The Constraint-Case Formulation of MLMs

Here's another way to write MLMs, which sometimes has advantages.

Consider the balanced one-way random effect model:

$$
\begin{aligned}
y_{ij} &= \theta_i + \epsilon_{ij}, \text{ where } \epsilon_{ij} \overset{iid}{\sim} N(0, \sigma_e^2) && (1) \\
\theta_i &= \mu + \delta_i, \text{ where } \delta_i \overset{iid}{\sim} N(0, \sigma_s^2) && (2) \\
\mu &= M + \xi, \text{ where } \xi \sim N(0, \sigma_p^2) && (3) \\
& \text{for } i = 1, \ldots, q \text{ and } j = 1, \ldots, m.
\end{aligned}
$$

M, σ_p^2 are known; $\theta_i, \epsilon_{ij}, \mu$, δ_i, σ_e^2, σ_s^2 are unknown.

Rewrite (2) and (3) as, respectively,

$$
\begin{aligned}
0 &= -\theta_i + \mu + \delta_i && (4) \\
M &= \mu - \xi. && (5)
\end{aligned}
$$

Equations (1), (4), and (5) now have the form of a linear model.

Equations (1), (4), and (5) have the form of a linear model.

$$\left[\begin{array}{c} \mathbf{y} \\ \hline \mathbf{0}_q \\ \hline M \end{array}\right] = \left[\begin{array}{c|c} \mathbf{I}_q \otimes \mathbf{1}_m & \mathbf{0}_{qm} \\ \hline -\mathbf{I}_q & \mathbf{1}_q \\ \hline \mathbf{0}_q' & 1 \end{array}\right] \left[\begin{array}{c} \theta_1 \\ \vdots \\ \theta_q \\ \hline \mu \end{array}\right] + \left[\begin{array}{c} \boldsymbol{\epsilon} \\ \hline \boldsymbol{\delta} \\ \hline -\xi \end{array}\right],$$

where $\otimes$ is the Kronecker product $\mathbf{A} \otimes \mathbf{B} = (a_{ij}\mathbf{B})$.

Left side: known $(qm + q + 1)$-vector.

Right side: $(qm + q + 1) \times (q + 1)$ design matrix times
$(q + 1)$-vector of coefficients, plus $(qm + q + 1)$-vector of errors.

The error vector has diagonal covariance matrix $\left[\begin{array}{ccc} \sigma_e^2 \mathbf{I}_{qm} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \sigma_s^2 \mathbf{I}_q & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \sigma_p^2 \end{array}\right]$

Nothing deep here; it's an "accounting identity" (Whittaker).

There's more than one way to do this

Write the balanced one-way RE model in the standard MLM form:

$\mathbf{X} = \mathbf{1}_{qm}$, $\boldsymbol{\beta} = \mu$, $\mathbf{Z} = \mathbf{I}_q \otimes \mathbf{1}_m$, $\mathbf{u} = (\delta_1, \ldots, \delta_q)'$, $\mathbf{G} = \sigma_s^2 \mathbf{I}_q$, $\mathbf{R} = \sigma_e^2 \mathbf{I}_{qm}$.

That implies these three equations:

$$\begin{aligned}
\mathbf{y} &= \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u} + \boldsymbol{\epsilon}, \text{ where } \boldsymbol{\epsilon} \sim N_{qm}(0, \sigma_e^2 \mathbf{I}_{qm}) && (6)\\
\mathbf{u} &= \boldsymbol{\delta}, \text{ where } \boldsymbol{\delta} \sim N_q(0, \sigma_s^2 \mathbf{I}_q) && (7)\\
\mu &= M + \xi, \text{ where } \xi \sim N(0, \sigma_p^2), && (8)
\end{aligned}$$

Using the same trick as above, rewrite (7) and (8) as

$$\begin{aligned}
\mathbf{0}_q &= -\mathbf{u} + \boldsymbol{\delta} && (9)\\
M &= \mu - \xi. && (10)
\end{aligned}$$

Now stack (6), (9), and (10).

Stack (6), (9), and (10):

$$\left[\begin{array}{c} \mathbf{y} \\ \hline \mathbf{0}_q \\ \hline M \end{array}\right] = \left[\begin{array}{c|c} \mathbf{1}_{qm} & \mathbf{I}_q \otimes \mathbf{1}_m \\ \hline \mathbf{0}_q & -\mathbf{I}_q \\ \hline 1 & \mathbf{0}_q' \end{array}\right] \left[\begin{array}{c} \mu \\ \hline \mathbf{u} \end{array}\right] + \left[\begin{array}{c} \boldsymbol{\epsilon} \\ \hline \boldsymbol{\delta} \\ \hline -\xi \end{array}\right].$$

Again, this has the form of a linear model with heteroscedastic errors.

