Quantifying the correlation between variance components: an extension to the double-hierarchical generalised linear model

Jessica G. King[1,2,3], Joel L. Pick[1], and Jarrod D. Hadfield[1]

[1]Institute of Ecology and Evolution, School of Biological Sciences, University of Edinburgh, Edinburgh, UK

[2]**Present address:** Centre for Ecology, Evolution and Environmental Changes (CE3C) & CHANGE – Global Change and Sustainability Institute, Faculty of Sciences of the University of Lisbon, Lisbon, Portugal

[3]**Corresponding author:** jaking@ciencias.ulisboa.pt

Abstract

1. The variational properties of biological systems are an increasing focus of current research, and statistical methods are required for drawing inferences about the processes that determine them.

2. Double-hierarchical generalised linear models (DHGLM) are ideally suited for studying variational properties since they provide a direct way of modelling the distribution of variances.

3. Although DHGLM have mainly been used to model heterogeneous residual variances over groups, models have been proposed that also allow heterogeneous random effect variances. However, these multi-way DHGLM make the assumption that the residual variance of a group is independent of its random effect variance. Here, using a Bayesian approach, we extend multi-way DHGLMs so that the correlation between residual and random-effect variances can be estimated.

4. Using simulated data, the performance of the model is compared with the non-DHGLM models that have traditionally been used to estimate such correlations. The proposed model is shown to perform well at estimating all model parameters, and in particular performs better than alternative models at estimating the correlation among variance components.

5. Numerical analyses are complemented with theoretical work showing the expected bias when using non-DHGLM models. In some cases, commonly-used non-DHGLM models are even expected to get the sign of the correlation wrong.

1 Introduction

For many questions in ecology and evolution, we want to make inferences about parameters that we do not directly observe. Historically, this has often involved estimating parameters in individual models and making inferences across model estimates. However, not accounting for uncertainty in these estimates can easily lead to the wrong conclusions and, indeed, there are several clear cases

of this in ecology and evolution (e.g. Morrissey & Hadfield, 2012, Morrissey, 2016). Currently, the most common way to deal with this uncertainty is to use meta-analysis, a method developed to make inferences from estimates scattered across the literature using reported sampling variances. However, this approach is limited by the loss of information in such summary statistics and should not be a replacement of appropriate methodology to analyse raw data.

Decomposing phenotypic variance into components caused by genetic and environmental factors is central to quantitative genetics and led to the development of the mixed model in the 1950's (Henderson *et al.*, 1959). Since its development, the mixed model (referred to as HGLM - hierarchical generalised linear model - henceforth) is now widely used across the sciences, and is one of the main modelling frameworks used by evolutionary biologists and ecologists (Nakagawa & Schielzeth, 2010; Bolker *et al.*, 2009). The variance decomposition of a quantitative trait often assumes that variance components are homogeneous, with a single variance being estimated for each set of random effects and/or the residuals. However, an increasing number of studies have shown that the variances of random effects or residuals can vary across groups. For example, between-individual (environmental) variances have been shown to vary over herds of livestock (Brotherstone & Hill, 1986), and more recent work in behavioural ecology has shown within-individual (residual) variances, or repeatability, of behaviour to vary over individuals (Schwagmeyer & Mock, 2003; Westneat *et al.*, 2013; Stamps *et al.*, 2012; Martin *et al.*, 2017). Appropriate methodology has been developed to estimate the variance of variances directly (see below). However, studies have also suggested that variance components that vary over the same groups might be correlated, yet currently there is no methodology for estimating this correlation directly. In lieu of an appropriate methodology, studies have instead estimated the correlation between *estimates* of variances. For example, mutational, genetic and/or environmental variances have been estimated for several groups (traits) using HGLM (e.g. Houle, 1992) or even non-HGLM (e.g. Landry *et al.*, 2007), and the correlations in the *estimates* across traits have been used to draw important conclusions about the determinants of genetic variation. However, these correlations are expected to reflect, in part, the sampling (co)variances of the estimates which will increase in magnitude as the information to estimate a variance component decreases. This problem can be acute in some high-throughput methodologies where it is cheap to

measure many groups, but the replication within a group necessary to estimate the variance may be modest. For example, microarray and RNA-seq technologies allow many groups (genes) to be measured simultaneously, but the level of within-group replication might be small: for example, in Landry *et al.* (2007) the mutational variance in expression for each gene was estimated from only five lines with four replicates per line. In contrast to other central inference problems in ecology and evolution (Morrissey, 2016) the implications of using estimates, rather than true values, to estimate the correlation in variances has not been well studied.

Several models and inference procedures have been put forward to deal with and estimate heterogeneous variances. Traditionally, heterogeneous variances were dealt with by estimating variance components separately for each group, such as herd or year, using hierarchical generalised linear models (HGLM, of which the linear mixed model is a particular case). This is essentially equivalent to modelling variances as fixed effects with no shrinkage to an underlying distribution (Hill, 1984). Multiple studies have used this approach to compare estimates of genetic and/or environmental variances across groups such as traits (e.g. Houle, 1992) or years (e.g. Nicolaus *et al.*, 2013) and the methodology is still advocated for studying heterogeneous variances (Royauté & Dochtermann, 2021). However, treating variance components as fixed effects is expected to give poor results when the sample size is small per group, with variation in variance components being overestimated due to sampling error. Hill (1984) suggested addressing the problems of the HGLM method by *shrinking* the variance component estimates towards their mean according to some (prior) distribution, i.e. treat the variance parameters as random effects. Gianola *et al.* (1992) developed an empirical Bayes approach using a scaled inverse chi-squared distribution for the variances, but only the mean of this distribution was estimated, not it's dispersion. Consequently, the degree of shrinkage was not informed by the data but set a priori. Using a log-normal distribution for the variances, Foulley *et al.* (1992) came up with a strategy for estimating both the mean and dispersion, and this model is now known as a double-hierarchical generalised linear model (DHGLM) (Lee & Nelder, 2006). Most DHGLM used in the literature only allow the residual variance to be heterogeneous over groups, although San Cristobal *et al.* (1993) extended Foulley (1992) *et al*'s method to also accommodate differences in random-effect (genetic) variances among groups.

Such DHGLM are hereon referred to as multi-way DHGLM. Related to, but independent of this work, stochastic variance models were developed in econometrics whereby the residual variance was allowed to vary over time according to an autoregressive process (Taylor, 1982). These models were later extended to deal with multiple assets, allowing correlations between the residual variances of different response variables at a given time (Harvey *et al.*, 1994).

While the models described above deal with and quantify the heterogeneity of variance, to the best of our knowledge no multi-way DHGLM has been implemented to accurately determine the correlation between sets of variances that vary over the same groups. Here, we extend the model of San Cristobal *et al.* (1993) so that the correlation between random-effect and residual variances can be estimated. Using simulated data we assess the accuracy of the multi-way DHGLM model and compare it to alternative non-HGLM and HGLM methods. A more general treatment of non-HGLM and HGLM methods is also given using analytical results.

2 Methods

Here, we describe a multi-way DHGLM, which includes a parameter for the covariance between variance components (on the log scale) that vary over the same groups. We then briefly describe the alternative non-HGLM and HGLM approaches and the results needed to derive theoretical expectations for their (co)variance estimates. Finally, we use a simulation scheme to evaluate the estimation accuracy of all modelling approaches.

2.1 Double-Hierarchical Generalised Linear Model with Covariance Structure

To give biological motivation to the model described, we could imagine a researcher would like to assess whether environmental and genetic variances covary across traits (i.e. do traits with high genetic variance also have high environmental variance?). To answer this question, the researcher could have made measurements of multiple traits (groups) from multiple individuals from different clones or inbred lines (subgroups). In this design, the genetic variance can be estimated as the

between-subgroup variance (e.g. Denver *et al.*, 2005, Landry *et al.*, 2007, Huang *et al.*, 2016, Lafuente *et al.*, 2018, Chinchilla-Ramírez *et al.*, 2020). The model used to analyse such data is described in two stages. First is the *Mean Model* that describes variation around the average value, and second is the *Dispersion Model* that describes how such variation (co)varies across groups. In the context of our example, the parameters of the Mean Model quantify the genetic and environmental variances for each trait, while those of the Dispersion Model quantify how genetic and environmental variances (co)vary over traits.

The Mean Model is given by

$$y_{ijk} = \mu + t_i + u_{ij} + e_{ijk}, \tag{1}$$

where y_{ijk} is the k^{th} observation made on subgroup j (e.g. line) of group i (e.g. trait). μ is the intercept, representing the expected value for an observation irrespective of which group was measured. t_i is the expected deviation of group i from the intercept (e.g. the average effect of trait i) and u_{ij} is the expected deviation caused by subgroup j from group i (e.g. average effect of line j on trait i). e_{ijk} is a residual effect associated with the specific observation. t, u and e are normally distributed random variables, and their distributions are given in Table 1.

The Dispersion Model is where a double-hierarchical model is developed in which the variance of subgroup effects (V_u, e.g. between-line variance) and residual effects (V_e, e.g. environmental variance) are drawn from a joint distribution across groups. Specifically, the subgroup and residual variances for the groups are assumed to follow a multivariate log-normal distribution, with the logarithm of the subgroup variance for group (trait) i given by

$$\log(V_{u(i)}) = \mu_{\log(V_u)} + d_{\log(V_u(i))} \tag{2}$$

where $\mu_{\log(V_u)}$ is the mean subgroup variance across groups (i.e. the mean between-line variance across traits) and $d_{\log(V_u(i))}$ is the deviation of the subgroup variance from the mean for group i (i.e. the degree to which the between-line variance of trait i deviates from that of the average trait), all

on the log-scale. Similarly, the logarithm of the residual variance for group i is given by

$$\log(V_{e(i)}) = \mu_{\log(V_e)} + d_{\log(V_e(i))}. \tag{3}$$

The deviations of the log variances for each group, $d_{\log(V_u(i))}$ and $d_{\log(V_e(i))}$, are assumed to come from a multivariate normal distribution,

$$\begin{bmatrix} d_{\log(V_u(i))} \\ d_{\log(V_e(i))} \end{bmatrix} \sim MVN(0, \mathbf{C}). \tag{4}$$

with zero mean and covariance matrix

$$\mathbf{C} = \begin{bmatrix} V_{\log(V_u)} & C_{\log(V_u),\log(V_e)} \\ C_{\log(V_u),\log(V_e)} & V_{\log(V_e)}. \end{bmatrix} \tag{5}$$

$V_{\log(V_u)}$ and $V_{\log(V_e)}$ represent the variances of the log subgroup (between-line) and residual (within-line) variances, respectively, over groups, and the covariance

$$C_{\log(V_u),\log(V_e)} = \rho_{\log(V_u),\log(V_e)} \sqrt{V_{\log(V_u)} V_{\log(V_e)}}, \tag{6}$$

where $\rho_{\log(V_u),\log(V_e)}$ is the log-scale correlation between the two variance components. In formulating such a relationship between variance components, it is assumed that they covary linearly on the log-scale. These log-scale parameters can be transformed to the arithmetic scale if required (see Equations 10-12 below).

Extensions to the Dispersion Model can be made to accommodate further sources of heterogeneity in variance components that may be considered important. For instance, if residual (within-line) variances are believed to vary over subgroups (lines) as well as over groups (traits), then the Dispersion Model should include subgroup as well as group random effects. Variance components may also vary systematically with respect to some classifying factor or continuous variable, in which case additional fixed effects other than the intercept can be included in the Dispersion Model. In

Table 1: Fixed and random effects, and their distributions

Model	Symbol	Parameter	Distribution
Mean	μ	Intercept (average value of y)	Constant
	t_i	Average effect of group i on y	Normal; mean 0 and variance V_t
	u_{ij}	Average effect of subgroup j from group i	Normal; mean 0 and variance $V_{u(i)}$ for each group i
	e_{ijk}	Residual effect for observation i from subgroup j from group i	Normal; mean 0 and variance $V_{e(i)}$ for each group i
Dispersion	$\mu_{\log(V_u)}$	Mean $\log(V_u)$	Constant
	$\mu_{\log(V_e)}$	Mean $\log(V_e)$	Constant
	$d_{\log(V_u(i))}$	Average effect of group i on $\log(V_u)$	Multivariate normal; mean 0 and covariance
	$d_{\log(V_e(i))}$	Average effect of group i on $\log(V_e)$	matrix $\mathbf{C}$ (Equation 5)

particular, t_i or $t_i + u_{ij}$ from the Mean Model may be included as (log-scaled) covariates in the Dispersion Model in order to accommodate any mean-variance coupling. These extensions to the basic model are detailed in the Supporting Information.

2.2 non-HGLM and HGLM

The basic architecture of non-HGLM and HGLM (ANOVA-based methods, with fixed and random effects respectively) is the same as that of the DHGLM Mean Model (Equation 1). The key differences lie in the distributions of t and u. For t, the DHGLM most naturally assumes them to be random effects drawn from a normal distribution with variance V_t. In both non-HGLM and HGLM, groups are analysed one at a time and the group effects are then the intercepts of the models and hence fixed rather than random. Since the total number of observations per group is often likely to be large, this distinction is likely to have little effect since group (trait) means will be well estimated. The subgroup effects, and their (co)variances, are however treated differently in the three approaches and this is likely to have consequences.

Non-HGLM (fixed-effect ANOVA): This model is similar to Equation 1 but the parameter u_{ij} is treated as fixed rather than random, and the estimate of the subgroup variance $V_u(i)$ is obtained by taking the variance of the u_{ij} *estimates*. The subgroup variance is expected to be overestimated due to sampling variance contributing to the variance of the u_{ij} estimates. Estimates of how the subgroup and residual variances covary are obtained from the (co)variance of the $V_u(i)$ and $V_e(i)$ *estimates* rather than through a model of how they (co)vary. An example where this type of analysis was used to calculate the correlation between variance components can be found in Landry *et al.*

(2007) (in which groups are traits and subgroups are lines), to determine whether traits with higher mutational variance were also more prone to environmental variation.

ANOVA-based HGLM (random-effect, or repeated-measures, ANOVA): This model is identical to Equation 1 and the subgroup effects are assumed to come from a distribution and hence treated as random. Estimates of the subgroup variance $V_u(i)$ are known to be unbiased in the balanced case analysed below. However, estimates of how the subgroup and residual variances covary are obtained from the (co)variance of the $V_u(i)$ and $V_e(i)$ *estimates* and so the covariance will be biased when the sampling errors on $V_u(i)$ and $V_e(i)$ are correlated. This approach was adopted by Denver *et al.* (2005) to ask whether traits with more standing genetic variation are also more prone to mutational variation.

REML-based HGLM: This model is identical to the ANOVA-based HGLM, although typically estimates of the variances are restricted to be non-negative and have better properties when the design is not balanced. Because estimates of the variances must be non-negative, the estimate of the subgroup variance is known to be upwardly biased when sample sizes are low (where large sampling error can generate negative variance estimates). It is currently the most widespread method for estimating variance components. An example of its application can be found in Hoffmann *et al.* (2016) where literature-derived estimates of (standardised) genetic and environmental variances were compared across livestock traits, and the majority of these estimates were obtained using REML.

2.3 Theoretical expectations of ANOVA-based non-HGLM and HGLM

The first and second moments of the sampling distribution for V_u and V_e can be obtained for ANOVA-based estimates when the design is balanced. Theoretical expectations are thus obtained for the expected values of the (co)variances of *estimated* variance components from ANOVA-based models (non-HGLM and HGLM). Throughout, HGLM estimates are denoted with hat symbols (e.g. $\widehat{\mu_{V_u}}$), while non-HGLM estimates are denoted with tilde symbols (e.g. $\widetilde{\mu_{V_u}}$).

For a variance component V_x (herein V_u or V_e) that varies over groups with mean μ_{V_x} and variance V_{V_x}, the expected mean and variance of the estimates from an ANOVA-based HGLM

(random-effect ANOVA) are given by

$$E[\widehat{V_x}] = \mu_{V_x} \tag{7}$$

and

$$E[\widehat{V_{V_x}}] = V_{V_x} + E[Var(\widehat{V_x})], \tag{8}$$

respectively, where $E[Var(\widehat{V_x})]$ is the expected sampling variance of V_x. Having C_{V_x,V_y} as the covariance between V_x and V_y, the covariance between HGLM *estimates* of V_x and V_y is:

$$E[\widehat{C_{V_x,V_y}}] = C_{V_x,V_y} + E[Cov(\widehat{V_x}, \widehat{V_y})], \tag{9}$$

where $E[Cov(\widehat{V_x}, \widehat{V_y})]$ is the expected sampling covariance of V_x and V_y. Even though the full sampling distribution is intractable, well-known expressions for the variance of sums of squares expressed as quadratic forms can be used to obtain analytical expressions for the sampling (co)variances (Crump, 1946; Searle, 1956).

Since the estimates of the variances in non-HGLM (fixed-effect ANOVA) are related to those of ANOVA-based HGLM ($\widetilde{V}_u = \widehat{V_u} + \frac{1}{n}\widehat{V_e}$ and $\widetilde{V}_e = \widehat{V_e}$ where n is the number of observations within subgroups) the expectations for non-HGLM estimates can be derived simply once the HGLM sampling (co)variances are obtained. The expected estimates from both the non-HGLM and HGLM are shown in the Results section with the full derivations provided in the Supporting Information.

2.4 Simulated data

Data (y_{ijk}) were simulated in R (R Core Team, 2022) according to the models described in Equations 1-5. For the main set of simulations there were $c = 4$ subgroups and $n = 5$ observations per subgroup giving a total of $N = nc = 20$ observations per group. The reasoning for this data structure is to test whether the DHGLM can fill the methodological gap for data with few observations per group but many groups (e.g. RNAseq data where there are many genes/traits but few replicates) that alternative methods are unable to cope with (shown theoretically in the Results). For studies whose focus is on addressing questions regarding patterns of variation, the unit of replication is primarily

group (e.g. herds, traits, genes) rather than subgroup (e.g. individuals or lines). In our first set of simulations we set the number of groups to be large (1000) such that the amount of information in the data to estimate the covariance between variances was substantial. 1000 simulated data sets were generated using the same model parameters. In our second set of simulations we varied the number of groups from 10 to 1000 in increments of 10 (from $10-500$) or 25 (from $500-1000$) in order to assess how the performance of each method changes as a function of the number of groups. 15 simulated data sets were generated for each group-size, again using the same model parameters.

Biologically realistic parameter values were used (Table 2) and taken from an analysis by Gianola *et al.* (1992) on pedigreed lamb weight data, where additive genetic variance (subgroup variance) and environmental variance (residual variance) estimates were obtained for each herd (group). In Gianola *et al.*'s (1992) study, the two sets of variances were assumed to come from independent scaled inverse chi-squared distributions with degrees of freedom equal to 5 ($\nu_e = \nu_u = 5$) and scale parameters $s_u^2 = 0.36$ and $s_e^2 = 0.32$ for the between-subgroup (V_u) and residual (V_e) variances, respectively. The expectation and variance of the variances for component x have the form $E[V_x] = s_x^2\nu_x/(\nu_x-2)$ and $Var(V_x) = 2(s_x^2\nu_x)^2/[(\nu_x-4)(\nu_x-2)^2]$, giving $E[V_u] = 0.6$, $E[V_e] = 0.533$, $Var(V_u) = 0.72$ and $Var(V_e) = 0.569$. Variances were simulated from a multivariate log-normal distribution with these means and variances, which are here denoted μ_{V_u}, μ_{V_e}, V_{V_u} and V_{V_e}, respectively, and a correlation of 0.467 on the arithmetic scale, denoted ρ_{V_u,V_e} (not given by Gianola *et al.* 1992 and chosen based on a preliminary analysis of gene expression traits in *Saccharomyces cerevisiae* (King et al. in prep.)). The intercept μ and between-group variance V_t were not given, and so to get realistic values a simple linear mixed model was fitted to the data from Gianola *et al.* (1992), with group as a random effect, where μ (4.97) is the estimated intercept and V_t the variance among group effects (0.143).

To further assess the behaviour of the model in other regions of the parameter space, we explored two extreme situations: 1) where the correlation between V_u and V_e is zero, and 2) where the mean and variance in V_u are extremely low ($\mu_{V_u} = V_{V_u} = 0.005$).

Finally, we also explored how different experimental designs affect the precision of estimates by simulating ten data sets from each of the possible 83 designs where the number of subgroups c

and the number of observations per subgroup c range between 2 and 40 and the total number of observations $N_g nc$ is fixed at 3200.

2.5 Model Fitting

Simulated data sets were analysed using the Bayesian implementation of the DHGLM described above, as well as alternative models for comparison, including the non-HGLM (ANOVA), and both ANOVA and REML implementations of the HGLM. All analyses were performed on the same simulated data sets in order to directly compare methods.

2.5.1 DHGLM implementation in STAN and Bayesian inference

Model parameters were estimated by Bayesian inference with Markov Chain Monte Carlo (MCMC) sampling, using the programming language STAN v2.26.0 (Stan Development Team, 2020b) interfaced with R v3.4.0 (R Core Team, 2022) with the *rstan* v2.26.1 package (Stan Development Team, 2020a) (See Supporting Information for code). To increase computational efficiency, the Dispersion Model was parameterised for the standard-deviations rather than the variances, although on the log-scale moving from the standard-deviation to the variance parameterisation simply rescales the distribution by a factor of 2 ($log(V_x) = 2log(\sqrt{V_x})$) and so the prior distributions are not expected to behave fundamentally differently under a variance parameterisation.

