

## Beispiel Shrinkage

Betrachte das Random-Intercept-Modell ohne Kovariablen:

$$y_{ij} = b_i + \varepsilon_{ij}, \quad b_i \stackrel{iid}{\sim} N(0, \tau^2), \quad \varepsilon_{ij} \stackrel{iid}{\sim} N(0, \sigma^2), \\ j = 1, \dots, n_i, i = 1, \dots, N.$$

Dann ist

$$\mathbf{Z} = \begin{pmatrix} 1 & & & & \\ \vdots & & & & \\ 1 & & & & \\ & \ddots & & & \\ & & & & 1 \\ & & & & \vdots \\ & & & & 1 \end{pmatrix}, \quad \mathbf{G} = \tau^2 \mathbf{I}_N, \quad \mathbf{R} = \sigma^2 \mathbf{I}_n.$$

### 1 KQ-Schätzer für die $b_i$

Behandeln wir die  $b_i$  wie feste Effekte, so ist  $\mathbf{X} = \mathbf{Z}$ ,  $\text{Cov}(\mathbf{y}) = \mathbf{R}$ , und es ergibt sich der KQ-Schätzer

$$\hat{\mathbf{b}}^{KQ} = (\mathbf{Z}' \mathbf{R}^{-1} \mathbf{Z})^{-1} \mathbf{Z}' \mathbf{R}^{-1} \mathbf{y} = \left( \frac{1}{\sigma^2} \text{diag}(n_i) \right)^{-1} \frac{1}{\sigma^2} \begin{pmatrix} n_1 \bar{y}_1 \\ \vdots \\ n_N \bar{y}_N \end{pmatrix} = \begin{pmatrix} \bar{y}_1 \\ \vdots \\ \bar{y}_N \end{pmatrix}$$

mit  $\bar{y}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} y_{ij}$ , also  $\hat{b}_i^{KQ} = \bar{y}_i$ . Damit ist

- $E(\hat{b}_i^{KQ}) = \frac{1}{n_i} \sum_{j=1}^{n_i} b_i = b_i$  (kein Bias)
- $\text{Var}(\hat{b}_i^{KQ}) = \frac{1}{n_i^2} \sum_{j=1}^{n_i} \sigma^2 = \frac{1}{n_i} \sigma^2$
- $\text{MSE}(\hat{b}_i^{KQ}) = E[(\hat{b}_i^{KQ} - b_i)^2] = \text{Bias}(\hat{b}_i^{KQ})^2 + \text{Var}(\hat{b}_i^{KQ}) = 0 + \frac{1}{n_i} \sigma^2 = \frac{1}{n_i} \sigma^2$ .

### 2 BLUPs für die $b_i$

Behandeln wir die  $b_i$  als zufällige Effekte mit Verteilungsannahme  $b_i \stackrel{iid}{\sim} N(0, \tau^2)$ , so ergibt sich der BLUP für  $\mathbf{b}$  als

$$\hat{\mathbf{b}}^{BLUP} = (\mathbf{Z}' \mathbf{R}^{-1} \mathbf{Z} + \mathbf{G}^{-1})^{-1} \mathbf{Z}' \mathbf{R}^{-1} \mathbf{y} = \left( \frac{1}{\sigma^2} \text{diag}(n_i) + \frac{1}{\tau^2} \mathbf{I}_N \right)^{-1} \frac{1}{\sigma^2} \begin{pmatrix} n_1 \bar{y}_1 \\ \vdots \\ n_N \bar{y}_N \end{pmatrix} = \text{diag} \left( \frac{n_i}{n_i + \sigma^2 / \tau^2} \right) \begin{pmatrix} \bar{y}_1 \\ \vdots \\ \bar{y}_N \end{pmatrix},$$

also  $\hat{b}_i^{BLUP} = \frac{n_i}{n_i + \sigma^2 / \tau^2} \bar{y}_i = c_i \bar{y}_i$  mit  $0 < c_i := \frac{n_i}{n_i + \sigma^2 / \tau^2} < 1$  (Shrinkage). Damit ist

