On the inefficiency
of the restricted maximum likelihood

Nicholas T. Longford,
SNTL and UPF, Barcelona, Spain

Summary

The restricted maximum likelihood is preferred by many to the full maximum likelihood for estimation with variance component and other random coefficient models, because the variance estimator is unbiased. It is shown that this unbiasedness is accompanied in some balanced designs by an inflation of the mean squared error. An estimator of the cluster-level variance that is uniformly more efficient than the full maximum likelihood is derived. Estimators of the variance ratio are also studied.

Key Words: *efficiency, random effects, truncation, variance component.*

Mathematical Sciences Classification: 62F10, 62J10

N. T. Longford is Company Director, SNTL, and Academic Visitor in the Department of Economics and Business, Universitat Pompeu Fabra, c/ Ramon Trias Fargas 25 – 27, 08005 Barcelona, Spain. Email: sntlnick@sntl.co.uk.

1 Introduction

Patterson and Thompson (1971) introduced the idea of estimating variance components in random coefficient models solely from the error contrasts. This method, restricted maximum likelihood (REML), has been widely adopted since, assisted by a simple adjustment of the likelihood derived by Harville (1974); see also Harville (1977). For background, theory and applications of REML, see Searle, Casella and McCulloch (1992), Verbyla (1993), Kenward and Roger (1997), Verbyla *et al.* (1999), Diggle *et al.* (2002) and Jiang (2007). This article compares analytically the efficiency of the maximum likelihood (ML) and REML estimators of the cluster-level variance in balanced linear models with one set of random effects. The estimators of the variance ratio are also compared. Some of our results are derived by Klotz, Milton and Zacks (1969) and Kubokawa (1995) for the variance and by Loh (1986) for the variance ratio. Swallow and Monahan (1984) studied the ML, REML and several other estimators of the variance by simulations for balanced and unbalanced one-way layout. Our results are in accord with theirs for the settings that we have in common. They undermine the often-made claim that no information is lost by restricting the analysis to the error contrasts.

In the model and the design we consider, there are m observations $(y_{1k}, \ldots, y_{mk})$ in each cluster $k = 1, \ldots, K$. They are conditionally independent and normally distributed with conditional expectation $\mathbf{x}_{ik}\boldsymbol{\beta} + \delta_k$ and variance σ_{W}^2 given δ_k; $y_{ik} = \mathbf{x}_{ik}\boldsymbol{\beta} + \delta_k + \varepsilon_{ik}$, where $\mathbf{x}_{ik}$ is the vector of the values of the covariates (including the intercept) for the elementary unit (i, k) and $\boldsymbol{\beta}$ is the vector of regression parameters. The cluster-level deviations δ_k are a random sample from a centered normal distribution with variance σ_{B}^2. The $Km + K$ random terms ε_{ik} and δ_k are mutually independent. The variance ratio is $\omega = \sigma_{\mathrm{B}}^2/\sigma_{\mathrm{W}}^2$. We consider $(\sigma_{\mathrm{W}}^2, \omega)$ as an alternative parametrization to $(\sigma_{\mathrm{W}}^2, \sigma_{\mathrm{B}}^2)$. We show that the estimators of ω are linear functions of F-distributed random variables and estimators of σ_{B}^2 are linear combinations of two independent χ^2- distributed variables. These distributional identities are key to our derivations.

In Sections $2-4$, only models with no covariates, in which $\mathbf{x}_{ik}\boldsymbol{\beta}$ is constant, are considered. Section 2 derives the ML and REML estimators of σ_{B}^2 as functions of the

within- and between-cluster sums of squares. Section 3 evaluates the bias and efficiency of the estimators of ω and shows that the ML estimators are uniformly more efficient than their REML counterparts. A simulation study in Section 4 compares the proper versions of these estimators, which are truncated to be nonnegative. It concludes that ML retains its advantage over REML while the truncated REML estimators are, of course, biased. Estimators of σ_B^2 and ω more efficient than ML are derived in respective Sections 2 and 3. The efficiencies of the estimators of σ_B^2 and ω are compared empirically for a particular unbalanced design in Section 5. Section 6 extends some of the results to models with covariates subject to some constraints related to their balance. Throughout, we regard the mean squared error (MSE) as the sole criterion for the quality of an estimator.

2 The estimators

With no covariates, the loglikelihood for the outcomes y_{ik} is

$$ l \;=\; -\frac{1}{2}\left\{ n\log\!\left(\sigma_W^2\right) + \log(\det \mathbf{W}) + \frac{1}{\sigma_W^2}\,\mathbf{e}_\bullet^\top \mathbf{W}^{-1}\mathbf{e}_\bullet\right\}, \tag{1}$$

where $\mathbf{W} = \sigma_W^{-2}\mathrm{var}(\mathbf{y})$ is the scaled variance matrix of the $n = Km$ observations and $\mathbf{e}_\bullet = \mathbf{y} - \mu\mathbf{1}_n$ is the vector of residuals ($\mathbf{1}_n$ is the vector of unities of length n). Let $\mathbf{e} = \mathbf{y} - \hat\mu\mathbf{1}_n$, where $\hat\mu = n^{-1}\mathbf{y}^\top\mathbf{1}_n$. Denote by S_W and S_B the within- and between-cluster sums of squares of the residuals:

$$
\begin{aligned}
S_W &= \sum_{k=1}^{K}\sum_{i=1}^{m}(e_{ik} - \bar{e}_k)^2 \\
S_B &= m\sum_{k=1}^{K}\bar{e}_k^2,
\end{aligned}
\tag{2}
$$

where $e_{ik} = y_{ik} - \hat\mu$ and $\bar{e}_k = (e_{1k} + \cdots + e_{mk})/m$ is the average residual in cluster k. We have the identity $S_W + S_B = \mathbf{e}^\top\mathbf{e}$ for the overall sum of squares of the residuals. With the balanced design, S_W and S_B are independent and have scaled χ^2 distributions;

$$
\begin{aligned}
\frac{S_W}{\sigma_W^2} &\sim \chi^2_{n-K} \\
\frac{S_B}{\sigma_W^2 + m\sigma_B^2} &\sim \chi^2_{K-1}.
\end{aligned}
\tag{3}
$$

2

In Appendix A, we derive the following expressions for the ML estimators:

$$\hat{\omega} = \frac{m-1}{m} \frac{S_{\mathrm{B}}}{S_{\mathrm{W}}} - \frac{1}{m}$$

$$\hat{\sigma}_{\mathrm{B}}^2 = \frac{S_{\mathrm{B}}}{n} - \frac{S_{\mathrm{W}}}{m(n-K)} . \tag{4}$$

Unlike their respective targets ω and σ_{B}^2, these estimators attain negative values with positive probabilities, and are therefore improper. Proper ML estimators are obtained by truncating them at zero.

An error contrast is defined as a linear combination of the observations y_{ik} that has zero expectation. Error contrasts define a linear space. In our case, its dimension is $n-1$. In REML, a basis of this linear space is used as the set of observations. The loglikelihood for such a basis is obtained by adjusting the loglikelihood l in (1) by

$$\Delta l = \frac{1}{2} \log\left(\frac{1}{\sigma_{\mathrm{W}}^2} \mathbf{1}_n^\top \mathbf{W}^{-1} \mathbf{1}_n \right)$$

(Harville, 1974), apart from an additive constant that is irrelevant for maximizing $l_{\mathrm{RE}} = l + \Delta l$, which yields the REML estimators. In Appendix B, we show that these estimators are

$$\hat{\omega}_{\mathrm{RE}} = \frac{n-K}{m(K-1)} \frac{S_{\mathrm{B}}}{S_{\mathrm{W}}} - \frac{1}{m}$$

$$\hat{\sigma}_{\mathrm{B,RE}}^2 = \frac{S_{\mathrm{B}}}{m(K-1)} - \frac{S_{\mathrm{W}}}{m(n-K)} , \tag{5}$$

although they are improper and have to be truncated at zero. We retain the notation in (4) and (5) for the improper estimators and indicate their truncation by adding 0 in the subscripts. Approximate or exact unbiasedness applies to the improper estimators.

The estimators $\hat{\sigma}_{\mathrm{B}}^2$ and $\hat{\sigma}_{\mathrm{B,RE}}^2$ have the same form $\tilde{\sigma}_{\mathrm{B}}^2(a) = a S_{\mathrm{B}} - S_{\mathrm{W}}/m/(n-K)$, with $a = a_{\mathrm{ML}} = 1/n$ and $a_{\mathrm{RE}} = 1/(n-m)$, respectively. An (improper) estimator of σ_{B}^2 more efficient than both $\hat{\sigma}_{\mathrm{B}}^2$ and $\hat{\sigma}_{\mathrm{B,RE}}^2$ is found by identifying the constant a for which the MSE of $\tilde{\sigma}_{\mathrm{B}}^2(a)$ is minimized. In Appendix C, we show that the (unique) minimum is attained for $a^* = 1/(n+m)$. Since $\mathrm{MSE}\{\tilde{\sigma}_{\mathrm{B}}^2(a)\}$ is a quadratic function of a, the inequality $a^* < a_{\mathrm{ML}} < a_{\mathrm{RE}}$ implies that $\hat{\sigma}_{\mathrm{B}}^2$ is uniformly more efficient than $\hat{\sigma}_{\mathrm{B,RE}}^2$ for all one-way balanced designs with no covariates.

In practice, the estimators $\hat{\sigma}_{\mathrm{B}}^2$ and $\hat{\sigma}_{\mathrm{B,RE}}^2$ are truncated, and therefore the appeal to unbiasedness prior to truncation should carry little weight. At the same time,

truncation makes our comparison of the MSEs incomplete and imperfect. Equations (4) and (5) imply that $\hat{\sigma}_B^2 < \hat{\sigma}_{B,RE}^2$; therefore $\hat{\sigma}_B^2 < 0$ whenever $\hat{\sigma}_{B,RE}^2 < 0$. A simulation study in Section 4 shows that the relative efficiency of $\hat{\sigma}_B^2$ is retained by truncation.

3 Efficiency of the estimators of ω

The improper ML and REML estimators of the variance ratio ω are linear functions of the ratio S_B/S_W which, for the balanced design with no covariates, has a scaled F distribution with $K-1$ and $n-K$ degrees of freedom:

$$(X =) \quad \frac{1}{1+m\omega}\frac{n-K}{K-1}\frac{S_B}{S_W} \sim F_{K-1,n-K}\,. \tag{6}$$

Denote $u_{RE} = 1/(1+m\omega)$ and $u = Ku_{RE}/(K-1)$, so that

$$\hat{\omega} = \frac{X-u}{mu}$$

$$\hat{\omega}_{RE} = \frac{X-u_{RE}}{mu_{RE}}\,.$$

In Appendix D, we derive the biases of these estimators, $-(m-3)D_1$ and $2D_1$, respectively, where $D_1 = (1/m+\omega)/(n-K-2)$, and their MSEs,

$$(M =) \quad \mathrm{MSE}(\hat{\omega}) = D_1^2\left\{\frac{2(K-1)(m-1)^2(n-3)}{n-K-4}+(m-3)^2\right\}$$

$$(R =) \quad \mathrm{MSE}(\hat{\omega}_{RE}) = D_1^2\left\{\frac{2(n-K)^2(n-3)}{(K-1)(n-K-4)}+4\right\}\,.$$

Let $D_2 = D_1^2(m-1)/\{(K-1)(n-K-4)\}$. The difference of the MSEs,

$$\begin{aligned} M-R &= D_1^2\left[\frac{2(m-1)^2(n-3)}{n-K-4}\left(K-1-\frac{K^2}{K-1}\right)+\left\{(m-3)^2-4\right\}\right]\\ &= D_2\left\{(K-1)(m-5)(n-K-4)-(4K-2)(m-1)(n-3)\right\}\,, \end{aligned}$$

is negative for all K and m because each factor of the first term in the braces in the second line is smaller than its counterpart in the second term. Thus, $\hat{\omega}$ is uniformly more efficient than $\hat{\omega}_{RE}$ in all one-way balanced designs. As an aside, we note that the absolute bias of $\hat{\omega}_{RE}$ is smaller than for $\hat{\omega}$ only for $m > 5$.

