Lecture 22: Weighted LSE and linear mixed effects models

The weighted LSE

In the linear model

$$X = Z\beta + \varepsilon, \tag{1}$$

the unbiased LSE of $l^\tau\beta$ may be improved by a slightly biased estimator when $V = \mathrm{Var}(\varepsilon)$ is not $\sigma^2 I_n$ and the LSE is not BLUE.

Assume that Z is of full rank so that every $l^\tau\beta$ is estimable.
If V is known, then the BLUE of $l^\tau\beta$ is $l^\tau\breve{\beta}$, where

$$\breve{\beta} = (Z^\tau V^{-1} Z)^{-1} Z^\tau V^{-1} X \tag{2}$$

(see the discussion after the statement of assumption A3 in §3.3.1).

If V is unknown and $\widehat{V}$ is an estimator of V, then an application of the substitution principle leads to a *weighted least squares estimator*

$$\widehat{\beta}_w = (Z^\tau \widehat{V}^{-1} Z)^{-1} Z^\tau \widehat{V}^{-1} X. \tag{3}$$

The weighted LSE is not linear in X and not necessarily unbiased for β.

If the weighted LSE $l^\tau\widehat{\beta}_w$ is unbiased, then the LSE $l^\tau\widehat{\beta}$ may not be a BLUE, since $\mathrm{Var}(l^\tau\widehat{\beta}_w)$ may be smaller than $\mathrm{Var}(l^\tau\widehat{\beta})$.

Asymptotic properties of the weighted LSE depend on the asymptotic behavior of $\widehat{V}$.

We say that $\widehat{V}$ is consistent for V iff

$$\|\widehat{V}^{-1}V - I_n\|_{\max} \to_p 0, \tag{4}$$

where $\|A\|_{\max} = \max_{i,j}|a_{ij}|$ for a matrix A whose (i,j)th element is a_{ij}.

Theorem 3.17

Consider model (1) with a full rank Z. Let $\breve{\beta}$ and $\widehat{\beta}_w$ be defined by (2) and (3), respectively, with a $\widehat{V}$ consistent in the sense of (4).

Under the conditions in Theorem 3.12,

$$l^\tau(\widehat{\beta}_w - \beta)/a_n \to_d N(0,1),$$

where $l \in \mathscr{R}^p$, $l \neq 0$, and

$$a_n^2 = \mathrm{Var}(l^\tau\breve{\beta}) = l^\tau(Z^\tau V^{-1}Z)^{-1}l.$$

Proof

Using the same argument as in the proof of Theorem 3.12, we obtain that

$$l^\tau(\breve{\beta} - \beta)/a_n \to_d N(0,1).$$

By Slutsky's theorem, the result follows from

$$l^\tau \widehat{\beta}_w - l^\tau \breve{\beta} = o_p(a_n).$$

Define

$$\xi_n = l^\tau (Z^\tau \widehat{V}^{-1} Z)^{-1} Z^\tau (\widehat{V}^{-1} - V^{-1})\varepsilon$$

and

$$\zeta_n = l^\tau [(Z^\tau \widehat{V}^{-1} Z)^{-1} - (Z^\tau V^{-1} Z)^{-1}] Z^\tau V^{-1} \varepsilon.$$

Then

$$l^\tau \widehat{\beta}_w - l^\tau \breve{\beta} = \xi_n + \zeta_n.$$

The result follows from $\xi_n = o_p(a_n)$ and $\zeta_n = o_p(a_n)$ (details are in the textbook).

- Theorem 3.17 shows that as long as $\widehat{V}$ is consistent in the sense of (4), the weighted LSE $\widehat{\beta}_w$ is asymptotically as efficient as $\check{\beta}$, which is the BLUE if V is known.
- By Theorems 3.12 and 3.17, the asymptotic relative efficiency of the LSE $l^\tau\widehat{\beta}$ w.r.t. the weighted LSE $l^\tau\widehat{\beta}_w$ is
$$\frac{l^\tau(Z^\tau V^{-1}Z)^{-1}l}{l^\tau(Z^\tau Z)^{-1}Z^\tau VZ(Z^\tau Z)^{-1}l},$$
which is always less than 1 and equals 1 if $l^\tau\widehat{\beta}$ is a BLUE ($\widehat{\beta} = \check{\beta}$).
- Finding a consistent $\widehat{V}$ is possible when V has a certain type of structure.