The prior distributions used are as follows: The fixed effects μ (Mean Model), and $\mu_{\log(\sqrt{V_u})}$ and $\mu_{\log(\sqrt{V_e})}$ (Dispersion Model) were assigned normal priors with mean zero and variance 100. Random effects t_i, u_{ij}, e_{ijk} from the Mean Model, and $d_{\log(V_u(i))}$ and $d_{\log(V_e(i))}$ from the Dispersion Model, were assigned priors with mean 0 and variances V_t, $V_{g(i)}$, $V_{e(i)}$, $V_{log(V_u)}$ and $V_{log(V_e)}$, respectively. Given that the dispersion random effects are on the log-scale, the recommendation by Gardini *et al.* (2021) was followed and priors were assigned to the dispersion variances ($V_{log(\sqrt{V_u})}$ and $V_{log(\sqrt{V_e})}$) that follow a Generalized Inverse Gaussian (GIG) distribution with parameters $\lambda = 1$, $\delta = 0.01$ and $\gamma = \sqrt{3 + 9/N_g}$ (where N_g is the number of groups) according to the notation of Gardini *et al.* (2021). According to the authors, GIG priors confer better behaviour than other commonly used priors when back-transforming parameter estimates from the logarithmic to the arithmetic scale (i.e.

when making inferences about μ_{V_u} or V_{V_u}). For comparison, the same data were analysed using a half-Cauchy prior distribution with location 0 and scale 5 for the dispersion standard deviations (i.e. $\sqrt{V_{log(\sqrt{V_u})}}$). Lastly, the correlation between $d_{\log(V_u(i))}$ and $d_{\log(V_e(i))}$ ($\rho_{\log(V_u),\log(V_e)}$) was assigned a Lewandowski-Kurowicka-Joe (LKJ) prior with shape parameter 1 (Lewandowski *et al.*, 2009), which means that the prior probability density function for the correlation is uniform between -1 and 1.

To obtain results, a single MCMC chain was run for 5000 iterations, with 2500 iterations of burn-in, with starting values randomly sampled from the priors. For real data we advocate running multiple chains and checking for any convergence/mixing issues. For our single chains, we diagnosed any issues with MCMC chain convergence by recording the number of divergent transitions and calculating Geweke's statistic - with few exceptions the algorithm seems to sample from the posterior density well (Supplementary Figure S2). While the Mean Model (Equation 1) related to outcomes on the arithmetic scale, the Dispersion Model (Equations 2-5) related to outcomes on the log-scale (i.e. $\log(V_u)$ and $\log(V_e)$). However, the distribution of arithmetic-scale variances (i.e. V_u and V_e) can be obtained using well-known results for the log-normal distribution. The mean and variance of variance component V_x are given by

$$\mu_{V_x} = \exp\left[\mu_{\log(V_x)} + \frac{V_{\log(V_x)}}{2}\right] \tag{10}$$

and

$$V_{V_x} = (\exp[V_{\log(V_x)}] - 1)\exp[2\mu_{\log(V_x)} + V_{\log(V_x)}]. \tag{11}$$

The covariance of variance components V_x and V_y is

$$C_{V_x,V_y} = (\exp[C_{\log(V_x),\log(V_y)}] - 1)\exp\left[\mu_{\log(V_x)} + \mu_{\log(V_y)} + \frac{V_{\log(V_x)} + V_{\log(V_y)}}{2}\right]. \tag{12}$$

The correlation on the arithmetic scale can be obtained as $\rho_{V_x,V_y} = C_{V_x,V_y}/\sqrt{V_{V_x}V_{V_y}}$.

2.5.2 non-HGLM, ANOVA-based HGLM and REML-based HGLM

In contrast to the DHGLM, which was implemented under a Bayesian framework encompassing all groups, non-HGLM, ANOVA-based HGLM and REML-based HGLM were implemented under a frequentist approach on a group-by-group basis. For each group, a linear model with intercept and subgroup effect was fitted using the function *lm* in *R* v3.4.0 (R Core Team, 2022). For non-HGLM, the variance components were estimated as

$$\begin{aligned} \widetilde{V_u} &= \frac{MSE_u}{n} \\ \widetilde{V_e} &= MSE_e \end{aligned} \tag{13}$$

where MSE_u is the mean squared error among subgroups, n the number of observations per subgroup, and MSE_e the mean squared error of the residuals. For ANOVA-based HGLM the variances were estimated as

$$\begin{aligned} \widehat{V_u} &= \frac{MSE_u - MSE_e}{n} \\ \widehat{V_e} &= MSE_e \end{aligned} \tag{14}$$

with REML-based HGLM being the same, except that negative values of $\widehat{V_u}$ were set to 0.

2.6 Intraclass correlation and coefficient of variation

Posterior distributions for the intraclass correlations (ICC) and coefficients of variation (CV) can be obtained simply from the posterior samples of the DHGLM. The ICC for group i is defined as the ratio of the between-subgroup variance component for group i to the total variance within group i:

$$ICC(i) = \frac{V_{u(i)}}{V_{p(i)}} = \frac{V_{u(i)}}{V_{u(i)} + V_{e(i)}}. \tag{15}$$

The CV is defined as the ratio of the between-subgroup standard deviation to the group mean, which for group i is

$$CV(i) = \frac{\sqrt{V_{u(i)}}}{\bar{p}_i} \tag{16}$$

where the denominator ($\bar{p}_i$) denotes the mean value of group i. Assuming no covariates, $\bar{p}_i = \mu + t_i$ (the first two terms of Equation 1).

However, neither the distribution of ICC nor CV can be obtained analytically from the inferred distribution of the variances and group means. Consequently, from the posterior predictive distribution, (co)variances and means were sampled for $10,000$ groups and Equations 15 and 16 were applied to obtain $10,000$ ICC and CV values. From these, summary statistics (mean, median and variance) were calculated.

For the alternative methods, estimates of the variance components and mean group value (given by the model intercepts on each group) were used to calculate ICC and CV of each group, again using Equations 15 and 16. From these the mean, median and variance were calculated for each data set.

3 Results

3.1 Consistent bias in non-HGLM and HGLM (co)variance estimates

Analytical expressions for the expected estimates of the mean and (co)variance of variance components that vary over groups (on the arithmetic scale; μ_{V_u}, μ_{V_e}, V_{V_u}, V_{V_e} and C_{V_u,V_e}) are given for non-HGLM and HGLM ANOVA. In the Methods, the expected estimates of these parameters are shown to depend on their sampling (co)variances. Here, we show how these sampling (co)variances depend on the amount of within- and between-subgroup replication and how this generates bias in estimates. While the theoretical results shown in this section refer to ANOVA-based methods, we do not expect REML-based HGLM to be very different to ANOVA-based HGLM.

Estimates of mean V_e are unbiased for both methods, i.e. on average they equal the true value regardless of the number of subgroups or observations per subgroup ($E[\widehat{\mu_{V_e}}] = E[\widetilde{\mu_{V_e}}] = \mu_{V_e}$). This is also true for the mean V_u estimated by ANOVA-based HGLM ($E[\widehat{\mu_{V_u}}] = \mu_{V_u}$). However for non-HGLM, the mean V_u is overestimated by a factor of $\frac{1}{n}\mu_{V_e}$, tending to the true value as n tends to infinity, independently of the number of subgroups c (Searle, 1971). Consequently, the expected mean between-subgroup variance only approaches its true value in a non-HGLM when the number

of observations per subgroup is large and/or the mean within-subgroup variance (μ_{V_e}) is small.

Using results in Crump (1946) and Searle (1956) (with errors corrected), the expected (co)variances for HGLM are (see Supporting Information):

$$E[\widehat{V_{V_u}}] = V_{V_u} + \frac{2}{c-1}\left[\frac{N-1}{n^2(N-c)}\left(\mu_{V_e}^2 + V_{V_e}\right) + \frac{2}{n}\left(\mu_{V_u}\mu_{V_e} + C_{V_u,V_e}\right) + \left(\mu_{V_u}^2 + V_{V_u}\right)\right], \quad (17)$$

$$E[\widehat{V_{V_e}}] = V_{V_e} + \frac{2}{N-c}\left(\mu_{V_e}^2 + V_{V_e}\right) \quad (18)$$

and

$$E[\widehat{C_{V_u,V_e}}] = C_{V_u,V_e} - \frac{2}{n(N-c)}\left(\mu_{V_e}^2 + V_{V_e}\right). \quad (19)$$

For a non-HGLM, the expected (co)variances are

$$E[\widetilde{V_{V_u}}] = V_{V_u} + \frac{1}{n}V_{V_e} + \frac{2}{c-1}\left[\frac{N-1}{n^2(N-c)}\left(\mu_{V_e}^2 + V_{V_e}\right) + \frac{2}{n}\left(\mu_{V_u}\mu_{V_e} + C_{V_u,V_e}\right) + \left(\mu_{V_u}^2 + V_{V_u}\right)\right] - \frac{1}{n^2}\left[V_{V_e} + \frac{2}{N-c}\left(\mu_{V_e}^2 + V_{V_e}\right)\right], \quad (20)$$

$$E[\widetilde{V_{V_e}}] = V_{V_e} + \frac{2}{N-c}\left(\mu_{V_e}^2 + V_{V_e}\right) \quad (21)$$

and

$$E[\widetilde{C_{V_u,V_e}}] = C_{V_u,V_e} + \frac{1}{n}V_{V_e}. \quad (22)$$

Equations 17-22 show that all (co)variance estimates are to some degree biased in non-HGLM and HGLM (and generally upwardly biased, except the downwardly biased (co)variance by HGLM). In each case, the degree to which their expectation deviates from the true value (the bias) is represented by the terms following the first, and depends on the number of subgroups c and/or observations per subgroup n, in addition to the true magnitude of the mean and (co)variance of variance components. Figure 1 shows how this bias tends to decrease as the number of subgroups and observations per subgroup increase (1A-D), since the variance components are then more precisely estimated.

Estimates of the variance in V_u (1A-B) only reach their true values in HGLM when c tends to infinity (green, 1A), since V_u is estimated perfectly for every group when the number of subgroups is infinite (conditional on there being at least two observations per subgroup). At the same limit, non-

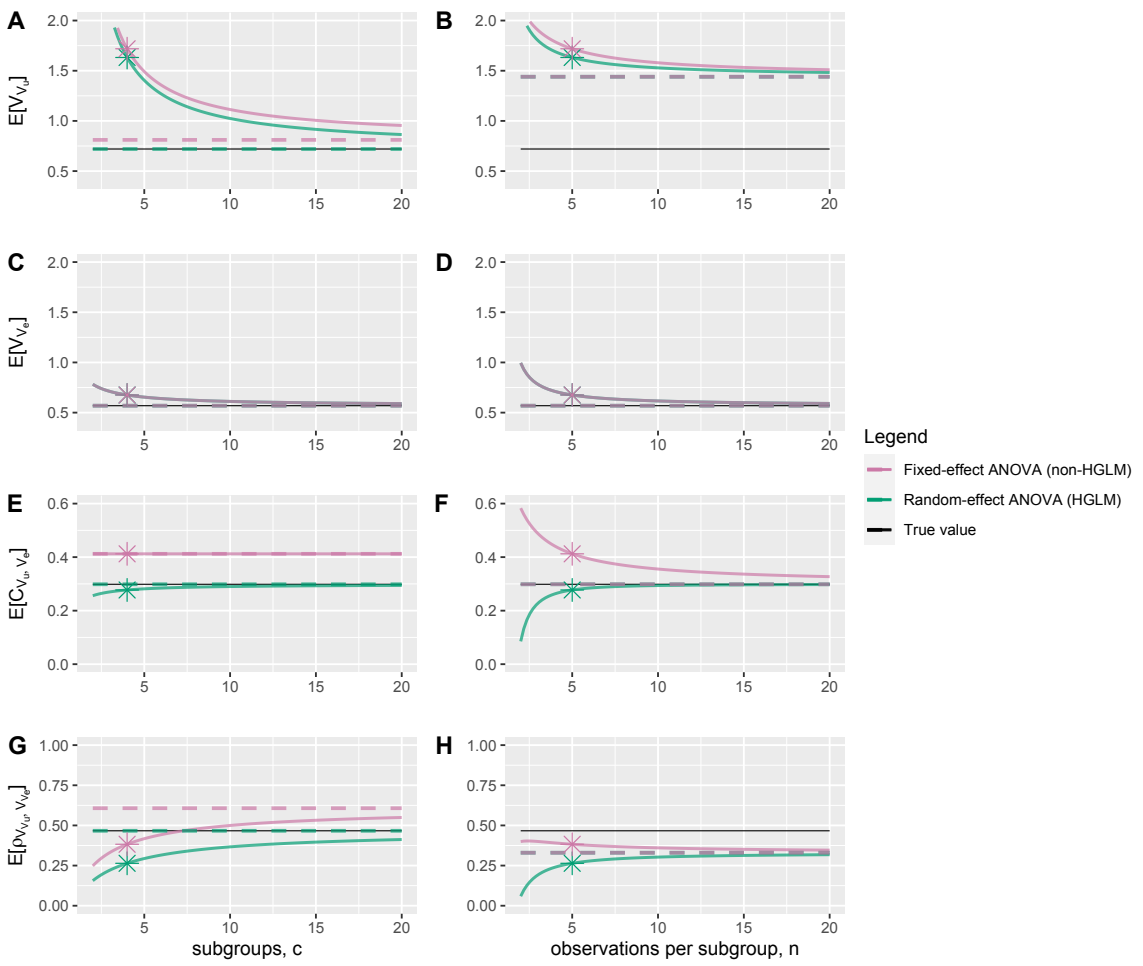


Figure 1: Theoretical expectations of the (co)variance of variance components, when estimated by HGLM (green) or non-HGLM (pink), as a function of the number of subgroups (left column) or the number of observations per subgroup (right column). Their limits, when c or n respectively, tend to infinity are represented by dashed lines. True values are represented by the black horizontal line: $\mu_{V_u} = 0.6$, $\mu_{V_e} = 0.533$,$V_{V_u} = 0.72$, $V_{V_e} = 0.569$ and $C_{V_u,V_e} = 0.299$ (Table 2). In the left column the number of observations per subgroup is $n = 5$ and in the right column the number of subgroups is $c = 4$. The star symbols indicate expectations when $n = 5$ *and* $c = 4$, as used in the simulations.

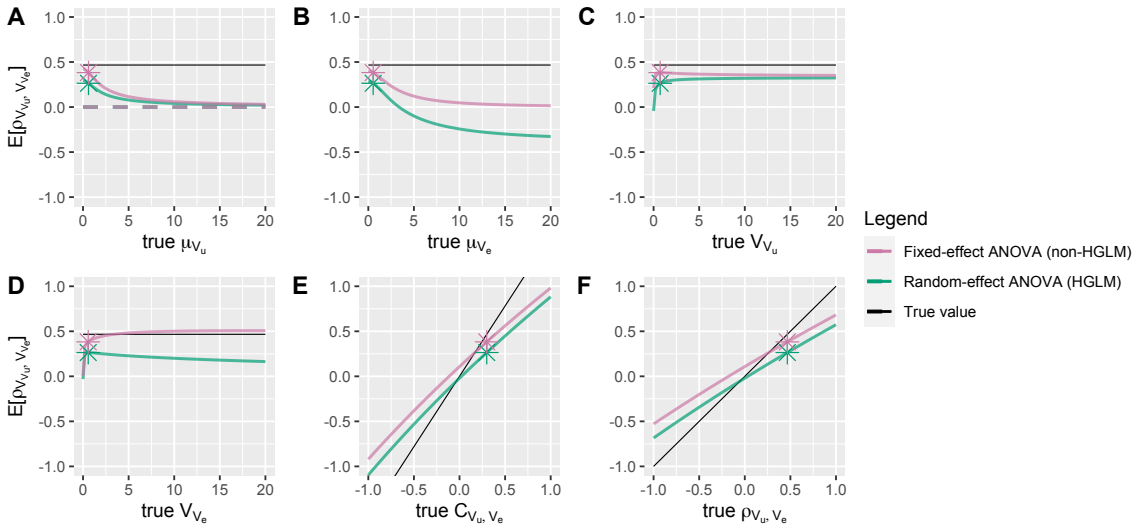


Figure 2: Expected correlation estimate among variance components as a function of model parameters. The theoretical correlation is calculated for ANOVA-based HGLM (green) and non-HGLM (pink), based on their expected (co)variances of variance components (Equation 23). In each panel (A-F) a single parameter is varying, while the remaining parameters are held constant at the values given in Table 2. The number of subgroups, c, is assumed to be 4 and the number of observations per subgroup, n, is 5. The star symbols indicate expectations for parameter values used in the simulations (see Table 2) and the black line indicates the true value of the correlation.

HGLM remains biased by a factor of $\frac{1-n}{n^2}V_{V_e}$, which may be large when the number of observations per subgroup is low and/or the variance in V_e is large. The different behaviour between HGLM and non-HGLM arises because in non-HGLM the estimate of V_u is essentially the HGLM estimate of V_u (i.e. $\widehat{V_u}$) plus the estimate of V_e divided by n (see Supporting Information for more detail). Consequently, estimation error in V_e further inflates estimates of V_{V_u} from non-HGLM and this only disappears when the number of observations per subgroup tends to infinity and the residual variance is perfectly estimated. At this limit, the bias in V_{V_u} arises solely from estimation error of V_u and is $\frac{2}{c-1}(\mu_{V_u}^2 + V_{V_u})$ for both non-HGLM and HGLM, and becomes larger when the number of subgroups, c, is low or when the mean and/or variance in V_u is large.

Estimates of the variance in V_e are identical for non-HGLM and HGLM (1C-D). In contrast to their means, the variances of V_e (V_{V_e}) are upwardly biased, with the bias depending on the data structure (decreasing with the number of subgroups and observations per subgroup) and the mean and variance of the residual variances (μ_{V_e} and V_{V_e} respectively). The bias only disappears when c

and/or n tend to infinity (conditional on there being some within-subgroup replication; $n > 1$) or the mean and variance in V_e tend to zero (Equations 18 and 21).

The covariance between V_u and V_e (1E-F) is overestimated in a non-HGLM, and tends to the true value when the V_e are estimated perfectly (i.e. as the number of observations per subgroup tends to infinity; 1F, pink), while remaining invariant to the number of subgroups and thus to the accuracy of V_u estimates (1E, pink). In contrast, the covariance is underestimated in a HGLM, and tends to the true value when either variance component is estimated perfectly (i.e. when the number of subgroups and/or observations per subgroup tend to infinity; 1E,J, green).

The expected estimate of the correlation between variance components, $E[\widehat{\rho_{V_u,V_e}}]$, cannot be obtained analytically, but can be approximated using the expected estimates of the (co)variances among variance components:

$$E[\widehat{\rho_{V_u,V_e}}] \approx \frac{E[\widehat{C_{V_u,V_e}}]}{\sqrt{E[\widehat{V_{V_u}}]E[\widehat{V_{V_e}}]}} \tag{23}$$

Simulations (see below) suggest that this approximation is accurate, and the approximation is shown for a range of values in Figure 1G-H. As the number of subgroups (c) tends to infinity, the estimate of the correlation from HGLM tends to the true value (1G, green) since the expectations of the corresponding estimates of the (co)variances reach their true values at this limit, assuming $n > 1$ (1A,C,E, green). This is not the case for non-HGLM (1G, pink), however, where it is estimated as

$$\frac{C_{V_u,V_e} + \frac{1}{n}V_{V_e}}{\sqrt{(V_{V_u} + \frac{1}{n}V_{V_e} - \frac{1}{n^2}V_{V_u})V_{V_e}}}. \tag{24}$$

When the number of observations per subgroup (n) tends to infinity, the correlation tends to

$$\frac{C_{V_u,V_e}}{\sqrt{(V_{V_u} + \frac{2}{c-1}(\mu^2_{V_u} + V_{V_u}))V_{V_e}}}, \tag{25}$$

for both non-HGLM and HGLM, which is an underestimate of the true value. As both the number of subgroups and observations per subgroup tend to infinity, the correlation tends to the true value in both non-HGLM and HGLM.

Figure 2 shows how the correlation between variance components varies as a function of the true means and (co)variances of variance components for the sampling design used in the simulations ($c = 4$ and $n = 5$). As the mean variances (μ_{V_u} and μ_{V_e}) increase, the expected correlation tends to decrease in magnitude, away from its true value, in both non-HGLM and HGLM (Figure 2A-B). In general, the magnitude of the correlation is underestimated by both non-HGLM and HGLM, although under certain parameter combinations the estimate of the correlation is expected to have the wrong sign. As the mean variances (μ_{V_u}, and μ_{V_e}) increase, the expected correlation estimate for non-HGLM tends towards zero. For HGLM, the same behaviour occurs for μ_{V_u}, but for μ_{V_e} the expected estimate actually becomes negative at large values (2A-B). In contrast, as the variances in variance components (V_{V_u} and V_{V_e}) increase, the correlation tends to get closer to the true value (2C-D) although they remain downwardly biased (except the non-HGLM method which is slightly upwardly biased at large values of V_{V_e}).

Panels 2E-F show that the magnitude of the correlation is generally underestimated by both methods, although for non-HGLM methods the estimate is expected to be more positive than for non-HGLM methods leading to estimates that are expected to have the wrong sign when the true correlation is negative and small in magnitude. Figure S1 (Supporting Information) shows that when the number of subgroups and observations per subgroup are high (100), only the magnitude of the true mean V_u has a considerable effect on the correlation estimate, decreasing from the vicinity of the true value and tending towards zero as V_u increases. The true mean V_e has a very slight effect, while the remaining parameters have almost no effect.