The improper estimators $\hat{\omega}$ and $\hat{\omega}_{RE}$ have the common lower bound of $-1/m$. This motivates our search for the efficient estimator of ω in the class of estimators $\tilde{\omega}(b) =$

4

$bS_{\mathrm{B}}/S_{\mathrm{W}} - 1/m$. In Appendix D, we show that the MSE of these estimators is minimized for $b^* = (n - K - 4)/(n + m)$, which is smaller than its counterparts for $\hat{\omega}$ and $\hat{\omega}_{\mathrm{RE}}$, equal to $b_{\mathrm{ML}} = (m-1)/m$ and $b_{\mathrm{RE}} = b_{\mathrm{ML}}K/(K-1)$. Neither $\hat{\omega}$ nor $\hat{\omega}_{\mathrm{RE}}$ is unbiased, but $\hat{\omega}$ is uniformly more efficient than $\hat{\omega}_{\mathrm{RE}}$. The estimator $\tilde{\omega}^* = \tilde{\omega}(b^*)$ can be interpreted as a shrinkage estimator with the focus of shrinkage at $-1/m$.

When σ_{W}^2 is known, ML and REML estimators of ω are obtained directly as the roots of the corresponding score functions. They are $\tilde{\omega}(c) = cS_{\mathrm{B}}/\sigma_{\mathrm{W}}^2 - 1/m$, with $c = 1/n$ for ML and $c = 1/(n - m)$ for REML; see Appendices A and B. The distributions of these estimators are linear transformations of χ_{K-1}^2, and so their biases and MSEs are derived straightforwardly. In general,

$$\mathrm{MSE}\left\{\tilde{\omega}(c); \sigma_{\mathrm{W}}^2\right\} = \left\{c^2\left(K^2 - 1\right) - 2c\,\frac{K-1}{m} + \frac{1}{m^2}\right\}\left(\frac{1}{m} + \omega\right)^2.$$

Therefore the REML estimator is unbiased with variance $2(1/m+\omega)^2/(K-1)$, the ML estimator has bias $-(1/m+\omega)/K$ and MSE $(1/m+\omega)^2(2K-1)/K^2$, and the minimum-MSE estimator, with $c = 1/(n + m)$, has MSE equal to $2(1/m + \omega)^2(K + 1)$. The ML estimator of ω is uniformly more efficient than REML. These results differ only slightly for their counterparts with σ_{W}^2 unknown, if S_{W} in them is replaced by its expectation $(n - K)\sigma_{\mathrm{W}}^2$. The ratio of MSEs of the minimum-MSE estimators, one assuming that σ_{W}^2 is estimated and the other that it is known, is $1 + (K-1)/(n-K-2)$. The fraction in excess of unity can be interpreted as the value of knowing σ_{W}^2 for estimating ω.

The truncated versions of the respective ML and REML estimators of ω are denoted by $\hat{\omega}_0$ and $\hat{\omega}_{\mathrm{RE},0}$. Their properties are summarized by the following proposition.

For $j = 1$ and 2, let

$$H_j = \frac{K-1}{K+2j-1}\,\frac{N-K+2j}{N-K}.$$

The respective biases of $\hat{\omega}_0$ and $\hat{\omega}_{\mathrm{RE},0}$ are

$$\frac{1}{m}\,F_{K-1,n-K}(u) + D_1\left\{-m + 3 - (K-1)(m-1)F_{K+1,n-K-2}(H_1 u)\right\}$$

$$\frac{1}{m}\,F_{K-1,n-K}(u_{\mathrm{RE}}) + D_1\left\{2 - (n-K)\,F_{K+1,n-K-2}(H_1 u_{\mathrm{RE}})\right\}$$

and their MSEs are

$$\mathrm{MSE}(\hat{\omega}_0) = \frac{(K^2-1)(m-1)^2}{(n-K-2)(n-K-4)}\left\{1 - F_{K+3,n-K-4}(H_2 u)\right\}\left(\frac{1}{m} + \omega\right)^2$$

5

$$-\frac{2(K-1)(m-1)}{n-K-2}\left\{1-F_{K+1,n-K-2}\left(H_1 u\right)\right\}\left(\frac{1}{m}+\omega\right)^2$$

$$+\left\{1-F_{K-1,n-K}(u)\right\}\left(\frac{1}{m}+\omega\right)^2+\omega^2 F_{K-1,n-K}(u)$$

$$\mathrm{MSE}(\hat{\omega}_{\mathrm{RE},0})\;=\;\frac{(K+1)(n-K)^2\left(\omega+\frac{1}{m}\right)^2}{(K-1)(n-K-2)(n-K-4)}\left\{1-F_{K+3,n-K-4}(H_2 u_{\mathrm{RE}})\right\}$$

$$-\frac{2(n-K)}{n-K-2}\left\{1-F_{K+1,n-K-2}(H_1 u_{\mathrm{RE}})\right\}\left(\frac{1}{m}+\omega\right)^2$$

$$+\left\{1-F_{K-1,n-K}(u_{\mathrm{RE}})\right\}\left(\frac{1}{m}+\omega\right)^2+\omega^2 F_{K-1,n-K}(u_{\mathrm{RE}})\ .$$

These identities are derived in Appendix E.

We do not have an analytical proof that $\mathrm{MSE}(\hat{\omega}_0)<\mathrm{MSE}(\hat{\omega}_{\mathrm{RE},0})$, but check it by direct evaluation. Figure 1 presents the contour plots of the relative efficiency, defined as $\sqrt{\mathrm{MSE}(\hat{\omega}_{\mathrm{RE},0})/\mathrm{MSE}(\hat{\omega}_0)}$, for $K\in(3,30)$ and $m\in(3,20)$ and ω set to zero, 0.05, 0.10 and 0.25. The diagram shows that $\hat{\omega}_0$ is uniformly more efficient than $\hat{\omega}_{\mathrm{RE},0}$. The relative efficiency depends more strongly on K than on m, and appears to converge to unity with increasing K. The relative efficiency is a decreasing function of ω.

The truncated version of estimator $\tilde{\omega}(b^*)$ has MSE

$$\frac{K-1}{K+1}\frac{n-K-2}{n-K-4}\left(\omega+\frac{1}{m}\right)^2\left\{1-F_{K+3,n-K-4}\left(\frac{K+1}{K+3}\frac{n-K-4}{n-K-2}\frac{1}{1+m\omega}\right)\right\}$$

$$-\frac{2(K-1)}{K+1}\left(\omega+\frac{1}{m}\right)^2\left\{1-F_{K+1,n-K-2}\left(\frac{1}{1+m\omega}\right)\right\}$$

$$+\left\{1-F_{K-1,n-K}(u^*)\right\}\left(\omega+\frac{1}{m}\right)^2+\omega^2 F_{K-1,n-K}(u^*)\ ,$$

where $u^*=(K+1)(n-K)/(K-1)/(n-K-4)/(1+m\omega)$. It is more efficient than $\hat{\omega}_0$ for K and m used in Figure 1, but not efficient in the class $\tilde{\omega}_0(b)$. For any pair (K,m), the value of b that minimizes the MSE can be found numerically, but it depends on ω.

The MSE of the truncated (ML or REML) estimator of σ_{B}^2 cannot be derived by this approach because the estimator is a linear combination of χ^2 distributed variables, and its density does not have a closed form.
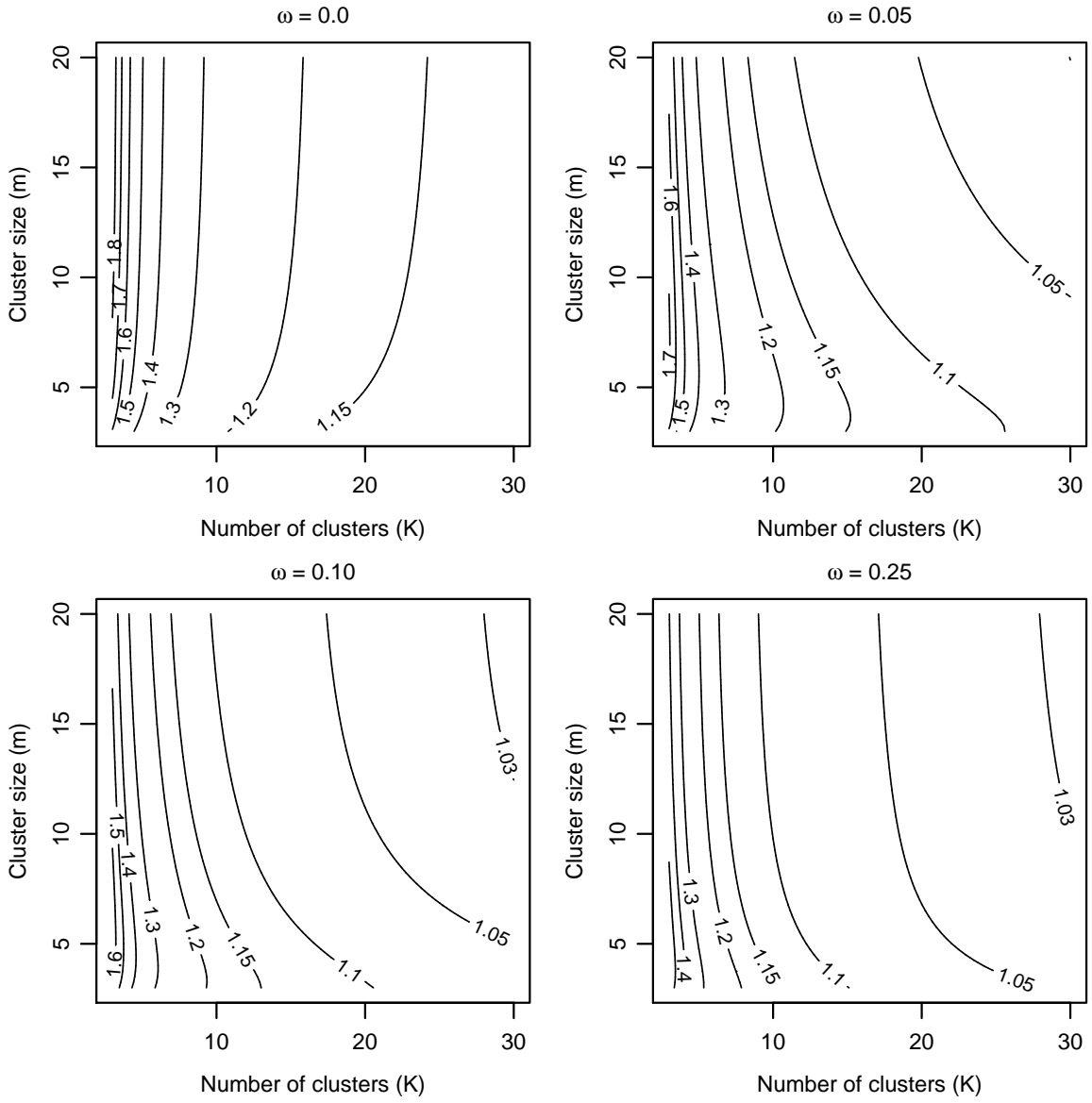
Figure 1: Contour plots of the relative efficiency of the proper (truncated) estimators of ω in one-way analysis of variance with random effects.