Example 3.29

Consider model (1).
Suppose that $V = \mathrm{Var}(\varepsilon)$ is a block diagonal matrix with the ith diagonal block
$$\sigma^2 I_{m_i} + U_i \Sigma U_i^\tau, \qquad i = 1,...,k, \tag{5}$$
where m_i's are integers bounded by a fixed integer m, $\sigma^2 > 0$ is an unknown parameter, Σ is a $q \times q$ unknown nonnegative definite matrix,

U_i is an $m_i \times q$ full rank matrix whose columns are in $\mathscr{R}(W_i)$, $q < \inf_i m_i$, and W_i is the $p \times m_i$ matrix such that $Z^\tau = (\ W_1\ \ W_2\ \dots\ W_k\ )$.

Under (5), a consistent $\widehat{V}$ can be obtained if we can obtain consistent estimators of σ^2 and Σ.

Let $X = (Y_1, ..., Y_k)$, where Y_i is an m_i-vector, and let R_i be the matrix whose columns are linearly independent rows of W_i.

If Y_i's are independent and $\sup_i E|\varepsilon_i|^{2+\delta} < \infty$ for some $\delta > 0$, then

$$\widehat{\sigma}^2 = \frac{1}{n-kq} \sum_{i=1}^{k} Y_i^\tau [I_{m_i} - R_i(R_i^\tau R_i)^{-1} R_i^\tau] Y_i$$

is an unbiased and consistent estimator of σ^2.

Let $r_i = Y_i - W_i^\tau \widehat{\beta}$ and

$$\widehat{\Sigma} = \frac{1}{k} \sum_{i=1}^{k} \left[(U_i^\tau U_i)^{-1} U_i^\tau r_i r_i^\tau U_i (U_i^\tau U_i)^{-1} - \widehat{\sigma}^2 (U_i^\tau U_i)^{-1} \right].$$

It can be shown (exercise) that $\widehat{\Sigma}$ is consistent for Σ in the sense that $\|\widehat{\Sigma} - \Sigma\|_{\max} \to_p 0$ or, equivalently, $\|\widehat{\Sigma} - \Sigma\| \to_p 0$ (see Exercise 116).

Linear mixed effects models

Adding random effects to a linear model

Consider linear model (1), $X = Z\beta + \varepsilon$.

In many applications we need to add random-effect terms, which leads to the linear mixed effects model

$$X = Z\beta + U_1\xi_1 + \cdots + U_k\xi_k + \varepsilon \tag{6}$$

where U_j's are fixed matrices and $\xi_1, ..., \xi_k$ are independent unobserved random effects (vectors), and ε and ξ_j's are independent.

The following are two main reasons for adding random effects.

- We want to model the correlation among the errors, since $U_1\xi_1 + \cdots + U_k\xi_k + \varepsilon$ can be viewed as error in a linear model.
- Random effects present unobserved variables of interests.

It is typically assumed that ξ_j's and ε have mean 0 and finite covariance matrices and $\mathrm{Var}(\varepsilon) = \sigma^2 I_n$ so that

$$E(X) = Z\beta \quad \text{and} \quad \mathrm{Var}(X) = U_1 \mathrm{Var}(\xi_1) U_1^\tau + \cdots + U_k \mathrm{Var}(\xi_k) U_k^\tau + \sigma^2 I_n$$

A special case is that $\mathrm{Var}(\xi_i) = \sigma_i^2 I_{m_i}$, where m_i is the dimension of ξ_i, $i = 1, ..., k$, in which case $\sigma_1^2, ..., \sigma_k^2$ are called variance components so that model (6) is also called variance components models.

Example: One-way random effects model

The one-way random effect model

$$Y_{ij} = \mu + A_i + e_{ij}, \qquad j = 1, ..., n_i,\ i = 1, ..., m,$$

discussed previously is a special case of model (6), $\xi_1 = (A_1, ..., A_m)$, $Z = J_n$, and U_1 is block diagonal whose ith diagonal block is J_{n_i}.

