Computational Statistics and Data Analysis 53 (2009) 2583–2595

Contents lists available at ScienceDirect

Computational Statistics and Data Analysis

journal homepage: www.elsevier.com/locate/csda

An improved approximation to the precision of fixed effects from restricted maximum likelihood

Michael G. Kenward [a,*], James H. Roger [b]

[a] Medical Statistics Unit, London School of Hygiene and Tropical Medicine, Keppel Street, London W1C 7HT, UK

[b] Research Statistics Unit, GlaxoSmithKline, Berkeley Avenue, Greenford, Middlesex UB6 0NN, UK

ARTICLE INFO

Article history:
Received 26 March 2008
Received in revised form 22 December 2008
Accepted 23 December 2008
Available online 19 January 2009

ABSTRACT

An approximate small sample variance estimator for fixed effects from the multivariate normal linear model, together with appropriate inference tools based on a scaled F pivot, is now well established in practice and there is a growing literature on its properties in a variety of settings. Although effective under linear covariance structures, there are examples of nonlinear structures for which it does not perform as well. The cause of this problem is shown to be a missing term in the underlying Taylor series expansion which accommodates the bias in the estimators of the parameters of the covariance structure. The form of this missing term is derived, and then used to adjust the small sample variance estimator. The behaviour of the resulting estimator is explored in terms of invariance under transformation of the covariance parameters and also using a simulation study. It is seen to perform successfully in the way predicted from its derivation.

1. Introduction

Kenward and Roger (1997) proposed an approximate small sample estimator for the variance of the fixed effects parameters in Restricted Maximum Likelihood. They also derive approximate degrees of freedom for pivots based on the t and F distributions, the latter being compatible with Hotelling's T^2 statistic in appropriate settings. This methodology has been implemented by the SAS Institute in their MIXED and GLIMMIX procedure as part of the DDFM = KenwardRoger option. Several books including Littell et al. (2006) and Brown and Prescott (2006) on mixed models recommend its use. As such the method has been used extensively and many papers have cited the method and several have explored its properties based on simulation, usually in restricted settings; some recent examples of the latter are Guiard et al. (2003), Savin et al. (2003), Chen and Wei (2003), Kowalchuk et al. (2004), Vallejo et al. (2004), Yutaka et al. (2004), and Spilke et al. (2005).

Kenward and Roger (1997) pointed out that the bias correction for the estimator of the variance of the fixed effects estimator, behaves well when the parameterizations of the covariance matrix is linear. However when it is not linear the bias correction can behave less well, in that the bias correction is less successful, but is often better than no correction. When a linear problem is re-parameterized in a nonlinear fashion, such as through replacement of a covariance parameter by a correlation parameter multiplied by a variance, then the bias correction changes more than one might expect. Here we look in detail at ways to improve this bias correction when the covariance matrix has a nonlinear parameterization.

The general setting is that we have n observations $\mathbf{Y}$ following a multivariate Gaussian distribution:

$$\mathbf{Y} \sim N(\mathbf{X}\boldsymbol{\beta};\ \boldsymbol{\Sigma}).$$

* Corresponding author.
E-mail addresses: mike.kenward@lshtm.ac.uk (M.G. Kenward), James.H.Roger@gsk.com (J.H. Roger).

0167-9473/$ – see front matter © 2009 Elsevier B.V. All rights reserved.
doi:10.1016/j.csda.2008.12.013

The fixed effects design matrix $\mathbf{X}$ is $n \times p$ with rank p. The elements of the variance–covariance matrix $\boldsymbol{\Sigma}$ are assumed to be functions of r parameters, $\boldsymbol{\sigma}(r \times 1)$, that are sufficiently well behaved for the resulting likelihood to be regular. In particular we assume that the first two partial derivatives with respect to $\boldsymbol{\sigma}$ exist. To simplify notation we show explicit dependence of $\boldsymbol{\Sigma}$ on $\boldsymbol{\sigma}$ only when we wish to emphasize it. The restricted maximum likelihood estimator of $\boldsymbol{\sigma}$, denoted $\hat{\boldsymbol{\sigma}}$, is the maximum likelihood estimator from the marginal likelihood of the variables $\mathbf{Z} = \mathbf{KY}$ where $\mathbf{K}$ is any $(n-p) \times n$ matrix of full rank satisfying $\mathbf{KX} = 0$. The marginal likelihood of $\mathbf{Z}$ can be expressed as:

$$2 \log \mathrm{L}(\boldsymbol{\sigma}) = \text{constant} - \log(|\boldsymbol{\Sigma}|) - \log(|\mathbf{X}^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}\mathbf{X}|) - \mathbf{Y}^{\mathrm{T}}\{\boldsymbol{\Sigma}^{-1} - \boldsymbol{\Sigma}^{-1}\mathbf{X}(\mathbf{X}^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}\}\mathbf{Y}.$$

The restricted maximum likelihood estimator of $\boldsymbol{\beta}$ is the generalized least squares estimator

$$\hat{\boldsymbol{\beta}} = \boldsymbol{\Phi}(\hat{\boldsymbol{\sigma}})\mathbf{X}^{\mathrm{T}}\boldsymbol{\Sigma}(\hat{\boldsymbol{\sigma}})^{-1}\mathbf{Y},$$

where $\boldsymbol{\Phi}(\boldsymbol{\sigma}) = \{\mathbf{X}^{\mathrm{T}}\boldsymbol{\Sigma}(\boldsymbol{\sigma})^{-1}\mathbf{X}\}^{-1}$ and its variance is most simply estimated by $\boldsymbol{\Phi}(\hat{\boldsymbol{\sigma}})$.

Our goal is to estimate the precision of $\hat{\boldsymbol{\beta}}$ from samples that are not sufficiently large for asymptotic results to apply acceptably well. In the original variance derivation of Kenward and Roger (1997), which was based on earlier work of Kackar and Harville (1984) and Jeske and Harville (1988), the approximations were based on Taylor series expansions which ignored the bias in the estimator of $\boldsymbol{\sigma}$. While unimportant for linear covariance structures, we see below that, for certain nonlinear structures, this does introduce bias of the same order as the correction that is being made, and so in such settings further correction must be made for this. In the following section we show how this should be done, and in the subsequent section show invariance of the adjustment under re-parameterization for a wide range of covariance structures. In Section 4 we use simulation studies to assess the performance of the resultant estimator of precision of the fixed effects under three different covariance structures with nonlinear parameterization. In Section 5 we apply the adjustment to two examples, and finish with some recommendations in Section 6.

2. The modified adjustment

When data are not independent the concept of sample size may be complicated by the covariance between observations. In order to rate the importance of individual terms in the Taylor series expansion we use the idea of an effective sample size N. Then each term can be identified as order N^r, indicated by $\mathrm{O}(N^r)$ for some value of r. This can be formalized by defining N as the number of blocks of observations, each block independent of each other block. Observations within a block may be highly correlated or even linearly related. With repeated measurements the block would be the subject, while in a hierarchical model it would be the top stratum unit. The purpose of N is to indicate the multiplicity in the data. Here we simply indicate the order of each term, allowing us to group terms which are of the same order. This indicates their relative importance when N is large. For a specific example the actual value of a term in a less important group (order is smaller) may be larger than a term in a more important group (order is larger). But it allows us to group together and respect all terms in a group whenever we already use a term from that group. The order of both $\boldsymbol{\Phi}(\boldsymbol{\sigma})$ and $\mathrm{E}\{\boldsymbol{\Phi}(\hat{\boldsymbol{\sigma}})\}$ is N^{-1}.

There are two sources of bias in the estimated variance $\boldsymbol{\Phi}(\hat{\boldsymbol{\sigma}})$ of the fixed effects estimator $\hat{\boldsymbol{\beta}}$ (Kenward and Roger, 1997). First there is the amount $\boldsymbol{\Lambda} = \mathrm{var}(\hat{\boldsymbol{\beta}} - \tilde{\boldsymbol{\beta}})$ where $\tilde{\boldsymbol{\beta}}$ is the estimator of $\boldsymbol{\beta}$ for known fixed parameters $\boldsymbol{\sigma}$. For restricted maximum likelihood it is known that $(\hat{\boldsymbol{\beta}} - \tilde{\boldsymbol{\beta}})$ is independent of $\tilde{\boldsymbol{\beta}}$. Then $\boldsymbol{\Lambda}$ represents the amount to which the asymptotic variance–covariance matrix underestimates, in a matrix sense, $\mathrm{var}(\hat{\boldsymbol{\beta}})$. A Taylor series expansion for $(\hat{\boldsymbol{\beta}} - \tilde{\boldsymbol{\beta}})$ in $\boldsymbol{\sigma}$ about $\hat{\boldsymbol{\sigma}}$ gives

$$(\hat{\boldsymbol{\beta}} - \tilde{\boldsymbol{\beta}}) \simeq \hat{\boldsymbol{\beta}} - \left\{ \hat{\boldsymbol{\beta}} + \sum_{i=1}^{r} \frac{\partial \tilde{\boldsymbol{\beta}}}{\partial \boldsymbol{\sigma}_i}(\boldsymbol{\sigma}_i - \hat{\boldsymbol{\sigma}}_i) + \frac{1}{2}\sum_{i=1}^{r}\sum_{j=1}^{r} \frac{\partial^2 \tilde{\boldsymbol{\beta}}}{\partial \boldsymbol{\sigma}_i \partial \boldsymbol{\sigma}_j}(\boldsymbol{\sigma}_i - \hat{\boldsymbol{\sigma}}_i)(\boldsymbol{\sigma}_j - \hat{\boldsymbol{\sigma}}_j) \right\}.$$

Then taking expectation of the square and noting that both $\hat{\boldsymbol{\beta}}$ and $\tilde{\boldsymbol{\beta}}$ are unbiased for REML

$$\boldsymbol{\Lambda} = \mathrm{E}\left(\frac{\partial \tilde{\boldsymbol{\beta}}}{\partial \boldsymbol{\sigma}} \mathbf{W} \frac{\partial \tilde{\boldsymbol{\beta}}}{\partial \boldsymbol{\sigma}}^{\mathrm{T}} \right) + O(N^{-2}),$$

for $\mathbf{W}$ the variance–covariance matrix of $\hat{\boldsymbol{\sigma}}$. The first term is of order N^{-1}. Note that this part of the approximation does not rely on any assumption about the unbiasedness of $\hat{\boldsymbol{\sigma}}$, and is invariant under re-parameterization of the covariances.