4 Empirical evaluation

We checked by simulations that $\hat{\sigma}_{B,0}^2$ is more efficient than $\hat{\sigma}_{B,RE,0}^2$. These simulations are conducted without generating the elementary outcomes y_{ik}, because the estimators, (4) and (5), depend on them only throught the scaled χ^2-distributed statistics S_W and S_B, which are simulated directly. The samples involved can be drawn also for fractional values of K and m. Little additional computational effort is required to compare also the proper estimators of ω, even though we have analytical expressions for their bias and MSE. We checked that the expressions in the previous section agree with their empirical counterparts.

The results are summarized by several quartets of contour plots in which we compare the biases and root-MSEs of the proper (truncated) ML and REML estimators of σ_B^2 and ω. Each quartet is for a separate value of ω and $\sigma_B^2 = \omega$, with σ_W^2 set to unity but treated as unknown. In every simulation we use $100\,000$ replications. The slight coarseness of the plots confirms that such a high number of replications is necessary.

The contour plots for $\omega = 0.05$ are displayed in Figure 2 for $K \in (3, 30)$ and $m \in (3, 20)$. The relative bias of the proper estimators $\hat{\sigma}_{B,0}^2$ and $\hat{\sigma}_{B,RE,0}^2$, denoted by Δbias, is defined as $|\operatorname{E}(\hat{\sigma}_{B,RE,0}^2) - \sigma_B^2| - |\operatorname{E}(\hat{\sigma}_{B,0}^2) - \sigma_B^2|$, so that its positive values indicate that the bias of the ML estimator is smaller in absolute value than the absolute bias of the REML estimator. The relative root-MSE of $\hat{\sigma}_{B,0}^2$ and $\hat{\sigma}_{B,RE,0}^2$ is defined as the ratio $\sqrt{\operatorname{MSE}(\hat{\sigma}_{B,RE,0}^2)/\operatorname{MSE}(\hat{\sigma}_{B,0}^2)}$; its values in excess of unity indicate that the ML estimator is more efficient than the REML estimator. The relative bias and relative root-MSE are defined for $\hat{\omega}_0$ and $\hat{\omega}_{RE,0}$ similarly.

The diagram shows that the proper ML estimators of σ_B^2 and ω are more efficient than their REML counterparts. For both σ_B^2 and ω, the relative bias decreases with K and m. For non-trivial m, it is negative for σ_B^2 when $K > 9$ and for ω when $K > 10$, in both cases by narrow margins. The relative MSE depends on m only slightly and decreases with K, but the ML estimators are $10-15\%$ more efficient (on the root-MSE scale) even for $K = 15$. Note that the bottom right-hand panel is the empirical version of the top right-hand panel of Figure 1.

Figure 3 displays the contour plots for $\sigma_{RE}^2 = \omega = 0.15$ using the same layout as Figure 2. The diagram confirms that ML estimation of σ_B^2 and ω is uniformly more
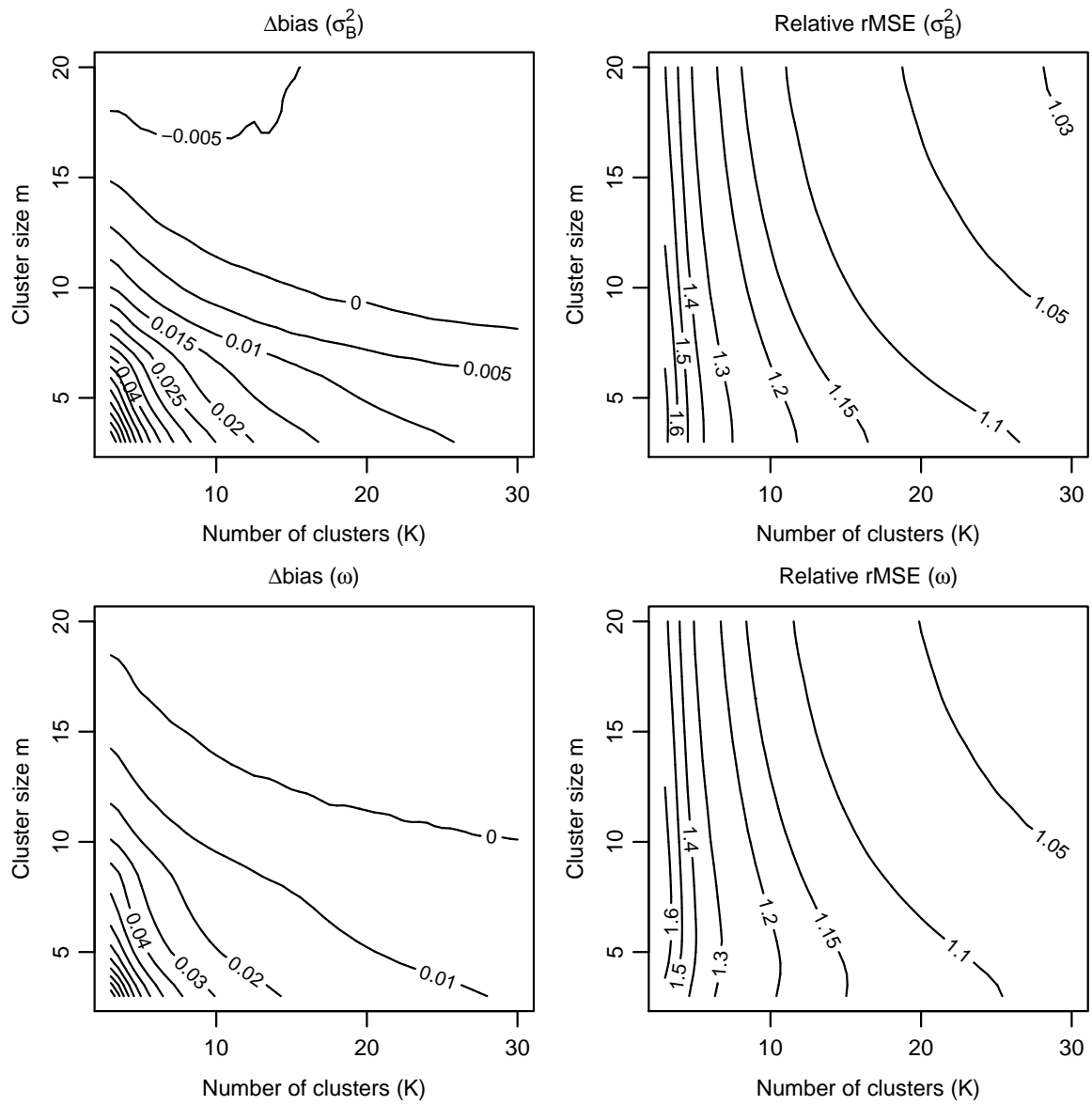
8

Figure 2: Contour plots of the relative bias and relative efficiency of the proper estimators of σ_{B}^2 and ω in one-way analysis of variance with random effects; $\sigma_{\mathrm{W}}^2 = 1$ and $\omega = 0.05$.
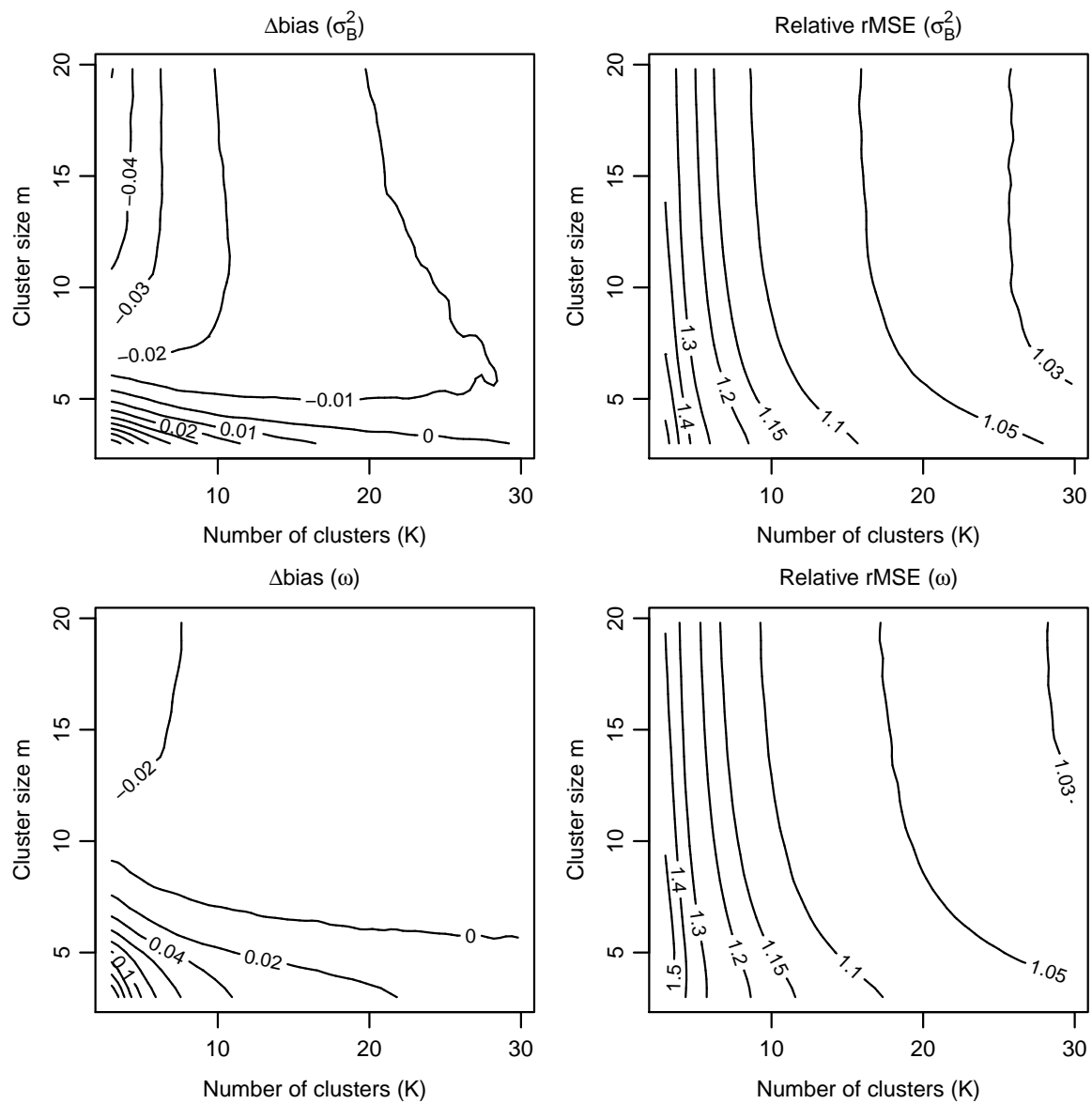
Figure 3: Contour plots of the relative bias and relative efficiency of the proper estimators of σ_{B}^2 and ω in one-way analysis of variance with random effects; $\sigma_{\mathrm{W}}^2 = 1$ and $\omega = 0.15$.

efficient, although the gains in efficiency are smaller than when $\omega = 0.05$. It seems that truncation confers additional advantage to ML estimation. For $\omega > 0.2$, the probability of truncation is very small, except for designs with very small K (and m), and the analytical results for the estimators that are not truncated provide a good approximation.