Parameter estimation

Besides β, parameters of interests in a linear mixed effects model are $\Sigma_i = \mathrm{Var}(\xi_i)$, $i = 1, ..., k$, and σ^2.

We carry out the estimation in two steps.

1. Obtain estimators $\widehat{\Sigma}_1, ..., \widehat{\Sigma}_k$, and $\widehat{\sigma}^2$.
2. Let $\Sigma = \mathrm{Var}(X)$ and $\widehat{\Sigma} = U_1 \widehat{\Sigma}_1 U_1^\tau + \cdots + U_k \widehat{\Sigma}_k U_k^\tau + \widehat{\sigma}^2 I_n$.
   We then estimate β by the weighted LSE
   $$\widehat{\beta}_W = (Z^\tau \widehat{\Sigma}^{-1} Z)^{-1} Z^\tau \widehat{\Sigma}^{-1} X$$

Main approaches for estimating variance components

1. The ANOVA method.
2. The MINQUE (developed by C.R. Rao)
3. The maximum likelihood estimation.
4. The restricted maximum likelihood estimation.

Three steps in estimating variance components

We consider the ANOVA method and the special case where $\Sigma_i = \sigma_i^2 I_{m_i}$, $i = 1, ..., k$.

(1) Treat ξ_i's as fixed effects and apply the ANOVA technique to obtain sums of squares.

(2) Treat ξ_i's as random and derive the expectations of the sums of squares, which are linear functions of variance components.

(3) Set each sum of squares equal to its expectation, and then follow the method of moments to estimate variance components.

To achieve (1), we need to get the decomposition

$$X^\tau X = SS_\beta + SS_{\xi_1} + \cdots + SS_{\xi_k} + SS_\varepsilon$$

Deriving sums of squares

To obtain SS_β, we consider model $X = Z\beta + \varepsilon$ and the sum of squares due to regression:

$$SS_\beta = RSS(\beta) = X^\tau Z(Z^\tau Z)^{-1} Z^\tau X$$

To obtain SS_{ξ_1}, we treat ξ_1 as fixed effect in the linear model after removing the β effect, i.e., consider model $X - Z\beta = U_1\xi_1 + \varepsilon$ and the sum of squares due to regression, which is equal to the SS due to regression in model $X = Z\beta + U_1\xi_1 + \varepsilon$ minus the SS due to regression in model $X = Z\beta + \varepsilon$, i.e.,

$$SS_{\xi_1} = RSS(\beta, \xi_1) - RSS(\beta)$$

Similarly, the SS due to regression in model $X = Z\beta + U_1\xi_1 + U_2\xi_2 + \varepsilon$ minus the SS due to regression in model $X = Z\beta + U_1\xi_1\varepsilon$ gives

$$SS_{\xi_2} = RSS(\beta, \xi_1, \xi_2) - RSS(\beta, \xi_1)$$

$\cdots\cdots$

$$SS_{\xi_k} = RSS(\beta, \xi_1, ..., \xi_k) - RSS(\beta, \xi_1, ..., \xi_{k-1})$$

Finally,

$$SS_\varepsilon = X^\tau X - RSS(\beta, \xi_1, ..., \xi_k)$$

For a square matrix M, M^- is its generalized inverse if $M = MM^-M$.
To derive a form for SS_{ξ_1}, we use the following result:

$$\begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}^- = \begin{pmatrix} A_{11}^{-1} + A_{11}^{-1}A_{12}B^-A_{21}A_{11}^{-1} & -A_{11}^{-1}A_{12}B^- \\ -B^-A_{21}A_{11}^{-1} & B^- \end{pmatrix}$$

where $B = A_{22} - A_{21}A_{11}^{-1}A_{12}$.
Under model $X = Z\beta + U_1\xi_1 + \varepsilon$,

$$RSS(\beta, \xi_1) = X^\tau (Z\ U_1) \begin{pmatrix} Z^\tau Z & Z^\tau U_1 \\ U_1^\tau Z & U_1^\tau U_1 \end{pmatrix}^- \begin{pmatrix} Z^\tau \\ U_1^\tau \end{pmatrix} X$$