Overall, HGLM is a better method than non-HGLM for estimating the mean and (co)variance of variance components. When the number of subgroups is extremely large the (co)variance of variance components are estimated with little bias on average, given that both between- and within-subgroup variances are well estimated for each group. As a consequence, there is potential for HGLM to return unbiased estimates of correlation between variance components when the number of subgroups is large. However, when the number of subgroups and observations per subgroup are low, the accuracy of both methods tends to be highly dependent on the true magnitude of variance component means and (co)variances.

3.2 Simulation results and comparison of methods

3.2.1 Accuracy of DHGLM estimation

In order to compare inferences from the Bayesian DHGLM model with the frequentist point estimates from non-DHGLM models, we use the posterior median as a point estimator, following the recommendation of Pick *et al.* (2023) (in the Supporting Information we confirm that the posterior median (and mode) have better properties than the posterior mean: Figures S4 and S5). 95% highest posterior density (HPD) intervals were chosen to assess coverage of the DHGLM.

The distribution of posterior medians for the DHGLM parameters estimated from data sets comprising 1000 groups, $c = 4$ subgroups and $n = 5$ observations per subgroup, are presented in Figure 3. With this data structure, both the log scale and arithmetic scale parameters are well estimated by the DHGLM with the mean of the point estimates coinciding with the true values. The method appears to have good coverage (see Table 2), with all parameters having a coverage probability close to 95%, although the average was slightly (and significantly; binomial test p-value: 0.00641) less: 94.6 $(94.3, 94.9)$%. The table also shows how often the HPD intervals fall below or above the true value, with true parameters more likely to fall below the interval than above, particularly the mean and (co)variances of arithmetic scale variance components (i.e. μ_{V_u}, μ_{V_e}, V_{V_u} , V_{V_e}, and C_{V_u,V_e}).

3.2.2 DHGLM outperforms alternative methods

For the 1000 simulated data sets comprising 1000 groups, the distribution of estimates from non-HGLM, HGLM (ANOVA and REML-based) and DHGLM are presented in Figure 4. Most methods are shown to estimate the mean variance components reasonably well (4A,C), with little bias and relatively narrow interquartile ranges. Notably, however, μ_{V_e} (4A) is slightly downwardly biased by REML-based HGLM, due to the restriction that the subgroup variance V_u must be non-negative, and hence μ_{V_u} is slightly overestimated to compensate (4C). In addition, V_u is upwardly biased in the non-HGLM by an amount V_e/n (see Supporting Information) which leads to a strong upward bias in μ_{V_u}. The (co)variances and correlation between variance components are overall better estimated

Table 2: Simulated parameter values, taken from Gianola *et al.* (1992), on both the log and arithmetic scales, and results of the DHGLM estimates. Results include the mean and interquartile range of point estimates (posterior medians) across simulated data sets, as well as the coverage probability (the proportion of analyses where the true parameter value falls inside the 95% higher posterior density (HPD) interval) and the probabilities that the true value falls either above (underestimates) or below (overestimates) the HPD interval. If the approach has good frequentist properties, the coverage probabilities are expected to be close to 95%, and when the posterior distribution is symmetric the remaining 5% should be evenly split between under and over-estimates. For details see the **Simulated data** and **Implementation in STAN and Bayesian inference** in the main text.

Model	**Symbol**	**True value**	**Point estimate** Mean (Interquartile range)	**Coverage** (%)	**Underestimates** (%)	**Overestimates** (%)
Mean	μ	4.97	4.97 (4.96, 4.98)	95.2	2.6	2.2
	V_t	0.0205	0.0205 (0.0169, 0.0239)	93.6	3.8	2.6
Dispersion	$\mu_{\log(V_u)}$	-1.06	-1.06 (-1.1, -1.02)	95.2	2.3	2.5
	$\mu_{\log(V_e)}$	-1.18	-1.18 (-1.2, -1.15)	94.3	2.1	3.6
	$V_{\log(V_u)}$	1.1	1.09 (1.04, 1.16)	93.1	4.3	2.6
	$V_{\log(V_e)}$	1.1	1.1 (1.06, 1.14)	94.9	2.9	2.2
	$C_{\log(V_u),\log(V_e)}$	0.659	0.654 (0.618, 0.694)	94	3.5	2.5
	$\rho_{\log(V_u),\log(V_e)}$	0.6	0.592 (0.572, 0.624)	94.1	2.6	3.3
	μ_{V_u}	0.6	0.6 (0.575, 0.623)	94.1	4.2	1.7
	μ_{V_e}	0.533	0.535 (0.519, 0.55)	96.1	1.7	2.2
	V_{V_u}	0.72	0.732 (0.611, 0.831)	94.2	4.9	0.9
	V_{V_e}	0.569	0.578 (0.514, 0.629)	95.4	2.8	1.8
	C_{V_u,V_e}	0.299	0.299 (0.263, 0.331)	93.2	4.9	1.9
	ρ_{V_u,V_e}	0.467	0.46 (0.438, 0.492)	93.7	3.5	2.8

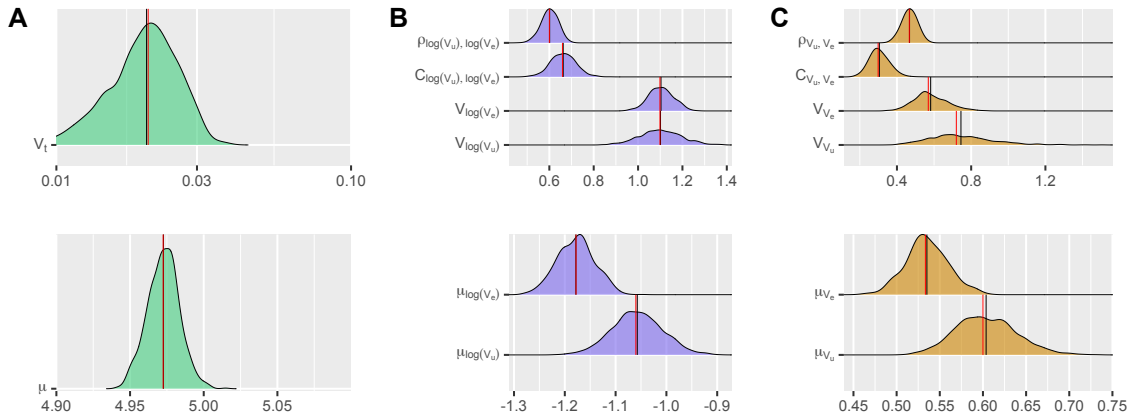


Figure 3: Model estimates from simulated data sets. The distributions of point estimates (posterior medians) are shown for 1000 data sets comprising 1000 groups, 4 subgroups and 5 observations per subgroup. Curves are the density distributions of posterior medians over data sets, black vertical lines are their means, and red vertical lines their corresponding true values (Table 2). Model parameters, as well as their arithmetic scale transformations, are divided into panels. **A)** Parameters from the Mean Model pertaining to the mean (μ) and variance between groups (V_t). **B)** mean, (co)variances and correlation among log-scale variance components, and **C)** the corresponding quantities on the arithmetic scale.

by the DHGLM, which shows little 'bias' and has narrower interquartile ranges (Figure 4B,D-F). With respect to the remaining methods, estimates of the variances in variance components (Figure 4B,D) are upwardly biased, particularly for the variance among subgroup variances (V_{V_u}). The estimates of the covariance (Figure 4E) are upwardly biased in the non-HGLM and slightly downwardly biased by the ANOVA-based HGLM. The estimates of the correlation (Figure 4F) have considerable downward bias for all non-DHGLM methods. Surprisingly, the non-HGLM appears to perform better at estimating the correlation than HGLMs, even though it does overall worse at estimating other parameters. However, according to the theoretical results (Figure 2F), this is expected when the true correlation between variance components is positive and is not a general result (when negative, the opposite would be observed). The accuracy of non-HGLM and HGLMs drops remarkably for the mean and variance of standardised variance components (CV and ICC; 4G-L), whereas the DHGLM performs very well. With respect to the residual CV (CV_e), both the estimates of the mean and variance are upwardly biased by all methods, particularly in non-HGLM and HGLM (the bias generated by DHGLM is very little in comparison). The estimate of the mean between-subgroup CV (CV_u) is biased upward in non-HGLM, downwardly in HGLM

and unbiased in DHGLM, while its variance is only slightly downwardly biased in DHGLM and extremely upwardly biased in the remaining methods. In addition, the distribution of estimates among simulated data sets for the mean and variance in CV have long tails, probably due to the mean group values not being constrained to be positive and generating extreme estimates of the CV as they approach zero. The mean intraclass correlation (ICC_g) is downwardly biased in all methods apart from the DHGLM, and the variance in intraclass correlations upwardly biased, particularly by the HGLM.

Figure 5 shows how the accuracy of the different methods in estimating variance components (on the arithmetic scale) changes as a function of the number of groups. When the number of groups is low, point estimates obtained by the DHGLM are sometimes biased and presumably driven by prior information. The influence of the prior is particularly strong on the variance in variance components (5B,D) and their correlation (5F), and to a lesser degree on the mean variance components (5A,C) and the variance in intraclass correlation (5L). When using a non-GIG prior, such as a half-Cauchy, for the variance in variance components, the bias when the number of groups is low is even more significant (Figure S10, S11). However, despite being biased at low sample sizes (with respect to the number of groups) the posterior median from the DHGLM is found to be a consistent estimator, with estimates converging to the true value at high sample sizes. This is also the case in more extreme scenarios such as the true correlation between variance components being zero (Figure S6, S7) or when the mean and variance in random-effect variances V_u are extremely low (on the arithmetic scale, Figure S9, although not on the log-scale where prior sensitivity is considerable, Figure S8). By contrast, the remaining methods do not show a decrease in bias, which is also predicted by the theoretical results obtained for ANOVA-based methods given that Equations 17-22 and Figures 1-2 are independent of the number of groups. Non-HGLM and HGLM are therefore biased and inconsistent estimators of variance components. These results also hold on the log-scale, except that the non-HGLM appears to be consistent for the mean log-scale subgroup variance component and the covariance among variance components, although this is probably due to the particular choice of parameter values (Figure S3).

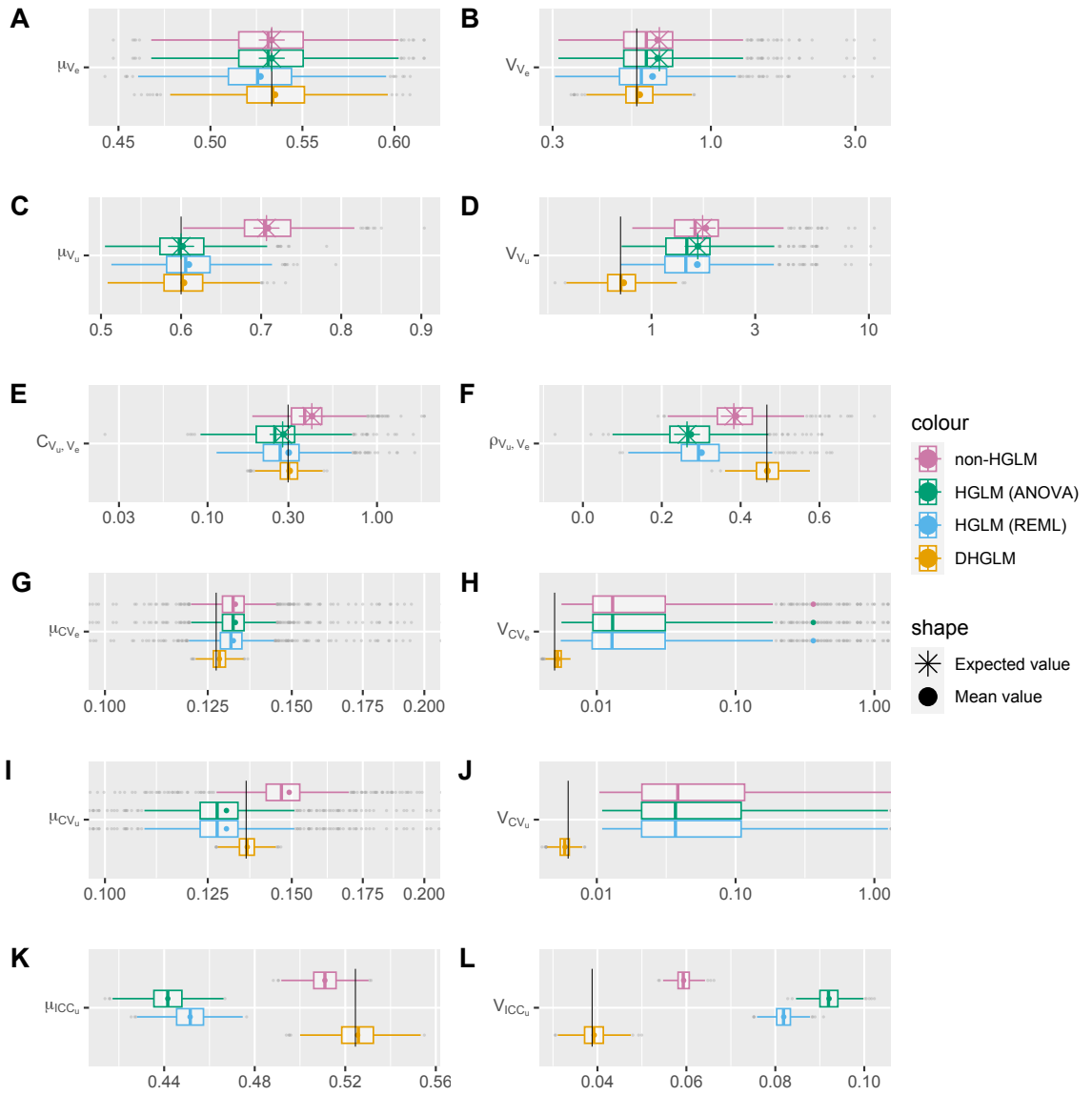


Figure 4: Performance of different methods in estimating heterogeneity and correlation, among variance components V_u (between-subgroup variances) and V_e (residual variances), as well as the between-subgroup intraclass correlations (ICC) and coefficients of variation (CV), on the arithmetic scale. A,C) Mean variance components. B,D) Variance in variance components. E) Covariance among variance components. F) Correlation between variance components. G,I) Mean coefficients of variation. H,J) Variance in coefficients of variation. K) Mean intraclass correlation. L) Variance in intraclass correlation. The true value of each parameter is represented by black vertical lines overlaying the corresponding boxplots. The results obtained from each method are shown with a unique colour (see legend on the right). Boxplots represent the sampling distribution obtained from 1000 simulated data sets, showing the quartiles, where box edges are the lower and upper quartiles (25 and 75% quantiles, respectively). The median (the 50% quantile) of each distribution is shown as a vertical bar within the boxplot, the mean value is depicted by a filled circle, and theoretical values obtained for ANOVA-based methods given by the star symbols, with direct correspondence to those in Figure 1.

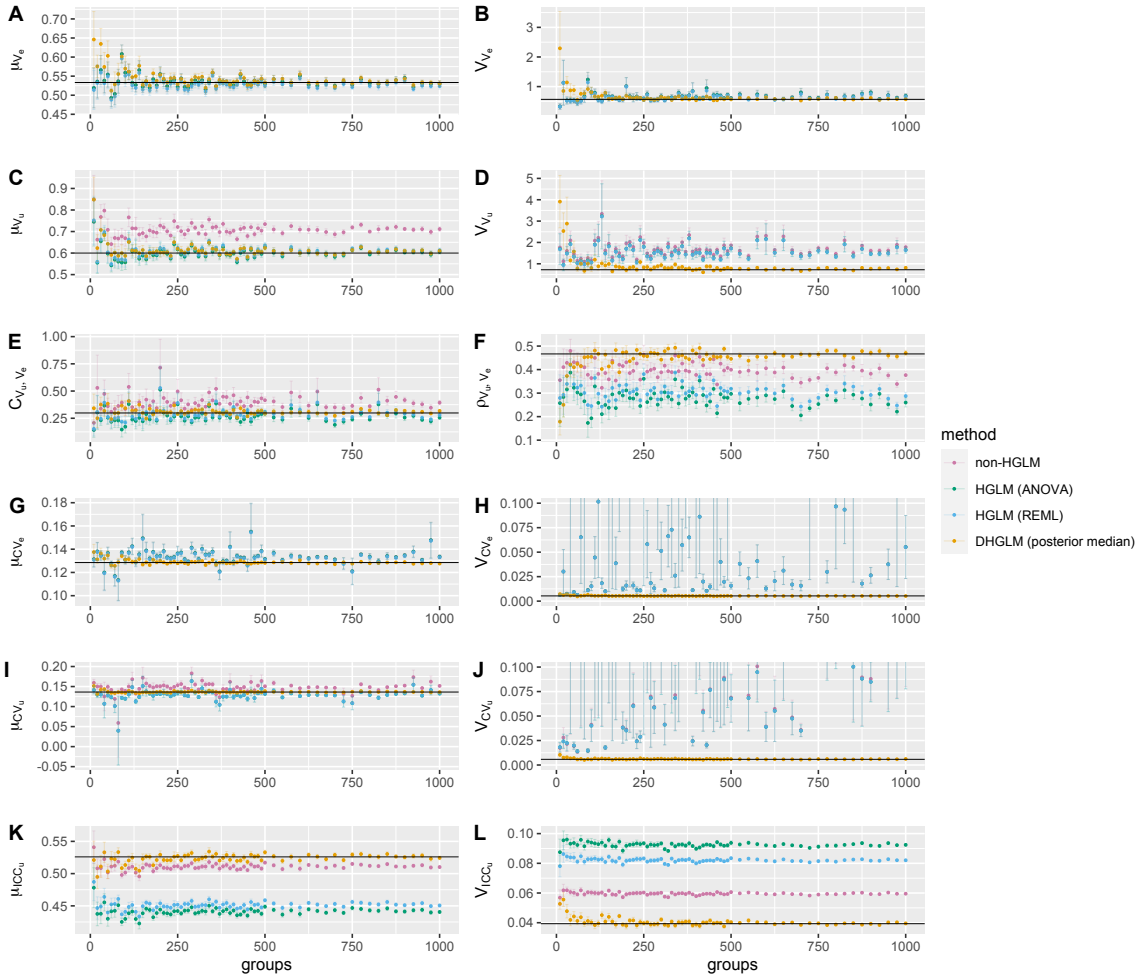


Figure 5: Accuracy of different methods in estimating the mean and variance of variance components and their standardisations (intraclass correlation and coefficients of variation), on the arithmetic scale, as a function of the number of groups. The number of subgroups and observations per subgroup are fixed at $c = 4$ and $n = 5$, respectively. A,C) Mean variance components. B,D) Variance in variance components. E) Covariance among variance components. F) Correlation between variance components. G,I) Mean coefficients of variation. H,J) Variance in coefficients of variation. K) Mean intraclass correlation. L) Variance in intraclass correlation. Each data point is the mean estimate obtained among 15 simulated data sets, with interquartile ranges shown by bars.

3.2.3 Optimal designs for DHGLM

The posterior standard deviation of the log-scale correlation between variance components ($\rho_{log(V_u),log(V_e)}$) was used as a metric of precision, with low values indicating greater precision. The optimal design has a modest number of observations within each group ($n = c = 5$) but the number of groups is large ($N_g = 128$). Although many designs have comparable precision, ensuring the number of groups is at least as large as the number of observations per group seems warranted. When deciding how observations are partitioned within a group it seems best to keep the number of subgroups c and the number of subgroups n roughly comparable, or to slightly favour n over c (Figure **??**). Although prioritising the number of groups is likely to be a general recommendation, the optimal design will vary as a function of the true model parameters and so the results here should be treated with some caution.

4 Discussion

Various studies in ecology and evolution have indicated that random-effect and residual variance components may vary over groups (e.g. Brotherstone & Hill, 1986, Westneat *et al.*, 2013). Methods have been developed to deal with these heterogeneous random-effect and residual variances, by directly estimating the parameters of the distribution of variance components (Gianola *et al.*, 1992; San Cristobal *et al.*, 1993; Foulley & Quaas, 1995; Lee & Nelder, 2006; Smyth, 2004), and these methods have been advocated (Cleasby *et al.*, 2015) and used (Westneat *et al.*, 2013) in ecological and evolutionary studies. Some studies have also suggested that random-effect and residual variances may also be correlated. For example, in quantitative genetics, the degree to which genetic, mutational and environmental variances vary and covary over traits has been explored (Price & Schluter, 1991; Houle, 1992, 1998; Hansen *et al.*, 2011). To our knowledge, there is currently no method for directly estimating the correlation between variance components. Instead, studies have estimated the correlation between *estimates* of variance components, which are typically downwardly biased by the sampling variance of estimates. Therefore, we suggest an alternative method where the correlation is modelled directly in a DHGLM framework.