5 Example

All our (analytical) results relate to balanced designs. The efficiencies of the alternative estimators can be compared by simulations for any particular unbalanced design and value of ω, but it is not feasible to represent the vast variety of designs in such a study. For an illustration, we conducted simulations for the design with within-cluster sample sizes $m_j = j + 4$, $j = 1, \ldots, K = 11$ ($n = 110$ observations), $\sigma_{\mathrm{W}}^2 = 1$ and $\omega = 0, 0.01$, $\ldots, 0.25$. For each value of ω, we generate a dataset according to the random-effect model and evaluate the ML and REML estimates of σ_{B}^2 and ω, and replicate these steps 500 times; σ_{W}^2 is estimated throughout. The empirical root-MSEs of the improper and truncated versions of these estimators are drawn in Figure 4 as functions of ω.

The diagram shows that ML estimators (solid lines) are uniformly more efficient than the corresponding REML estimators (dashes of the same colour). The ratios of their root-MSE's decrease with ω, from 1.26 for proper estimation of σ_{B}^2 (black curves in the left-hand panel) when $\sigma_{\mathrm{B}}^2 = 0$ to 1.05 when $\sigma_{\mathrm{B}}^2 = 0.25$. The corresponding ratios for estimating ω are slightly higher. In a balanced design, these ratios do not depend on ω. For the studied design, truncation is rare and has next to no impact for $\omega > 0.18$.

The algorithms for ML and REML estimation are described in Appendix F.

6 Models with covariates

We consider next the regression model

$$\mathbf{y}_k \;=\; \mathbf{X}_k \boldsymbol{\beta} + \delta_k + \boldsymbol{\varepsilon}_k$$

for $m \times 1$ vectors of observations $\mathbf{y}_k$, $k = 1, \ldots, K$. We assume that $\mathbf{X} = (\mathbf{X}_1^\top, \ldots, \mathbf{X}_K^\top)^\top$ comprises only variables that are balanced within clusters ($\mathbf{Z}$) and variables defined
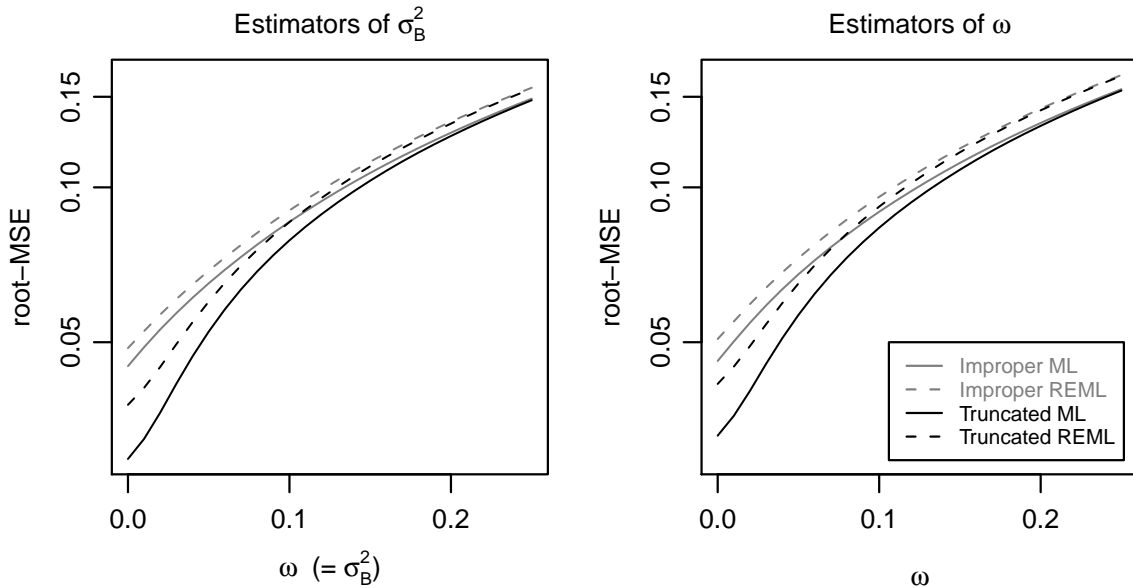
11

Figure 4: The root-MSE's of the estimators of σ_B^2 and ω for the design with a cluster each of size 5, 6, ..., 15 and $\sigma_W^2 = 1$.

for clusters ($\mathbf{U}$), which in their expanded forms are constant within clusters; $\mathbf{X} = (\mathbf{1}_n, \mathbf{Z}, \mathbf{U})$. The intercept is represented by the vector of unities, $\mathbf{1}_n$. Denote by r_Z and r_U the respective numbers of columns of $\mathbf{Z}$ and $\mathbf{U}$. We assume that $\mathbf{X}$ is of full rank equal to $r = 1 + r_Z + r_U < n$. Denote by $\mathbf{Z}_k$ and $\mathbf{U}_k$ the respective submatrices of $\mathbf{Z}$ and $\mathbf{U}$ that correspond to the units in cluster k; $\mathbf{X}_k = (\mathbf{1}_m, \mathbf{Z}_k, \mathbf{U}_k)$. No generality is lost by assuming that every variable in $\mathbf{Z}$ and $\mathbf{U}$ has zero mean. With these assumptions, the variables in $\mathbf{Z}$ and $\mathbf{U}$ are orthogonal: $\mathbf{Z}_k^\top \mathbf{U}_k = \mathbf{0}$, the matrix of zeros, with dimensions implied by the context. Further, $\mathbf{Z}_k^\top \mathbf{1}_m = \mathbf{0}$ and $\mathbf{U}_k = \mathbf{1}_m \mathbf{u}_k$, where $\mathbf{u}_k$ is the row vector of the values of $\mathbf{U}$ in cluster k. That implies that $\sum_k \mathbf{u}_k \mathbf{u}_k^\top$ is nonsingular. The dimensions of the linear spaces spanned by the variables that qualify for $\mathbf{U}$ and $\mathbf{Z}$ are $K - 1$ and $n - K$, respectively. Together with $\mathbf{1}_n$ they span the entire n-dimensional linear space. Thus, the only restriction we impose is that there are no covariates that have both within and between-cluster variation.

Let $\mathbf{W}$ be the scaled variance matrix, as in (1). The ML and REML estimators of $\boldsymbol{\beta}$ are both generalized least squares,

$$\hat{\boldsymbol{\beta}} = \left(\mathbf{X}^\top \hat{\mathbf{W}}^{-1} \mathbf{X}\right)^{-1} \mathbf{X}^\top \hat{\mathbf{W}}^{-1} \mathbf{y},$$

12

with the ML or REML estimator of ω substituted for it in $\mathbf{W}$. In Appendix G, we show that the subvectors of $\hat{\boldsymbol{\beta}}$ that correspond to $\mathbf{Z}$ and $\mathbf{U}$ are uncorrelated and they do not depend on ω. Hence the sum-of-squares statistics S_{W} and S_{B} are independent, even after the residuals e_{ik} are adjusted for $\mathbf{X}$, and

$$\frac{S_{\mathrm{W}}}{\sigma_{\mathrm{W}}^2} \sim \chi_{n-K-r_{\mathrm{Z}}}^2$$

$$\frac{S_{\mathrm{B}}}{\sigma_{\mathrm{W}}^2 + m\sigma_{\mathrm{B}}^2} \sim \chi_{K-1-r_{\mathrm{U}}}^2.$$

The REML estimators of σ_{W}^2 and σ_{B}^2 coincide with their counterparts in (5) with $K-1$ and $n-K$ replaced by $K-1-r_{\mathrm{U}}$ and $n-K-r_{\mathrm{Z}}$, respectively:

$$\hat{\omega}_{\mathrm{RE}} = \frac{n-K-r_{\mathrm{Z}}}{m(K-1-r_{\mathrm{U}})} \frac{S_{\mathrm{B}}}{S_{\mathrm{W}}} - \frac{1}{m}$$

$$\hat{\sigma}_{\mathrm{B,RE}}^2 = \frac{S_{\mathrm{B}}}{m(K-1-r_{\mathrm{U}})} - \frac{S_{\mathrm{W}}}{m(n-K-r_{\mathrm{Z}})},$$

whereas $\hat{\omega}$ and $\hat{\sigma}_{\mathrm{B}}^2$ are defined by these expressions with r_{U} and r_{Z} both set to zero. In brief, REML counts the degrees of freedom in a natural way; every variable defined for clusters reduces the effective number of clusters by one, and every variable defined for elements (and balanced within clusters) reduces the effective number of elementary units by one, from $n-K$ to $n-K-r_{\mathrm{Z}}$. As a consequence, the results for the model with no covariates carry over to the model with covariates of the two types with no changes other than an adjustment of the effective numbers of clusters and observations. However, the differences between ML and REML estimators have to be reviewed because more degrees of freedom are lost.

