\Sigma}$ is biased, $E[\hat{\Sigma}] = \frac{1}{n} E[E] = \frac{n-r}{n} \Sigma$

Corollary: $\hat{B}$ is the maximum likelihood estimate for B $\quad (\Theta = XB, \ \text{rank}\, X = p)$

Consider $Y = \Theta + U = XB + U$ when $H_0: AB = C$ is true, where

$\text{rank}\, X = r < p$, $\text{rank}\, A = q \leq r$, $A = MX$, $n - r + q \geq d$

Equivalent model: $\tilde{Y} = \Phi + U$ when $H_0: A\Lambda = O$ is true

$\underset{\sim}{\varphi}^{(j)} \in \omega = \Omega \cap \mathcal{N}(M)$, $j = 1, \dots, d$

$\hat{\Phi}_H$ and $\frac{E_H}{n}$ are the maximum likelihood estimates for Φ and Σ, $\hat{\Theta}_H = \hat{\Phi}_H + XB_0$ is the maximum likelihood estimates for Θ independent of the choice of B_0 *

For $\text{rank}\, X = p$, $\hat{B}_H$ is the maximum likelihood estimate for B

The residual matrix $\hat{U}$

$$\hat{U} = Y - \hat{\Theta} = (I_n - P_\Omega) Y = (I_n - P_\Omega) U$$

$$\hat{\underset{\sim}{u}}^{(j)} = (I_n - P_\Omega)\, \underset{\sim}{y}^{(j)}, \quad j = 1, \dots, d$$

$$\text{Cov}(\hat{\underset{\sim}{u}}^{(j)}, \hat{\underset{\sim}{u}}^{(k)}) = \sigma_{jk}(I_n - P_\Omega) \approx \sigma_{jk} I_n$$

Univariate plotting methods can be applied to $\hat{\underset{\sim}{u}}^{(j)}$

$\hat{u}_i \sim N_d(\underset{\sim}{0}, \Sigma)$, $i = 1, \dots, n$, approximative independent

Constraints on the residuals

$$P_\Omega \hat{U} = P_\Omega (I_n - P_\Omega) Y = O$$

$$* \quad \hat{\Theta}_H = \hat{\Phi}_H + XB_0 = P_\omega \tilde{Y} + XB_0 = (P_\Omega - P_{\omega^\perp \cap \Omega})(Y - XB_0) + XB_0$$

$$= P_\Omega Y - P_\Omega X B_0 - P_{\omega^\perp \cap \Omega} Y + P_{\omega^\perp \cap \Omega} XB_0 + XB_0 = P_\omega Y + P_{\omega^\perp \cap \Omega} XB_0 \qquad (P_\Omega X = X)$$

$$= P_\omega Y + X(X^T X)^- A^T (A(X^T X)^- A^T)^{-1} A (X^T X)^- X^T X B_0$$

$$= P_\omega Y + X(X^T X)^- A^T (A(X^T X)^- A^T)^{-1} C, \text{ as}$$

$$A(X^T X)^- X^T X B_0 = MX(X^T X)^- X^T X B_0 = M P_\Omega X B_0 = MXB_0 = AB_0 = C$$