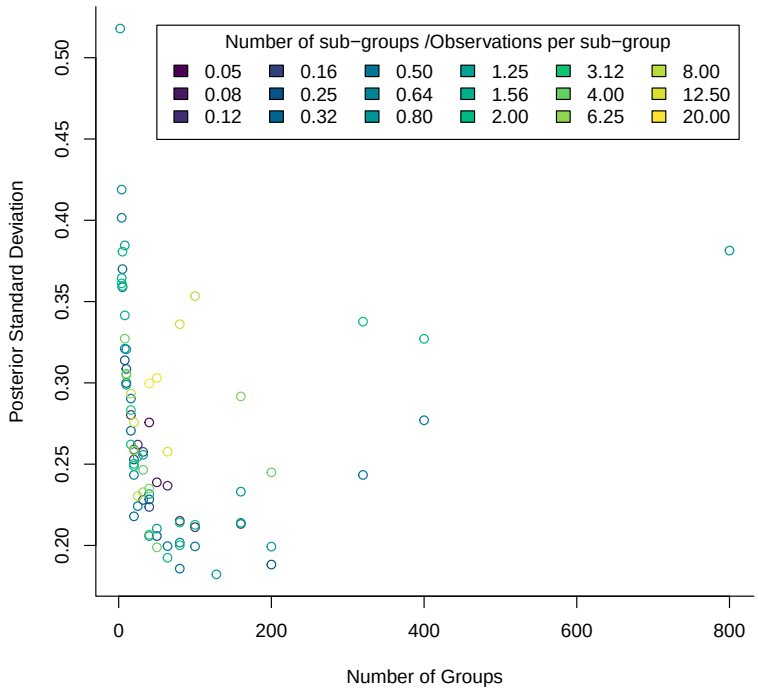


Figure 6: Precision of estimates of $r_{log(V_u),log(V_e)}$ (log-scale correlation between variances) from 3200 observations but with varying combinations of N_g (number of groups), c (number of subgroups) and n (number of observations per subgroup). Precision is measured as the average (over the 10 data sets simulated for each design) posterior standard deviation with high values indicating poor precision.

The issue of estimating variance components for each group, and then using the (co)variances of the *estimates* as an estimator of the true (co)variances are identified. First, single-hierarchical (HGLM) and non-hierarchical (non-HGLM) methods are shown, both theoretically and empirically, to carry systematic bias (Figures 1-5) whereby estimates of the variances of variance components are upwardly biased and the covariance between variance components are either upwardly (HGLM) or downwardly (non-HGLM) biased (Eq 17-22). Second, a somewhat surprising result from the theory suggests that, even though HGLM generally outperform non-HGLM (Figure 4A-E), whether one model or the other is a better estimator of the *correlation* among variance components depends on the true value (Figure 2F), partly arising due to the opposite directions of bias of non-HGLM and HGLM for the covariance, and consequently the correlation.

In an extension to the DHGLM proposed by San Cristobal *et al.* (1993), we allow the variance components to follow a *multivariate* log-normal distribution which gives unbiased and precise estimates of the mean and (co)variances of the variance components when the number of groups is sufficiently large. This is achieved even when the number of subgroups and/or observations per subgroup are small (Figure 4 and 5) because, rather than estimating variance components with large sampling variance (due to the small sample sizes), which is carried into the mean and variance of variance components, the distribution of variance components is estimated directly.

The typical study from the quantitative genetics and behavioural ecology literature obtains variance component estimates from HGLM (e.g. de Villemereuil *et al.*, 2013; Stoffel *et al.*, 2017). Often, these studies have data on a large number of subgroups, such as genotypes or individuals, and so the sampling variances might be expected to be small and the results accurate (the left column of Figure 1). However, such studies often fit other, partly confounded, random effects which may result in a much lower effective number of subgroups. For example, data may be collected on a large number of families in order to estimate the genetic variance, but genetic effects are often partly confounded with maternal or common-environment effects such that fewer observations are useful for estimating the genetic variance (Kruuk & Hadfield, 2007). Moreover, such studies often work with a small number of groups (e.g. 8 traits in Houle, 1998) and so even in cases where estimates of the correlation in variance components have little bias, the lack of replication at the appropriate

level tends to make the estimates very imprecise.

For those studies that explicitly seek to test whether variances (co)vary over groups, the number of subgroups or observations per subgroup is often modest. For example, Westneat *et al.* (2013) considered heterogeneous residual variances in food provisioning among 27 female red-winged blackbirds (*Agelaius phoeniceus*) with an average of only 20 observations per group (female). Similarly, Landry *et al.* (2007) and Denver *et al.* (2005) estimate genetic and environmental variances in gene expression traits for thousands of genes (groups) in *Saccharomyces cerevisiae* and *Caenorhabditis elegans* respectively, yet only have a maximum of 6 subgroups (lines) and 9 observations per subgroup. In these cases, we expect substantial sampling error in their estimates, as obtained by the non-HGLM (Landry *et al.*, 2007) and HGLM (Denver *et al.*, 2005) methodology used, and as consequence substantial bias when estimating how variance components are likely to vary and covary. In contrast, the DHGLM performs well in designs which prioritise group replication over replication at lower levels, and these designs are better suited to getting precise estimates of how variance components vary over groups. Indeed, empirical Bayes DHGLM procedures have been developed in the context of gene expression microarray/RNAseq analyses, albeit with the aim of increasing the power to detect differential expression at specific genes, rather than characterising patterns of (co)variation (Smyth, 2004). However, to our knowledge, these methods have only modelled the distribution of the residual variances across gene expression traits and do not accommodate other sources of dispersion variation or covariation.

The DHGLM gains much of its estimation power from the number of groups and a dramatic drop in DHGLM performance was observable for certain parameters when the number of groups was low, presumably due to prior sensitivity. Careful selection of appropriate priors for the variance in variance components is therefore important when replication is low. A commonly used prior for covariance matrices is the inverse-Wishart distribution. However, as summarised by Alvarez *et al.* (2014) these have been shown to have several problems, including *a priori* dependencies between variances and correlations (Tokuda *et al.*, 2011), marginal distributions for the variances (inverse-Gamma) with high density around zero (Gelman, 2006), and a single degree of freedom controlling all variances and correlations (Gelman *et al.*, 2013). A separation strategy was proposed by Barnard

et al. (2000) that decomposes the covariance matrix so that priors can be placed on variances (or standard deviations) and correlations separately. Using the separation strategy, Huang & Wand (2013) suggested that the half-t family of priors (including half-Cauchy) are used as a prior for standard deviations, following Gelman's (2006) recommendation. While these recommendations are likely to perform well when considering the (co)variances of the log-scale variances, results in Gardini *et al.* (2021) suggest that problems may occur if inferences are to be drawn about the distribution of arithmetic-scale variances. In particular, the posterior moments may be undefined for some parameters (e.g. the mean variance) when using the half-t family of prior distributions due to their very long right-tails on the arithmetic scale. As an alternative, Gardini *et al.* (2021) suggested the use of a Generalised Inverse Gaussian (GIG) prior distribution (Fabrizi & Trivisano, 2012, 2016) which is a flexible family of distributions that performs well when variances are small (as the half-t family) while placing conditions that guarantee posterior moments for some aspects of the distribution of arithmetic-scale variances. Indeed, we found the GIG prior to outperform other priors in our simulations. However, when the mean and variance in random-effect variances was very low we showed that log-scale estimates behaved poorly (Figure S9) despite very good estimation on the arithmetic-scale (Figure S8). This probably arises in this extreme case because large shifts on the log-scale equate to extremely small effects on the arithmetic-scale such that there is little information in the data to distinguish large log-scale shifts. Whether priors can be found that work well on both scales for cases such as these remains an open question. With respect to priors on correlations matrices, the general approach is to have priors which either result in uniform (-1 to 1) marginal priors for each correlation (Barnard *et al.*, 2000), or more recently, are uniform on the space of the complete correlation matrix (LKJ prior; Lewandowski *et al.* (2009); Stan Development Team (2022)). These two priors are equivalent in the 2-dimensional case presented here, but with the exception of very high dimensional problems we recommend the LKJ prior.

In our simple model, the residual variance is assumed constant across subgroups within a group, although this assumption could be relaxed by allowing subgroup random effects in the Dispersion Model. Indeed, if not dealt with, any heterogeneity in the residual variance between subgroups is likely to upwardly bias any estimates of the between-group variation in the residual/random-

effect variance. Similarly, heterogeneity in the residual variance at the level of the observation may also be present and would manifest itself as excess kurtosis in the residuals. This could be accommodated by including observation-level random effects in the Dispersion model, although the addition of a large number of weakly identified parameters may present computational difficulties. Switching from a log-normal to an inverse-gamma distribution for the observation-level random effects would solve this issue as the random effects can then be analytically marginalised by assigning the residuals a scaled-t rather than a normal distribution, with the estimated degree-of-freedom parameter controlling the amount of kurtosis. In addition, our basic model assumed that group means and variances are independent of each other, although parameters of the Mean Model could be included as predictors in the Dispersion Model in order to model any mean-variance coupling. For example, variances could be modelled as log-linear (or power law) functions of the mean by including group or subgroup means (or their logarithm) as predictors in the Dispersion Model. Such an extension would be necessary for those that believe that heterogeneity in variance is only interesting if it cannot be explained by scaling relationships, although this majority belief has been questioned (Wagner, 2023). In the Supporting Information we discuss these extensions in more detail, and provide code for model fitting.

Our model also assumes that the subgroup and residual variances are themselves identically and independently distributed over groups, an assumption that may not be met. Many strategies exist for modelling dependency between random effects in standard HGLM, and these could in theory be applied to log-scale variances. Indeed, ARCH (Engle, 1982) and GARCH (Bollerslev, 1986) models use standard autoregressive models to model changes in variance over time. Dealing with more general and arbitrary patterns of dependence between variances would be more challenging, although latent variable or factor analytic approaches may prove feasible (Warton *et al.*, 2015; Runcie & Mukherjee, 2013). However, in both cases it would seem preferential to allow some relationship between the dependency structures for the means and the variances. For example, we might expect the expression levels of two co-regulated genes (groups h and i) to covary over genotypes (subgroups, j) due to polymorphism in the binding affinity of their shared transcription factor (i.e. $COV(u_{hj}, u_{ij}) \neq 0$), and for the same reason we might also expect the genetic variances

for the two gene-expression traits (groups) to be more similar than the genetic variances of two randomly picked genes (i.e. $COV(V_{u(h)}, V_{u(i)}) > 0$). Whether suitable low-parameter modelling strategies that allow dependency at the mean-level to mirror dependency at the variance-level can be found remains an open question, and any solutions may well prove to be computational prohibitive for many problems.

Despite these issues, we recommend the use of the proposed model for studies interested in how two or more variance components covary over groups, especially in cases where replication within groups is limited but there are many groups. Present methods usually rely on precise variance component estimation achieved by large sample sizes within groups. However this might not be feasible, and might even be ill-advised if replication within groups comes at the cost of sampling fewer groups (the level of replication that is most important when assessing differences among groups in their variances). Specific examples of cases where it could be used include studies interested in the correlation between genetic and environmental variances across multiple classes of traits or multiple environmental conditions (Hansen *et al.*, 2011). Other examples apply to behavioural ecology, where interest might be in how within-individual variation in behaviour can be partitioned into permanent environmental and genetic effects (Martin *et al.*, 2017; Prentice *et al.*, 2020). Given that the quantification of such correlations among components of variation can be important starting points for understanding the mechanistic causes of variation (Geiler-Samerotte *et al.*, 2020), we hope that this model can facilitate future research in a wide range of fields.

Acknowledgements: We thank Enrico Fabrizi with help implementing a GIG prior in STAN. We also thank Lars Rönengård, Luis-Miguel Chevin and Nick Colegrave for discussions about this work. This work was funded by a Principal's Career Development PhD Studentship to Jessica King and a Royal Society Research Fellowship (UF150696) to Jarrod Hadfield.

Author Contributions: Code was written by Jessica King (JK) with help from Jarrod Hadfield (JH) & Joel Pick (JP) and the simulations ran by JK. Theoretical results were obtained by JH. JK, JH and JP wrote the manuscript.

Data availability statement: Data sharing is not applicable to this article as no new data were created or analysed in this study.

Conflicts of interest: No conflicts of interest to declare.

5 References

Alvarez, I., Niemi, J. & Simpson, M. (2014) Bayesian inference for a covariance matrix. W. Song, ed., *Proceedings of the 26th Annual Conference on Applied Statistics in Agriculture*, pp. 71–82. New Prairie Press.

Barnard, J., McCulloch, R. & Meng, X.L. (2000) Modeling covariance matrices in terms of standard deviations and correlations, with application to shrinkage. *Statistica Sinica*, **10**, 1281–1311.

Bolker, B., Brooks, M., Clark, C., Geange, S., Poulsen, J., Stevens, M. & White, J. (2009) Generalized linear mixed models: a practical guide for ecology and evolution. *Trends in ecology & evolution*, **24**, 127–135. https://dx.doi.org/10.1016/J.TREE.2008.10.008.

Bollerslev, T. (1986) Generalized autoregressive conditional heteroskedasticity. *Journal of econometrics*, **31**, 307–327.

Brotherstone, S. & Hill, W.G. (1986) Heterogeneity of variance amongst herds for milk production. *Animal Production*, **42**, 297–303. https://dx.doi.org/10.1017/S0003356100018067.

Chinchilla-Ramírez, M., Pérez-Hedo, M., Pannebakker, B.A. & Urbaneja, A. (2020) Genetic variation in the feeding behavior of isofemale lines of *Nesidiocoris tenuis*. *Insects*, **11**, 1–13. https://dx.doi.org/10.3390/insects11080513.

Cleasby, I.R., Nakagawa, S. & Schielzeth, H. (2015) Quantifying the predictability of behaviour: statistical approaches for the study of between-individual variation in the within-individual variance. *Methods in Ecology and Evolution*, **6**, 27–37.

Crump, S.L. (1946) The estimation of variance components in analysis of variance. *Biometrics Bulletin*, **2**, 7–11.

de Villemereuil, P., Gimenez, O. & Doligez, B. (2013) Comparing parent-offspring regression with frequentist and Bayesian animal models to estimate heritability in wild populations: A simulation study for Gaussian and binary traits. *Methods in Ecology and Evolution*, **4**, 260–275. https://dx.doi.org/10.1111/2041-210X.12011.

Denver, D.R., Morris, K., Streelman, J.T., Kim, S.K., Lynch, M. & Thomas, W.K. (2005) The transcriptional consequences of mutation and natural selection in *Caenorhabditis elegans*. *Nature Genetics*, **37**, 544–548. https://dx.doi.org/10.1038/ng1554.

Engle, R.F. (1982) Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica: Journal of the econometric society*, pp. 987–1007.

Fabrizi, E. & Trivisano, C. (2012) Bayesian estimation of log-normal means with finite quadratic expected loss. *Bayesian Analysis*, **7**, 975–996. https://dx.doi.org/10.1214/12-BA733.

Fabrizi, E. & Trivisano, C. (2016) Bayesian conditional mean estimation in log-normal linear regression models with finite quadratic expected loss. *Scandinavian Journal of Statistics*, **43**, 1064–1077. https://dx.doi.org/https://doi.org/10.1111/sjos.12229.

Foulley, J.L. & Quaas, R.L. (1995) Heterogeneous variances in Gaussian linear mixed models. *Genet Sel Evol*, **27**, 211–228.

Foulley, J.L., San Cristobal, M., Gianola, D. & Im, S. (1992) Marginal likelihood and Bayesian approaches to the analysis of heterogeneous residual variances in mixed linear Gaussian models. *Computational Statistics & Data Analysis*, **13**, 291–305. https://dx.doi.org/https://doi.org/10.1016/0167-9473(92)90137-5.

Gardini, A., Trivisano, C. & Fabrizi, E. (2021) Bayesian analysis of ANOVA and mixed models on the log-transformed response variable. *Psychometrika*, **86**, 619–641. https://dx.doi.org/10.1007/s11336-021-09769-y.

Geiler-Samerotte, K.A., Li, S., Lazaris, C., Taylor, A., Ziv, N., Ramjeawan, C., Paaby, A.B. &

Siegal, M.L. (2020) Extent and context dependence of pleiotropy revealed by high-throughput single-cell phenotyping. *PLoS biology*, **18**, e3000836.

Gelman, A. (2006) Prior distributions for variance parameters in hierarchical models. *Bayesian Analysis*, **1**, 515–533.

Gelman, A., Carlin, J.B., Stern, H.S., Dunson, D.B., Vehtari, A. & Rubin, D.B. (2013) *Bayesian Data Analysis*. Chapman and Hall/CRC, 3rd edition.

Gianola, D., Foulley, J.L., Fernando, R.L., Henderson, C.R. & Weigel, K.A. (1992) Estimation of heterogeneous variances using Empirical Bayes methods: Theoretical considerations. *Journal of Dairy Science*, **75**, 2805–2823. https://dx.doi.org/10.3168/jds.S0022-0302(92)78044-8.

Hansen, T.F., Pélabon, C. & Houle, D. (2011) Heritability is not evolvability. *Evolutionary Biology*, **38**, 258–277. https://dx.doi.org/10.1007/s11692-011-9127-6.

Harvey, A., Ruiz, E. & Shephard, N. (1994) Multivariate stochastic variance models. *The Review of Economic Studies*, **61**, 247–264. https://dx.doi.org/10.2307/2297980.

Henderson, C.R., Kempthorne, O., Searle, S.R. & Von Krosigk, C.M. (1959) The estimation of environmental and genetic trends from records subject to culling. *Biometrics*, **15**, 192–218.

Hill, W.G. (1984) On selection among groups with heterogeneous variance. *Animal Production*, **39**, 473–477. https://dx.doi.org/10.1017/S0003356100032220.

Hoffmann, A.A., Merilä, J. & Kristensen, T.N. (2016) Heritability and evolvability of fitness and nonfitness traits: Lessons from livestock. *Evolution*, **70**, 1770–1779.

Houle, D. (1992) Comparing evolvability and variability of quantitative traits. *Genetics*, **130**, 195–204.

Houle, D. (1998) How should we explain variation in the genetic variance of traits? *Genetica*, **102/103**, 241–253. https://dx.doi.org/10.1023/A:1017034925212.

719 Huang, A. & Wand, M.P. (2013) Simple marginally noninformative prior distributions for covariance
720 matrices. *Bayesian Analysis*, **8**, 439–452. https://dx.doi.org/10.1214/13-BA815.

721 Huang, W., Lyman, R.F., Lyman, R.A., Carbone, M.A., Harbison, S.T., Magwire, M.M. & Mackay,
722 T.F.C. (2016) Spontaneous mutations and the origin and maintenance of quantitative genetic
723 variation. *eLife*, **5**, e14625. https://dx.doi.org/10.7554/eLife.14625.

724 Kruuk, L.E.B. & Hadfield, J.D. (2007) How to separate genetic and environmental causes of simi-
725 larity between relatives? *Journal of Evolutionary Biology*, **20**, 1890–1903.

726 Lafuente, E., Duneau, D. & Beldade, P. (2018) Genetic basis of thermal plastic-
727 ity variation in *Drosophila melanogaster* body size. *PLoS Genetics*, **14**, 1–24.
728 https://dx.doi.org/10.1371/journal.pgen.1007686.

729 Landry, C.R., Lemos, B., Rifkin, S.A., Dickinson, W.J. & Hartl, D.L. (2007) Genetic properties
730 influencing the evolvability of gene expression. *Science*, **317**, 118–121.

731 Lee, Y. & Nelder, J.A. (2006) Double hierarchical generalized linear models (with discus-
732 sion). *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, **55**, 139–185.
733 https://dx.doi.org/10.1111/J.1467-9876.2006.00538.X.

734 Lewandowski, D., Kurowicka, D. & Joe, H. (2009) Generating random correlation matrices based
735 on vines and extended onion method. *Journal of Multivariate Analysis*, **100**, 1989–2001.
736 https://dx.doi.org/10.1016/j.jmva.2009.04.008.

737 Martin, J.G., Pirotta, E., Petelle, M.B. & Blumstein, D.T. (2017) Genetic basis of between-
738 individual and within-individual variance of docility. *Journal of Evolutionary Biology*, **30**, 796–
739 805.

740 Morrissey, M.B. (2016) Meta-analysis of magnitudes, differences and variation in evolutionary pa-
741 rameters. *Journal of Evolutionary Biology*, **29**, 1882–1904. https://dx.doi.org/10.1111/jeb.12950.

Morrissey, M.B. & Hadfield, J.D. (2012) Directional selection in temporally replicated studies is remarkably consistent. *Evolution*, **66**, 435–442. https://dx.doi.org/10.1111/j.1558-5646.2011.01444.x.

Nakagawa, S. & Schielzeth, H. (2010) Repeatability for Gaussian and non-Gaussian data: A practical guide for biologists. *Biological Reviews*, **85**, 935–956. https://dx.doi.org/10.1111/j.1469-185X.2010.00141.x.

Nicolaus, M., Brommer, J.E., Ubels, R., Tinbergen, J.M. & Dingemanse, N.J. (2013) Exploring patterns of variation in clutch size-density reaction norms in a wild passerine bird. *Journal of Evolutionary Biology*, **26**, 2031–2043. https://dx.doi.org/10.1111/jeb.12210.

Pick, J.L., Kasper, C., Allegue, H., Dingemanse, N.J., Dochtermann, N.A., Laskowski, K.L., Lima, M.R., Schielzeth, H., Westneat, D.F., Wright, J. & Araya-Ajoy, Y.G. (2023) Describing posterior distributions of variance components: Problems and the use of null distributions to aid interpretation. *Methods in Ecology and Evolution*, **14**, 2557–2574. https://dx.doi.org/10.1111/2041-210X.14200.

Prentice, P.M., Houslay, T.M., Martin, J.G. & Wilson, A.J. (2020) Genetic variance for behavioural 'predictability' of stress response. *Journal of Evolutionary Biology*, **33**, 642–652.

Price, T. & Schluter, D. (1991) On the low heritability of life-history traits. *Evol*, **45**, 853–861.

R Core Team (2022) R: A Language and Environment for Statistical Computing.

Royauté, R. & Dochtermann, N.A. (2021) Comparing ecological and evolutionary variability within datasets. *Behavioral Ecology and Sociobiology*, **75**, 127.