The ratio $S_{\mathrm{B}}/S_{\mathrm{W}}$ has scaled F distribution with $K' = K-1-r_{\mathrm{U}}$ and $n' = n-K-r_{\mathrm{Z}}$ degrees of freedom:

$$(X =) \quad \frac{1}{1+m\omega} \frac{n'}{K'} \frac{S_{\mathrm{B}}}{S_{\mathrm{W}}} \sim F_{K',n'}. \tag{7}$$

The estimators $\hat{\omega}$ and $\hat{\omega}_{\mathrm{RE}}$ have the form $\tilde{\omega}(b) = bS_{\mathrm{B}}/S_{\mathrm{W}} - 1/m$ for $b = (m-1)/m$ and $b_{\mathrm{RE}} = n'/K'/m$. In Appendix H, we prove that the coefficient b for which $\tilde{\omega}(b)$ attains its minimum MSE is

$$b^* = \frac{n'-4}{(K'+2)m}$$

and the minimum attained is

$$\mathrm{MSE}\{\tilde{\omega}(b^*)\} = \frac{2(n'+K'-2)}{(n'-2)(K'+2)} \left(\frac{1}{m} + \omega \right)^2.$$

13

Further,

$$\mathrm{MSE}(\hat{\omega}) = \frac{1}{n'-2}\left(\frac{1}{m}+\omega\right)^2\left\{\frac{K'(K'+2)(m-1)^2}{n'-4}+n'-2K'm+2K'-2\right\}$$

$$\mathrm{MSE}(\hat{\omega}_{\mathrm{RE}}) = \frac{1}{n'-2}\left(\frac{1}{m}+\omega\right)^2\left\{\frac{(K'+2)\,n'^2}{K'\,(n'-4)}-n'-2\right\}.$$

By simple rearranging we obtain the identity

$$\mathrm{MSE}(\hat{\omega}_{\mathrm{RE}})-\mathrm{MSE}(\hat{\omega}) = \frac{K'(K'+2)}{(n'-2)(n'-4)}\left(\frac{1}{m}+\omega\right)^2\{n'+K'(m-1)-2\}$$
$$\times\;\{n'-K'(m-1)\}\;.$$

It has the same sign as the concluding factor, equal to $(r_{\mathrm{U}}+1)(m-1)-r_{\mathrm{Z}}$. Hence $\hat{\omega}$ is more efficient than $\hat{\omega}_{\mathrm{RE}}$, unless there are at least $m-1$ variables in $\mathbf{Z}$ and none in $\mathbf{U}$, or at least $2m-2$ variables in $\mathbf{Z}$ and one in $\mathbf{U}$, and so on. These settings are rarely realistic.

Estimation of σ_{B}^2

The properties of the estimators of σ_{B}^2 are derived directly from their expressions as linear combinations of two independent F-distributed variates. Thus, $\hat{\sigma}_{\mathrm{RE}}^2$ is unbiased, with variance (and MSE)

$$\mathrm{var}\big(\hat{\sigma}_{\mathrm{B,RE}}^2\big) = 2\sigma_{\mathrm{W}}^4\left\{\frac{1}{K'}\left(\frac{1}{m}+\omega\right)^2+\frac{1}{m^2n'}\right\}.$$

The bias and MSE of the ML estimator are

$$\mathrm{E}\big(\hat{\sigma}_{\mathrm{B}}^2\big)-\sigma_{\mathrm{B}}^2 = \sigma_{\mathrm{W}}^2\left\{\frac{K'}{K}\left(\frac{1}{m}+\omega\right)-\frac{n'}{m(n-K)}-\omega\right\}$$

$$\mathrm{MSE}\big(\hat{\sigma}_{\mathrm{B}}^2\big) = \sigma_{\mathrm{W}}^4\left\{\frac{2K'}{K^2}\left(\frac{1}{m}+\omega\right)^2+\frac{2n'}{m^2(n-K)^2}\right\}+\big\{\mathrm{E}\big(\hat{\sigma}_{\mathrm{B}}^2\big)-\sigma_{\mathrm{B}}^2\big\}^2\;.$$

The efficiency of $\hat{\sigma}_{\mathrm{B}}^2$ and $\hat{\sigma}_{\mathrm{B,RE}}^2$ is difficult to compare analytically because the terms involving ω cannot be forced into factors of $1/m+\omega$ when $r_{\mathrm{Z}}>0$. For $r_{\mathrm{Z}}=0$, when $n'=n-K$, substantial simplification takes place and

$$\mathrm{var}\big(\hat{\sigma}_{\mathrm{B,RE}}^2\big)-\mathrm{MSE}\big(\hat{\sigma}_{\mathrm{B}}^2\big) = 2\sigma_{\mathrm{W}}^4\left(\frac{1}{m}+\omega\right)^2\frac{K-K'}{K'K^2}\left\{K'^2-K'(K-2)+2K\right\}$$

The expression in braces is a quadratic function of K'. From its roots we conclude that the difference is negative when

$$\frac{K}{2} - 1 - \sqrt{\left(\tfrac{1}{2}K - 3\right)^2 - 8} < K' < \frac{K}{2} - 1 + \sqrt{\left(\tfrac{1}{2}K - 3\right)^2 - 8}$$

and $K \geq 12$. For $K < 12$, $\hat{\sigma}_B^2$ is more efficient. For $K = 12$, $\hat{\sigma}_{B,RE}^2$ is more efficient than $\hat{\sigma}_B^2$ only for $K' = 5$, that is, $r_U = 6$, and their MSEs are identical for $K' = 4$ and $K' = 6$. For greater K, the range in which $\hat{\sigma}_{B,RE}^2$ is more efficient widens, converging to $(4, K - 4)$, that is, $r_U \in (5, K - 3)$. Only the lower bound for r_U is important in practice. Note that the differences in MSEs diminish for large K. It can be checked by direct evaluation that r_Z has only a slight impact on the MSEs of $\hat{\sigma}_B^2$ and $\hat{\sigma}_{B,RE}^2$, and can in all realistic settings be ignored. For example, the root-MSE of $\hat{\sigma}_{B,RE}^2$ decreases from 0.06940 to 0.06850 for $K = 20$, $m = 10$, $r_U = 2$, $\sigma_W^2 = 1$ and $\sigma_B^2 = \omega = 0.1$ as r_Z is increased from zero to 20. In the same setting, the root-MSE of $\hat{\sigma}_B^2$ decreases from 0.06642 to 0.06209.

Proper estimators of ω

The proper estimators, $\hat{\sigma}_{B,0}^2$ and $\hat{\sigma}_{B,RE,0}^2$ can be compared only by simulations, although these can be based on draws from pairs of F distributions, as described in Section 4. In general, truncation accords some (additional) advantage to the ML estimator, especially for small values of ω.

The derivations in Appendix E carry over directly to the regression model by changing all references of $F_{K-1,n-K}$ distribution to $F_{K',n'}$ distribution. For the MSE of the estimator $\tilde{\omega}(b)$ truncated at zero, we have the expression

$$
\begin{aligned}
\text{MSE}\left\{\tilde{\omega}_0(b)\right\} \;=\; & \frac{b^2 m^2 K'(K'+2)}{(n'-2)(n'-4)} \left(\frac{1}{m} + \omega\right)^2 \left\{1 - F_{K'+4,n'-4}\left(\frac{dH_2'}{b}\right)\right\} \\
& - \frac{2bmK'}{n'-2} \left(\frac{1}{m} + \omega\right)^2 \left\{1 - F_{K'+2,n'-2}\left(\frac{dH_1'}{b}\right)\right\} \\
& + \omega^2 F_{k',n'}\left(\frac{d}{b}\right) + \left(\frac{1}{m} + \omega\right)^2 \left\{1 - F_{K',n'}\left(\frac{d}{b}\right)\right\},
\end{aligned}
$$

where

$$d \;=\; \frac{n'}{K'm}\frac{1}{1+m\omega}$$

15

Table 1: The relative efficiency of the ML and REML estimators of ω; $K = 12$, $m = 20$, $\omega = 0.25$.

r_U	r_Z						
	0	1	2	3	4	5	6
0	1.151	1.145	1.139	1.133	1.127	1.120	1.113
1	1.232	1.231	1.230	1.228	1.226	1.224	1.222
2	1.234	1.237	1.240	1.243	1.245	1.247	1.250
3	1.185	1.190	1.195	1.200	1.204	1.209	1.214
4	1.122	1.127	1.132	1.137	1.142	1.148	1.153
5	1.073	1.077	1.082	1.087	1.092	1.097	1.101
6	1.060	1.064	1.068	1.072	1.076	1.080	1.084

$$H'_j \;=\; \frac{K'}{n'} \frac{n' - 2j}{K' - 2j}\,,$$

$j = 1, 2$. General analytical comparisons of the MSEs of $\hat{\omega}_0$ and $\hat{\omega}_{RE}$ are not feasible. Table 1 displays the relative root-MSEs, $\sqrt{\mathrm{MSE}(\hat{\omega}_{RE})/\mathrm{MSE}(\hat{\omega})}$, for $K = 12$, $m = 20$, $\omega = 0.25$ and $0 \leq r_U, r_Z \leq 6$. The relative efficiency depends on r_Z only slightly because the relative reduction of the degrees of freedom associated with it, from $n - K$ to $n - K - r_Z$ is quite modest when compared to the reduction from $K - 1$ to $K - 1 - r_U$. The ML estimator is more efficient in all settings in the table, although its advantage over REML decreases with r_U for $r_U > 2$.

For smaller values of ω, the relative efficiency increases with r_U. For example, with $K = 12$ and $m = 20$, as in Table 1, but with $\omega = 0.05$, it is equal to 1.20 for $r_U = r_Z = 0$ and 1.77 for $r_U = 6$ and $r_Z = 0$. The relative efficiency increases also with r_Z, but much more slowly; it is equal to 1.24 for $r_U = 0$ and $r_Z = 6$ and 1.80 for $r_U = r_Z = 6$; compare these figures with 1.20 and 1.77, respectively. One might regard a comparison of $\hat{\omega}_0$ and $\hat{\omega}_{RE,0}$ for small ω as unfair because $\hat{\omega} < \hat{\omega}_{RE}$, and therefore $\hat{\omega}$ is truncated with greater probability and more radically to a value very close to the target ω.

7 Saturated cluster-level variation

Some parallels of the results in Sections 2 and 6 can be drawn with estimating the variance of a random sample $y_1, \ldots, y_n$ from the normal distribution $\mathcal{N}(\mu, \sigma^2)$. Let $S = \sum_i (y_i - \bar{y})^2$. The conventional estimator $\hat{\sigma}^2_{\mathrm{RE}} = S/(n-1)$ is unbiased, with variance $2\sigma^4/(n-1)$, but $\hat{\sigma}^2 = S/n$ is more efficient, with MSE equal to $(2n-1)\sigma^4/n^2$, and $\tilde{\sigma}^2 = S/(n+1)$, with MSE equal to $2\sigma^4/(n+1)$, is efficient among estimators that are scalar multiples of S. See Stuart (1969) for these results and their generalizations.

In ordinary regression, $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$, with $n \times r$ matrix $\mathbf{X}$ or full rank r, the residual variance is conventionally estimated by $(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})^\top (\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})/(n-r)$, what is the REML estimator. The reference to scaled χ^2 indicates that $n - r + 2$ is the optimal divisor, whilst the ML estimator has the divisor n. Simple evaluations yield the following comparison. The REML estimator of σ^2 is more efficient than ML when $r > 4$ and $n > r(r-2)/4$. Note that the second condition is not particularly restrictive.

The multivariate normal regression is defined by the model

$$\mathbf{y}_j = \mathbf{x}_j \mathbf{B} + \boldsymbol{\gamma}_j,$$

where $\mathbf{B}$ is a $(r_{\mathrm{U}} + 1) \times m$ matrix of covariates and $\boldsymbol{\gamma}_j$, $j = 1, \ldots, K$, are a random sample from a centered m-variate normal distribution with variance matrix $\boldsymbol{\Sigma}$. It is equivalent to the random coefficient model with m categories involved in $\mathbf{Z}$ and each category observed once in every cluster, that is, with saturated cluster-level variation. In this model, $r_{\mathrm{Z}} = m - 1$.