Letting $H_Z = Z(Z^\tau Z)^{-1}Z^\tau$, $D = I - H_Z$, $B = U_1^\tau D U_1$, and applying the generalized inverse formula, we obtain that

$$\begin{aligned} RSS(\beta, \xi_1) &= X^\tau[H_Z + H_Z U_1 B^- U_1^\tau H_Z - H_Z U_1 B^- U_1^\tau \\ &\quad - U_1 B^- U_1^\tau H_Z + U_1 B^- U_1^\tau]X \\ &= RSS(\beta) + X^\tau D U_1 B^- U_1^\tau D X \end{aligned}$$

Hence

$$SS_{\xi_1} = X^\tau D U_1 (U_1^\tau D U_1)^- U_1^\tau D X = X^\tau (D - D_1) X$$

where $D_1 = D - DU_1(U_1^\tau D U_1)^- U_1^\tau D$

Similarly, we can obtain

$$SS_{\xi_2} = X^\tau(D_1 - D_2)X, \quad D_2 = D_1 - D_1U_2(U_2^\tau D_1 U_2)^- U_2^\tau D_1$$

......

$$SS_{\xi_k} = X^\tau(D_{k-1} - D_k)X, \quad D_k = D_{k-1} - D_{k-1}U_k(U_k^\tau D_{k-1} U_k)^- U_k^\tau D_{k-1}$$

Furthermore, $SS_\varepsilon = X^\tau D_k X$ so that

$$X^\tau X = SS_\beta + SS_{\xi_1} + \cdots + SS_{\xi_k} + SS_\varepsilon$$

Expectations of SS

To derive the expectation of SS_{ξ_i}, we use the following result.

Lemma. For a random vector $X \in \mathscr{R}^n$, if $E(X) = \mu$, $\mathrm{Var}(X) = \Sigma$, and A is an $n \times n$ symmetric matrix, then

$$E(X^\tau AX) = \mu^\tau A\mu + \mathrm{tr}(A\Sigma)$$

For SS_{ξ_1}, since $DZ = D_1Z = 0$,

$$\begin{aligned} E(SS_{\xi_1}) &= E[X^\tau(D - D_1)X] \\ &= \beta^\tau Z^\tau(D - D_1)Z\beta + \mathrm{tr}[(D - D_1)\mathrm{Var}(X)] \\ &= \mathrm{tr}[(D - D_1)(U_1\mathrm{Var}(\xi_1)U_1^\tau + \cdots + U_k\mathrm{Var}(\xi_k)U_k^\tau) + \sigma^2 I_n] \end{aligned}$$

Since $\Sigma_i = \sigma_i^2 I_{m_i}$, $i = 1, ..., k$,

$$E(SS_{\xi_1}) = \mathrm{tr}[(D - D_1)(\sigma_1^2 U_1 U_1^\tau + \cdots + \sigma_k^2 U_k U_k^\tau) + \sigma^2 I_n]$$
$$= \sigma_1^2 \mathrm{tr}[(D - D_1)U_1 U_1^\tau] + \cdots + \sigma_k^2 \mathrm{tr}[(D - D_1)U_k U_k^\tau] + \sigma^2 \mathrm{tr}(D - D_1)$$

Note that

$$\mathrm{tr}(D - D_1) = \mathrm{tr}[DU_1(U_1^\tau DU_1)^- U_1^\tau D] = \mathrm{rank}(U_1^\tau DU_1) = r_1$$

$$U_1^\tau D_1 U_1 = U_1^\tau DU_1 - U_1^\tau DU_1(U_1^\tau DU_1)^- U_1^\tau DU_1 = U_1^\tau DU_1 - U_1^\tau DU_1 = 0$$

Hence

$$E(SS_{\xi_1}) = \sigma_1^2 \mathrm{tr}(U_1^\tau DU_1) + \sigma_2^2[\mathrm{tr}(U_2^\tau DU_2) - \mathrm{tr}(U_2^\tau D_1 U_2)] + \cdots$$
$$\cdots + \sigma_k^2[\mathrm{tr}(U_k^\tau DU_k) - \mathrm{tr}(U_k^\tau D_1 U_k)] + r_1\sigma^2$$

which is a linear function of variance components.