All MLMs can be written in constraint-case form as

$$\left[\begin{array}{c} \mathbf{y} \\ \hline \mathbf{0}_q \\ \hline M \end{array}\right] = \left[\begin{array}{c|c} \mathbf{X} & \mathbf{Z} \\ \hline \mathbf{0}_q & -\mathbf{I}_q \\ \hline \mathbf{I}_p & \mathbf{0}_{p\times q} \end{array}\right] \left[\begin{array}{c} \boldsymbol{\beta} \\ \hline \mathbf{u} \end{array}\right] + \left[\begin{array}{c} \boldsymbol{\epsilon} \\ \hline \boldsymbol{\delta} \\ \hline -\boldsymbol{\xi} \end{array}\right]$$

$$\text{cov}\left(\begin{array}{c} \boldsymbol{\epsilon} \\ \hline \boldsymbol{\delta} \\ \hline -\boldsymbol{\xi} \end{array}\right) = \left[\begin{array}{ccc} \mathbf{R} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{G} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \boldsymbol{\Sigma} \end{array}\right].$$

Some jargon for the constraint-case formulation

An MLM written in constraint-case form:

$$\left[\begin{array}{c} \mathbf{y} \\ \hline \mathbf{0}_q \\ \hline M \end{array}\right] = \left[\begin{array}{c|c} \mathbf{X} & \mathbf{Z} \\ \hline \mathbf{0}_q & -\mathbf{I}_q \\ \hline \mathbf{I}_p & \mathbf{0}_{p\times q} \end{array}\right] \left[\begin{array}{c} \boldsymbol{\beta} \\ \hline \mathbf{u} \end{array}\right] + \left[\begin{array}{c} \boldsymbol{\epsilon} \\ \hline \boldsymbol{\delta} \\ \hline -\boldsymbol{\xi} \end{array}\right] \quad \begin{array}{l} \text{data cases} \\ \text{constraint cases} \\ \text{prior cases} \end{array}$$

$$\text{cov}\left(\begin{array}{c} \boldsymbol{\epsilon} \\ \hline \boldsymbol{\delta} \\ \hline -\boldsymbol{\xi} \end{array}\right) = \left[\begin{array}{ccc} \mathbf{R} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{G} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \boldsymbol{\Sigma} \end{array}\right].$$

This formulation, conditioning on **R** and **G**, makes some derivations easy in the conventional theory.

It's also been used to speed computing, by Henderson et al (1959) and in lme4 (Bates & DebRoy 2004).

Sometimes it's easier to write a model this way

Plots in a field are in one long row, labeled $i = 1, ..., n$.

Two treatments are allocated randomly to plots, $T_i = 0$ or 1.

F_i is plot i's unobserved fertility: $F_i = F_{i-1} + \delta_i$, where $\delta_i \overset{iid}{\sim} N(0, \sigma_s^2)$.

Model the yield in plot i as $y_i = T_i\beta + F_i + \epsilon_i$, where $\epsilon_i \overset{iid}{\sim} N(0, \sigma_e^2)$.

Rewrite the model for F_i as $0 = -F_i + F_{i-1} + \delta_i, i = 2, \ldots, n$.

Put a $N(M, \sigma_p^2)$ prior on β and stack these "cases":

$$\begin{bmatrix} \mathbf{y} \\ \hline \mathbf{0}_{n-1} \\ \hline M \end{bmatrix} = \left[\begin{array}{c|cccccc} \mathbf{T} & & & \mathbf{I}_n & & & \\ \hline & 1 & -1 & 0 & \ldots & 0 & 0 \\ & 0 & 1 & -1 & \ldots & 0 & 0 \\ \mathbf{0}_{n-1} & & \vdots & & \ddots & & \vdots \\ & 0 & 0 & 0 & \ldots & 1 & -1 \\ \hline 1 & & & \mathbf{0}_{1\times n} & & & \end{array}\right] \begin{bmatrix} \boldsymbol{\beta} \\ \hline F_1 \\ \vdots \\ F_n \end{bmatrix} + \begin{bmatrix} \boldsymbol{\epsilon} \\ \hline \boldsymbol{\delta} \\ \hline -\boldsymbol{\xi} \end{bmatrix},$$

Much simpler than the MLM formulation.