The variance matrix $\boldsymbol{\Sigma}$ is estimated by the multivariate version of the ordinary least squares, and the results about unbiasedness and efficiency carry over from the univariate case. Let $\mathbf{S}$ be the matrix version of S;

$$\mathbf{S} = \sum_{j=1}^{K} \left(\mathbf{y}_j - \mathbf{x}_j \hat{\mathbf{B}} \right) \left(\mathbf{y}_j - \mathbf{x}_j \hat{\mathbf{B}} \right)^\top.$$

Then $n^{-1}\mathbf{S}$ is the ML estimator of $\boldsymbol{\Sigma}$, $(n - r_{\mathrm{U}} - 1)^{-1}\mathbf{S}$ the REML estimator, which is unbiased, but $(n - r_{\mathrm{U}} + 1)^{-1}\mathbf{S}$ is the efficient estimator among the scalar multiples of $\mathbf{S}$. Here the properties of no bias and efficiency are interpreted elementwise; they also apply to linear combinations the elements of $\boldsymbol{\Sigma}$.

8 Conclusion

We derived expressions for the bias and MSE of the ML and REML estimators of the cluster-level variance σ_B^2 and the variance ratio ω in one-way designs with normally distributed outcomes. In some settings, the ML estimators are uniformly more efficient. Their advantage is retained by truncation, which increases the efficiency of the estimators of both parameters. Of course, the truncated REML estimator of σ_B^2 is biased, so the rationale for REML estimation is difficult to sustain. In some settings, the proper ML estimator has smaller absolute bias than the proper REML estimator. The REML estimator of ω is biased even without truncation. The results for estimating σ_B^2 and $\omega = \sigma_B^2/\sigma_W^2$ differ only slightly because there is (relatively) little uncertainty about σ_W^2, except for the designs with very small K and m. The discord between unbiasedness and efficiency for estimating the within-cluster variance σ_W^2 is implied by Markowitz (1968) and Stuart (1969).

The analytical results cannot be extrapolated to designs without balance nor to all models with covariates, but we hope that they will encourage a reevaluation of the uncritical preference for REML. We derived shrinkage estimators of σ_B^2 and ω that are more efficient than improper ML. Their superiority over ML is retained by truncation, although the truncated (proper) estimators are unlikely to be efficient in their respective classes of estimators. There is no straightforward extension of the shrinkage estimators to designs without balance. Extensions of our results to more complex designs, with several sets of random effects, are an open problem. We note that the returns on such extensions may be diminishing, because the condition of balance becomes more onerous with complexity of the design and model.

We conclude by pointing out the fragile nature of unbiasedness and efficiency. For estimators derived from finite samples, these properties are not retained by nonlinear transformations. Thus, the (improper) REML estimator of σ_B^2 is unbiased, but the REML estimators of $\omega = \sigma_B^2/\sigma_W^2$ or of σ_B are not. Of course, unbiasedness is not retained by truncation either.

References

Diggle, P.J., Heagerty, P.J., Liang, K.-Y., and Zeger, S. (2002). *Analysis of Longitudinal Data*, 2nd ed. Oxford, UK: Oxford University Press.

Harville, D.A. (1974). Bayesian inference for variance components using only error contrasts. *Biometrika* 61:383–385.

Harville, D.A. (1977). Maximum likelihood approaches to variance component estimation and related problems. *Journal of the American Statistical Association* 72:320–340.

Jiang, J. (2007). *Linear and Generalized Linear Mixed Models and Their Applications.* New York: Springer-Verlag.

Kenward, M.G., and Roger, J.H. (1997). Small sample inference for fixed effects from restricted maximum likelihood. *Biometrics* 53:983–997.

Klotz, J.H., Milton, J.H., and Zacks, S. (1969). Mean square efficiency of estimators of variance components. *Journal of the American Statistical Association* **64**, 1383–1402.

Kourouklis, S. (2012). A new estimator of the variance based on minimizing mean squared error. *The American Statistician* 66:234–236.

Kubokawa, T. (1995). Estimation of variance components in mixed linear models. *Journal of Multivariate Analysis* **53**, 210–236.

Loh, W. (1986). Improved estimators of ratios of variance components. *Journal of the American Statistical Association* **81**, 699–702.

Longford, N.T. (1993). *Random Coefficient Models.* Oxford, UK: Oxford University Press.

Markowitz, E. (1968). Minimum mean-square-error of estimation of the standard deviation of normal distribution. *The American Statistician* 22:26–26.

Patterson, H.D., and Thompson, R. (1971). Recovery of inter-block information when block sizes are unequal. *Biometrika* 58:545–554.

Searle, S.R., Casella, G., and McCulloch, C.E. (1992). *Variance Components.* New York: Springer-Verlag.

Simon, M.K. (2004). *Probability Distributions Involving Gaussian Random Variables.* New York: Springer-Verlag.

Stuart, A. (1969). Reduced mean-square-error estimation of σ^p in normal samples. *The American Statistician* 23:27–28.

Swallow, W.H., and Monahan, J.F. (1984). Monte Carlo comparison of ANOVA, MIVQUE, REML ad ML estimators of variance components. *Technometrics* 26:47–57.

Verbyla, A.P. (1993). Modelling variance heterogeneity: residual maximum likelihood and diagnostics. *Journal of the Royal Statistical Society* Series B 55:493–508.

Verbyla, A.P., Cullis, B.R., Kenward, M.G., and Welham, S.J. (1999). The analysis of designed experiments and longitudinal data by using smoothing splines. *Applied Statistics* 48:269–311.

Appendices

Appendix A. Maximum likelihood

We derive expressions for the ML estimators of σ_{B}^2 and ω. The log-likelihood is given by equation (1). The scaled variance matrix $\mathbf{W} = \sigma_{\mathrm{W}}^{-2}\mathrm{var}(\mathbf{y})$ is block-diagonal, with K identical blocks equal to $\mathbf{W}_k = \mathbf{I}_m + \omega\mathbf{J}_m$, $k = 1, \ldots, K$. Here $\mathbf{I}_m$ is the $m \times m$ identity matrix and $\mathbf{J}_m$ the $m \times m$ matrix comprising unities. We have the identities

$$
\begin{aligned}
\det(\mathbf{W}_k) &= 1 + m\omega \\
\mathbf{W}_k^{-1} &= \mathbf{I}_m - \frac{\omega}{1 + m\omega}\mathbf{J}_m
\end{aligned}
$$

(Longford, 1993, Section 2.1). The quadratic form $\mathbf{e}^\top \mathbf{W}^{-1}\mathbf{e}$ can be expressed in terms of S_{W} and S_{B}, defined in (2) as

$$
\mathbf{e}^\top\mathbf{e} - \frac{m^2\omega}{1 + m\omega}\sum_{k=1}^{K}\bar{e}_k^2 = S_{\mathrm{W}} + \frac{1}{1 + m\omega}S_{\mathrm{B}},
$$

after substituting $S_W + S_B$ for $\mathbf{e}^\top \mathbf{e}$. The improper ML estimators are found as the roots of the first-order partial derivatives. From the respective identities

$$\frac{\partial l}{\partial \sigma_W^2} = -\frac{1}{2\sigma_W^4}\left(n\sigma_W^2 - S_W - \frac{1}{1+m\omega}S_B\right)$$

$$\frac{\partial l}{\partial \omega} = -\frac{1}{2}\frac{m}{1+m\omega}\left(K - \frac{1}{\sigma_W^2}\frac{1}{1+m\omega}S_B\right)$$

we have the equations for $\hat{\sigma}_W^2$:

$$\hat{\sigma}_W^2 = \frac{1}{n}\frac{1}{1+m\hat{\omega}}S_B + \frac{1}{n}S_W$$

$$\hat{\sigma}_W^2 = \frac{1}{K}\frac{1}{1+m\hat{\omega}}S_B . \tag{8}$$

Matching the right-hand sides of (8) yields the expression

$$\hat{\omega} = \frac{m-1}{m}\frac{S_B}{S_W} - \frac{1}{m}, \tag{9}$$

and by its substitution to the second equation in (8) we obtain for $\hat{\sigma}_B^2 = \hat{\sigma}_W^2\hat{\omega}$

$$\hat{\sigma}_B^2 = \frac{S_B}{n} - \frac{S_W}{m(n-K)} .$$

Appendix B. Restricted maximum likelihood

The adjustment of the log-likelihood for REML is

$$\Delta l = -\frac{1}{2}\log\left(\sigma_W^2\right) + \frac{1}{2}\log\left(\frac{n}{1+m\omega}\right) .$$

The adjusted log-likelihood is

$$l_{RE} = -\frac{1}{2}\left\{(n-1)\log\left(\sigma_W^2\right) + (K-1)\log(1+m\omega) + \frac{1}{\sigma_W^2}\mathbf{e}_\bullet^\top\mathbf{W}^{-1}\mathbf{e}_\bullet\right\} .$$

Its derivatives,

$$\frac{\partial l_{RE}}{\partial \sigma_W^2} = -\frac{1}{2\sigma_W^4}\left\{(n-1)\sigma_W^2 - S_W - \frac{1}{1+m\omega}S_B\right)$$

$$\frac{\partial l_{RE}}{\partial \omega} = -\frac{1}{2}\frac{m}{1+m\omega}\left(K-1 - \frac{1}{\sigma_W^2}\frac{1}{1+m\omega}S_B\right),$$

yield the REML counterparts of equations (8),

$$\hat{\sigma}_{W,RE}^2 = \frac{1}{n-1}\frac{1}{1+m\hat{\omega}_{RE}}S_B + \frac{1}{n-1}S_W$$

$$\hat{\sigma}_{W,RE}^2 = \frac{1}{K-1}\frac{1}{1+m\hat{\omega}_{RE}}S_B ,$$

from which we obtain the expressions for the REML estimators:

$$\hat{\omega}_{\mathrm{RE}} = \frac{n-K}{n-m}\frac{S_{\mathrm{B}}}{S_{\mathrm{W}}} - \frac{1}{m}$$

$$\hat{\sigma}^2_{\mathrm{B,RE}} = \frac{S_{\mathrm{B}}}{n-m} - \frac{S_{\mathrm{W}}}{m(n-K)}\,. \tag{10}$$

Appendix C. An efficient estimator

Having expressed the estimators of σ^2_{B} in terms of S_{B} and S_{W}, their bias and MSE can be derived by referring to the moments of the χ^2 distribution; see equation (3).

We evaluate the MSE of the improper estimator $\tilde{\sigma}^2_{\mathrm{B}}(a) = aS_{\mathrm{B}} - S_{\mathrm{W}}/m/(n-K)$. The respective values of a for ML and REML are $1/n$ and $1/(n-m)$. Then we find the coefficient a for which the MSE attains its minimum. The bias of $\tilde{\sigma}^2_{\mathrm{B}}(a)$ is

$$\mathrm{E}\left\{aS_{\mathrm{B}} - \frac{1}{m(n-K)}S_{\mathrm{W}} - \sigma^2_{\mathrm{B}}\right\} = a(K-1)\left(\sigma^2_{\mathrm{W}} + m\sigma^2_{\mathrm{B}}\right) - \frac{1}{m}\sigma^2_{\mathrm{W}} - \sigma^2_{\mathrm{B}}$$

$$= \left\{a(K-1) - \frac{1}{m}\right\}\left(\sigma^2_{\mathrm{W}} + m\sigma^2_{\mathrm{B}}\right)\,.$$

Thus, $\hat{\sigma}^2_{\mathrm{B,RE}}$ is unbiased and the bias of $\hat{\sigma}^2_{\mathrm{B}}$ is $-(\sigma^2_{\mathrm{W}} + m\sigma^2_{\mathrm{B}})/n$.