Since $D_i Z = 0$ and $D_i U_j = 0$, $i = 2, ..., k$, $i \geq j$, for $i = 2, ..., k$,

$$E(SS_{\xi_i}) = \mathrm{tr}[(D_{i-1} - D_i)(\sigma_i^2 U_i U_i^\tau + \cdots + \sigma_k^2 U_k U_k^\tau) + \sigma^2 I_n]$$
$$= \sigma_i^2 \mathrm{tr}(U_i^\tau D_{i-1} U_i) + \sigma_{i+1}^2[\mathrm{tr}(U_{i+1}^\tau D_{i-1} U_{i+1}) - \mathrm{tr}(U_{i+1}^\tau D_i U_{i+1})] +$$
$$\cdots + \sigma_k^2[\mathrm{tr}(U_k^\tau D_{i-1} U_k) - \mathrm{tr}(U_k^\tau D_i U_k)] + r_i\sigma^2$$

$$E(SS_\varepsilon) = E(X^\tau D_k X) = \sigma^2 \mathrm{tr}(D_k) = (n - p - r_1 - \cdots - r_k)\sigma^2$$

where $r_i = \mathrm{rank}(D_{i-1} - D_i) = \mathrm{rank}(Z, U_1, ..., U_i) - \mathrm{rank}(Z, U_1, ..., U_{i-1})$.

Estimation of variance components by ANOVA

Set

$$SS_{\xi_i} = \sigma_i^2 \mathrm{tr}(U_i^\tau D_{i-1} U_i) + \sigma_{i+1}^2 [\mathrm{tr}(U_{i+1}^\tau D_{i-1} U_{i+1}) - \mathrm{tr}(U_{i+1}^\tau D_i U_{i+1})] + \cdots + \sigma_k^2 [\mathrm{tr}(U_k^\tau D_{i-1} U_k) - \mathrm{tr}(U_k^\tau D_i U_k)] + r_i \sigma^2$$

$$i = 1, ..., k$$

$$SS_\varepsilon = (n - p - r_1 - \cdots - r_k)\sigma^2$$

These equations can be easily solved by first obtaining

$$\widehat{\sigma}^2 = \frac{SS_\varepsilon}{n - p - r_1 - \cdots - r_k}$$

then $\widehat{\sigma}_k^2$, then $\widehat{\sigma}_{k-1}^2$, ..., then $\widehat{\sigma}_1^2$.

- Advantage: estimators can be easily computed and are unbiased.
- Disadvantage: except for $\widehat{\sigma}^2$, each $\widehat{\sigma}_i^2$ may be negative.

Example: one-way random effects model

The one-way random effects model is

$$X_{ij} = \mu + A_i + e_{ij}, \qquad j = 1,...,n_i,\ i = 1,...,m,$$

where $\mu \in \mathscr{R}$ is an unknown parameter, A_i's are iid unobserved random variables having mean 0 and variance $\sigma_1^2 = \sigma_a^2$, e_{ij}'s are iid unobserved random errors with mean 0 and variance σ^2, and A_i's and e_{ij}'s are independent.

This is a special case of (6) with $k = 1$, X and ε being vectors of X_{ij}'s and e_{ij}'s, $Z = J_n$, $n = n_1 + \cdots + n_m$, U_1 being the block diagonal matrix whose ith block is J_{n_i}, $i = 1,...,m$, $\xi_1 = (A_1,...,A_m)$, $p = 1$, and $\beta = \mu$.

It is easy to see that $RSS(\beta) = n^{-1}\bar{X}^2$, where $\bar{X}$ is the mean of all X_{ij}'s.

It can be shown that

$$SS_\varepsilon = \sum_{i=1}^m \sum_{j=1}^{n_i} (X_{ij} - \bar{X}_i)^2, \quad SS_{\xi_1} = \sum_{i=1}^m n_i(\bar{X}_i - \bar{X})^2, \quad \bar{X}_i = \frac{1}{n_i}\sum_{j=1}^{n_i} X_{ij}$$