Runcie, D.E. & Mukherjee, S. (2013) Dissecting high-dimensional phenotypes with Bayesian sparse factor analysis of genetic covariance matrices. *Genetics*, **194**, 753–767. https://dx.doi.org/10.1534/genetics.113.151217.

San Cristobal, M., Foulley, J. & Manfredi, E. (1993) Inference about multiplicative heteroskedastic

components of variance in a mixed linear Gaussian model with an application to beef cattle breeding. *Genetics Selection Evolution*, **25**, 3–30. https://dx.doi.org/10.1186/1297-9686-25-1-3.

Schwagmeyer, P.L. & Mock, D.W. (2003) How consistently are good parents good parents? Repeatability of parental care in the House Sparrow, *Passer domesticus*. *Ethology*, **109**, 303–313. https://dx.doi.org/10.1046/j.1439-0310.2003.00868.x.

Searle, S.R. (1956) Matrix methods in components of variance and covariance analysis. *The Annals of Mathematical Statistics*, **27**, 737–748.

Searle, S.R. (1971) *Linear models*. Wiley, New York.

Smyth, G.K. (2004) Linear models and empirical Bayes methods for assessing differential expression in microarray experiments. *Statistical applications in genetics and molecular biology*, **3**.

Stamps, J.A., Briffa, M. & Biro, P.A. (2012) Unpredictable animals: Individual differences in intraindividual variability (IIV). *Animal Behaviour*, **83**, 1325–1334. https://dx.doi.org/10.1016/j.anbehav.2012.02.017.

Stan Development Team (2020a) RStan: the R interface to Stan.

Stan Development Team (2020b) Stan Modeling Language Users Guide and Reference Manual, 2.21.0.

Stan Development Team (2022) Stan Modeling Language Users Guide and Reference Manual, 2.26.1.

Stoffel, M.A., Nakagawa, S. & Schielzeth, H. (2017) rptR: repeatability estimation and variance decomposition by generalized linear mixed-effects models. *Methods in Ecology and Evolution*, **8**, 1639–1644. https://dx.doi.org/10.1111/2041-210X.12797.

Taylor, S.J. (1982) Financial returns modelled by the product of two stochastic processes-a study of the daily sugar prices 1961-75. *Time series analysis: theory and practive*, **1**, 203–226.

Tokuda, T., Goodrich, B., van Mechelen, I., Gelman, A. & Tuerlinckx, F. (2011) Visualizing distributions of covariance matrices. Technical report, Columbia University.

Wagner, G.P. (2023) Models of contingent evolvability suggest dynamical instabilities in body shape evolution. T.F. Hansen, D. Houle, M. Pavlicev & C. Pélabon, eds., *Evolvability: A unifying concept in evolutionary biology?*, chapter 10, pp. 199–219. MIT Press.

Warton, D.I., Blanchet, F.G., O'Hara, R.B., Ovaskainen, O., Taskinen, S., Walker, S.C. & Hui, F.K. (2015) So many variables: joint modeling in community ecology. *Trends in ecology & evolution*, **30**, 766–779.

Westneat, D.F., Schofield, M. & Wright, J. (2013) Parental behavior exhibits among-individual variance, plasticity, and heterogeneous residual variance. *Behavioral Ecology*, **24**, 598–604. https://dx.doi.org/10.1093/beheco/ars207.

6 Supplementary tables and figures

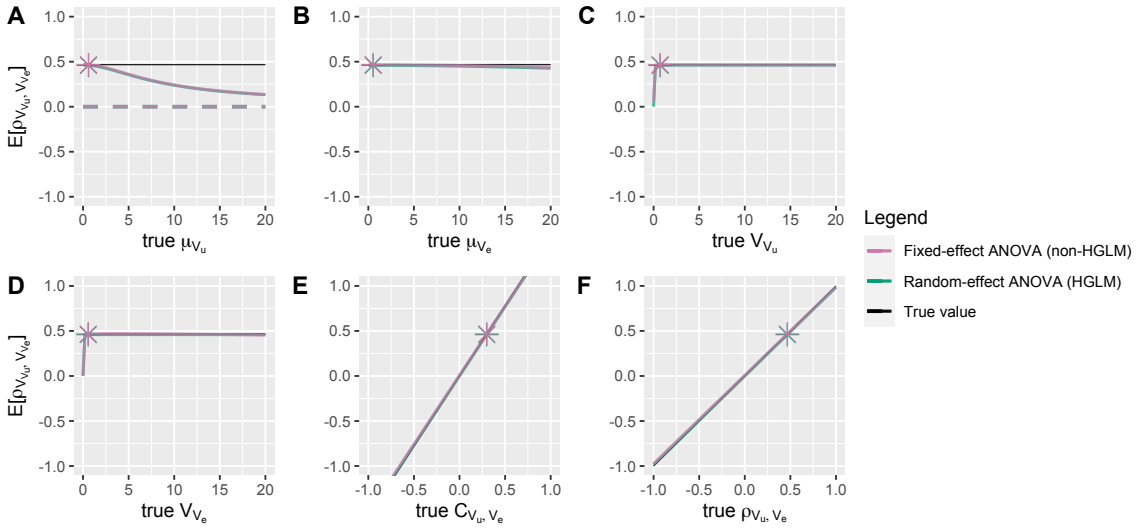


Figure S1: Theoretical correlation among variance components as a function of model parameters. The theoretical correlation is calculated for HGLM (green) and non-HGLM (pink), based on their expected (co)variances of variance components (Equations 17-19 and 20-22, respectively), $\rho_{E[V_{V_g}],E[V_{V_e}]}$. In each panel (A-F) a single parameter is varying, while the remaining are held constant according to the true values in Table 2. The number of subgroups, c, is assumed to be 100 and the number of observations per subgroup, n, is 100.

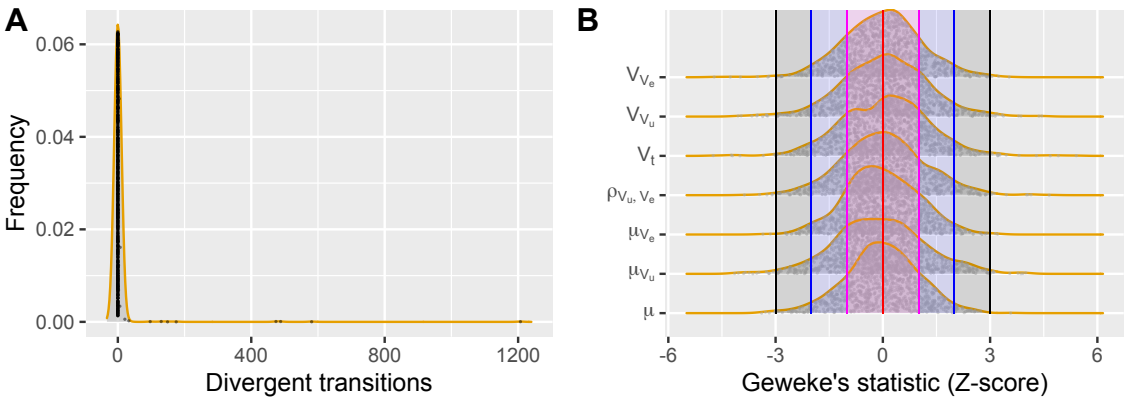


Figure S2: MCMC convergence diagnostics of 1000 simulated data sets, each with a single MCMC chain. **A)** Number of divergent transitions per MCMC chain (black dots). The yellow curve shows its distribution over 1000 MCMC chains. **B)** Geweke's statistics (Z-score), based on the comparison of means of the first 10% and the latter 50% of each MCMC chain. The yellow curves show, for each parameter, the density distributions of Geweke's statistics over 1000 MCMC chains (1 per simulated dataset; grey dots). The red line marks a Z-score of zero, the magenta lines a Z-score of -1 and 1 (1 standard deviation) and the blue lines a Z-score of -2 and 2 (2 standard deviations) and the black lines a Z-score of -3 and 3 (3 standard deviation).

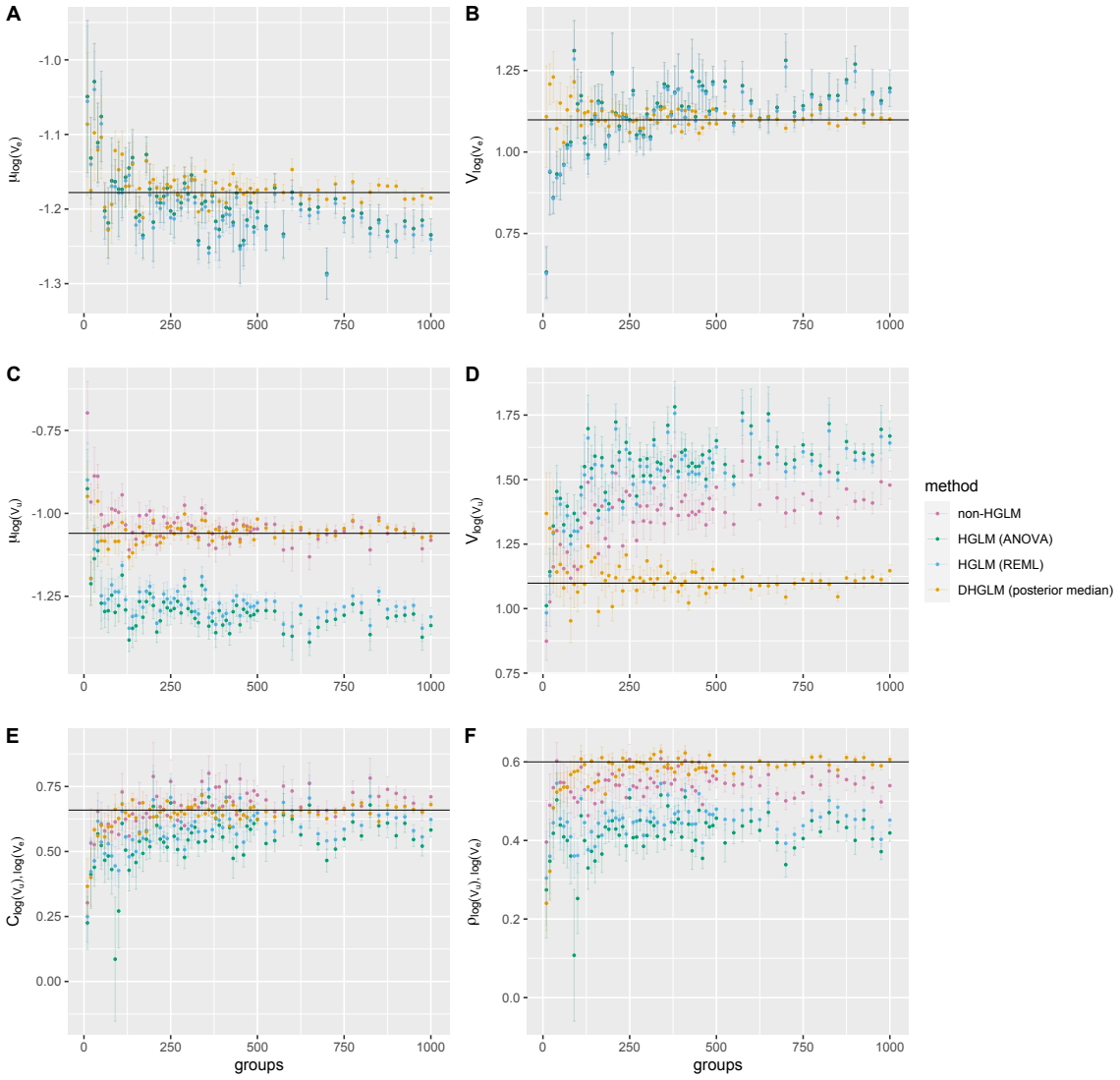


Figure S3: Accuracy of different methods in estimating mean and variance of variance components, on the logarithmic scale, as a function of the number of groups. The number of subgroups and observations per subgroup are fixed at $c = 4$ and $n = 5$, respectively. A,C) Mean variance components. B,D) Variance in variance components. E) Covariance among variance components. F) Correlation between variance components. Each data point is the mean estimate obtained among 15 simulated data sets, with interquartile ranges shown by bars. True values are shown by horizontal black lines.

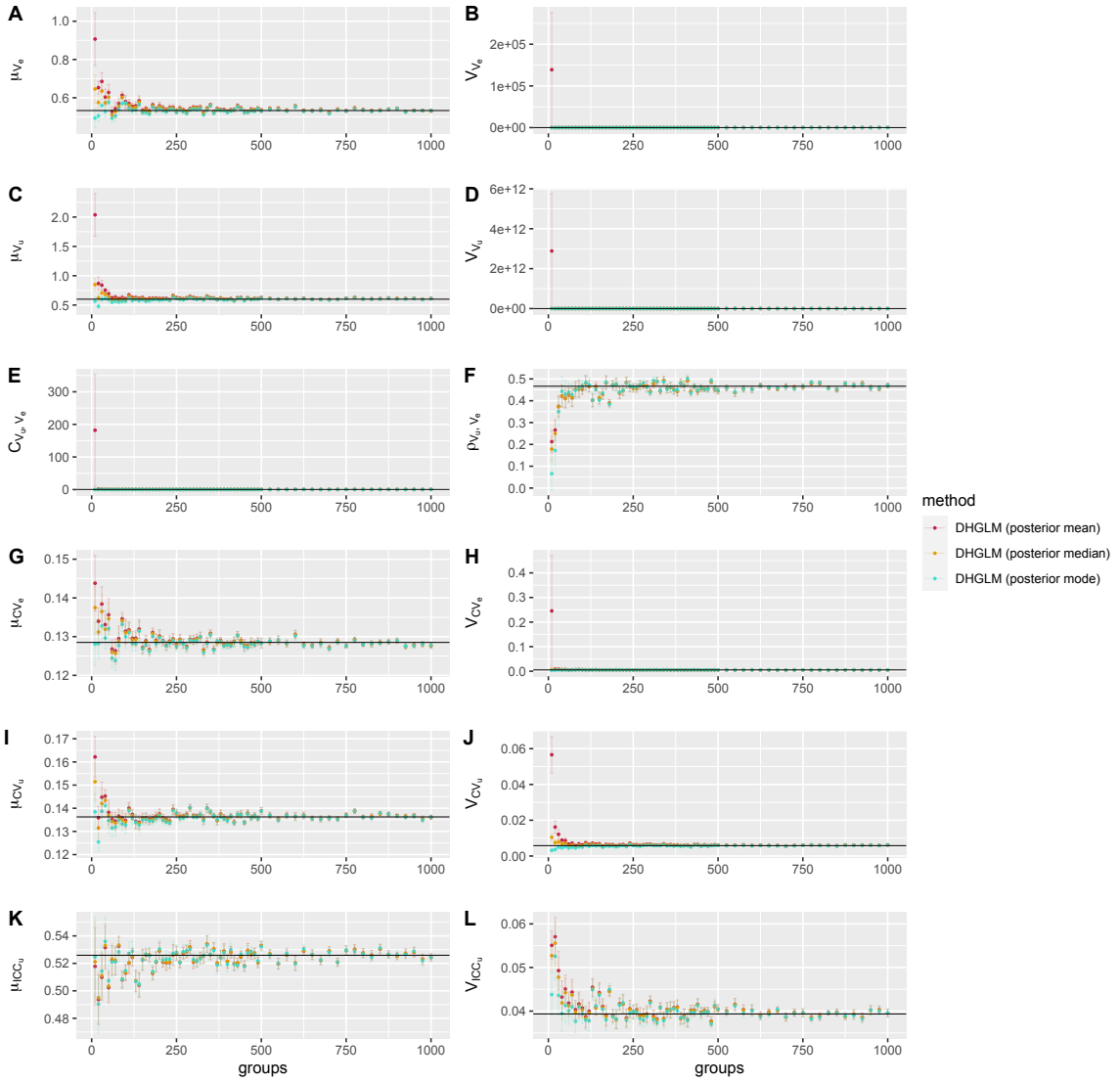


Figure S4: Accuracy of different central tendency measures of posterior distributions (posterior mean, median and mode) in estimating the mean and variance of variance components and their standardisations (coefficients of variation and intraclass correlation), on the arithmetic scale, as a function of the number of groups. The number of subgroups and observations per subgroup are fixed at $c = 4$ and $n = 5$, respectively. A,C) Mean variance components. B,D) Variance in variance components. E) Covariance among variance components. F) Correlation between variance components. G,I) Mean coefficients of variation. H,J) Variance in coefficients of variation. K) Mean intraclass correlation. L) Variance in intraclass correlation. Each data point is the mean estimate obtained among 15 simulated data sets, with interquartile ranges shown by bars. True values are shown by horizontal black lines.

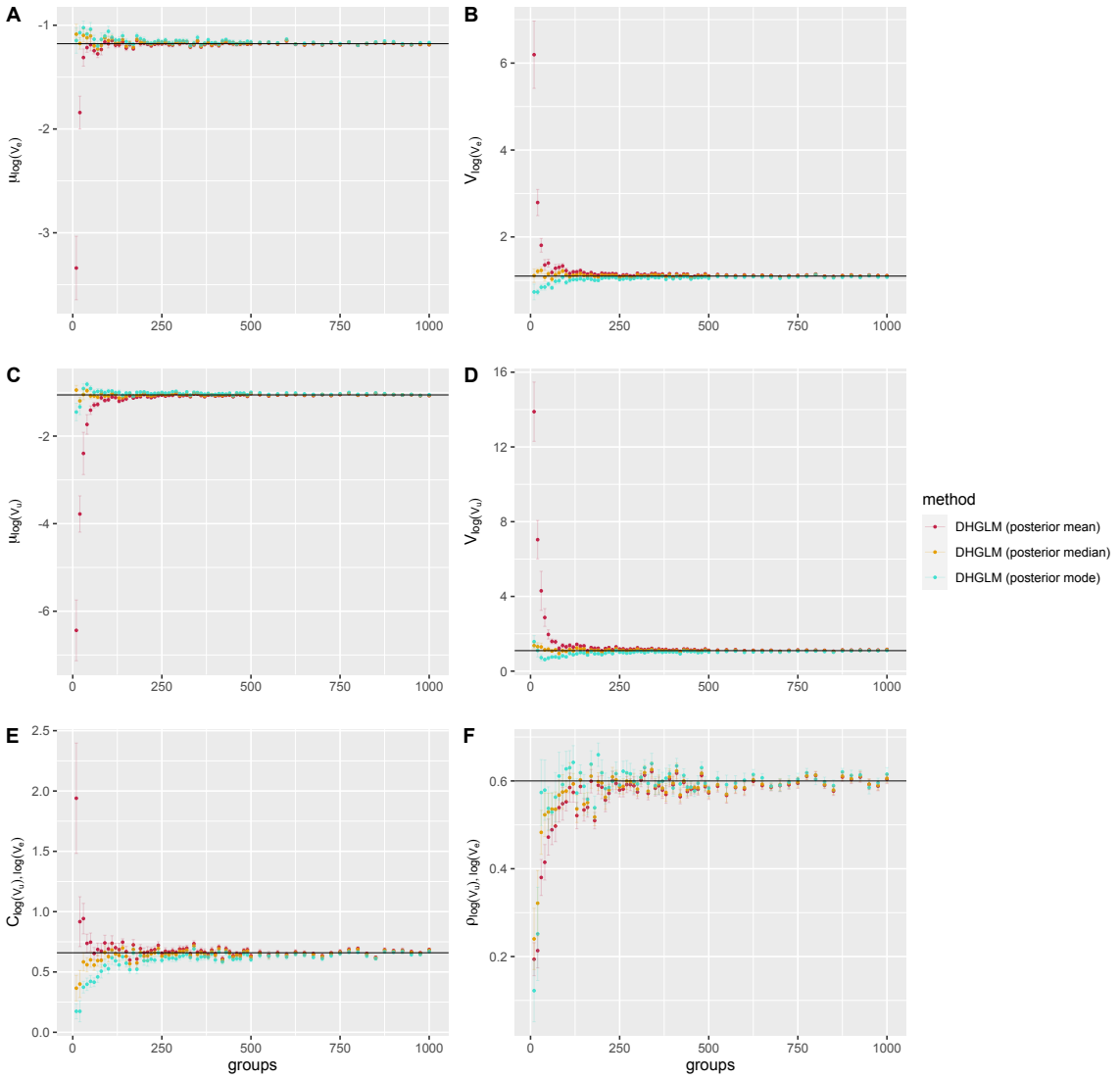


Figure S5: Accuracy of different central tendency measures of posterior distributions (posterior mean, median and mode) in estimating the mean and variance of of variance components, on the logarithmic scale, as a function of the number of groups. The number of subgroups and observations per subgroup are fixed at $c = 4$ and $n = 5$, respectively. A,C) Mean variance components. B,D) Variance in variance components. E) Covariance among variance components. F) Correlation between variance components. Each data point is the mean estimate obtained among 15 simulated data sets, with interquartile ranges shown by bars. True values are shown by horizontal black lines.