We note that $\boldsymbol{\Lambda}$ can then be written succinctly in terms of the first derivatives of $\boldsymbol{\Sigma}^{-1}$:

$$\boldsymbol{\Lambda} = \boldsymbol{\Phi} \left\{ \sum_{i=1}^{r}\sum_{j=1}^{r} W_{ij}(\mathbf{Q}_{ij} - \mathbf{P}_i \boldsymbol{\Phi} \mathbf{P}_j) \right\} \boldsymbol{\Phi} + O(N^{-2}),$$

where

$$\mathbf{P}_i = \mathbf{X}^{\mathrm{T}} \frac{\partial \boldsymbol{\Sigma}^{-1}}{\partial \sigma_i} \mathbf{X} \quad \text{and} \quad \mathbf{Q}_{ij} = \mathbf{X}^{\mathrm{T}} \frac{\partial \boldsymbol{\Sigma}^{-1}}{\partial \sigma_i} \boldsymbol{\Sigma} \frac{\partial \boldsymbol{\Sigma}^{-1}}{\partial \sigma_j} \mathbf{X}.$$

The second part of the bias in the estimator of the variance–covariance for the estimator of the fixed effects comes from the fact that $\hat{\boldsymbol{\Phi}}$ is biased as an estimator of $\boldsymbol{\Phi}$.

Expanding $\boldsymbol{\Phi}$ about $\boldsymbol{\sigma}$ and evaluating at $\hat{\boldsymbol{\sigma}}$ we have

$$\hat{\boldsymbol{\Phi}} \simeq \boldsymbol{\Phi} + \sum_{i=1}^{r}(\hat{\sigma}_i - \sigma_i)\frac{\partial \boldsymbol{\Phi}}{\partial \sigma_i} + \frac{1}{2}\sum_{i=1}^{r}\sum_{j=1}^{r}(\hat{\sigma}_i - \sigma_i)(\hat{\sigma}_j - \sigma_j)\frac{\partial^2 \boldsymbol{\Phi}}{\partial \sigma_i \partial \sigma_j} + \cdots.$$

Taking expectations we find

$$\mathrm{E}(\hat{\boldsymbol{\Phi}}) = \boldsymbol{\Phi} + \sum_{i=1}^{r}\mathrm{E}\{(\hat{\sigma}_i - \sigma_i)\}\frac{\partial \hat{\boldsymbol{\Phi}}}{\partial \sigma_i} + \frac{1}{2}\sum_{i=1}^{r}\sum_{j=1}^{r}W_{ij}\frac{\partial^2 \boldsymbol{\Phi}}{\partial \sigma_i \partial \sigma_j} + O(N^{-3}). \tag{1}$$

Kenward and Roger (1997) ignore the second term on the grounds that they "ignore possible bias in $\hat{\boldsymbol{\sigma}}$" and use the third term as the required bias in using $\hat{\boldsymbol{\Phi}}$ as an estimator for $\boldsymbol{\Phi}$. This leads to the adjusted formula

$$\hat{\boldsymbol{\Phi}}_A^{\mathrm{OLD}} = \hat{\boldsymbol{\Phi}} + 2\hat{\boldsymbol{\Phi}}\left\{\sum_{i=1}^{r}\sum_{j=1}^{r}W_{ij}\left(\mathbf{Q}_{ij} - \mathbf{P}_i\hat{\boldsymbol{\Phi}}\mathbf{P}_j - \frac{1}{4}\mathbf{R}_{ij}\right)\right\}\hat{\boldsymbol{\Phi}}, \tag{2}$$

where

$$\mathbf{R}_{ij} = \mathbf{X}^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}\frac{\partial^2 \boldsymbol{\Sigma}}{\partial \sigma_i \partial \sigma_j}\boldsymbol{\Sigma}^{-1}\mathbf{X}.$$

and the variance–covariance matrix $\mathbf{W}$ of $\hat{\boldsymbol{\sigma}}$ is obtained from the inverse of the expected information

$$2\mathbf{W}_{ij} = \mathrm{tr}\left(\frac{\partial \boldsymbol{\Sigma}^{-1}}{\partial \sigma_i}\boldsymbol{\Sigma}\frac{\partial \boldsymbol{\Sigma}^{-1}}{\partial \sigma_j}\boldsymbol{\Sigma}\right) - \mathrm{tr}\left(2\boldsymbol{\Phi}\mathbf{Q}_{ij} - \boldsymbol{\Phi}\mathbf{P}_i\boldsymbol{\Phi}\mathbf{P}_j\right).$$

Alternatively one might use the observed or average information rather than the expected information for this.

All the terms except that involving $\mathbf{R}_{ij}$ are invariant under re-parameterization of the covariances. For this reason, the SAS Institute have implemented a *FirstOrder* modifier on the DDFM = KenwardRoger option in the GLIMMIX procedure, which ignores the $\mathbf{R}_{ij}$ term. This is useful for parameterizations, which are effectively equivalent to a linear parameterization, where the $\mathbf{R}_{ij}$ is know to be zero. A fully unstructured matrix expressed in terms of variances and correlations is a classic example.

The first term in expansion (1) is of order N^{-1} while the third term is of order N^{-2}. But, as we will see, unless the covariance matrix $\boldsymbol{\Sigma}$ is linear in $\boldsymbol{\sigma}$, then the second term is also of order N^{-2}. By ignoring this additional term the small sample adjustment fails to work well with some nonlinear parameterized covariance matrices. We now construct the appropriate correction for this. We first derive an approximation for the bias in $\hat{\boldsymbol{\sigma}}$ as an estimate of $\boldsymbol{\sigma}$.

We define $\mathbf{U}_r(\boldsymbol{\sigma}) = \partial\ell(\boldsymbol{\sigma})/\partial\sigma_r$ and $\mathbf{U}_{rs}(\boldsymbol{\sigma}) = \partial^2\ell(\boldsymbol{\sigma})/\partial\sigma_r\partial\sigma_s$, and so on, where $\ell(\boldsymbol{\sigma}) = \log\{\mathrm{L}(\boldsymbol{\sigma})\}$. The joint null cumulants are $\kappa_{r,s} = \mathrm{E}(\mathbf{U}_r\mathbf{U}_s)$, $\kappa_{r,s,t} = \mathrm{E}(\mathbf{U}_r\mathbf{U}_s\mathbf{U}_t)$, $\kappa_{r,st} = E(\mathbf{U}_r\mathbf{U}_{st})$ and so on. Note that the standard $1/n$ term does not appear here as we are using the whole log-likelihood. We know that

$$\kappa_{rs} + \kappa_{r,s} = 0 \quad \text{and} \quad \kappa_{rst} + \kappa_{r,st} + \kappa_{s,rt} + \kappa_{t,rs} + \kappa_{r,s,t} = 0.$$

Cox and Snell (1968) show that the first-order bias of $\hat{\boldsymbol{\sigma}}_s$ is given by

$$-\sum_{ijt}\kappa^{i,j}\kappa^{s,t}(\kappa_{i,j,t} + \kappa_{t,ij})/2 = \sum_{ijt}\kappa^{i,j}\kappa^{s,t}(\kappa_{i,jt} + \kappa_{j,it} + \kappa_{ijt})/2,$$

where $\kappa^{i,j}$ is the (i,j) element of the inverse of the Fisher information matrix of elements $\kappa_{i,j}$.

Now $2\mathrm{U}(\boldsymbol{\sigma})_t = \mathrm{tr}(\mathbf{A}_t) + \mathbf{Y}^{\mathrm{T}}\mathbf{B}_t\mathbf{Y}$ where

$$\mathbf{A}_t = \boldsymbol{\Sigma}\mathbf{D}\frac{\partial \boldsymbol{\Sigma}^{-1}}{\partial \sigma_t}, \qquad \mathbf{B}_t = -\mathbf{D}\frac{\partial \boldsymbol{\Sigma}^{-1}}{\partial \sigma_t}\mathbf{D}^{\mathrm{T}} \quad \text{and} \quad \mathbf{D} = \mathbf{I} - \boldsymbol{\Sigma}^{-1}\left(\mathbf{X}\boldsymbol{\Phi}\mathbf{X}^{\mathrm{T}}\right), \quad \text{so } \mathbf{X}^{\mathrm{T}}\mathbf{D} = \mathbf{0}.$$

We know $\mathrm{E}[2\mathrm{U}(\boldsymbol{\sigma})_t] = \mathrm{tr}(\mathbf{A}_t) + \mathrm{tr}(\boldsymbol{\Sigma}\mathbf{B}_t) = 0$ which implies

$$\mathrm{tr}\left(\frac{\partial \mathbf{A}_t}{\partial \sigma_j}\right) + \mathrm{tr}\left(\boldsymbol{\Sigma}\frac{\partial \mathbf{B}_t}{\partial \sigma_j}\right) + \mathrm{tr}\left(\frac{\partial \boldsymbol{\Sigma}}{\partial \sigma_j}\mathbf{B}_t\right) = 0,$$

and so

$$2\kappa_{tj} = -2\kappa_{t,j} = E\left\{2\frac{\partial \mathbf{U}(\boldsymbol{\sigma})_t}{\partial \sigma_j}\right\} = \mathrm{tr}\left(\frac{\partial \mathbf{A}_t}{\partial \sigma_j}\right) + \mathrm{tr}\left(\boldsymbol{\Sigma}\frac{\partial \mathbf{B}_t}{\partial \sigma_j}\right) = -\mathrm{tr}\left(\frac{\partial \boldsymbol{\Sigma}}{\partial \sigma_j}\mathbf{B}_t\right).$$

Repeating this with $\mathbf{U}$ differentiated twice we find that

$$2\kappa_{ijt} = E\left\{2\frac{\partial^2 \mathbf{U}(\boldsymbol{\sigma})_t}{\partial\sigma_i\partial\sigma_j}\right\} = \mathrm{tr}\left(\frac{\partial^2\mathbf{A}_t}{\partial\sigma_i\partial\sigma_j}\right) + \mathrm{tr}\left(\frac{\boldsymbol{\Sigma}\partial^2\mathbf{B}_t}{\partial\sigma_i\partial\sigma_j}\right)$$
$$= -\mathrm{tr}\left(\frac{\partial\boldsymbol{\Sigma}}{\partial\sigma_i}\frac{\partial\mathbf{B}_t}{\partial\sigma_j}\right) - \mathrm{tr}\left(\frac{\partial\boldsymbol{\Sigma}}{\partial\sigma_j}\frac{\partial\mathbf{B}_t}{\partial\sigma_i}\right) - \mathrm{tr}\left(\frac{\partial^2\boldsymbol{\Sigma}}{\partial\sigma_i\partial\sigma_j}\mathbf{B}_t\right).$$