The MSE of $\tilde{\sigma}^2_{\mathrm{B}}(a)$ is

$$\left[2a^2m^2(K-1) + \{a(n-m) - 1\}^2\right]\left(\sigma^2_{\mathrm{B}} + \frac{1}{m}\sigma^2_{\mathrm{W}}\right)^2 + \frac{2(n-K)}{n^2(m-1)^2}\sigma^2_{\mathrm{W}}$$

$$= \left\{a^2m^2(K-1)(K+1) - 2a(n-m) + 1\right\}\left(\sigma^2_{\mathrm{B}} + \frac{1}{m}\sigma^2_{\mathrm{W}}\right)^2 + C\sigma^2_{\mathrm{W}}\,,$$

where the constant $C = 2K/(m-1)/n^2$ does not involve a, so its value is immaterial for minimizing the MSE. The MSE is minimized when the expression in the braces, a quadratic function of a, attains its minimum. That occurs for

$$a^* = \frac{n-m}{m^2(K-1)(K+1)} = \frac{1}{n+m}\,.$$

The minimum MSE is

$$\frac{2}{K+1}\left(\sigma^2_{\mathrm{B}} + \frac{1}{m}\sigma^2_{\mathrm{W}}\right)^2 + C\sigma^4_{\mathrm{W}}\,.$$

The MSEs of the ML and REML estimators differ only in their leading factors, which are $(2K-1)/K^2$ and $2/(K-1)$ for ML and REML, respectively, both greater than $2/(K+1)$. See Kfrouklis (2012) for a similar approach to minimum MSE estimation in a related context.

Appendix D. Bias and MSE of the estimators of ω

The statistic $S_{\mathrm{B}}/S_{\mathrm{W}}$ has scaled F distribution with $K' = K-1$ and $n' = n-K$ degrees of freedom. The F distributions have the densities

$$f_{K',n'}(x) \;=\; \frac{\Gamma_2(K'+n')}{\Gamma_2(K')\Gamma_2(n')}\left(\frac{K'}{n'}\right)^{K'/2}\frac{x^{K'/2-1}}{(1+K'x/n')^{(K'+n')/2}}\;; \tag{11}$$

the half-gamma function $\Gamma_2(k) = \Gamma(k/2)$ is introduced for typographical reasons. The expectation of this distribution is $n'/(n'-2)$ when $n' > 2$ and its variance is

$$\frac{2n'^2(K'+n'-2)}{K'(n'-2)^2(n'-4)}$$

when $n' > 4$ (Simon, 2004). We assume throughout that $n > K+4$.

The estimators $\hat{\omega}$ and $\hat{\omega}_{\mathrm{RE}}$ have the form $\tilde{\omega}(b) = bS_{\mathrm{B}}/S_{\mathrm{W}} - 1/m$, with $b = b_{\mathrm{ML}} = (m-1)/m$ for $\hat{\omega}$ and $b_{\mathrm{RE}} = b_{\mathrm{ML}}K/(K-1)$ for $\hat{\omega}_{\mathrm{RE}}$; see equations (9) and (10). The bias of $\tilde{\omega}(b)$ is

$$\left(\frac{bmK'}{n'-2}-1\right)\left(\frac{1}{m}+\omega\right)$$

and its MSE is

$$D_1^2\left[2b^2m^2\,\frac{K'(n'+K'-2)}{n'-4}+\{bmK'-(n'-2)\}^2\right]$$
$$=\; D_1^2\left[2b^2m^2\left\{\frac{2K'(n'+K'-2)}{n'-4}+K'^2\right\}-2bmK'(n'-2)+(n'-2)^2\right], \tag{12}$$

where $D_1 = (1/m+\omega)/(n'-2)$. Hence the respective biases of $\hat{\omega}$ and $\hat{\omega}_{\mathrm{RE}}$ are

$$\mathrm{E}(\hat{\omega})-\omega \;=\; \left\{\frac{(m-1)K'}{n'-1}-1\right\}\left(\frac{1}{m}+\omega\right)=-\frac{m-3}{n'-2}\left(\frac{1}{m}+\omega\right).$$

$$\mathrm{E}(\hat{\omega}_{\mathrm{RE}})-\omega \;=\; \left\{\frac{(m-1)K}{n'-2}-1\right\}\left(\frac{1}{m}+\omega\right)\;=\;\frac{2}{n'-2}\left(\frac{1}{m}+\omega\right).$$

The MSEs of $\hat{\omega}$ and $\hat{\omega}_{\mathrm{RE}}$ are

$$\mathrm{MSE}(\hat{\omega}) \;=\; D_1^2\left\{\frac{2K'(m-1)^2(n'+K'-2)}{n'-4}+(m-3)^2\right\}$$

$$\mathrm{MSE}(\hat{\omega}_{\mathrm{RE}}) \;=\; D_1^2\left\{\frac{2n'^2(n'+K'-2)}{K'(n'-4)}+4\right\}.$$

The MSE in (12) is a quadratic function of b. It attains its minimum for

$$b^* = \frac{1}{m} \frac{(n'-2)(n'-4)}{2(n'+K'-2)+K'(n'-4)}$$

$$= \frac{n'-4}{m(K'+2)} = 1 - \frac{m+K+4}{m+n}$$

and the minimum attained is

$$\left\{1 - \frac{K'}{K'+2} \frac{n'-4}{n'-2}\right\} \left(\frac{1}{m} + \omega\right)^2 = \frac{2(n'+K'-2)}{(n'-2)(K'+2)} \left(\frac{1}{m} + \omega\right)^2.$$

Appendix E. Proper estimators of ω

In this section, we derive the MSEs of the proper estimators $\hat{\omega}_0$ and $\hat{\omega}_{\mathrm{RE},0}$ which are formed by truncating $\hat{\omega}$ and $\hat{\omega}_{\mathrm{RE}}$ at zero. We use the following identities that link the densities of F distributions:

$$x f_{K',n'}(x) = \frac{K'}{K'+2} f_{K'+2,n'-2}(H_1 x)$$

$$x^2 f_{K',n'}(x) = \frac{K'}{H_1(K'+4)} f_{K'+4,n'-4}(H_2 x) , \tag{13}$$

where

$$H_j = \frac{K'}{n'} \frac{n'-2j}{K'+2j} , \tag{14}$$

$j = 1, 2$. The first identity is obtained by matching the expression for $x f_{K',n'}(x)$ with another density of form (11). First, the factor $x^{K'/2}$ implies $K'+2$ degrees of freedom instead of K'; then the term $1 + yK'/n'$ in the denominator implies the argument $H_1 x$ instead of x, and its exponent $(K'+n')/2$ implies the change from n' degrees of freedom to $n'-2$. The constant factors remaining from the match with $f_{K'+2,n'-2}(H_1 x)$ reduce to $K'/(K'+2)$. The second identity in (13) is obtained by reusing the first, with $K'+2$ and $K'-2$ in place of K' and k_2, respectively.

Consider the class of estimators $\tilde{\omega}_0(b)$ defined by truncating $bS_{\mathrm{B}}/S_{\mathrm{W}} - 1/m$ at zero; $b > 0$ is a scalar. They include $\hat{\omega}_0$, for which $b = 1 - 1/m$, and $\hat{\omega}_{\mathrm{RE},0}$, for which $b = (n-K)/(n-m)$.

Denote

$$d = \frac{n-K}{n-m} \frac{1}{1+m\omega} .$$

We have the identities

$$
\begin{aligned}
\mathrm{E}\{\tilde{\omega}_0(b)\} &= \frac{b}{dm} \int_{d/b}^{+\infty} x f_{K',n'}(x)\,\mathrm{d}x - \frac{1}{m}\left\{1 - F_{K',\,n'}\left(\frac{d}{b}\right)\right\} \\
&= \frac{b}{dm} \frac{K'}{K'+2} \int_{d/b}^{+\infty} x f_{K'+2,n'-2}(H_1 x)\,\mathrm{d}x - \frac{1}{m}\left\{1 - F_{K',\,n-K}\left(\frac{d}{b}\right)\right\}, \\
&= \frac{b}{dm} \frac{n'}{n'-2}\left\{1 - F_{K'+2,\,n'-2}\left(\frac{dH_1}{b}\right)\right\} - \frac{1}{m}\left\{1 - F_{K',\,n'}\left(\frac{d}{b}\right)\right\}.
\end{aligned}
$$

The biases of $\hat{\omega}_0$ and $\hat{\omega}_{\mathrm{RE},0}$ are obtained by substituting $d/b_{\mathrm{ML}} = u$ and $d/b_{\mathrm{RE}} = u_{\mathrm{RE}}$ for d/b.

For the MSE we use the general identity

$$
\mathrm{MSE}(\tilde{\omega}) = \mathrm{E}\left\{\left(\tilde{\omega} + \frac{1}{m}\right)^2\right\} - 2\left(\omega + \frac{1}{m}\right)\mathrm{E}\left(\tilde{\omega} + \frac{1}{m}\right) + \left(\omega + \frac{1}{m}\right)^2.
$$

The first term for $\tilde{\omega}_0 = \tilde{\omega}_0(b)$ is

$$
\begin{aligned}
\mathrm{E}\left\{\left(\tilde{\omega}_0 + \frac{1}{m}\right)^2\right\} &= \frac{b^2}{d^2 m^2} \int_{d/b}^{+\infty} x^2 f_{K',n'}(x)\,\mathrm{d}x + \frac{1}{m^2} F_{K',n'}\left(\frac{d}{b}\right) \\
&= \frac{b^2}{d^2 m^2} \frac{K'+2}{K'+4} \frac{n'}{n'-2} \int_{d/b}^{+\infty} f_{K'+4,n'-4}\left(H_2 x\right)\,\mathrm{d}x \\
&\quad + \frac{1}{m^2} F_{K',n'}\left(\frac{d}{b}\right).
\end{aligned}
$$

After consolidating the terms involving $F_{K',n'}$ and partly substituting for d we obtain

$$
\begin{aligned}
\mathrm{MSE}(\tilde{\omega}_0) &= \frac{b^2}{d^2 m^2} \frac{K'+2}{K'} \frac{n'^2}{(n'-2)(n'-4)}\left\{1 - F_{K'+4,n'-4}\left(\frac{dH_2}{b}\right)\right\} \\
&\quad - \frac{2b}{dm} \frac{n'}{n'-2}\left(\omega + \frac{1}{m}\right)\left\{1 - F_{K'+2,n'-2}\left(\frac{dH_1}{b}\right)\right\} \\
&\quad + \omega^2 F_{K',n'}\left(\frac{d}{b}\right) + \left(\omega + \frac{1}{m}\right)^2 \left\{1 - F_{K',n'}\left(\frac{d}{b}\right)\right\} \\
&= \frac{b^2 K'(K'-1)m^2}{(n'-2)(n'-4)}\left(\omega + \frac{1}{m}\right)^2\left\{1 - F_{K'+4,n'-4}\left(\frac{dH_2}{b}\right)\right\} \\
&\quad - \frac{2bK'm}{n'-2}\left(\omega + \frac{1}{m}\right)^2\left\{1 - F_{K'+2,n'-2}\left(\frac{dH_1}{b}\right)\right\} \\
&\quad + \omega^2 F_{K',n'}\left(\frac{d}{b}\right) + \left(\omega + \frac{1}{m}\right)^2\left\{1 - F_{K',n'}\left(\frac{d}{b}\right)\right\}.
\end{aligned}
$$