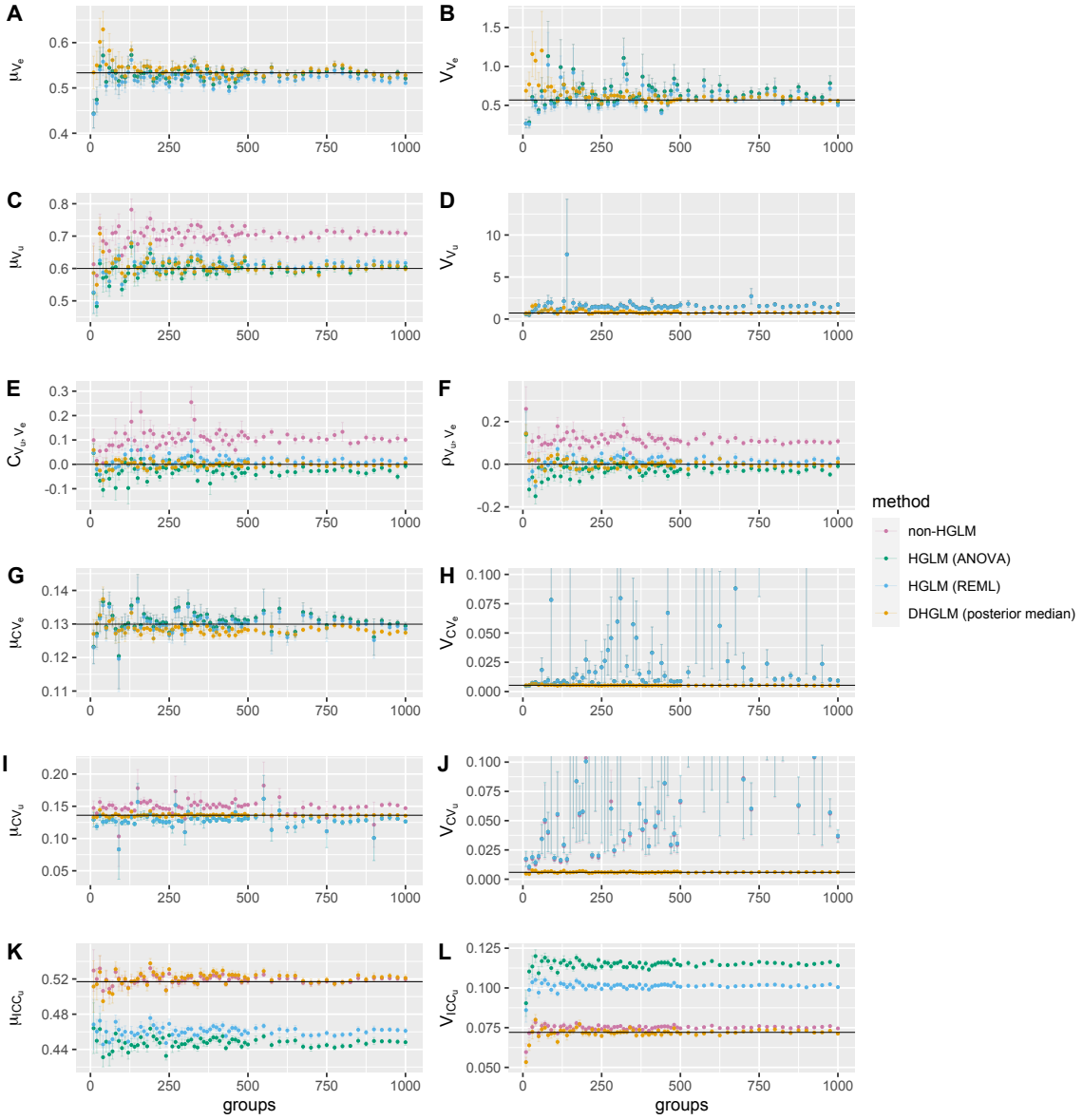


Figure S6: Accuracy of different methods in estimating the mean and variance of variance components and their standardisations (coefficients of variation and intraclass correlation), on the arithmetic scale, as a function of the number of groups, when their correlation is zero. The number of subgroups and observations per subgroup are fixed at $c = 4$ and $n = 5$, respectively. A,C) Mean variance components. B,D) Variance in variance components. E) Covariance among variance components. F) Correlation between variance components. G,I) Mean coefficients of variation. H,J) Variance in coefficients of variation. K) Mean intraclass correlation. L) Variance in intraclass correlation. Each data point is the mean estimate obtained among 15 simulated data sets, with interquartile ranges shown by bars. True values are shown by horizontal black lines.

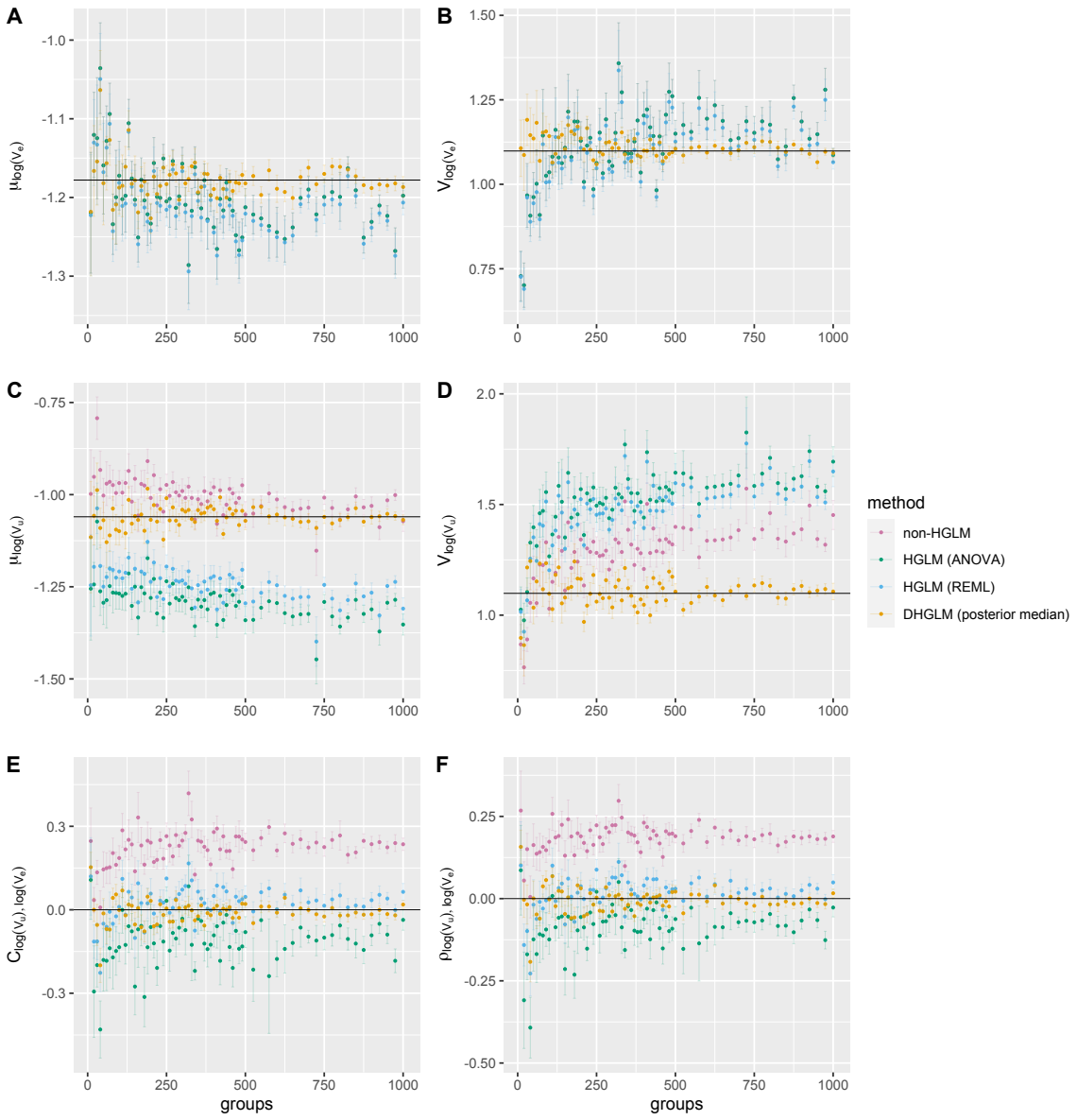


Figure S7: Accuracy of different methods in estimating the mean and variance of variance components, on the logarithmic scale, as a function of the number of groups, when their correlation is zero. The number of subgroups and observations per subgroup are fixed at $c = 4$ and $n = 5$, respectively. A,C) Mean variance components. B,D) Variance in variance components. E) Covariance among variance components. F) Correlation between variance components. Each data point is the mean estimate obtained among 15 simulated data sets, with interquartile ranges shown by bars. True values are shown by horizontal black lines.

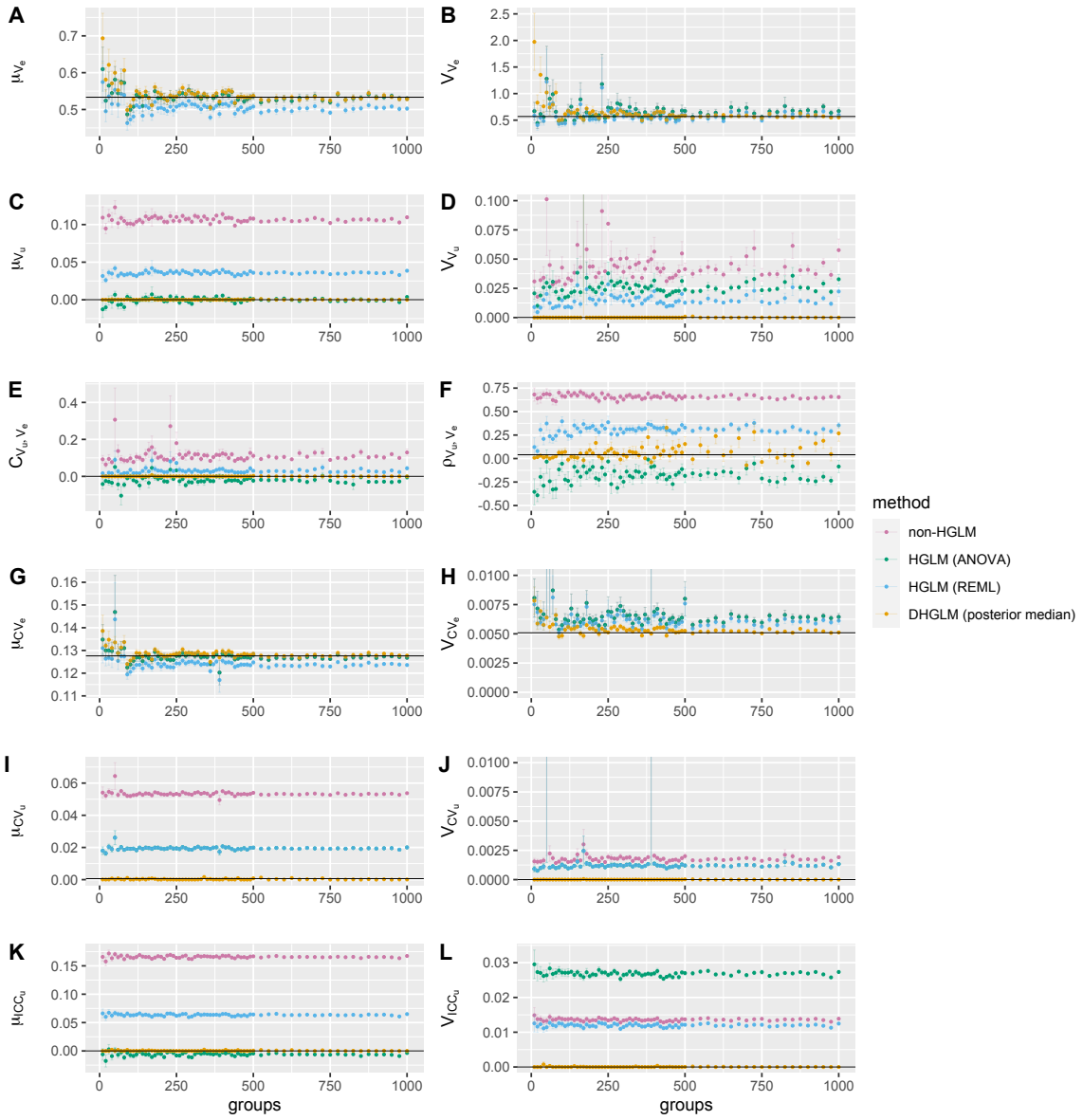


Figure S8: Accuracy of different methods in estimating the mean and variance of variance components and their standardisations (coefficients of variation and intraclass correlation), on the arithmetic scale, as a function of the number of groups, when the mean and variance in V_u are $\mu_{V_u} = V_{V_u} = 0.005$. The number of subgroups and observations per subgroup are fixed at $c = 4$ and $n = 5$, respectively. A,C) Mean variance components. B,D) Variance in variance components. E) Covariance among variance components. F) Correlation between variance components. G,I) Mean coefficients of variation. H,J) Variance in coefficients of variation. K) Mean intraclass correlation. L) Variance in intraclass correlation. Each data point is the mean estimate obtained among 15 simulated data sets, with interquartile ranges shown by bars. True values are shown by horizontal black lines.

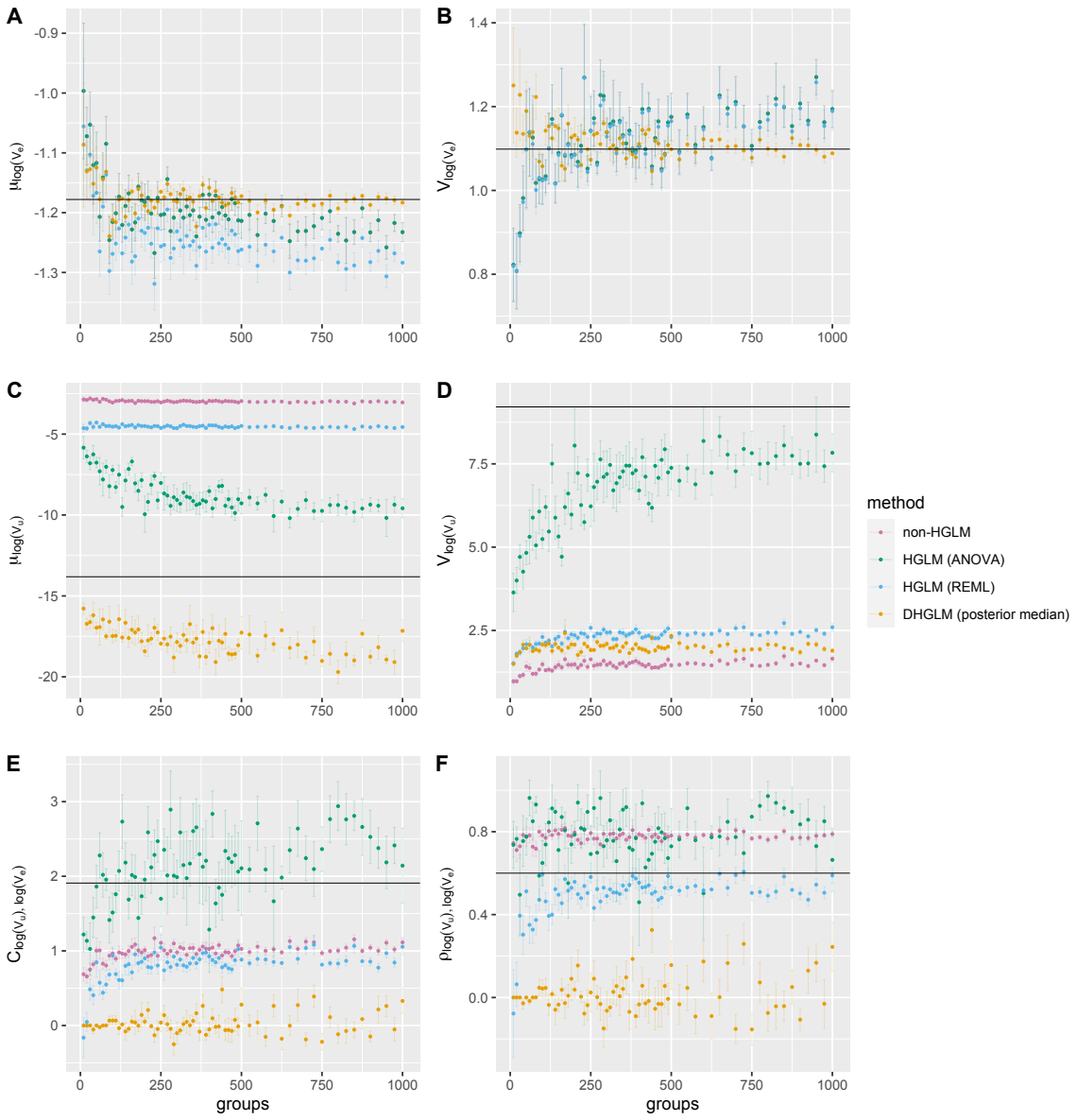


Figure S9: Accuracy of different methods in estimating the mean and variance of variance components, on the logarithmic scale, as a function of the number of groups, when the mean and variance in V_u are $\mu_{V_u} = V_{V_u} = 0.005$. The number of subgroups and observations per subgroup are fixed at $c = 4$ and $n = 5$, respectively. A,C) Mean variance components. B,D) Variance in variance components. E) Covariance among variance components. F) Correlation between variance components. Each data point is the mean estimate obtained among 15 simulated data sets, with interquartile ranges shown by bars. True values are shown by horizontal black lines.

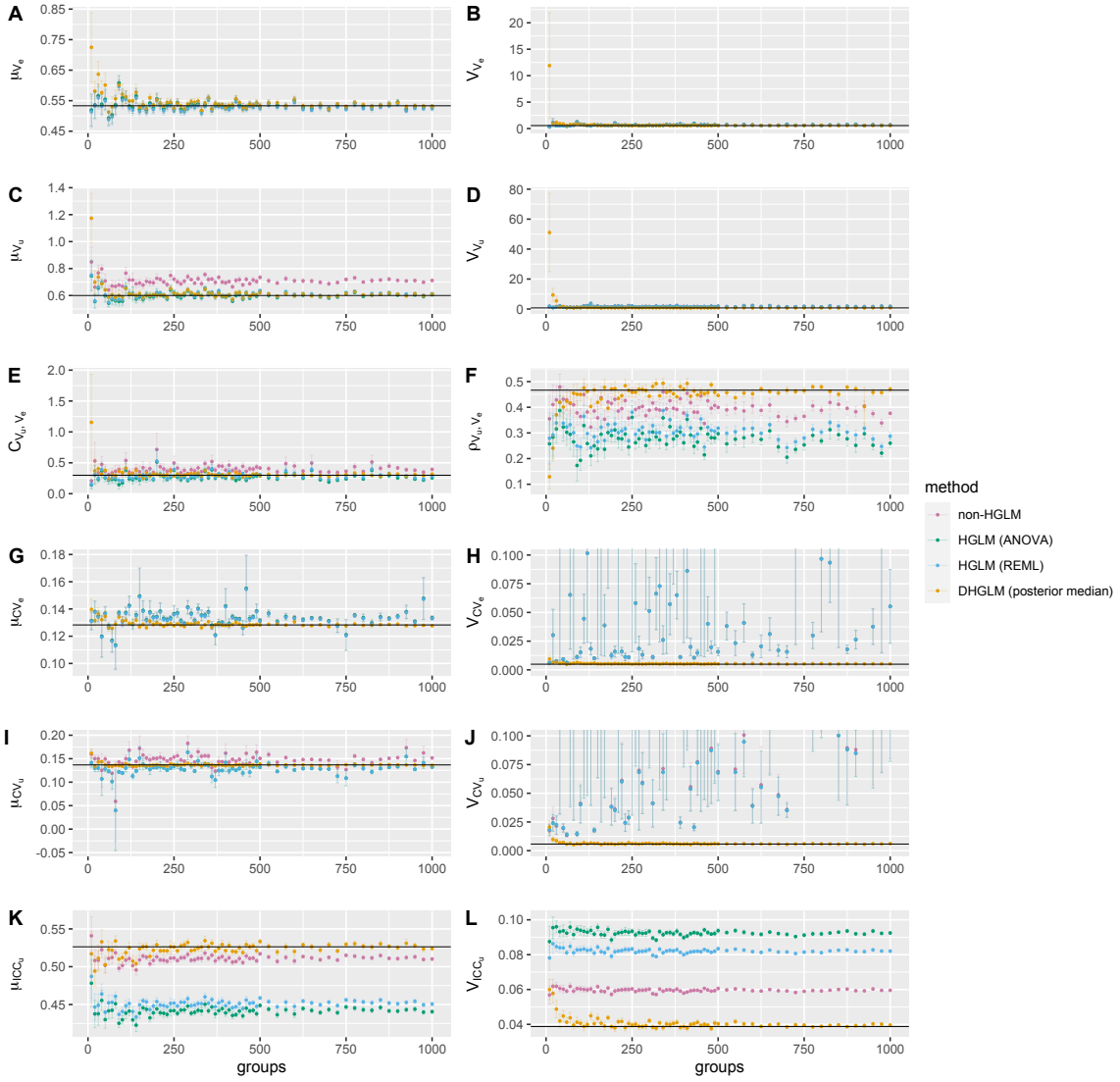


Figure S10: Accuracy of different methods in estimating the mean and variance of variance components and their standardisations (coefficients of variation and intraclass correlation), on the arithmetic scale, as a function of the number of groups, when using a half-Cauchy prior for the dispersion standard deviations (rather than the GIG prior on the dispersion variances). The number of subgroups and observations per subgroup are fixed at $c = 4$ and $n = 5$, respectively. A,C) Mean variance components. B,D) Variance in variance components. E) Covariance among variance components. F) Correlation between variance components. G,I) Mean coefficients of variation. H,J) Variance in coefficients of variation. K) Mean intraclass correlation. L) Variance in intraclass correlation. Each data point is the mean estimate obtained among 15 simulated data sets, with interquartile ranges shown by bars. True values are shown by horizontal black lines.