Finally,

$$4E\left\{\mathrm{U}(\boldsymbol{\sigma})_i\frac{\partial\mathrm{U}(\boldsymbol{\sigma})_t}{\partial\sigma_j}\right\} = E\left[\left\{\mathrm{tr}(\mathbf{A}_i) + \mathbf{Y}^{\mathrm{T}}\mathbf{B}_i\mathbf{Y}\right\}\left\{\mathrm{tr}\left(\frac{\partial\mathbf{A}_t}{\partial\sigma_j}\right) + \mathbf{Y}^{\mathrm{T}}\frac{\partial\mathbf{B}_t}{\partial\sigma_j}Y\right\}\right]$$
$$= \mathrm{tr}(\mathbf{A}_i)\mathrm{tr}\left(\frac{\partial\mathbf{A}_t}{\partial\sigma_j}\right) + \mathrm{tr}(\mathbf{A}_i)\mathrm{tr}\left(\frac{\partial\mathbf{B}_t}{\partial\sigma_j}\boldsymbol{\Sigma}\right) + \mathrm{tr}\left(\frac{\partial\mathbf{A}_t}{\partial\sigma_j}\right)\mathrm{tr}(\mathbf{B}_i\boldsymbol{\Sigma}) + E\left(\mathbf{Y}^{\mathrm{T}}\mathbf{B}_i\mathbf{Y}\mathbf{Y}^{\mathrm{T}}\frac{\partial\mathbf{B}_t}{\partial\sigma_j}\mathbf{Y}\right),$$

and using

$$E\left(\mathbf{Y}^{\mathrm{T}}\mathbf{B}_i\mathbf{Y}\,\mathbf{Y}^{\mathrm{T}}\frac{\partial\mathbf{B}_t}{\partial\sigma_j}\mathbf{Y}\right) = \mathrm{tr}\left(\frac{\partial\mathbf{B}_t}{\partial\sigma_j}\boldsymbol{\Sigma}\right)\mathrm{tr}\left(\mathbf{B}_i\boldsymbol{\Sigma}\right) + 2\mathrm{tr}\left(\frac{\partial\mathbf{B}_t}{\partial\sigma_j}\boldsymbol{\Sigma}\mathbf{B}_i\boldsymbol{\Sigma}\right),$$

we have

$$4\kappa_{i,jt} = 4E\left\{U(\boldsymbol{\sigma})_i\frac{\partial\mathrm{U}(\boldsymbol{\sigma})_t}{\partial\sigma_j}\right\} = 2\mathrm{tr}\left(\frac{\partial\mathbf{B}_t}{\partial\sigma_j}\boldsymbol{\Sigma}\mathbf{B}_i\boldsymbol{\Sigma}\right).$$

The approximate bias in $\hat{\sigma}_s$ is then

$$\frac{1}{4}\sum_{t=1}^{r}\sum_{i=1}^{r}\sum_{j=1}^{r}\kappa^{i,j}\kappa^{s,t}\left\{\mathrm{tr}\left(\frac{\partial\mathbf{B}_t}{\partial\sigma_i}\boldsymbol{\Sigma}\mathbf{B}_j\boldsymbol{\Sigma}\right) + \mathrm{tr}\left(\frac{\partial\mathbf{B}_t}{\partial\sigma_j}\boldsymbol{\Sigma}\mathbf{B}_i\boldsymbol{\Sigma}\right) - \mathrm{tr}\left(\frac{\partial\boldsymbol{\Sigma}}{\partial\sigma_i}\frac{\partial\mathbf{B}_t}{\partial\sigma_j}\right) - \mathrm{tr}\left(\frac{\partial\boldsymbol{\Sigma}}{\partial\sigma_j}\frac{\partial\mathbf{B}_t}{\partial\sigma_i}\right) - \mathrm{tr}\left(\frac{\partial^2\boldsymbol{\Sigma}}{\partial\sigma_i\partial\sigma_j}\mathbf{B}_t\right)\right\}.$$

Now we can show that

$$\mathrm{tr}\left(\frac{\partial\mathbf{B}_t}{\partial\sigma_j}\boldsymbol{\Sigma}\mathbf{B}_i\boldsymbol{\Sigma}\right) = \mathrm{tr}\left(\frac{\partial\mathbf{B}_t}{\partial\sigma_j}\frac{\partial\boldsymbol{\Sigma}}{\partial\sigma_i}\right),$$

based on the facts that $\partial\mathbf{B}_t/\partial\sigma_j = \mathbf{D}\mathbf{M}_{jt}\mathbf{D}^{\mathrm{T}}$ where,

$$\mathbf{M}_{jt} = \frac{\partial\boldsymbol{\Sigma}^{-1}}{\partial\sigma_j}\mathbf{X}\boldsymbol{\Phi}\mathbf{X}^{\mathrm{T}}\frac{\partial\boldsymbol{\Sigma}^{-1}}{\partial\sigma_t} + \frac{\partial\boldsymbol{\Sigma}^{-1}}{\partial\sigma_t}\mathbf{X}\boldsymbol{\Phi}\mathbf{X}^{\mathrm{T}}\frac{\partial\boldsymbol{\Sigma}^{-1}}{\partial\sigma_j} - \frac{\partial^2\boldsymbol{\Sigma}^{-1}}{\partial\sigma_j\sigma_t},$$

and so

$$\mathrm{tr}\left\{\frac{\partial\mathbf{B}_t}{\partial\sigma_j}\left(\frac{\partial\boldsymbol{\Sigma}}{\partial\sigma_i} - \boldsymbol{\Sigma}\mathbf{B}_i\boldsymbol{\Sigma}\right)\right\} = -\mathrm{tr}\left\{\mathbf{D}\mathbf{M}_{jt}\mathbf{D}^{\mathrm{T}}\left(\boldsymbol{\Sigma}\frac{\partial\boldsymbol{\Sigma}^{-1}}{\partial\sigma_i}\boldsymbol{\Sigma} - \boldsymbol{\Sigma}\mathbf{D}\frac{\partial\boldsymbol{\Sigma}^{-1}}{\partial\sigma_i}\mathbf{D}^{\mathrm{T}}\boldsymbol{\Sigma}\right)\right\}$$
$$= -\mathrm{tr}\left\{\mathbf{D}M_{jt}\mathbf{D}^{\mathrm{T}}\left(\boldsymbol{\Sigma}\frac{\partial\boldsymbol{\Sigma}^{-1}}{\partial\sigma_i}\mathbf{X}\boldsymbol{\Phi}\mathbf{X}^{\mathrm{T}} + \mathbf{X}\boldsymbol{\Phi}\mathbf{X}^{\mathrm{T}}\frac{\partial\boldsymbol{\Sigma}^{-1}}{\partial\sigma_i}\boldsymbol{\Sigma} - \mathbf{X}\boldsymbol{\Phi}\mathbf{X}^{\mathrm{T}}\frac{\partial\boldsymbol{\Sigma}^{-1}}{\partial\sigma_i}\mathbf{X}\boldsymbol{\Phi}\mathbf{X}^{\mathrm{T}}\right)\right\} = 0,$$

as every term effectively includes $\mathbf{X}^{\mathrm{T}}\mathbf{D} = \mathbf{0}$.

This leads to an expression for the bias in $\hat{\sigma}_s$ of

$$-\frac{1}{4}\sum_{t=1}^{r}\sum_{i=1}^{r}\sum_{j=1}^{r}\kappa^{i,j}\kappa^{s,t}\,\mathrm{tr}\left(\frac{\partial^2\boldsymbol{\Sigma}}{\partial\sigma_i\partial\sigma_j}\mathbf{B}_t\right).$$

The trace term is of order N^{-1}, as a result of the effective summation. So the bias in $\hat{\sigma}$ is of order N^{-1}, as suggested previously. This is the source of the problem in the current version of the adjustment $\hat{\boldsymbol{\Phi}}_A$ that is seen with certain nonlinear covariance structures.

This approximate bias in $\hat{\sigma}_s$ can now be used to obtain an additional bias adjustment term for the variance of the fixed effect estimator $\hat{\boldsymbol{\beta}}$ added to those described by Kenward and Roger (1997)

$$\frac{1}{4}\sum_{s,t=1}^{r}W_{st}\sum_{i,j=1}^{r}W_{ij}\mathrm{tr}\left(\frac{\partial^2\boldsymbol{\Sigma}}{\partial\sigma_i\partial\sigma_j}\mathbf{B}_t\right)\frac{\partial\boldsymbol{\Phi}}{\partial\sigma_s}, \tag{3}$$

or

$$\frac{1}{4}\sum_{s,t=1}^{r}W_{st}\left[\sum_{i,j=1}^{r}W_{ij}\mathrm{tr}\left\{\frac{\partial^2\boldsymbol{\Sigma}}{\partial\sigma_i\partial\sigma_j}\left(\mathbf{I} - \boldsymbol{\Sigma}^{-1}\,\mathbf{X}\boldsymbol{\Phi}\mathbf{X}^{\mathrm{T}}\right)\frac{\partial\boldsymbol{\Sigma}^{-1}}{\partial\sigma_t}\left(\mathbf{I} - \boldsymbol{\Sigma}^{-1}\mathbf{X}\boldsymbol{\Phi}\mathbf{X}^{\mathrm{T}}\right)^{\mathrm{T}}\right\}\right]\boldsymbol{\Phi}\mathbf{P}_s\boldsymbol{\Phi}.$$