Appendix F. ML and REML estimation with unbalanced designs

The log-likelihood for the design with within-cluster sample sizes m_j, $j = 1, \ldots, K$, is given by equation (1) in which $\mathbf{W}$ is block-diagonal with $m_j \times m_j$ matrices $\mathbf{W}_j = \mathbf{I}_{m_j} + \omega \mathbf{J}_{m_j}$ as its blocks. We have the identities $\det(\mathbf{W}_j) = g_j$ and $\mathbf{W}_j^{-1} = \mathbf{I}_{m_j} - \omega/g_j \, \mathbf{J}_{m_j}$, where $g_j = 1 + m_j \omega$. The ML estimator of the overall mean μ is the generalized least squares estimator

$$
\begin{aligned}
\hat{\mu} &= \left(\mathbf{1}_n^\top \hat{\mathbf{W}}^{-1} \mathbf{1}_n \right)^{-1} \mathbf{1}_n^\top \hat{\mathbf{W}}^{-1} \mathbf{y} \\
&= \sum_{j=1}^{K} \frac{\mathbf{y}_j^\top \mathbf{1}_{m_j}}{g_j} \bigg/ \sum_{j=1}^{K} \frac{m_j}{g_j} \,,
\end{aligned}
\tag{15}
$$

where $n = m_1 + \cdots + m_K$ is the overall sample size and $\hat{\omega}$, involved in $\hat{\mathbf{W}}$ and g_j, is specified below. The ML estimator of σ_{W}^2 is the root of the score,

$$
\hat{\sigma}_{\mathrm{W}}^2 = \frac{1}{n} \mathbf{e}^\top \hat{\mathbf{W}}^{-1} \mathbf{e}
\tag{16}
$$

($\mathbf{e}$ is the vector of residuals, $\mathbf{y} - \hat{\mu} \mathbf{1}_n$), and the quadratic form is evaluated as

$$
\mathbf{e}^\top \hat{\mathbf{W}}^{-1} \mathbf{e} = \mathbf{e}^\top \mathbf{e} - \hat{\omega} \sum_{j=1}^{K} \frac{1}{g_j} \left(\mathbf{e}_j^\top \mathbf{1}_{m_j} \right)^2 .
$$

The score and Hessian for ω are

$$
\begin{aligned}
(s =) \quad \frac{\partial l}{\partial \omega} &= -\frac{1}{2} \sum_{j=1}^{K} \left\{ \mathrm{tr} \left(\mathbf{W}_j^{-1} \frac{\partial \mathbf{W}_j}{\partial \omega} \right) - \frac{1}{\sigma_{\mathrm{W}}^2} \mathbf{e}_j^\top \mathbf{W}_j^{-1} \frac{\partial \mathbf{W}_j}{\partial \omega} \mathbf{W}_j^{-1} \mathbf{e}_j \right\} \\
&= \frac{1}{2} \sum_{j=1}^{K} \left\{ -\frac{m_j}{g_j} + \frac{1}{g_j^2 \sigma_{\mathrm{W}}^2} \left(\mathbf{e}_j^\top \mathbf{1}_{m_j} \right)^2 \right\}
\end{aligned}
$$

$$
(H =) \quad -\mathrm{E} \left(\frac{\partial^2 l}{\partial \omega^2} \right) = \frac{1}{2} \sum_{j=1}^{K} \frac{m_j^2}{g_j^2} ,
\tag{17}
$$

after substituting $\mathrm{E}\{(\mathbf{e}_j^\top \mathbf{1}_{m_j})^2\} = \mathrm{var}(\mathbf{e}_j^\top \mathbf{1}_{m_j}) = m_j g_j \sigma_{\mathrm{W}}^2$.

The Fisher scoring algorithm starts with the initial (provisional) fit $\hat{\mu} = n^{-1} \mathbf{y}^\top \mathbf{1}_n$, $\hat{\sigma}_{\mathrm{W}}^2 = \widehat{\mathrm{var}}(\mathbf{y})$, the sample variance of the elements of $\mathbf{y}$, and $\hat{\omega} = 0.1$. Its iterations evaluate the expressions (15)–(17) and update $\hat{\omega}$ as $\hat{\omega}_{\mathrm{new}} = \hat{\omega}_{\mathrm{old}} + s/H$. The iterations are stopped when the absolute deviations of the parameter estimates and the change

in the log-likelihood become smaller than 10^{-8} in total. In the example in Section 6, this takes between 10 and 50 iterations, with the average around 20. The values of $\hat{\omega}$ are 'cut back' if in any iteration they fall below $-1/\max(m_j)$ by halving the value of s/H as many times as necessary. For $\hat{\omega} < -1/\max(m_j)$, some of the matrices $\hat{\mathbf{W}}_j$ are not positive definite.

For REML, the divisor in (16) is changed from n to $n-1$ and the score function is adjusted to

$$\frac{1}{2} \sum_{j=1}^{K} \left\{ -\frac{m_j}{g_j} + \frac{1}{g_j^2 \sigma_{\mathrm{W}}^2} \left(\mathbf{e}_j^\top \mathbf{1}_{m_j}\right)^2 \right\} + \frac{1}{2} \sum_{j=1}^{K} \frac{m_j^2}{g_j^2} \Big/ \sum_{j=1}^{K} \frac{m_j}{g_j}$$

The Hessian could also be adjusted, but this is not necessary.

In a replication, the ML and REML estimates are evaluated by a single function, using the same dataset. The sets of replications for the distinct values of ω use the same set of random numbers — they start with the same value of the random seed.

The code, compiled in R, is available from the author on request.

Appendix G. Estimation of $\hat{\boldsymbol{\beta}}$

The inverse of $\mathbf{W}$ is given in Appendix A. We have the identities

$$\begin{aligned}
\mathbf{U}_k^\top \mathbf{W}_k^{-1} &= \mathbf{U}_k^\top - \frac{\omega}{1 + m\omega} \mathbf{U}_k^\top \mathbf{1}_m \mathbf{1}_m^\top \\
&= \frac{1}{1 + m\omega} \mathbf{u}_k^\top \mathbf{1}_m^\top,
\end{aligned}$$

where $\mathbf{u}_k$ is the row vector of values of $\mathbf{U}$ in cluster k; $\mathbf{U}_k = \mathbf{1}_m \mathbf{u}_k$. Further,

$$\begin{aligned}
\mathbf{Z}_k^\top \mathbf{W}_k^{-1} &= \mathbf{Z}_k^\top - \frac{\omega}{1 + m\omega} \mathbf{Z}_k^\top \mathbf{1}_m \mathbf{1}_m^\top \\
&= \mathbf{Z}_k^\top,
\end{aligned}$$

because $\mathbf{Z}$ is balanced and centered within clusters. Hence $\mathbf{Z}_k^\top \mathbf{W}_k^{-1} \mathbf{U}_k = \mathbf{0}$; $\hat{\boldsymbol{\beta}}_{\mathrm{Z}}$ and $\hat{\boldsymbol{\beta}}_{\mathrm{U}}$, the vectors of (ML or REML) estimators that correspond to $\mathbf{Z}$ and $\mathbf{U}$, respectively, are uncorrelated. Further,

$$\mathbf{U}^\top \mathbf{W}^{-1} \mathbf{U} = \frac{K}{1 + m\omega} \sum_{k=1}^{K} \mathbf{u}_k^\top \mathbf{u}_k$$

$$\mathbf{U}^\top \mathbf{W}^{-1} \mathbf{y} = \frac{K}{1 + m\omega} \sum_{k=1}^{K} \mathbf{u}_k^\top \mathbf{y}_k$$

27

and

$$\mathbf{Z}^{\top}\mathbf{W}^{-1}\mathbf{Z} = \sum_{k=1}^{K} \mathbf{Z}_k^{\top}\mathbf{Z}_k$$

$$\mathbf{Z}^{\top}\mathbf{W}^{-1}\mathbf{y} = \sum_{k=1}^{K} \mathbf{Z}_k^{\top}\mathbf{y}_k.$$

Therefore neither $\hat{\boldsymbol{\beta}}_{\mathrm{U}}$ nor $\hat{\boldsymbol{\beta}}_{\mathrm{Z}}$ depends on ω. The log-likelihood has the same form as for the model with no covariates, when $\mathbf{X}_k\boldsymbol{\beta} = \mu$, and therefore the ML estimators of σ_{B}^2 and ω are the same as in (9), with S_{B} and S_{W} based on appropriately corrected values of the residuals e_{ik}.

The REML adjustment of the log-likelihood is

$$\begin{aligned}
\Delta l &= \frac{1}{2}\log\left\{\det\left(\frac{1}{\sigma_{\mathrm{W}}^2}\mathbf{X}^{\top}\mathbf{W}^{-1}\mathbf{X}\right)\right\} \\
&= \frac{1}{2}\left\{-(1+r_{\mathrm{U}}+r_{\mathrm{Z}})\log\left(\sigma_{\mathrm{W}}^2\right) - (1+r_{\mathrm{U}})\log(1+m\omega)\right\} + C
\end{aligned}$$

where C is a constant that involves no parameters and is therefore immaterial for maximizing the likelihood. Hence the derivatives of l_{RE} given in Appendix B apply in the current setting with $n' = n - K$ replaced by $n - K - r_{\mathrm{Z}}$ and $K' = K - 1$ replaced by $K - 1 - r_{\mathrm{U}}$.

Appendix H. Estimation of ω

Using the identities for the moments of the F distribution yields the expressions

$$\begin{aligned}
\mathrm{E}\{\tilde{\omega}(b)\} &= \left(\frac{bK'm}{n'-2} - 1\right)\left(\frac{1}{m} + \omega\right) \\
\mathrm{var}\{\tilde{\omega}(b)\} &= \frac{2b^2 K'm^2 (n'+K'-2)}{(n'-2)(n'-4)}\left(\frac{1}{m} + \omega\right)^2
\end{aligned}$$

for $\tilde{\omega}(b) = bS_{\mathrm{B}}/S_{\mathrm{W}} - 1/m$. The MSE of $\tilde{\omega}(b)$ is a quadratic function of b. It attains its minimum for

$$b^* = \frac{n'-4}{(K'+2)\,m}$$

and the attained minimum is

$$\mathrm{E}\{\tilde{\omega}(b^*)\} = \frac{2(n'+K'-2)}{(K'+2)(n'-4)}\left(\frac{1}{m} + \omega\right)^2.$$