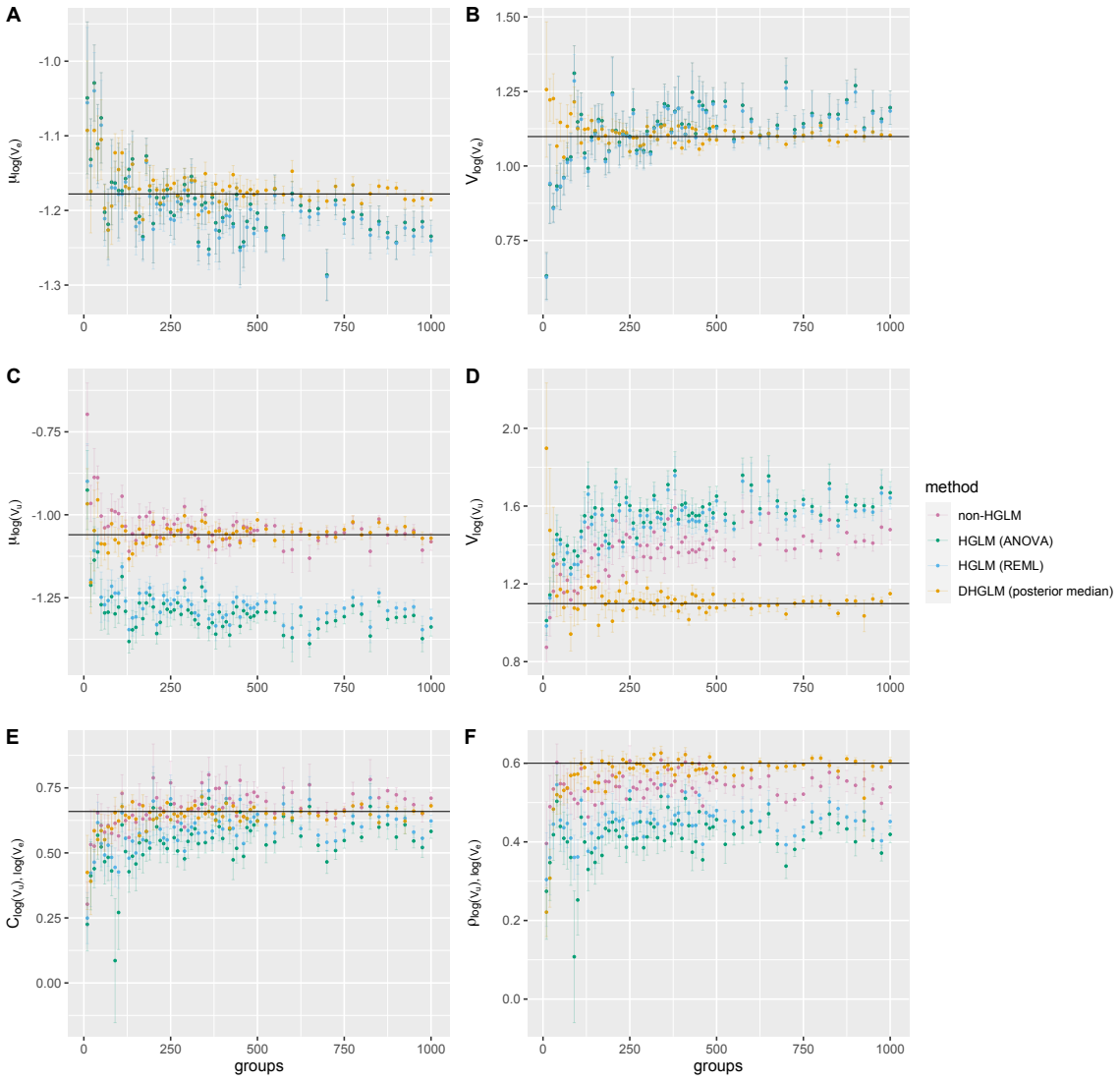


Figure S11: Accuracy of different methods in estimating the mean and variance of variance components, on the logarithmic scale, as a function of the number of groups, when using a half-Cauchy prior for the dispersion standard deviations (rather than the GIG prior on the dispersion variances). The number of subgroups and observations per subgroup are fixed at $c = 4$ and $n = 5$, respectively. A,C) Mean variance components. B,D) Variance in variance components. E) Covariance among variance components. F) Correlation between variance components. Each data point is the mean estimate obtained among 15 simulated data sets, with interquartile ranges shown by bars. True values are shown by horizontal black lines.

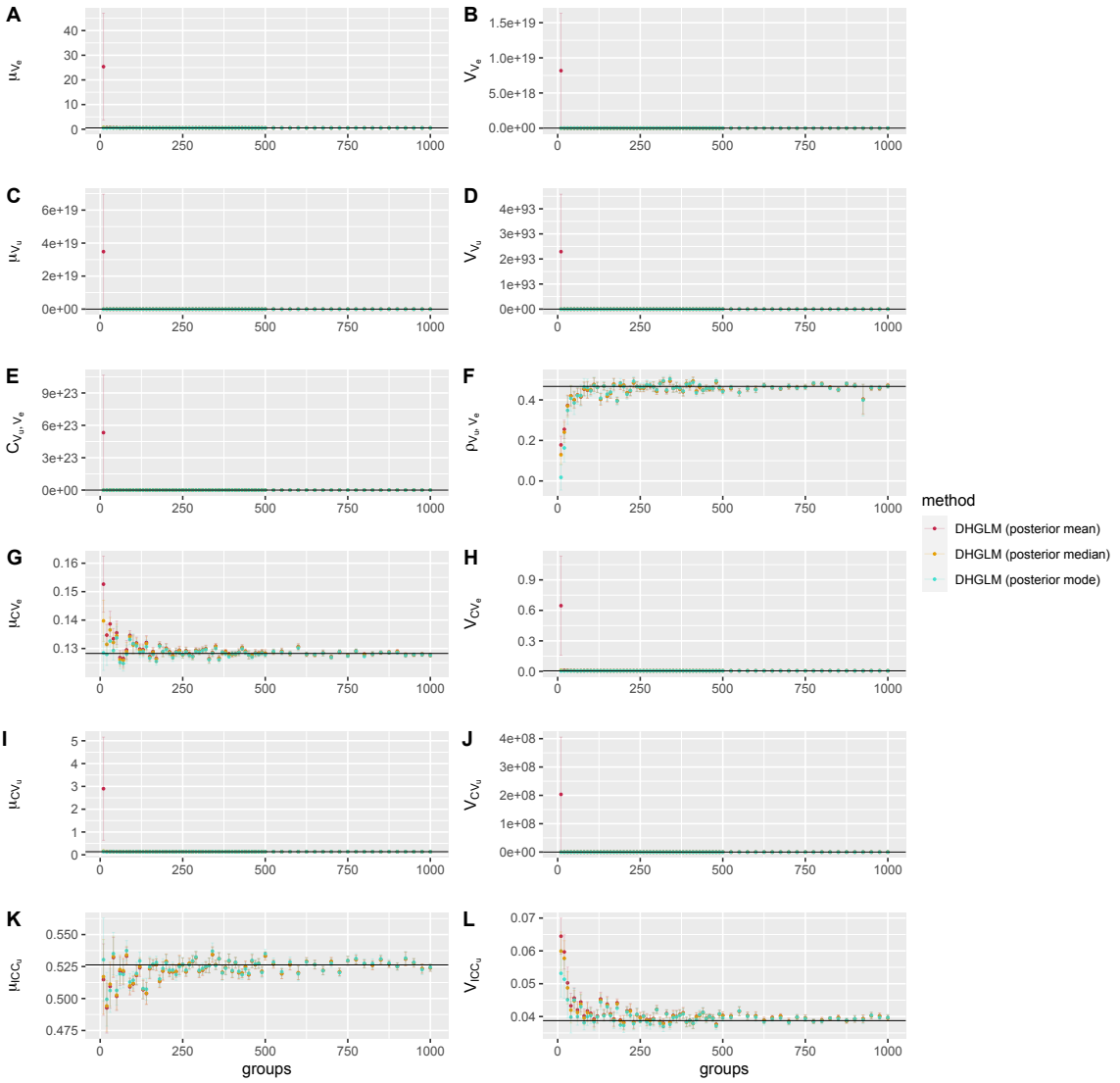


Figure S12: Accuracy of different central tendency measures of posterior distributions (posterior mean, median and mode) in estimating mean and variance of variance components and their standardisations (coefficients of variation and intraclass correlation), on the arithmetic scale, as a function of the number of groups, when using a half-Cauchy prior for the dispersion standard deviations (rather than the GIG prior on the dispersion variances). The number of subgroups and observations per subgroup are fixed at $c = 4$ and $n = 5$, respectively. A,C) Mean variance components. B,D) Variance in variance components. E) Covariance among variance components. F) Correlation between variance components. Each data point is the mean estimate obtained among 15 simulated data sets, with interquartile ranges shown by bars. True values are shown by horizontal black lines.

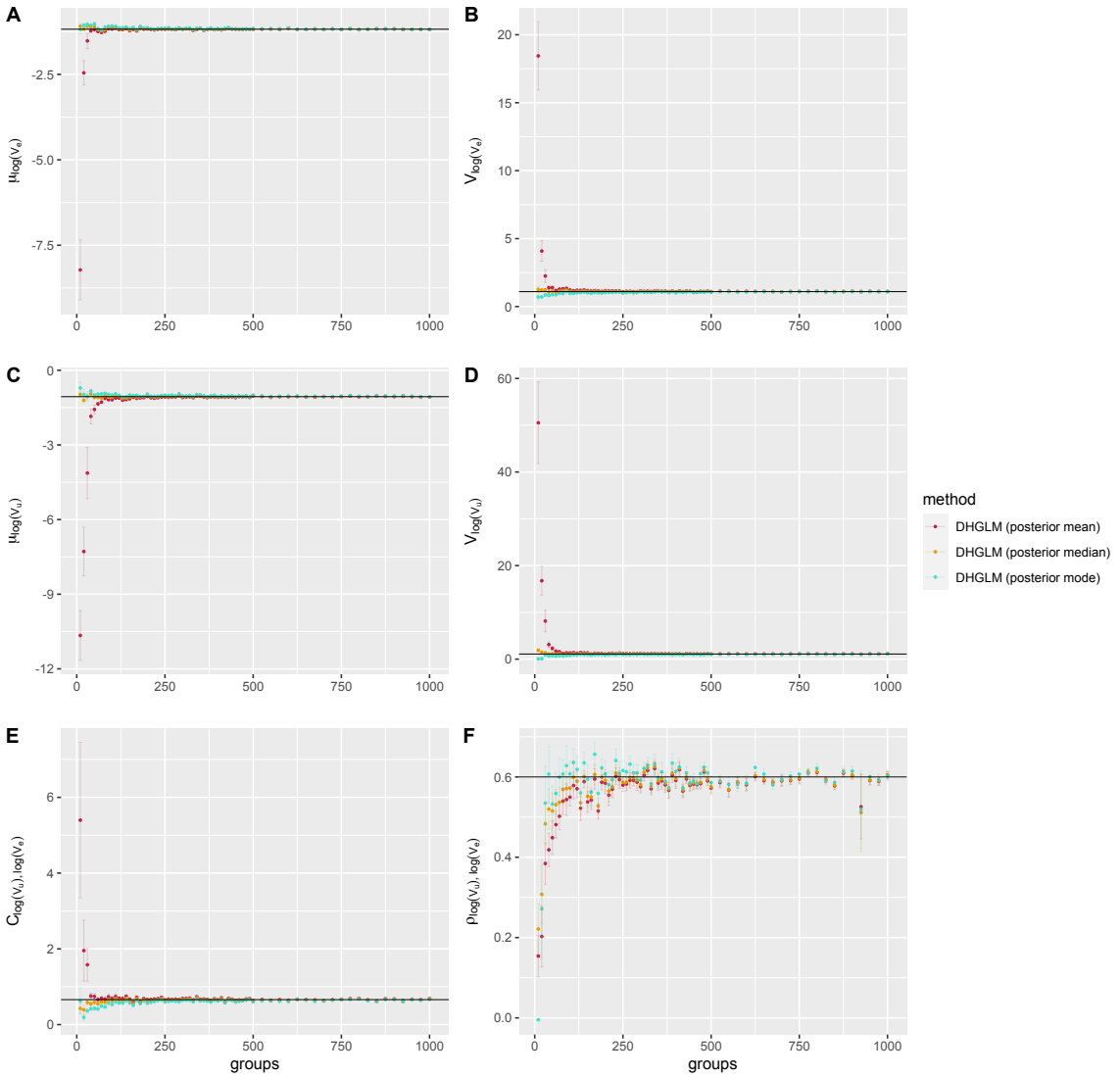


Figure S13: Accuracy of different central tendency measures of posterior distributions (posterior mean, median and mode) in estimating the mean and variance of variance components, on the logarithmic scale, as a function of the number of groups, when using a half-Cauchy prior for the dispersion standard deviations (rather than the GIG prior on the dispersion variances). The number of subgroups and observations per subgroup are fixed at $c = 4$ and $n = 5$, respectively. A,C) Mean variance components. B,D) Variance in variance components. E) Covariance among variance components. F) Correlation between variance components. Each data point is the mean estimate obtained among 15 simulated data sets, with interquartile ranges shown by bars. True values are shown by horizontal black lines.

Bias when estimating the variances and covariances of variance components

```
library(MASS)
```

In this notebook we consider properties of the sampling distribution for estimates of the between-subgroup (V_u) and within-subgroup (V_e) variance using fixed-effects (non-HGLM) and random-effects (HGLM) ANOVA. When these variances themselves are believed to vary over groups, the mean and variance of the *estimated* variances are often used as estimates of the mean and variance of the *true* variances. We go on to derive expressions for the bias in these estimates under a balanced design with c subgroups and n observations per subgroup.

Sampling (co)variances of variance components in random-effect ANOVA (from Searle (1956))

Sums of squares can be expressed as quadratic forms and well-known expressions for the variance of quadratic forms can be used to obtain the sampling (co)variances of ANOVA-based variance component estimates, even though the full sampling distribution is intractable (Crump 1946; Searle 1956). Searle (1956) derives the expressions for random-effect ANOVA:

$$VAR(\widehat{V_u}) = \frac{1}{f^2}\left[\frac{2V_e^2(N-1)}{(c-1)(N-c)} + \frac{2V_eV_u(N^2-S_2)}{N(c-1)^2} + \frac{2V_u^2(N^2S_2+S_2^2-2NS_3)}{N^2(c-1)^2}\right]$$

,

$$VAR(\widehat{V_e}) = \frac{2V_e^2}{N-c}$$

and

$$COV(\widehat{V_u},\widehat{V_e}) = (-1/n)\frac{2V_e^2}{N-c}$$

When the design is balanced $N = nc$, $f = n$, $S_2 = Nn$ and $S_3 = Nn^2$. This leads to

$$\begin{aligned}
VAR(\widehat{V_u}) &= \tfrac{1}{n^2}\left[\tfrac{2V_e^2(N-1)}{(c-1)(N-c)} + \tfrac{2V_eV_u(N^2-Nn)}{N(c-1)^2} + \tfrac{2V_u^2(N^3n+N^2n^2-2N^2n^2)}{N^2(c-1)^2}\right] \\
&= \tfrac{1}{n^2}\left[\tfrac{2V_e^2(N-1)}{(c-1)(N-c)} + \tfrac{2V_eV_u(N-n)}{(c-1)^2} + \tfrac{2V_u^2(Nn+n^2-2n^2)}{(c-1)^2}\right] \\
&= \tfrac{1}{n^2}\left[\tfrac{2V_e^2(N-1)}{(c-1)(N-c)} + \tfrac{2V_eV_u(N-n)}{(c-1)^2} + \tfrac{2V_u^2n(N-n)}{(c-1)^2}\right] \\
&= \tfrac{2}{n^2}\left[\tfrac{V_e^2(N-1)}{(c-1)(N-c)} + \tfrac{V_eV_un}{c-1} + \tfrac{V_u^2n^2}{c-1}\right] \\
&= \tfrac{2}{(c-1)n^2}\left[\tfrac{V_e^2(N-1)}{N-c} + V_eV_un + V_u^2n^2\right]
\end{aligned}$$

We implement these expressions in the function `SVCV.vc` which returns a 2x2 matrix with the sampling variances of $\widehat{V_u}$ and $\widehat{V_e}$ along the diagonal and the sampling covariance on the off-diagonal:

```
SVCV.vc<-function(Vu, Ve, n,c){
  N<-n*c
  V<-matrix(NA, 2, 2)
  V[1,1]<-(2/n^2)*((Ve^2)*(N-1)/((c-1)*(N-c))+Ve*Vu*(N-n)/((c-1)^2)+(Vu^2)*n*(N-n)/((c-1)^2))
  V[1,2]<-V[2,1]<-(-2/n)*(Ve^2)/(N-c)
  V[2,2]<-2*(Ve^2)/(N-c)
  return(V)

}
```

Simulation to check the equations for the sampling (co)variances

To check the results in Searle (1956), we can simulate data, obtain estimates of V_u and V_e using random-effect ANOVA (`Vu.est` and `Ve.est`) and fixed-effect ANOVA (see below: `Vu.est.fixed` and `Ve.est.fixed`) and compare their variances to the predicted sampling variances. We can specify the variance parameters and a specific design from which they are estimated:

```
c<-4            # number of subgroups
n<-7            # number of observations per subgroup
N<-n*c

Vu<-1           # variance in subgroup effects
Ve<-3           # residual variance
```

We can then simulate data under this design to obtain the distribution of estimates

```
n_sim<-100000  # number of simulations

fac<-gl(c,n)   # subgroup factors

Vu.est<-Ve.est<-1:n_sim
Vu.est.fixed<-Ve.est.fixed<-1:n_sim
# vectors for storing estimates

for(i in 1:n_sim){

  y<-rnorm(c,0,sqrt(Vu))[fac]+rnorm(N,0,sqrt(Ve))
  # simulate observations

  m1<-summary(aov(y~fac))
  # fit linear model and get sum-of-squares

  Vu.est[i]<-(m1[[1]]$`Mean Sq`[1]-m1[[1]]$`Mean Sq`[2])/n
  Ve.est[i]<-Ve.est.fixed[i]<-m1[[1]]$`Mean Sq`[2]
  # estimates from random-effect ANOVA

  Vu.est.fixed[i]<-m1[[1]]$`Mean Sq`[1]/n
  # estimates of Vu from fixed-effect ANOVA
  # Ve estimate is the same as random-effect ANOVA

}
```

We can then compare the distribution to what we expect

```
SVCV.vc(Vu, Ve, n, c)
```

```
##             [,1]       [,2]
## [1,]  1.0901361 -0.1071429
## [2,] -0.1071429  0.7500000
```

```
# predicted sampling (co)variances

cov(cbind(Vu.est, Ve.est))
```

```
##            Vu.est     Ve.est
## Vu.est  1.3849855 -0.1070139
## Ve.est -0.1070139  0.7468500
```

```
# observed sampling (co)variances
```

The sampling variance of $\widehat{V_u}$ seems larger than that predicted by the Equation in Searle (1956).

Rederivation of the results in Searle (1956): a factor of 2 is missing

From first principals, the estimate of V_e has the form:

$$\widehat{V_e} = (\mathbf{y} - \bar{\mathbf{y}})^\top \mathbf{F}_1 (\mathbf{y} - \bar{\mathbf{y}})/(N - c)$$

where $\bar{\mathbf{y}}$ is a vector of expected values of $\mathbf{y}$ (with the subgroup effects marginalised). The matrix, $\mathbf{F}_1$, is fixed such that

$$VAR(\widehat{V_e}) = 2Tr(\mathbf{V}\mathbf{F}_1\mathbf{V}\mathbf{F}_1)/(N - c)^2$$

where $\mathbf{V}$ is the covariance matrix of $(\mathbf{y} - \bar{\mathbf{y}})$ and can be expressed as the direct sum

$$\mathbf{V} = \oplus^c \left(\mathbf{I}_n(V_u + V_e) + \mathbf{J}_n V_u\right)$$

where observations in the same subgroup are consecutive. $\mathbf{V}$ is referred to as C-type matrix in Searle (1956) with $a = V_u + V_e$ and $b = V_u$, and $\mathbf{F}_1$ is also a C-type matrix with $a = (1 - 1/n)$ and $b = -1/n$. This gives

$$VAR(\widehat{V_e}) = \frac{2V_e^2}{N - c}$$

as given in Searle (1956).