This is more efficiently computed as

$$-\frac{1}{4}\sum_{s,t=1}^{r} W_{st}V_t\boldsymbol{\Phi}\mathbf{P}_s\boldsymbol{\Phi},$$

where the scalars V_t are defined as

$$V_t = \mathrm{tr}\left\{\mathbf{S}\frac{\partial\boldsymbol{\Sigma}}{\partial\sigma_t}\right\} - 2\mathrm{tr}\left\{\left(\mathbf{X}^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}\frac{\partial\boldsymbol{\Sigma}}{\partial\sigma_t}\mathbf{S}\mathbf{X}\right)\boldsymbol{\Phi}\right\} + \mathrm{tr}\left\{\left(\mathbf{X}^{\mathrm{T}}\mathbf{S}\mathbf{X}\right)\boldsymbol{\Phi}\left(\mathbf{X}^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}\frac{\partial\boldsymbol{\Sigma}}{\partial\sigma_t}\boldsymbol{\Sigma}^{-1}\mathbf{X}\right)\boldsymbol{\Phi}\right\},$$

with

$$\mathbf{S} = \boldsymbol{\Sigma}^{-1}\sum_{i,j=1}^{r} W_{ij}\left(\frac{\partial^2\boldsymbol{\Sigma}}{\partial\sigma_i\partial\sigma_j}\right)\boldsymbol{\Sigma}^{-1}.$$

In conclusion, the new adjusted variance estimator can be written

$$\hat{\boldsymbol{\Phi}}_A^{\mathrm{NEW}} = \hat{\boldsymbol{\Phi}} + 2\hat{\boldsymbol{\Phi}}\left\{\sum_{i=1}^{r}\sum_{j=1}^{r} W_{ij}\left(\mathbf{Q}_{ij} - \mathbf{P}_i\hat{\boldsymbol{\Phi}}\mathbf{P}_j - \frac{1}{4}\mathbf{R}_{ij}\right)\right\}\hat{\boldsymbol{\Phi}} - \frac{1}{4}\sum_{s,t=1}^{r} W_{st}V_t\boldsymbol{\Phi}\mathbf{P}_s\boldsymbol{\Phi}. \tag{4}$$

From a computational point of view, matrices such as $\mathbf{X}^{\mathrm{T}}\mathbf{S}\mathbf{X}$ are typically comparatively small and can be calculated efficiently taking advantage of the block diagonal structure of $\mathbf{S}$, $\boldsymbol{\Sigma}$ and its derivatives. In practice the values of W_{ij}, the variances and covariances of the estimated covariance parameters $\hat{\boldsymbol{\sigma}}$, can be estimated via the information matrix in the conventional manner. However it should be noted that although the observed, expected, or average information matrices can be used to estimate these quantities in the part of the adjustment taken from the original Kenward and Roger (1997) paper, this is not true for new additional component (3). For this, the inner values of W_{ij} should be estimated using the expected information, while for the outer values, W_{st}, any of the three alternatives can be used.

Example 1. Suppose that we have a sample of n independent observations from a normal distribution with mean μ and variance σ^2. The simplest situation for which the original variance formula fails is for the variance of the usual fixed effect estimator $\bar{y}$ of the mean μ, where the variance is parameterized through the standard deviation, *i.e.* as $\sigma^2 = \sigma_1^2$, rather than through the variance for which $\sigma^2 = \sigma_1$. With the conventional variance parameterization, which is linear, the adjustment terms are zero, and the variance estimator for the mean is $\hat{\sigma}^2/n$, where $\hat{\sigma}^2$ is the usual sum of squares about the mean divided by the sample degrees of freedom $(n-1)$. However, with the standard deviation parameterization the original adjustment term is non-zero of order $1/n^2$.

For the new term we have $\mathbf{X} = \mathbf{j}_n$ the unit vector of length n, $\boldsymbol{\Sigma} = \left(\sigma_1^2\right)\otimes\mathbf{I}_n$, $\boldsymbol{\Phi} = \left(\sigma_1^2/n\right)$, $\mathbf{P}_1 = \left(-2n/\sigma_1^3\right)$, $\mathbf{Q}_{11} = \left(4n/\sigma_1^4\right)$ and $\mathbf{R}_{11} = \left(2n/\sigma_1^4\right)$. Now

$$\mathbf{W} = \frac{\sigma_1^2}{2(n-1)}, \qquad \frac{\partial^2\boldsymbol{\Sigma}}{\partial\sigma_1\partial\sigma_1} = (2)\otimes\mathbf{I}_n, \qquad \mathbf{B} = \frac{2}{\sigma^3}\left(\mathbf{I}_n - \frac{1}{n}\mathbf{J}_n\right), \quad \text{and} \quad \frac{\partial\boldsymbol{\Phi}}{\partial\sigma_1} = \left(\frac{2\sigma_1}{n}\right),$$

where $\mathbf{I}_n$ and $\mathbf{J}_n$ are $n\times n$ matrices with the diagonal elements of $\mathbf{I}$ and all elements of $\mathbf{J}$ set to unity. Following the original adjustment formulae this leads to an adjustment for the variance $-(\hat{\sigma}^2{}_1)^2/2n^2$ while the new additional term is $(\hat{\sigma}^2{}_1)^2/2n^2$. These two terms cancel, recovering the conventional estimator for the variance of a sample mean.

Example 2. Several of the practical problems that occur when using the previous version of the adjustment are associated with nonlinear covariance structures that involve products of parameters. In this example we consider the case of a simple clustered structure in which two observation are taken on each of n individuals. When expressed in terms of the two variance components, *i.e.* in a linear form, the previous version works well, making no adjustment. Suppose instead that the same structure is parameterized in terms of an overall (common) variance and within-subject correlation, which we label σ_1 and σ_2 respectively in keeping with the previous notation. The off-diagonal term of the covariance matrix is the product of these two parameters. With this nonlinear parameterization the previous adjustment introduces an erroneous correction term. We now show how the proposed adjustment makes the required correction.

In the original version of the adjustment, the P and Q terms cancel, leaving an erroneous adjustment based on the R term

$$\frac{-(2n\sigma_2 + 1 - \sigma_2)(1-\sigma_2^2)\sigma_1}{4n^2(n-1)}.$$

The additional term described here cancels exactly with this previous adjustment term, leaving the correct unadjusted estimator.

These two simple examples illustrate one of two more general results about the behaviour of the new adjustment under re-parameterization of the covariances, which are explored in the next section.

3. Re-parameterization of the covariance matrices

3.1. Definitions and general results

We now show that the introduction of the new term into the adjustment (3) leads to important invariance properties of the resulting variance estimator under re-parameterization of the covariance structure. We begin by defining two classes of covariance structure. Suppose that we have a covariance structure $\boldsymbol{\Sigma}$ with r parameters $\boldsymbol{\sigma} = (\sigma_1, \ldots, \sigma_r)^{\mathrm{T}}$. Then we can define the same structure using any bijective re-parameterization in terms of a new set of r parameters $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_r)^{\mathrm{T}}$ say. It follows that the matrix $\left[\frac{\partial \boldsymbol{\lambda}}{\partial \boldsymbol{\sigma}}\right]$ is of full rank throughout.

The first class, which we call *intrinsically linear* has, for some choice of $\boldsymbol{\lambda}$, a *linear* structure, *i.e.*

$$\boldsymbol{\Sigma} = \sum_{i=1}^{r} \lambda_i \mathbf{A}_i,$$

for fixed matrices $\{\mathbf{A}_1, \ldots, \mathbf{A}_r\}$ of the same dimension as $\boldsymbol{\Sigma}$.

The second class, which we call *intrinsically linear inverse* has, for some parameterization $\boldsymbol{\lambda}$, an inverse which has a linear structure, *i.e.*,

$$\boldsymbol{\Sigma}^{-1} = \sum_{i=1}^{r} \lambda_i \mathbf{A}_i,$$

again for fixed matrices $\{\mathbf{A}_1, \ldots, \mathbf{A}_r\}$ of the same dimension as $\boldsymbol{\Sigma}$.

It follows in general that, for any parameterization $\boldsymbol{\sigma}$ and re-parameterization $\boldsymbol{\lambda}$,

$$\mathbf{P}_i = \mathbf{X}^{\mathrm{T}} \frac{\partial \boldsymbol{\Sigma}^{-1}}{\partial \sigma_i} \mathbf{X} = -\sum_s \frac{\partial \lambda_s}{\partial \sigma_i} \left\{ \mathbf{X}^{\mathrm{T}} \boldsymbol{\Sigma}^{-1} \frac{\partial \boldsymbol{\Sigma}}{\partial \lambda_s} \boldsymbol{\Sigma}^{-1} \mathbf{X} \right\} = \sum_s \frac{\partial \lambda_s}{\partial \sigma_i} \mathbf{P}^*_s,$$

and similarly

$$\mathbf{Q}_{ij} = \sum_{st} \frac{\partial \lambda_s}{\partial \sigma_i} \frac{\partial \lambda_t}{\partial \sigma_j} \mathbf{Q}^*_{st},$$

where $\mathbf{P}^*_s$ and $\mathbf{Q}^*_{st}$ are equivalent to $\mathbf{P}$ and $\mathbf{Q}$ but expressed in terms of the $\boldsymbol{\lambda}$ parameterization. There is no similar result however for the $\mathbf{R}_{ij}$ term.