Measuring model complexity: Degrees of freedom (DF)

DF are used to describe the complexity of an MLM fit.

For mixed linear models, DF are used to:

- Specify F-tests.
- Describe a model's size to penalize it in a model-selection criterion.
- Specify a prior distribution on ϕ, the unknowns in $\mathbf{G}$ and $\mathbf{R}$.

I'll emphasize using DF to specify priors for the unknowns in $\mathbf{G}$ and $\mathbf{R}$.

DF can also be used to measure the complexity of *parts* of a fit.

Motivation

Consider again the balanced one-way random effects model:

$$\begin{aligned} y_{ij} &= \theta_i + \epsilon_{ij}, \text{ where } \epsilon_{ij} \overset{iid}{\sim} N(0, \sigma_e^2) \\ \theta_i &= \mu + \delta_i, \text{ where } \delta_i \overset{iid}{\sim} N(0, \sigma_s^2) \\ & \quad \text{for } i = 1, \ldots, q \text{ and } j = 1, \ldots, m. \end{aligned}$$

The fitted values are $\hat{y}_{ij} = \hat{\mu} + \hat{\delta}_i$. What is the fit's complexity?

If $\hat{\sigma}_s^2 \to \infty$ for fixed $\hat{\sigma}_e^2$, then $\hat{y}_{ij} = \bar{y}_{i.}$ and this fit has q DF.

If $\hat{\sigma}_s^2 \to 0$ for fixed $\hat{\sigma}_e^2$, then $\hat{y}_{ij} = \bar{y}_{..}$ and this fit has 1 DF.

It seems awkward to suggest that the fit's complexity changes discontinuously at either extreme.

We'll define a continuous complexity measure instead.

Motivating a more general DF measure

OLS regression: $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$: the fitted values are $\hat{\mathbf{y}} = \mathbf{H}\mathbf{y}$ for $\mathbf{H} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$. The DF in the fit is rank($\mathbf{X}$) = trace($\mathbf{H}$).

Linear smoother: $\hat{\mathbf{y}} = \mathbf{S}_\lambda \mathbf{y}$, where λ is a known tuning parameter. By analogy with linear models, the DF in the fit is trace($\mathbf{S}_\lambda$).

An MLM is a linear smoother with $\mathbf{C} = [\mathbf{X}|\mathbf{Z}]$, $\lambda = (\phi_G, \phi_R)$, and

$$\mathbf{S}_\lambda = \mathbf{C}\left[\mathbf{C}'\mathbf{R}^{-1}\mathbf{C} + \begin{pmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{G}^{-1} \end{pmatrix}\right]^{-1}\mathbf{C}'\mathbf{R}^{-1}$$

Thus the DF in an MLM fit is

$$\text{trace}(\mathbf{S}_\lambda) = \text{trace}\left(\mathbf{C}\left[\mathbf{C}'\mathbf{R}^{-1}\mathbf{C} + \begin{pmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{G}^{-1} \end{pmatrix}\right]^{-1}\mathbf{C}'\mathbf{R}^{-1}\right)$$

$\Rightarrow$ DF is a function of ϕ_G and ϕ_R.

Example: Balanced one-way RE model (BOWREM)

BOWREM in standard form:

$\mathbf{X} = \mathbf{1}_{qm}$, $\boldsymbol{\beta} = \mu$, $\mathbf{Z} = \mathbf{I}_q \otimes \mathbf{1}_m$, $\mathbf{u} = (\delta_1, \ldots, \delta_q)'$, $\mathbf{G} = \sigma_s^2 \mathbf{I}_q$, $\mathbf{R} = \sigma_e^2 \mathbf{I}_{qm}$.

For $\mathbf{C} = [\mathbf{X}|\mathbf{Z}]$, the DF in the BOWREM fit is

$$\begin{aligned}
&= \text{trace}\left(\mathbf{C}\left[\mathbf{C}'\mathbf{R}^{-1}\mathbf{C} + \begin{pmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{G}^{-1} \end{pmatrix}\right]^{-1}\mathbf{C}'\mathbf{R}^{-1}\right) \\
(\textit{homework}) &= \left[\begin{pmatrix} qm & m\mathbf{1}_q' \\ m\mathbf{1}_q & m\mathbf{I}_q \end{pmatrix} + \begin{pmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_q \sigma_e^2/\sigma_s^2 \end{pmatrix}\right]^{-1} \begin{pmatrix} qm & m\mathbf{1}_q' \\ m\mathbf{1}_q & m\mathbf{I}_q \end{pmatrix} \\
(\textit{homework}) &= 1 + (q-1)m/(m+r) \text{ for } r = \sigma_e^2/\sigma_s^2
\end{aligned}$$

Example: Balanced one-way RE model (BOWREM)

DF in BOWREM fit: $= 1 + (q-1)m/(m+r)$ for $r = \sigma_e^2/\sigma_s^2$

This has some features that are true about DF much more generally.