- $E(\hat{b}_i^{BLUP}|b_i) = c_i \frac{1}{n_i} \sum_{j=1}^{n_i} b_i = c_i b_i \neq b_i$  mit  $|c_i b_i| < |b_i|$  (Bias durch Shrinkage)  
Die BLUPs streuen also weniger als die KQ-Schätzer.
- $\text{Var}(\hat{b}_i^{BLUP}|b_i) = c_i^2 \frac{1}{n_i^2} \sum_{j=1}^{n_i} \sigma^2 = c_i^2 \frac{1}{n_i} \sigma^2 < \frac{1}{n_i} \sigma^2$ . Die Varianz ist also kleiner als beim KQ-Schätzer.
- Der MSE betrachtet einen Kompromiss aus Varianz und Bias:

$$\begin{aligned}
\text{MSE}(\hat{b}_i^{BLUP}) &= E_y[(\hat{b}_i^{BLUP} - b_i)^2] = E_b E_{y|b}[(\hat{b}_i^{BLUP} - b_i)^2] \\
&= E_b E_{y|b}[(\hat{b}_i^{BLUP} - E(\hat{b}_i^{BLUP}|b_i) + E(\hat{b}_i^{BLUP}|b_i) - b_i)^2] \\
&\stackrel{(*)}{=} E_b E_{y|b}[(\hat{b}_i^{BLUP} - E(\hat{b}_i^{BLUP}|b_i))^2 + (E(\hat{b}_i^{BLUP}|b_i) - b_i)^2] \\
&= E_b[\text{Var}(\hat{b}_i^{BLUP}|b_i) + (E(\hat{b}_i^{BLUP}|b_i) - b_i)^2] \\
&= c_i^2 \frac{1}{n_i} \sigma^2 + E_b[(1 - c_i)^2 b_i^2] = c_i^2 \frac{1}{n_i} \sigma^2 + (1 - c_i)^2 \tau^2 \\
&= \frac{n_i \sigma^2}{(n_i + \sigma^2/\tau^2)^2} + \frac{(\sigma^2/\tau^2)^2 \tau^2}{(n_i + \sigma^2/\tau^2)^2} \\
&= \frac{\sigma^2(n_i + \sigma^2/\tau^2)}{(n_i + \sigma^2/\tau^2)^2} = \frac{\sigma^2}{n_i + \sigma^2/\tau^2} < \frac{1}{n_i} \sigma^2,
\end{aligned}$$

da  $\sigma^2/\tau^2 > 0$ . (\*) gilt dabei wegen  $E_{y|b}[(\hat{b}_i^{BLUP} - E(\hat{b}_i^{BLUP}|b_i))(E(\hat{b}_i^{BLUP}|b_i) - b_i)] = 0$ .  
Im MSE wird also der Bias des BLUP durch die kleinere Varianz mehr als aufgehoben.

### 3 Zusammenfassung

Für den BLUP  $\hat{\mathbf{b}}^{BLUP}$  und den KQ-Schätzer  $\hat{\mathbf{b}}^{KQ}$ , der  $\mathbf{b}$  als feste Effekte schätzt, gilt

	$\hat{\mathbf{b}}^{BLUP}$	$\hat{\mathbf{b}}^{KQ}$	Kommentar zu $\hat{\mathbf{b}}^{BLUP}$
$\hat{b}_i$	$\frac{n_i}{n_i + \sigma^2/\tau^2} \bar{y}_i$	$\bar{y}_i$	Shrinkage
$E(\hat{b}_i b_i)$	$\frac{n_i}{n_i + \sigma^2/\tau^2} b_i$	$b_i$	Bias
$\text{Var}(\hat{b}_i b_i)$	$(\frac{n_i}{n_i + \sigma^2/\tau^2})^2 \frac{1}{n_i} \sigma^2$	$\frac{1}{n_i} \sigma^2$	Varianz kleiner
$\text{MSE}(\hat{b}_i)$	$\frac{\sigma^2}{n_i + \sigma^2/\tau^2}$	$\frac{\sigma^2}{n_i}$	MSE kleiner trotz Bias

Dabei ist der Unterschied zwischen den beiden Schätzern größer, je kleiner  $n_i$  ist und je größer  $\sigma^2/\tau^2$ . Für diese Fälle ist die Information in den  $y_{ij}$  pro  $i$  am geringsten und man gewinnt am meisten durch den Shrinkage-Effekt. Der KQ-Schätzer ( $c_i \rightarrow 1$ ) ergibt sich als Grenzfall des BLUPs für  $\tau^2 \rightarrow \infty$ .