Now the W_{ij} term based on the expected information is

$$\begin{aligned} 2W^{ij} &= \mathrm{tr}\left(\boldsymbol{\Sigma}^{-1} \frac{\partial \boldsymbol{\Sigma}}{\partial \sigma_i} \boldsymbol{\Sigma}^{-1} \frac{\partial \boldsymbol{\Sigma}}{\partial \sigma_j} \right) - \mathrm{tr}(2\boldsymbol{\Phi}\mathbf{Q}_{ij} - \boldsymbol{\Phi}\mathbf{P}_i\boldsymbol{\Phi}P_j) \\ &= \sum_{st} \frac{\partial \lambda_s}{\partial \sigma_i} \frac{\partial \lambda_t}{\partial \sigma_j} \left[\mathrm{tr}\left(\boldsymbol{\Sigma}^{-1} \frac{\partial \boldsymbol{\Sigma}}{\partial \lambda_s} \boldsymbol{\Sigma}^{-1} \frac{\partial \boldsymbol{\Sigma}}{\partial \lambda_t} \right) - \mathrm{tr}(2\boldsymbol{\Phi}\mathbf{Q}^*_{st} - \boldsymbol{\Phi}\mathbf{P}^*_s\boldsymbol{\Phi}P^*_t) \right]. \end{aligned}$$

So

$$W^{ij} = \sum_{st} \frac{\partial \lambda_s}{\partial \sigma_i} \frac{\partial \lambda_t}{\partial \sigma_j} (W^*)^{st},$$

where W^{ij} are terms in the inverse of $\mathbf{W}$. So we have

$$\mathbf{W}^{-1} = \left[\frac{\partial \boldsymbol{\lambda}}{\partial \boldsymbol{\sigma}}^{\mathrm{T}} \right] (\mathbf{W}^*)^{-1} \left[\frac{\partial \boldsymbol{\lambda}}{\partial \boldsymbol{\sigma}} \right],$$

and

$$\mathbf{W} = \left[\frac{\partial \boldsymbol{\lambda}}{\partial \boldsymbol{\sigma}} \right]^{-1} \mathbf{W}^* \left[\frac{\partial \boldsymbol{\lambda}}{\partial \boldsymbol{\sigma}}^{\mathrm{T}} \right]^{-1}.$$

This means that

$$\begin{aligned} \sum_{ij} W_{ij} \sum_{st} \frac{\partial \lambda_s}{\partial \sigma_i} \frac{\partial \lambda_t}{\partial \sigma_j} \mathbf{M}_{st} &= \sum_{ab} \sum_{ij} \sum_{st} \left[\frac{\partial \lambda}{\partial \sigma}^{\mathrm{T}} \right]^{ia} W^*_{ab} \left[\frac{\partial \lambda}{\partial \sigma} \right]^{bj} \frac{\partial \lambda_s}{\partial \sigma_i} \frac{\partial \lambda_t}{\partial \sigma_j} \mathbf{M}_{st} \\ &= \sum_{ab} W^*_{ab} \mathbf{M}_{ab}, \end{aligned}$$

for any set of matrices $\{\mathbf{M}_{jk}\}$ of appropriate dimensions.

This demonstrates the invariance of the correction term under re-parameterization as long as we exclude the $\mathbf{R}_{ij}$ term, because

$$\sum_{ij} W_{ij}\left(\mathbf{Q}_{ij} - \mathbf{P}_i\boldsymbol{\Phi}\mathbf{P}_j\right) = \sum_{ij} W^*_{ij}\left(\mathbf{Q}^*_{ij} - \mathbf{P}^*_i\boldsymbol{\Phi}\mathbf{P}^*_j\right).$$

We now show that the new expression for the covariance matrix estimator is invariant under re-parameterization within the classes of intrinsically linear, and intrinsically linear inverse, covariance matrices. A useful general result that we use below is

$$2\mathbf{W}^{ij} = \mathrm{tr}\left\{\frac{\partial\boldsymbol{\Sigma}}{\partial\sigma_i}(\mathrm{I} - \boldsymbol{\Sigma}^{-1}\mathbf{X}\boldsymbol{\Phi}X^{\mathrm{T}})\boldsymbol{\Sigma}^{-1}\frac{\partial\boldsymbol{\Sigma}}{\partial\sigma_j}\boldsymbol{\Sigma}^{-1}(\mathrm{I} - \boldsymbol{\Sigma}^{-1}\mathbf{X}\boldsymbol{\Phi}X^{\mathrm{T}})^{\mathrm{T}}\right\},$$

and $\mathbf{W}^*$ can be expressed similarly in terms of $\boldsymbol{\lambda}$ rather than $\boldsymbol{\sigma}$.

3.2. Intrinsically linear covariance structures

The first thing we note is that

$$\mathbf{R}^*_{ij} = X^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}\frac{\partial^2\boldsymbol{\Sigma}}{\partial\lambda_i\partial\lambda_j}\boldsymbol{\Sigma}^{-1}\mathbf{X} = 0,$$

for all i and j, while this does not hold for $\mathbf{R}_{ij}$. Also

$$\frac{\partial\boldsymbol{\Sigma}}{\partial\sigma_i} = \sum_s \frac{\partial\lambda_s}{\partial\sigma_i}\frac{\partial\boldsymbol{\Sigma}}{\partial\lambda_s} \quad \text{and} \quad \frac{\partial^2\boldsymbol{\Sigma}}{\partial\sigma_i\partial\sigma_j} = \sum_s \frac{\partial^2\lambda_s}{\partial\sigma_i\partial\sigma_j}\frac{\partial\boldsymbol{\Sigma}}{\partial\lambda_s}.$$

Now the trace component of the new term, when parameterized in terms of $\boldsymbol{\sigma}$, is

$$\begin{aligned}
&-\mathrm{tr}\left\{\frac{\partial^2\boldsymbol{\Sigma}}{\partial\sigma_i\partial\sigma_j}(\mathrm{I} - \boldsymbol{\Sigma}^{-1}\mathbf{X}\boldsymbol{\Phi}\mathbf{X}^{\mathrm{T}})\boldsymbol{\Sigma}^{-1}\frac{\partial\boldsymbol{\Sigma}}{\partial\sigma_t}\boldsymbol{\Sigma}^{-1}(\mathrm{I} - \boldsymbol{\Sigma}^{-1}\mathbf{X}\boldsymbol{\Phi}\mathbf{X}^{\mathrm{T}})^{\mathrm{T}}\right\}\\
&= -\sum_{ab}\frac{\partial^2\lambda_a}{\partial\sigma_i\partial\sigma_j}\frac{\partial\lambda_b}{\partial\sigma_t}\mathrm{tr}\left\{\frac{\partial\boldsymbol{\Sigma}}{\partial\lambda_a}(\mathrm{I} - \boldsymbol{\Sigma}^{-1}\mathbf{X}\boldsymbol{\Phi}X^{\mathrm{T}})\boldsymbol{\Sigma}^{-1}\frac{\partial\boldsymbol{\Sigma}}{\partial\lambda_b}\boldsymbol{\Sigma}^{-1}(\mathrm{I} - \boldsymbol{\Sigma}^{-1}\mathbf{X}\boldsymbol{\Phi}\mathbf{X}^{\mathrm{T}})^{\mathrm{T}}\right\}\\
&= -2\sum_{ab}\frac{\partial^2\lambda_a}{\partial\sigma_i\partial\sigma_j}\frac{\partial\lambda_b}{\partial\sigma_t}(\mathbf{W}^*)^{ab}.
\end{aligned}$$

Note how this simplification relies crucially on the intrinsic linearity of the parameterization. It follows that the new term can be expressed

$$-\frac{1}{2}\sum_{st}^{r} W_{st}\left[\sum_{ij}^{r} W_{ij}\sum_{ab}\frac{\partial^2\lambda_a}{\partial\sigma_i\partial\sigma_j}\frac{\partial\lambda_b}{\partial\sigma_t}(W^*)^{ab}\right]\boldsymbol{\Phi}\mathbf{P}_s\boldsymbol{\Phi} = -\frac{1}{2}\sum_{ij}^{r}\sum_{k}^{r} W_{ij}\frac{\partial^2\lambda_k}{\partial\sigma_i\partial\sigma_j}\boldsymbol{\Phi}\mathbf{P}^*_k\boldsymbol{\Phi}.$$

While the $\mathbf{R}$ term gives

$$\frac{1}{2}\sum_{ij}^{r} W_{ij}\boldsymbol{\Phi}\mathbf{R}_{ij}\boldsymbol{\Phi} = \frac{1}{2}\sum_{ij}^{r}\sum_{k}^{r} W_{ij}\boldsymbol{\Phi}\frac{\partial^2\lambda_k}{\partial\sigma_i\partial\sigma_j}\mathbf{P}^*_k\boldsymbol{\Phi}.$$

So, in this case, the two terms cancel exactly. This shows that as long as we have a intrinsically linear covariance parameterization then one can simply use the original correction formula in Kenward and Roger (1997), discarding the R_{ij} term. This is what the FirstOrder modifier on the DDFM = KR option does when using the GLIMMIX procedure in SAS. So the revised formula is not needed under any parameterization which is equivalent to a linear one, for instance an unstructured covariance matrix parameterized through variances and correlations.

3.3. Intrinsically linear inverse covariance structures

An example of covariance structures with an intrinsically linear inverse is the antidependence class (Kenward, 1987). As before, we assume that we have parameterizations in terms of $\boldsymbol{\sigma}$ and $\boldsymbol{\lambda}$. We can then express $\mathbf{R}_{ij}$ in the following form

$$\begin{aligned}
&-\sum_s \frac{\partial^2\lambda_s}{\partial\sigma_i\partial\sigma_j}\mathbf{X}^{\mathrm{T}}\frac{\partial\boldsymbol{\Sigma}^{-1}}{\partial\lambda_s}\mathbf{X} + \sum_{st}\frac{\partial\lambda_s}{\partial\sigma_i}\frac{\partial\lambda_t}{\partial\sigma_i}\left(\mathbf{X}^{\mathrm{T}}\frac{\partial\boldsymbol{\Sigma}^{-1}}{\partial\lambda_s}\boldsymbol{\Sigma}\frac{\partial\boldsymbol{\Sigma}^{-1}}{\partial\lambda_t}\mathbf{X} + \mathbf{X}^{\mathrm{T}}\frac{\partial\boldsymbol{\Sigma}^{-1}}{\partial\lambda_t}\boldsymbol{\Sigma}\frac{\partial\boldsymbol{\Sigma}^{-1}}{\partial\lambda_s}\mathbf{X}\right)\\
&= -\sum_s \frac{\partial^2\lambda_s}{\partial\sigma_i\partial\sigma_j}\mathbf{P}^*_s + \sum_{st}\frac{\partial\lambda_s}{\partial\sigma_i}\frac{\partial\lambda_t}{\partial\sigma_i}(\mathbf{Q}^*_{st} + \mathbf{Q}^*_{ts}).
\end{aligned}$$