Also, the matrix $(Z, U_1) = (J_n, U_1)$ has rank m, so $r_1 = m - 1$, and

$$U_1^\tau D U_1 = U_1^\tau [I_n - J_n(J_n^\tau J_n)^{-1} J_n^\tau] U_1$$

$$= U_1^\tau U_1 - n^{-1} U_1^\tau J_n J_n^\tau U_1$$

$$= \begin{pmatrix} n_1 & & \\ & \ddots & \\ & & n_m \end{pmatrix} - n^{-1} \begin{pmatrix} n_1 \\ \vdots \\ n_m \end{pmatrix} (n_1 \cdots n_m)$$

Hence, $q = \mathrm{tr}(U_1^\tau D U_1) = n - (n_1^2 + \cdots + n_m^2)/n$ and

$$\widehat{\sigma}^2 = (n-m)^{-1} SS_\varepsilon, \quad \widehat{\sigma}_a^2 = [SS_{\xi_1} - (m-1)\widehat{\sigma}^2]/q$$

In this example, let's find out what the estimator of $\beta = \mu$ is.

First, we find $\widehat{\mu}_{V^{-1}} = (J_n^\tau V^{-1} J_n)^{-1} J_n^\tau V^{-1} X$ with $V = \sigma_a^2 U_1 U_1^\tau + \sigma^2 I_n$.

V is a block diagonal matrix with the ith diagonal block $\sigma^2 I_{n_i} + \sigma_a^2 J_{n_i} J_{n_i}^\tau$.

Using the formula

$$(A + BB^\tau)^{-1} = A^{-1} - A^{-1} B (I + B^\tau A^{-1} B)^{-1} B^\tau A^{-1}$$

we obtain that each block has the inverse

$$\left(\sigma^2 I_{n_i} + \sigma_a^2 J_{n_i} J_{n_i}^\tau\right)^{-1} = \frac{1}{\sigma^2} I_{n_i} - \frac{\sigma_a^2}{\sigma^2(\sigma^2 + n_i \sigma_a^2)} J_{n_i} J_{n_i}^\tau$$

Then, V^{-1} is the block diagonal matrix whose ith block diagonal is given by the previous expression, and

$$J_n^\tau V^{-1} J_n = \sum_{i=1}^m J_{n_i}^\tau \left[\frac{1}{\sigma^2} I_{n_i} - \frac{\sigma_a^2}{\sigma^2(\sigma^2 + n_i\sigma_a^2)} J_{n_i} J_{n_i}^\tau \right] J_{n_i}$$

$$= \left(\frac{n}{\sigma^2} - \frac{\sigma_a^2}{\sigma^2} \sum_{i=1}^m \frac{n_i^2}{\sigma^2 + n_i\sigma_a^2} \right) = \sum_{i=1}^m \frac{n_i}{\sigma^2 + n_i\sigma_a^2}$$

Writing $X_i = (X_{i1}, ..., X_{in_i})$, we obtain

$$J_n^\tau V^{-1} X = \sum_{i=1}^m J_{n_i}^\tau \left[\frac{1}{\sigma^2} I_{n_i} - \frac{\sigma_a^2}{\sigma^2(\sigma^2 + n_i\sigma_a^2)} J_{n_i} J_{n_i}^\tau \right] X_i$$

$$= \left(\frac{n\bar{X}}{\sigma^2} - \frac{\sigma_a^2}{\sigma^2} \sum_{i=1}^m \frac{n_i^2 \bar{X}_i}{\sigma^2 + n_i\sigma_a^2} \right) = \sum_{i=1}^m \frac{n_i \bar{X}_i}{\sigma^2 + n_i\sigma_a^2}$$

Thus, the WLSE of μ is

$$\widehat{\mu}_W = (J_n^\tau V^{-1} J_n)^{-1} J_n^\tau V^{-1} X \quad \sigma^2 = \widehat{\sigma}^2, \; \sigma_a^2 = \widehat{\sigma}_a^2$$

$$= \left(\sum_{i=1}^m \frac{n_i \bar{X}_i}{\widehat{\sigma}^2 + n_i \widehat{\sigma}_a^2} \right) \Bigg/ \left(\sum_{i=1}^m \frac{n_i}{\widehat{\sigma}^2 + n_i \widehat{\sigma}_a^2} \right)$$

Asymptotic normality of the WLSE $\widehat{\mu}_W$, $\widehat{\sigma}^2$ and $\widehat{\sigma}_a^2$ can be proved.