- DF $\in [1, q]$ and increases continuously with σ_s^2 for given σ_e^2, as our motivation suggested it should.

- For models with normal errors and random effects, DF is a function of the *ratio* of variances $r = \sigma_s^2/\sigma_e^2$, not the individual variances.

Example: Plots in a field

Yield in plot i is $y_i = T_i\beta + F_i + \epsilon_i,$ where $\epsilon_i \overset{iid}{\sim} N(0, \sigma_e^2),$ $T_i = 0$ or 1

$F_i = F_{i-1} + u_i$, where $u_i \overset{iid}{\sim} N(0, \sigma_s^2)$, so $F_i = F_1 + u_2 + \cdots + u_i$, $i \geq 2$

Thus $y_i = F_1 + T_i\beta + \sum_{j=2}^{i} u_j + \epsilon_i$

In the standard form:

$$\begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} 1 & T_1 \\ \vdots & \\ 1 & T_n \end{bmatrix} \begin{bmatrix} F_1 \\ \beta \end{bmatrix} + \begin{bmatrix} 0 & 0 & 0 & & 0 \\ 1 & 0 & 0 & \dots & 0 \\ 1 & 1 & 0 & & 0 \\ 1 & 1 & 1 & & 0 \\ & \vdots & & \ddots & \vdots \\ 1 & 1 & 1 & \dots & 1 \end{bmatrix} \begin{bmatrix} u_2 \\ \vdots \\ u_n \end{bmatrix} + \boldsymbol{\epsilon}$$

with $\mathbf{G} = \sigma_s^2 \mathbf{I}_{n-1}$ and $\mathbf{R} = \sigma_e^2 \mathbf{I}_n$.

Example: Plots in a field (2)

$$\begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} 1 & T_1 \\ \vdots & \\ 1 & T_n \end{bmatrix} \begin{bmatrix} F_1 \\ \beta \end{bmatrix} + \begin{bmatrix} 0 & 0 & 0 & & 0 \\ 1 & 0 & 0 & \dots & 0 \\ 1 & 1 & 0 & & 0 \\ 1 & 1 & 1 & & 0 \\ & \vdots & & \ddots & \vdots \\ 1 & 1 & 1 & \dots & 1 \end{bmatrix} \begin{bmatrix} u_2 \\ \vdots \\ u_n \end{bmatrix} + \boldsymbol{\epsilon}$$

with $\mathbf{G} = \sigma_s^2 \mathbf{I}_{n-1}$ and $\mathbf{R} = \sigma_e^2 \mathbf{I}_n$.

The DF in this fit is $2 + \sum_{j=1}^{n-2} \left[1 + \frac{\sigma_e^2}{\sigma_s^2} \frac{1}{d_j} \right]^{-1}$

where the d_j are the eigenvalues of $\mathbf{Z}'(\mathbf{I} - \mathbf{P}_X)\mathbf{Z}$.

<u>Intuition</u>: Along the j^{th} singular vector of $(\mathbf{I} - \mathbf{P}_X)\mathbf{Z}$, the fit is shrunk to

$\left[1 + \frac{\sigma_e^2}{\sigma_s^2} \frac{1}{d_j} \right]^{-1}$ of its original length.

DF is a convenient way to put a prior on (ϕ_G, ϕ_R)

The idea:

- Put a prior on DF $\equiv$ DF(ϕ_G, ϕ_R), about which you have intuition;
- This induces a prior on (ϕ_G, ϕ_R), at least partly.

Example: 1-way RE model, q groups, m observations/group

DF$(r) = 1 + (q-1)m/(m+r)$ for $r = \sigma_e^2/\sigma_s^2$

Flat prior on DF: $F(\text{DF} \leq x) = x/(q-1)$ for $x \in [1, q]$

$\Rightarrow \text{Prob}(r \leq \xi) = \xi/(m+\xi)$ for $\xi \in (0, \infty)$.

Interpretable alternative to a prior on (σ_s^2, σ_e^2):

Re-parameterize to (DF, σ_e^2), put independent priors on DF and σ_e^2.

Cui et al (2010) treats this much more generally and has cool examples.