In this case the trace term in the new term is

$$\begin{aligned}
&\mathrm{tr}\left\{\frac{\partial^2\boldsymbol{\Sigma}}{\partial\sigma_i\partial\sigma_j}(\mathrm{I}-\boldsymbol{\Sigma}^{-1}\mathbf{X}\boldsymbol{\Phi}\mathbf{X}^{\mathrm{T}})\frac{\partial\boldsymbol{\Sigma}^{-1}}{\partial\sigma_t}(\mathrm{I}-\boldsymbol{\Sigma}^{-1}\mathbf{X}\boldsymbol{\Phi}\mathbf{X}^{\mathrm{T}})^{\mathrm{T}}\right\}\\
&\quad=-\sum_{ab}\frac{\partial^2\lambda_a}{\partial\sigma_i\partial\sigma_j}\frac{\partial\lambda_b}{\partial\sigma_t}\mathrm{tr}\left\{\boldsymbol{\Sigma}\frac{\partial\boldsymbol{\Sigma}^{-1}}{\partial\lambda_a}\boldsymbol{\Sigma}(\mathrm{I}-\boldsymbol{\Sigma}^{-1}\mathbf{X}\boldsymbol{\Phi}\mathbf{X}^{\mathrm{T}})\frac{\partial\boldsymbol{\Sigma}^{-1}}{\partial\lambda_b}(\mathrm{I}-\boldsymbol{\Sigma}^{-1}\mathbf{X}\boldsymbol{\Phi}\mathbf{X}^{\mathrm{T}})^{\mathrm{T}}\right\}\\
&\qquad+\sum_{abc}\frac{\partial\lambda_a}{\partial\sigma_i}\frac{\partial\lambda_b}{\partial\sigma_j}\frac{\partial\lambda_c}{\partial\sigma_t}\mathrm{tr}\left\{\left(\frac{\partial\boldsymbol{\Sigma}}{\partial\lambda_a}\boldsymbol{\Sigma}^{-1}\frac{\partial\boldsymbol{\Sigma}}{\partial\lambda_b}+\frac{\partial\boldsymbol{\Sigma}}{\partial\lambda_b}\boldsymbol{\Sigma}^{-1}\frac{\partial\boldsymbol{\Sigma}}{\partial\lambda_a}\right)(\mathrm{I}-\boldsymbol{\Sigma}^{-1}\mathbf{X}\boldsymbol{\Phi}\mathbf{X}^{\mathrm{T}})\frac{\partial\boldsymbol{\Sigma}^{-1}}{\partial\lambda_c}(\mathrm{I}-\boldsymbol{\Sigma}^{-1}\mathbf{X}\boldsymbol{\Phi}\mathbf{X}^{\mathrm{T}})^{\mathrm{T}}\right\}.
\end{aligned}$$

Now the first term is identical to that for the previous case of the intrinsically linear class, and as before cancels with the first part of the revised R_{ij} term. We are therefore left in the new adjustment formula with a term

$$\frac{1}{2}\sum_{ij}^{r}W_{ij}\boldsymbol{\Phi}\sum_{st}\frac{\partial\lambda_s}{\partial\sigma_i}\frac{\partial\lambda_t}{\partial\sigma_i}(\mathbf{Q}^*_{st}+\mathbf{Q}^*_{ts})\boldsymbol{\Phi},$$

coming from the $\mathbf{R}_{ij}$ term, which is expressible as

$$\frac{1}{2}\sum_{ij}^{r}W_{ij}\boldsymbol{\Phi}(\mathbf{Q}_{ij}+\mathbf{Q}_{ij})\boldsymbol{\Phi}=\frac{1}{2}\sum_{st}^{r}W^*_{st}\boldsymbol{\Phi}(\mathbf{Q}^*_{st}+\mathbf{Q}^*_{ts})\boldsymbol{\Phi},$$

which is invariant under re-parameterization. Then there is a part coming from the new term in the expression

$$\begin{aligned}
&\frac{1}{4}\sum_{st}^{r}\sum_{ij}^{r}W_{st}W_{ij}\left[\sum_{abc}\frac{\partial\lambda_a}{\partial\sigma_i}\frac{\partial\lambda_b}{\partial\sigma_j}\frac{\partial\lambda_c}{\partial\sigma_t}\right.\\
&\quad\left.\mathrm{tr}\left\{\left(\frac{\partial\boldsymbol{\Sigma}}{\partial\lambda_a}\boldsymbol{\Sigma}^{-1}\frac{\partial\boldsymbol{\Sigma}}{\partial\lambda_b}+\frac{\partial\boldsymbol{\Sigma}}{\partial\lambda_b}\boldsymbol{\Sigma}^{-1}\frac{\partial\boldsymbol{\Sigma}}{\partial\lambda_a}\right)(\mathrm{I}-\boldsymbol{\Sigma}^{-1}\mathbf{X}\boldsymbol{\Phi}X^{\mathrm{T}})\frac{\partial\boldsymbol{\Sigma}^{-1}}{\partial\lambda_c}(\mathrm{I}-\boldsymbol{\Sigma}^{-1}\mathbf{X}\boldsymbol{\Phi}\mathbf{X}^{\mathrm{T}})^{\mathrm{T}}\right\}\right]\boldsymbol{\Phi}\mathbf{P}_s\boldsymbol{\Phi},
\end{aligned}$$

which can be re-expressed as

$$=\frac{1}{4}\sum_{st}^{r}\sum_{ij}^{r}W^*_{st}W^*_{ij}\left[\mathrm{tr}\left\{\left(\frac{\partial\boldsymbol{\Sigma}}{\partial\lambda_i}\boldsymbol{\Sigma}^{-1}\frac{\partial\boldsymbol{\Sigma}}{\partial\lambda_j}+\frac{\partial\boldsymbol{\Sigma}}{\partial\lambda_j}\boldsymbol{\Sigma}^{-1}\frac{\partial\boldsymbol{\Sigma}}{\partial\lambda_i}\right)(\mathrm{I}-\boldsymbol{\Sigma}^{-1}\mathbf{X}\boldsymbol{\Phi}\mathbf{X}^{\mathrm{T}})\frac{\partial\boldsymbol{\Sigma}^{-1}}{\partial\lambda_t}(\mathrm{I}-\boldsymbol{\Sigma}^{-1}\mathbf{X}\boldsymbol{\Phi}\mathbf{X}^{\mathrm{T}})^{\mathrm{T}}\right\}\right]\boldsymbol{\Phi}\mathbf{P}^*_s\boldsymbol{\Phi}.$$

So this part is also invariant under re-parameterization. We have shown that the whole of the modified adjustment is indeed invariant under transformation within the intrinsically linear inverse class. In contrast to the intrinsically linear class however, the additional terms are not zero, and so must be retained.

4. A simulation study

To explore the behaviour of the new variance estimator we provide some results from a small simulation study. A repeated measurements design has been used for this, with T equally spaced times ($T = 6, 12$), and no between-subject factors. A very simple linear model has been used with two parameters: β_A, an average over all the T times and β_C, a contrast between the first $T/2$ times and the last $T/2$ times. In this way the behaviour of the variance estimator can be explored for both a between-subject and a within-subject term. We consider three covariance structures as follows.

(1) AR(1): a stationary first order autoregressive structure, which has constant variance and correlation ρ^t between times t units apart. This two parameter structure belongs to neither of the intrinsically linear classes.
(2) AD(1): an antidependence covariance structure of order 1. This is defined as the class of covariance structures with an unconstrained positive-definite symmetric tri-diagonal inverse. This belongs to the intrinsically linear inverse class. It has $2T-1$ parameters (Kenward, 1987).
(3) UN: an unstructured form in which the covariance matrix is required only to be positive-definite symmetric. This has $T(T+1)/2$ parameters and is a member of the intrinsically linear class.

Sample sizes have been chosen to illustrate a range of performance of the new adjusted estimator, from sizes too small for the adjustment to be effective up to the point at which bias becomes negligible.

The choice of covariance structure for the generated data does not have a major influence on the properties of the estimators, so we here assume an AR(1) structure with $\rho = 0.7$. Other values for ρ produce very similar results. We note that the three structures are nested, in the sense that the AR(1) is a special case of the AD(1) which in turn is necessarily a special case of the unstructured. For each combination of T, sample size N, and structure, we present the percentage relative bias in the estimated variance of each parameter β_A and β_C:

$$100\left(\frac{\bar{V}_{\mathrm{ADJ}}-V_{\mathrm{SIM}}}{V_{\mathrm{SIM}}}\right),$$

Table 1
Percentage relative error in the variance estimators for the average and contrast parameters under a fitted stationary first order (AR(1)) covariance matrix. Simulation size: 10,000; T: number of times, N sample size.

T	N	Parameter							
		Average, β_A				Contrast, β_C			
		Unadj	Lin	Old	New	Unadj	Lin	Old	New
6	6	3	4	−6	0	−7	−3	8	−1
	12	1	1	−4	−1	−3	−1	4	0
	24	1	1	−2	0	−1	0	2	0
12	6	3	5	−5	0	−7	−1	3	0
	12	2	3	−2	0	−4	−2	1	−1
	24	1	1	−1	0	−2	0	1	0

Unadj: unadjusted asymptotic estimator, Lin and Old: the original adjusted estimator (2) respectively excluding and including R_{ij}, and New: the new estimator (4).

for $\bar{V}_{\text{ADJ}}$ the average over the simulations of the adjusted variance estimate, and V_{SIM} the variance over the simulations of the estimated parameter. The main source of variability in the empirical calculation of these quantities is in the denominator V_{SIM}, but this is relatively quick to calculate as it does not involve the adjustment factor. So, for results presented here, this has been calculated using 1,000,000 simulations while the numerator term $\bar{V}_{\text{ADJ}}$ has been calculated using 10,000 simulations.

The percentage relative bias is presented for each of four variance estimators: the unadjusted asymptotic estimator $\Phi(\hat{\sigma})$, the original adjusted estimator $\hat{\boldsymbol{\Phi}}_A^{\text{OLD}}$ (Eq. (2)) both excluding and including respectively the nonlinear term R_{ij}, and the new adjusted estimator $\hat{\boldsymbol{\Phi}}_A^{\text{NEW}}$ (Eq. (4)).

To illustrate the invariance properties derived in Section 3, alternative parameterizations are used for the AD(1) and UN structures. We consider each of the three covariance structures in turn.

The percentage relative biases are presented for the AR(1) structure in Table 1. Because this is a very parsimonious structure we expect relatively little bias anyway in the variance estimators. In contrast to the other two structures considered, additional times represent a form of replication, and so we expect smaller bias with $T = 12$ as opposed to $T = 6$. We consider first the behaviour of the variance estimator of the between-subject effect, β_A. Counter-intuitively, the uncorrected asymptotic variance estimator *over-estimates* the variance. It is important to remember that there are two main sources of bias in the unadjusted variance estimator (Kenward and Roger, 1997). Only one of these automatically leads to under-estimation of the variance on average, and it is this one that tends to influence common intuition. With intrinsically linear structures the second source also leads to under-estimation, and is in fact the same size as the first, up to the order of expansion considered here. For some other structures, including the AR(1), the second source acts in the opposite direction, and in some particular settings like the one considered here, may dominate the first source. While not a common occurrence in practice, it is important that any adjustment method used can accommodate such behaviour. Indeed, we see that the original adjustment of Kenward and Roger (1997) reverses the sign of the bias and appears to make it worse in absolute size. The new adjustment removes the bias while this is apparent in the other estimators. The picture is similar with the within-subject comparison, β_C, except that the signs of the bias in the unadjusted and old adjusted estimators are reversed.

Two alternative parameterizations have been used for the AD(1) structure. Recall that this belongs to the intrinsically linear inverse class. The first parameterization, which might be termed an autoregressive type, is defined in terms of generating equations for the covariance structure. Suppose that $E_1, \ldots, E_T$ are independent random variables, each with zero mean and unit variance, then the transformed variables $F_1, \ldots, F_T$, defined as follows

$$F_t = \begin{cases} \sigma_1 E_1 & t = 1 \\ \lambda_{t-1} F_{t-1} + \sigma_t E_t & t = 2, \ldots, T, \end{cases}$$

have an AD(1) covariance structure with $2T + 1$ parameters $\sigma_1^2, \ldots, \sigma_T^2$ and $\lambda_1, \ldots, \lambda_{T-1}$. The second parameterization, which we term the leading tri-diagonal, takes as parameters the T variances, and $T - 1$ covariances on the immediate off diagonal of the covariance matrix. For balanced complete data with a saturated means model, the restricted maximum likelihood estimators of the leading tri-diagonal parameters are unbiased, and in general will be expected to have relatively small bias. The percentage relative errors under these two parameterizations of the AD(1) structure are presented in Table 2.

For all estimators, and both contrasts, large negative bias is seen in all the estimators for the smaller sample sizes. This is smallest for the new adjustment and effectively removed by this method for the moderate to large sample sizes. As predicted, in both parameterizations the new adjustment has exactly the same performance, indeed the resulting estimates were identical. Note how the bias of the old estimator is not invariant in the same way although, as expected, for the tri-diagonal parameterization, for which we know the covariance parameters have negligible bias, the old and new adjustments perform in a very similar way. It is interesting that in this example the old adjustment with R_{ij} removed, *i.e.* with the so-called "linear" adjustment, is seen to give a very similar bias to the new estimator. There are numerical differences in the corresponding estimates, but too small to be apparent given the significant figures in the presentation of the results.

Three alternative parameterizations have been used with the unstructured covariance matrix. The first uses the T variances and $T(T-1)/2$ covariances as parameters, and this represents a directly *linear* parameterization. The second

Table 2
Percentage relative error in the variance estimators for the average and contrast parameters under a fitted first order antidependence (AD(1)) covariance matrix, with (1) the autoregressive, and (2), the leading tri-diagonal parameterizations. Simulation size: 10,000; T: number of times, N sample size.

T	N	Unadj	Parameterization					
			Autoregressive			Tri-diagonal		
			Lin	Old	New	Lin	Old	New
Average: β_A								
6	12	−43	−17	−20	−17	−17	−17	−17
	24	−20	−2	−4	−2	−2	−2	−2
	48	−11	0	−1	0	0	0	0
	96	−6	−1	−2	−1	−1	−1	−1
12	12	−59	−30	−33	−30	−30	−30	−30
	24	−33	−9	−11	−9	−9	−9	−9
	48	−15	1	−1	1	1	1	1
	96	−11	−2	−3	−2	−2	−2	−2
Contrast: β_C								
6	12	−21	−3	−9	−3	−3	−3	−3
	24	−12	−2	−6	−2	−2	−2	−2
	48	−4	1	−1	1	1	1	1
	96	−6	−1	−2	−1	−1	−1	−1
12	12	−40	−12	−17	−12	−12	−12	−12
	24	−16	1	−3	1	1	1	1
	48	−9	0	−1	0	0	0	0
	96	−5	−1	−2	−1	−1	−2	−1

Unadj: unadjusted asymptotic estimator, Lin and Old: the original adjusted estimator (2) respectively excluding and including R_{ij}, and New: the new estimator (4).

Table 3
Percentage relative error in the variance estimators for the average and contrast parameters under a fitted unstructured (UN) covariance matrix, with (1) linear, (2) correlation, and (3) Cholesky parameterizations. Simulation size: 10,000; T: number of times, N sample size.

T	N	Unadj	Parameterization								
			Linear			Correlation			Cholesky		
			Lin	Old	New	Lin	Old	New	Lin	Old	New
Average: β_A											
6	12	−62	−34	−34	−34	−34	−34	−34	−34	−42	−34
	24	−33	−9	−9	−9	−9	−11	−9	−9	−15	−9
	48	−15	−1	−1	−1	−1	−2	−1	−1	−4	−1
	96	−10	0	0	0	0	0	0	0	−3	0
Contrast: β_C											
6	12	−62	−34	−34	−34	−34	−34	−34	−34	−45	−34
	24	−34	−11	−11	−11	−11	−11	−11	−11	−20	−11
	48	−18	−4	−4	−4	−4	−4	−4	−4	−10	−4
	96	−8	0	0	0	0	0	0	0	−3	0

Unadj: unadjusted asymptotic estimator, Lin and Old: the original adjusted estimator (2) respectively excluding and including R_{ij}, and New: the new estimator (4).

parameterization keeps the variances, but replaces the covariances by the correlations, and so is nonlinear. The third parameterization uses the elements of the Cholesky decomposition of the covariance matrix, and is again nonlinear. Because of computational demands, the simulations have been done only for $T = 6$ times of measurement. As this is an intrinsically linear covariance structure we expect identical results with the so-called "linear" and new adjustments, and this is indeed what is seen across the alternative parameterizations. We also expect similar results from both contrasts, and again this is what is seen from the results. The lack of invariance for the old estimator is also clearly seen, with marked differences in bias across the parameterizations. We see that the new adjustment, which is in this case equivalent to the simple "linear" form, successfully removes the bias in all but the smallest samples, irrespective of the particular covariance parameterization in the very smallest samples (see Table 3).

5. Two examples

Here we apply the new variance adjustment to two examples. The first is a setting in which a non-standard parameterization is used in order to induce a constraint on the covariance structure in the model, resulting in a nonlinear parameterization. The new term in the small sample adjustment means that the standard errors are identical for two alternative parameterizations. The second example shows, in addition to the behaviour of the new adjustment, the

Table 4
Standard errors (approximate degrees of freedom) for the sensitivity (β_1) and specificity (β_2) of the telomerase marker from the meta-analysis of Riley et al. (2007). Covariance parameterizations at study level: (1) unstructured (no constraint), (2) Cholesky, and (3) unstructured with correlation.

Method	Expected information			Observed information		
	UN	Cholesky	Correlation	UN	Cholesky	Correlation
Sensitivity:	$\beta_1 = 1.180$	$\beta_1 = 1.166$	$\beta_1 = 1.166$	$\beta_1 = 1.180$	$\beta_1 = 1.166$	$\beta_1 = 1.166$
Standard	0.181	0.186	0.186	0.181	0.186	0.186
Old correction	0.195 (8.2)	0.179 (9.7)	0.187 (9.7)	0.209 (10)	0.180 (11)	0.187 (11)
New correction	0.195 (8.2)	0.183 (9.7)	0.183 (9.7)	0.209 (10)	0.184 (11)	0.184 (11)
Specificity:	$\beta_2 = 2.082$	$\beta_2 = 2.058$	$\beta_2 = 2.058$	$\beta_2 = 2.082$	$\beta_2 = 2.058$	$\beta_2 = 2.058$
Standard	0.556	0.554	0.554	0.556	0.554	0.554
Old correction	0.572 (8.7)	0.537 (8.8)	0.555 (8.8)	0.590 (8.3)	0.536 (8.6)	0.555 (8.6)
New correction	0.572 (8.7)	0.555 (8.8)	0.555 (8.8)	0.590 (8.3)	0.555 (8.6)	0.555 (8.6)

importance of using the small sample correction when including baseline as part of a repeated measures analysis of variance rather than as a covariate. Note that for both examples the observed information matrix is used in the variance calculations.

EXAMPLE 1

Riley et al. (2007) present a method for bivariate random effects meta-analysis that they apply to two examples, one of which involves a diagnostic marker, and is taken from Glas et al. (2003). We fit their model to the same data, and use our modified variance estimator for the estimated effects. Sensitivity and specificity for diagnosing primary bladder cancer are presented for a particular marker, telomerase, from 10 studies. The logit transformed sensitivity and specificity are jointly modelled by Riley et al. as follows. The cell counts for study 7 have been increased by 0.5 conforming to their *continuity correction*. Denote by $\mathbf{Y}_i = (Y_{i1}, Y_{i2})^{\mathrm{T}}$ the logit sensitivity and specificity from the ith study, $i = 1, \ldots, 10$. Suppose that $\boldsymbol{\beta} = (\beta_1, \beta_2)$ represents the population average for Y_i, with random study effects $\mathbf{U}_i = (U_{i1}, U_{i2})^{\mathrm{T}}$ representing deviations from this, *i.e.* assuming normality for the random variables, we can write:

$$\mathbf{Y}_i \mid \mathbf{U}_i \sim \mathrm{N}(\boldsymbol{\beta} + \mathbf{U}_i, \Gamma_i),$$
$$\mathbf{U}_i \sim \mathrm{N}(0, \boldsymbol{\Lambda}), \quad i = 1, \ldots, 10,$$

for Γ_i the ith within-study covariance matrix, and Λ the between-study covariance matrix. Marginally, this implies $\boldsymbol{\beta}_i \sim \mathrm{N}(\boldsymbol{\beta}, \boldsymbol{\Lambda} + \Gamma_i)$. Because the sensitivity and specificity are calculated using different observations in each study, Riley et al. (2007) assume that the covariance in each Γ_i is zero. The remaining two variances are treated as known, although in practice obtained as estimates from each study. The elements of Λ therefore are the only covariance parameters to be estimated in the meta-analysis.

This type of model is relatively easy to fit using software such as the MIXED procedure in SAS/STAT® software (*SAS 9.1.3 Help and Documentation*, Cary, NC: SAS Institute Inc., 2000–2004). In the following code the three parameters in the matrix $\boldsymbol{\Lambda}$ are determined by the random statement, while the known variances and zero covariance are specified through the repeated statement, and held at their fixed known value using the hold facility on the parms statement.

```
proc mixed data=riley;
class study type;
model logity=type/noint;
random type / subject=study type=UN ddfm=KR;
repeated type / subject=study group=study type=UN;
parms / parmsdata=pdata hold=4 to 33;
run;
```

When this model is fitted to these data in this way the restricted maximum likelihood estimator for the matrix $\boldsymbol{\Lambda}$ is not positive definite. The UN column in Table 4 shows the results for this model. Here the new term described in this paper is having no effect since the parameterization of the covariance matrix is linear. But we see that even though much of the variation is within trial and hence regarded as fixed, the estimated variance for β_1 is inflated by about 10%.

The easiest way to constrain $\boldsymbol{\Lambda}$ in the MIXED procedure is to use on the random statement either a correlation parameterization of the unstructured (UNR), or the second order factor analytic structure (FA0(2)), which is effectively the Cholesky decomposition. This is important as the restricted maximum likelihood estimator for these data is on the boundary with a correlation of minus one, as indicated by Riley et al. (2007). The results for these two parameterizations are shown in the remaining columns of Table 4. Those in the *standard* row are identical to those for the BRMA method in Riley et al. (2007). As expected these parameter estimates are different to those obtained with the unconstrained model.

For the purposes of the small sample approximations, the constrained parameter is excluded from calculation and is assumed known and fixed at the boundary value. This means that the correlation parameterization of the unstructured matrix has diagonal elements σ_1 and σ_2 with off-diagonal $-\sqrt{\sigma_1\sigma_2}$, while the Cholesky decomposition has diagonals σ_1^2 and σ_2^2 with off-diagonal $-\sigma_1\sigma_2$. The results are very similar for expected and observed information. When the new adjustment term is included, the estimated variances for $\boldsymbol{\beta}$ are identical to four decimal places between the two different parameterizations.

Table 5
Average afternoon body temperature at Baseline and 2 days after vaccination in two groups B and C of seven ferrets.

Group B		Group C	
Baseline	Response	Baseline	Response
38.3905	38.4499	38.3083	39.5305
38.2982	38.4071	38.0473	38.6537
38.0390	38.5397	38.2175	39.1029
38.4314	38.2692	38.7405	39.4100
38.1079	38.0871	38.6047	39.2516
38.5819	38.8815	38.1381	38.9442
38.2760	38.7505	38.0743	39.2386

Table 6
Standard errors (approximate degrees of freedom) for the simple two group analysis of variance with baseline treated either (a) as a covariate, or (b) as an outcome variable with no associated treatment effect. Covariance parameterizations: (1) unstructured, (2) Cholesky, and (3) unstructured with correlation.

Method	UN	Cholesky	Correlation
(a) ANOVA	0.133 (11)		
(b) Asymptotic	0.127 (12)	0.127 (12)	0.127 (12)
(b) Old correction	0.137 (12)	0.130 (12)	0.135 (12)
(b) New correction	0.137 (12)	0.137 (12)	0.137 (12)

EXAMPLE 2

The second example illustrates the importance of using the small sample correction. Table 5 presents selected data from a GlaxoSmithKline study on the effect of immunization with flu vaccine on ferret body temperature. Body temperature was recorded every ten minutes from 14 ferrets in two equally sized randomised groups (B and C), and here temperature is averaged across the afternoon. A baseline value on Day 16, was taken before immunization, and the response to vaccination was measured on the afternoon of Day 18. The seven ferrets in Group B were primed and vaccinated while the seven in group C were primed but not vaccinated.

The primary comparison of interest is the difference in group means, adjusted for baseline, and a conventional analysis for this would use a two-way analysis of variance with baseline included as a fixed covariate. We consider now an alternative way of approaching the analysis in which the two temperature measurements are treated as bivariate repeated measurements with no treatment difference at day 16 but a treatment difference at day 18. In the bivariate normal model the two-by-two covariance matrix has three elements; two variances (before and after immunization) and a covariance which, after scaling by the baseline variance, is equal to the regression parameter on baseline covariate in the analysis of variance model.

Table 6 presents the standard errors for the difference in mean temperature between groups B and C under the usual analysis of variance model and the bivariate repeated measures model using the three parameterizations, unstructured, Cholesky, and unstructured with correlations.

As expected the estimated difference 0.677 is identical using all methods. The only difference is in the standard errors. We expect the analysis of variance standard error to be slightly different from that from the bivariate repeated measures, as it is based on the conditional distribution of the second measurement given the first. This is also reflected in the associated degrees of freedom, 11 from the analysis of covariance and 12 from the joint model.

The results using observed and expected information are very similar so we present only those from the expected information. Because the unstructured covariance matrix is intrinsically linear, we know from the invariance results in Section 3 that the new term reduces the adjustment to the simple "linear" form, the same variance estimate will arise under the three parameterizations. These are all confirmed from the results. By contrast, the old adjustment shows increasing differences from the new with the increasing nonlinearity of the covariance parameterization.

6. Conclusion

Kenward and Roger (1997) derived a small sample variance estimator of fixed effects in the multivariate normal linear model. We have identified a problem with this that is apparent under certain nonlinear covariance structures. This is caused by a failure to adjust for bias in the parameter estimators of the covariance structure, an issue that does not arise under linear structures. The required missing term has been derived using a result of Cox and Snell (1968), and incorporated into the variance estimator. We have shown that this new variance estimator is invariant under re-parameterization of the covariance structure within two important and commonly used classes of covariance matrix. For the first class, termed *intrinsically linear*, the new estimator reduces to the simple form that is appropriate for directly linear structures, the so-called "linear" or "first-order" form, leading to great simplification in the calculation of adjusted variance estimator. In the second class, termed *intrinsically linear inverse*, the new adjustment does not reduce to this simple form.

A small simulation study with three covariance structures confirmed that the proposed modification behaved as intended in more complex settings, both in terms of invariance and reduction in bias. In particular, in a counter-intuitive setting in

which the unadjusted variance estimator leads to *over-estimation* of the variance, the new estimator corrects appropriately for this, while both the old and "linear" forms do not.

It is proposed that this new estimator be used in place of that developed in Kenward and Roger (1997). This will be of particular benefit for those mixed model computational algorithms that are based upon variance/correlation parameterization of covariance structures such as the lme function in *R*.

The main thrust of Kenward and Roger (1997) was the construction of approximate pivotal quantities using small sample *F* approximations on which inference for fixed effects from REML could be based. The adjusted variance estimator is an important component of this pivot, and we will consider elsewhere whether the current developments have implications for this approach under nonlinear covariance structures. It can however be shown that two terms, A_1 and A_2 from the original paper (page 987), that drive both the scale factor λ and the denominator degrees of freedom, are both invariant under transformation of the covariance parameters, which suggests that the impact is unlikely to be significant in most settings.

We have not addressed the issue of covariance parameter estimators lying on a boundary of the parameter space, usually, but not necessarily, zero. This is most common with variance components that are assumed to be positive. The series expansions on which the results above are based do not apply in such settings. The main problem is usually the lack of information in the data at hand on the relevant variance components. Methods using formal sensitivity analyses, or incorporating external information directly through, for example, Bayesian methods, would then seem more appropriate than classical analyses such as those considered here.

Acknowledgments

We are grateful to David Gower of GlaxoSmithKline PLC for access to the ferret immunization data. We thank Oliver Schabenberger and an anonymous referee for stimulating the investigation of invariance under re-parameterization.

References

Brown, H., Prescott, R., 2006. Applied MIXED Models in Medicine, second ed. Wiley, New York.
Chen, X., Wei, L., 2003. A comparison of recent methods for the analysis of small-sample cross-over studies. Statist. Med. 22, 2821–2833.
Cox, D.R., Snell, E.J., 1968. A general definition of residuals (with discussion). J. R. Statist. Soc. B 30, 248–275.
Glas, A.S., Roos, D., Deutekom, M., Zwindermann, A.H., Bossuyt, P.M., Kurth, K.H., 2003. Tumor markers in the diagnosis of primary bladder cancer. A systematic review. J. Urol. 169, 1975–1982.
Guiard, V., Spilke, J., Dänicke, S., 2003. Evaluation and interpretation of results for three cross-over designs. Arch. Animal Nutr. 57, 177–195.
Jeske, D.R., Harville, D.A., 1988. Prediction-interval procedures and (fixed-effects) confidence interval procedures for linear mixed models. Commun. Stat. Theory 17, 1053–1087.
Kackar, A.N., Harville, D.A., 1984. Approximations for standard errors of estimators of fixed and random effects in mixed linear models. J. Amer. Statist. Assoc. 79, 853–862.
Kenward, M.G., 1987. A method for comparing profiles of repeated measurements. Appl. Stat.-J. Roy. St. C. 36, 296–308.
Kenward, M.G., Roger, J.H., 1997. Small sample inference for fixed effects estimators from restricted maximum likelihood. Biometrics 53, 983–997.
Kowalchuk, R.K., Keselman, H.J., Algina, J., Wolfinger, R.D., 2004. The analysis of repeated measurements with mixed-model adjusted F tests. Educ. Psychol. Meas. 64, 224–242.
Littell, R.C., Milliken, G.A., Stroup, W.W., Wolfinger, R.D., Schabenberger, O., 2006. SAS System for MIXED Models, second ed. SAS Institute Inc., Cary, NC.
Riley, R.D., Abrams, K.R., Sutton, A.J., Lambert, P.C., Thompson, J.R., 2007. Bivariate random-effects meta-analysis and the estimation of between-subject correlation. BMC Med. Res. Methodol. 7, 3.
Savin, A., Wimmer, G., Witkovsky, V., 2003. On Kenward–Roger confidence intervals for common mean in interlaboratory trials. Meas. Sci. Rev. 3, 53–56.
Spilke, J., Piepho, H.-P., Hu, X., 2005. A simulation study on tests of hypotheses and confidence intervals for fixed effects in mixed models for blocked experiments with missing data. J. Agr. Biol. Envir. St. 10, 374–389.
Vallejo, G., Fernandez, P., Herrero, F.J., Conejo, N.M., 2004. Alternative procedures for testing fixed effects in repeated measures designs when assumptions are violated. Psicothema 16, 498–508.
Yutaka, U., Feng, Z., Diehr, P., Mclerran, D., Beresford, S.A.A., McCulloch, C.E., 2004. Evaluation of community-intervention trials via generalized linear mixed models. Biometrics 60, 1043–1052.