The estimate of V_u has the form:

$$\widehat{V_u} = (\mathbf{y} - \bar{\mathbf{y}})^\top \mathbf{F}_2 (\mathbf{y} - \bar{\mathbf{y}})/n$$

such that

$$VAR(\widehat{V_u}) = 2Tr(\mathbf{V}\mathbf{F}_2\mathbf{V}\mathbf{F}_2)/n^2$$

where $\mathbf{F}_2$ is a C-type matrix with $a = 0$ and $b = 1/(N - c)$ but with elements outside of the diagonal blocks (i.e. elements corresponding to different subgroups) equal to $k = -1/N(c - 1)$. This gives

$$
\begin{aligned}
VAR(\widehat{V_u}) &= \frac{2N}{n^2}\left[\left(\frac{(n-1)V_u}{N-c}\right)^2 + (n-1)\left(\frac{V_e+(n-1)V_u}{N-c}\right)^2 + (N-n)\left(-\frac{V_e+nV_u}{N(c-1)}\right)^2\right] \\
&= \frac{2N}{n^2}\left[\frac{(n-1)^2V_u^2}{(N-c)^2} + (n-1)\frac{V_e^2+(n-1)^2V_u^2+2(n-1)V_eV_u}{(N-c)^2} + (N-n)\frac{V_e^2+n^2V_u^2+2nV_eV_u}{N^2(c-1)^2}\right] \\
&= \frac{2N}{n^2}\left[\frac{(n-1)V_e^2}{(N-c)^2} + \frac{(N-n)V_e^2}{N^2(c-1)^2} + \frac{2(n-1)^2V_eV_u}{(N-c)^2} + \frac{2(N-n)nV_eV_u}{N^2(c-1)^2} + \frac{(n-1)^2V_u^2}{(N-c)^2} + \frac{(n-1)^3V_u^2}{(N-c)^2} + \frac{(N-n)n^2V_u^2}{N^2(c-1)^2}\right] \\
&= \frac{2N}{n^2}\left[V_e^2\left(\frac{(n-1)}{(N-c)^2} + \frac{(N-n)}{N^2(c-1)^2}\right) + V_eV_u\left(\frac{2(n-1)^2}{(N-c)^2} + \frac{2(N-n)n}{N^2(c-1)^2}\right) + V_u^2\left(\frac{(n-1)^2}{(N-c)^2} + \frac{(n-1)^3}{(N-c)^2} + \frac{(N-n)n^2}{N^2(c-1)^2}\right)\right] \\
&= \frac{2N}{n^2}\left[V_e^2\left(\frac{1}{c^2(n-1)} + \frac{1}{c^2(N-n)}\right) + V_eV_u\left(\frac{2}{c^2} + \frac{2n}{c^2(N-n)}\right) + V_u^2\left(\frac{1}{c^2} + \frac{(n-1)}{c^2} + \frac{n^2}{c^2(N-n)}\right)\right] \\
&= \frac{2N}{n^2}\left[V_e^2\left(\frac{(N-n)+(n-1)}{c^2(n-1)(N-n)}\right) + V_eV_u\left(\frac{2(N-n)+2n}{c^2(N-n)}\right) + V_u^2\left(\frac{(N-n)+(n-1)(N-n)+n^2}{c^2(N-n)}\right)\right] \\
&= \frac{2}{n^2}\left[V_e^2\left(\frac{(N-1)}{(N-c)(c-1)}\right) + 2V_eV_u\left(\frac{n}{(c-1)}\right) + V_u^2\left(\frac{n^2}{(c-1)}\right)\right] \\
&= \frac{2}{n^2}\left[\frac{V_e^2(N-1)}{(N-c)(c-1)} + \frac{2V_eV_un}{c-1} + \frac{V_u^2n^2}{c-1}\right] \\
&= \frac{2}{n^2(c-1)}\left[\frac{N-1}{N-c}V_e^2 + 2nV_eV_u + n^2V_u^2\right]
\end{aligned}
$$

since $N-c = c(n-1)$ and $N(c-1) = c(N-n)$. Consequently, it seems Searle (1956) missed a factor of 2 from the second term in the sum. Consequently, we reimplement the function `SVCV.vc` with the correct expressions (and with an additional argument `random` which we discuss latter):

```
SVCV.vc<-function(Vu, Ve, n,c, random=TRUE){
  N<-n*c
  V<-matrix(NA, 2, 2)
  if(random){
    V[1,1]<-(2/((c-1)*n^2))*((Ve^2)*(N-1)/((N-c))+2*Ve*Vu*n+(Vu^2)*n^2)
    V[1,2]<-V[2,1]<-(-2/n)*(Ve^2)/(N-c)
    V[2,2]<-2*(Ve^2)/(N-c)
  }else{
    V[1,1]<-(2/((c-1)*n^2))*((Ve^2)*(N-1)/((N-c))+2*Ve*Vu*n+(Vu^2)*n^2)-2*(Ve^2)/((n^2)*(N-c))
    V[1,2]<-V[2,1]<-0
    V[2,2]<-2*(Ve^2)/(N-c)
  }
  return(V)
}
```

This new function agrees with the simulations:

```
SVCV.vc(Vu, Ve, n, c)
```

```
##            [,1]       [,2]
## [1,]  1.3758503 -0.1071429
## [2,] -0.1071429  0.7500000
```

```
cov(cbind(Vu.est, Ve.est))
```

```
##            Vu.est     Ve.est
## Vu.est  1.3849855 -0.1070139
## Ve.est -0.1070139  0.7468500
```

Sampling (co)variances of variance components in fixed-effect ANOVA

The above sampling variances are from a random-effects ANOVA. If a fixed effects ANOVA was used, the corresponding terms (denoted with a tilde) are easily derived since $\widetilde{V_e} = \widehat{V_e}$ and

$$\widetilde{V_u} = \widehat{V_u} + \frac{1}{n}\widehat{V_e}$$

Since $\widehat{V_u}$ and $\widehat{V_e}$ are unbiased, then $E[\widetilde{V_u}] = V_u + \frac{1}{n}V_e$. The (co)variances are

$$\begin{aligned} VAR(\widetilde{V_u}) = &\ VAR(\widehat{V_u}) + \tfrac{1}{n^2}VAR(\widehat{V_e}) + \tfrac{2}{n}COV(\widehat{V_u}, \widehat{V_e}) \\ = &\ VAR(\widehat{V_u}) + \tfrac{1}{n^2}VAR(\widehat{V_e}) - \tfrac{2}{n^2}VAR(\widehat{V_e}) \\ = &\ VAR(\widehat{V_u}) - \tfrac{1}{n^2}VAR(\widehat{V_e}) \end{aligned}$$

and

$$\begin{aligned} COV(\widetilde{V_u}, \widetilde{V_e}) = &\ COV(\widehat{V_u} + \tfrac{1}{n}\widehat{V_e}, \widehat{V_e}) \\ = &\ COV(\widehat{V_u}, \widehat{V_e}) + \tfrac{1}{n}VAR(\widehat{V_e}) \\ = &\ 0 \end{aligned}$$

This is implemented in the function `SVCV.vc` but with `random=FALSE` and agrees with the simulation results

```
Vu+Ve/n
```

```
## [1] 1.428571
```

```
mean(Vu.est.fixed)
```

```
## [1] 1.429488
```

```
SVCV.vc(Vu, Ve, n, c, random=FALSE)
```

```
##          [,1] [,2]
## [1,] 1.360544 0.00
## [2,] 0.000000 0.75
```

```
cov(cbind(Vu.est.fixed, Ve.est.fixed))
```

```
##                Vu.est.fixed  Ve.est.fixed
## Vu.est.fixed   1.3696519163 -0.0003210977
## Ve.est.fixed -0.0003210977  0.7468499652
```

Expected estimates of the (co)variance of variance components as estimated from random-effect ANOVA

If the variance components themselves vary over groups with variance component x having mean μ_{V_x} and variance V_{V_x}, we can work out the expected values for the estimates of these quantities by noting that estimates of variance components from random-effect ANOVA are unbiased:

$$E[\widehat{\mu_{V_x}}] = \mu_{V_x}$$

and

$$E[\widehat{V_{V_x}}] = V_{V_x} + E[VAR(\widehat{V_x})]$$

$E[VAR(\widehat{V_x})]$ involves the expectations of V_x^2 or V_xV_j which are equal to $\mu_{V_x}^2 + V_{V_x}$ and $\mu_{V_x}\mu_{V_j} + C_{V_x,V_j}$ respectively. Consequently, we have

$$E[\widehat{V_{V_u}}] = V_{V_u} + \frac{2}{n^2(c-1)}\left[\frac{N-1}{N-c}(\mu_{V_e}^2 + V_{V_e}) + 2n(\mu_{V_e}\mu_{V_u} + C_{V_e,V_u}) + n^2(\mu_{V_u}^2 + V_{V_u})\right]$$

$$E[\widehat{V_{V_e}}] = V_{V_e} + \frac{2}{N-c}(\mu_{V_e}^2 + V_{V_e})$$

and

$$E[\widehat{C_{V_u,V_e}}] = C_{V_e,V_u} - \frac{2}{n(N-c)}(\mu_{V_e}^2 + V_{V_e})$$

We implement these in the function `Evarhat` which takes arguments `mu.vu` and `mu.ve` for the mean subgroup and residual variance respectively, `v.vu` and `v.ve` for the variance of the subgroup and residual variance respectively, and `c.vuve` for the covariance between the two variances:

```
Evarhat<-function(mu.vu, mu.ve, v.vu, v.ve, c.vuve, n, c, random=TRUE){

  N<-n*c

  if(random){

    Ev.vu<-v.vu+(2/((n^2)*(c-1)))*(((N-1)/(N-c))*(mu.ve^2+v.ve)
      +2*n*(mu.ve*mu.vu+c.vuve)+(n^2)*(mu.vu^2+v.vu))
    # expected variance in subgroup variance estimates (random-effect ANOVA)

    Ec.vuve<-c.vuve-2*(mu.ve^2+v.ve)/(n*(N-c))
    # expected covariance between subgroup and residual variance estimates (random-effect ANOVA)

  }else{

    Ev.vu<-v.vu+(v.ve/n)+(2/((n^2)*(c-1)))*(((N-1)/(N-c))*(mu.ve^2+v.ve)
      +2*n*(mu.ve*mu.vu+c.vuve)+(n^2)*(mu.vu^2+v.vu))-(v.ve+2*(mu.ve^2+v.ve)/(N-c))/(n^2)
    # expected variance in subgroup variance estimates (fixed-effect ANOVA)

    Ec.vuve<-c.vuve+(v.ve/n)
    # expected covariance between subgroup and residual variance estimates (fixed-effect ANOVA)

  }

  Ev.ve<-v.ve+2*(mu.ve^2+v.ve)/(N-c)
  # expected variance in residual variance estimates

  return(c(Ev.vu, Ec.vuve, Ev.ve))
}
```

Simulations to check expressions for expected estimates of the (co)variance of variance components

To check whether our expectations for the (co)variance of variance component estimates is correct, we can simulate data for multiple groups, obtain estimates of V_u and V_e for reach group and compute their (co)variances. By doing this a number of times we can then calculate the means of these (co)variances and compare them to our expectations.

First we implement a function that takes the means and (co)varinaces of the variance components on the arithmetic scale, and returns the means and (co)varinaces of the bivariate log-normal that are consistent with this.

```
lognormal_par<-function(mu.vu, mu.ve, v.vu, v.ve, c.vuve){

  l.mu.vu<-2*log(mu.vu)-0.5*log(mu.vu^2+v.vu)
  l.v.vu<-log(mu.vu^2+v.vu)-2*log(mu.vu)
```

```
  # obtain parameters of the log-normal from which the subgroup variances are drawn

  l.mu.ve<-2*log(mu.ve)-0.5*log(mu.ve^2+v.ve)
  l.v.ve<-log(mu.ve^2+v.ve)-2*log(mu.ve)
  # obtain parameters of the log-normal from which the residual variances are drawn

  c.l.vuve  = log(1+c.vuve/exp(l.mu.vu+l.mu.ve+l.v.vu/2+l.v.ve/2))
  # obtain the covariance on the log-scale

  l.mu.v<-c(l.mu.vu, l.mu.ve)
  l.v.v<-cbind(c(l.v.vu, c.l.vuve), c(c.l.vuve,l.v.ve))

  return(list(mu=l.mu.v, Sigma=l.v.v))
}
```

We can then specify the mean and (co)variances of the variances on the arithmetic scale together with the experimental design:

```
Ng<-1000    # number of groups
c<-4        # number of subgroups
n<-7        # number of observations per subgroup
N<-n*c

mu.vu<-1    # mean variance in subgroup effects
v.vu<-1     # variance in variance in subgroup effects

mu.ve<-3    # mean residual variance
v.ve<-2     # variance in residual variance

c.vuve<-0.5 # covariance between subgroup and residual variance
```

Then we can simulate data by drawing each group's variance components from the log-normal and then simulating Gaussian data according to the design and the variance components.

```
n_sim<-1000 # number of simulations

par<-lognormal_par(mu.vu, mu.ve, v.vu, v.ve, c.vuve)

fac<-gl(c,n)  # subgroup factors

Var.V.est<-Var.V.est.fixed<-matrix(NA, n_sim,3)

# matrices for storing parameter estimates

for(i in 1:n_sim){

  Vv<-exp(MASS::mvrnorm(Ng, par$mu, par$Sigma))

  # simulate variances

  Y<-matrix(rnorm(c*Ng,0,rep(sqrt(Vv[,1]), each=c)), c,Ng)[fac,]+matrix(rnorm(N*Ng,0,rep(sqrt(Vv[,2]),
  # simulate observations (each column is a group)

  Vest<-apply(Y,2,function(x){summary(aov(x~fac))[[1]]$`Mean Sq`})
  # get mean-squares for each group
```

```
  Vest[1,]<-Vest[1,]/n
  # estimate of the subgroup variance for each group (fixed-effect ANOVA)
  # Vest[2,] is an estimate of the residual variance

  Var.V.est.fixed[i,]<-cov(t(Vest))[c(1,2,4)]
  # variance in subgroup variance estimates
  # covariance between subgroup and residual variance estimates
  # variance in residual variance estimates

  Vest[1,]<-Vest[1,]-Vest[2,]/n
  # estimate of the subgroup variance for each group (random-effect ANOVA)
  # Vest[2,] is an estimate of the residual variance

  Var.V.est[i,]<-cov(t(Vest))[c(1,2,4)]
  # variance in subgroup variance estimates
  # covariance between subgroup and residual variance estimates
  # variance in residual variance estimates
}
```

Comparing

```
c(v.vu, c.vuve, v.ve)
```

```
## [1] 1.0 0.5 2.0
```

```
# true (co)variances

colMeans(Var.V.est)
```

```
## [1] 3.1687140 0.3689505 2.9100788
```

```
# mean (co)variance estimates

varhat_random<-Evarhat(mu.vu, mu.ve, v.vu, v.ve, c.vuve, n, c, random=TRUE)
varhat_random
```

```
## [1] 3.1683673 0.3690476 2.9166667
```

```
# predicted mean (co)variance estimates
```

Expected estimates of the (co)variance of variance components as estimated from fixed-effect ANOVA

The equivalent expressions for the fixed-effect ANOVA are:

$$\begin{aligned} E[\widetilde{V_{V_u}}] = \quad & V_{V_u} + \tfrac{1}{n}V_{V_e} + \tfrac{2}{n^2(c-1)}\left[\tfrac{N-1}{N-c}(\mu^2_{V_e} + V_{V_e}) + 2n(\mu_{V_e}\mu_{V_u} + C_{V_e,V_u}) + n^2(\mu^2_{V_u} + V_{V_u})\right] \\ & -\tfrac{1}{n^2}\left[V_{V_e} + \tfrac{2}{N-c}(\mu^2_{V_e} + V_{V_e})\right] \end{aligned}$$

$$E[\widetilde{V_{V_e}}] = V_{V_e} + \frac{2}{N-c}(\mu^2_{V_e} + V_{V_e})$$

and

$$E[\widetilde{C_{V_e,V_u}}] = C_{V_e,V_u} + \frac{1}{n}V_{V_e}$$

.

It may seem surprising that $E[\widetilde{C_{V_e,V_u}}]$ does not simply equal C_{V_e,V_u} given $COV(\widetilde{V_u},\widetilde{V_e}) = 0$. However, the bias in the fixed-effect ANOVA estimates causes covariances between the sampling errors and the true values that contribute to the expected covariance of estimates when the variances vary.

These equations can be evaluated using the function `Evarhat` but with `random=FALSE`.

```
c(v.vu, c.vuve, v.ve)
```

```
## [1] 1.0 0.5 2.0
```

```
colMeans(Var.V.est.fixed)
```

```
## [1] 3.3335178 0.7846761 2.9100788
```

```
varhat_fixed<-Evarhat(mu.vu, mu.ve, v.vu, v.ve, c.vuve, n, c, random=FALSE)
varhat_fixed
```

```
## [1] 3.3945578 0.7857143 2.9166667
```

Expected estimates of the correlation between variance components

Obtaining the expected estimate for the correlation is more difficult, but simply using the expectations of the component parts seems to be reasonably accurate. For example, for random-effect ANOVA:

```
c.vuve/sqrt(v.vu*v.ve)
```

```
## [1] 0.3535534
```

```
mean(Var.V.est[,2]/sqrt(Var.V.est[,1]*Var.V.est[,3]))
```

```
## [1] 0.1214923
```

```
varhat_random[2]/sqrt(prod(varhat_random[c(1,3)]))
```

```
## [1] 0.1214007
```

and for fixed-effect ANOVA:

```
c.vuve/sqrt(v.vu*v.ve)
```

```
## [1] 0.3535534
```

```
mean(Var.V.est.fixed[,2]/sqrt(Var.V.est.fixed[,1]*Var.V.est.fixed[,3]))
```

```
## [1] 0.2543438
```

```
varhat_fixed[2]/sqrt(prod(varhat_fixed[c(1,3)]))
```

```
## [1] 0.2497064
```


Crump, S Lee. 1946. "The Estimation of Variance Components in Analysis of Variance." *Biometrics Bulletin* 2 (1): 7–11.

Searle, SR. 1956. "Matrix Methods in Components of Variance and Covariance Analysis." *The Annals of Mathematical Statistics*, 737–48.

Fitting multi-way DHGLM in Stan

```
library(rstan)
library(coda)
library(MASS)
library(tidyverse)
```

In this workbook we implement Stan code for fitting a simple multi-way DHGLM. The multi-way DHGLM can be envisaged as a series of standard linear mixed models applied to subsets (groups) of the data. For each group, a single set of random effects (subgroup effects) are fitted, leading to a subgroup variance and a residual variance. The linear mixed models are linked in two ways: group means (the intercepts of the standard linear mixed models) are treated as random over groups, and the pair of variances for each group (residual and subgroup) are assumed to be drawn from a bivariate log-normal distribution over groups, the parameters of which are estimated. We also provide a function for simulating data under this model assuming a balanced design, and then fit the model to data generated using this function.

Stan code for fitting DHGLM

The data structure consists of integers specifying the total number of observations `Nt`, the number of groups `Ng` and the number subgroups `c`, a real vector of observations `y`, and integer vectors `group` and `subgroup` specifying the group and subgroup identifier for each observation. The data do not need to be balanced (i.e. all subgroups present for all groups with equal replication) but the `group` and `subgroup` indices must be integers in the sequence `1:Ng` and `1:c` respectively. `muvar` is an integer indicating whether a mean-variance relationship over groups should be fitted (`1`) or not (`0`). The Mean Model is

$$y_{ijk} = \mu + t_i + u_{ij} + e_{ijk},$$

where μ is the global intercept (`beta`) and t_i is the group i effect (`egroup`), u_{ij} is the subgroup j effect in group i (`egroup_by_subgroup`) and e_{ijk} is a residual. t are normally distributed with zero mean and standard deviation $\sqrt{V_t}$ (`sgroup`). The $u_{i\bullet}$ and $e_{i\bullet\bullet}$ are normally distributed with zero mean and group specific standard deviations $\sqrt{V_{u(i)}}$ (`sds[1,i]`)and $\sqrt{V_{e_i}}$ (`sds[2,i]`), respectively.

The Dispersion Model for the subgroup standard deviations is

$$log(\sqrt{V_{u(i)}}) = \mu_u + \beta_u t_i + d_{u(i)}$$

where μ_u and β_u specify the intercept and slope for the logged standard deviation $\log(\sqrt{V_{u(i)}})$ regressed on the group effect t_i and $d_{u(i)}$ is the residual. An equivalent Dispersion Model is fitted for the residual standard deviations as

$$log(\sqrt{V_{e(i)}}) = \mu_e + \beta_e t_i + d_{e(i)}.$$

When `muvar` is `0` then `beta_lsds`=$[\mu_u,\ \mu_e]^{'}$ and the slopes are set to zero. When `muvar` is `1` then `beta_lsds`=$[\mu_u,\ \mu_e, \beta_u, \beta_e]^{'}$. Alternative models for the mean-variance relationship might be considered. For example, rather than assuming the residual standard deviations are constant across sub-groups within a group, the residual standard deviations could be regressed on the sub-group locations $t_i + u_{ij}$ rather than

those of the group t_i. Additionally, in many cases it might be more suitable to fit the log group means ($log(\mu + t_i)$) as a covariate allowing the variances to follow a power law (Wagner 2023) of the form (for the residual standard deviation):

$$\sqrt{V_{e(i)}} = exp(\mu_e + d_{e(i)})(\mu + t_i)^{\beta_e}$$

$d_{u(i)}$ and $d_{e(i)}$ are assumed to follow a bivariate normal distribution with zero mean and covariance matrix parameterised in terms of a correlation `r_lsds` and a vector of standard deviations `sigma_lsds`. Note that when mean-variance relationships are modelled, `r_lsds` measures the log scale correlation in variance components after controlling for any mean-variance relationship. Calculating the unconditional log-scale correlation under a log-linear mean-variance relationship would be straightforward: $\beta_e\beta_u V_t$ would have to be added to the covariance, and $\beta_e^2 V_t$ and $\beta_u^2 V_t$ added to the variances, respectively, before recalculating the correlation. However, calculating the unconditional log-scale correlation under a power-law relationship would be more difficult since V_t in the above expressions would have to be replaced by $Var(log(\mu + t))$ which could only be approximated: using the Delta method, $Var(log(\mu + t)) \approx V_t/\mu^2$).

As with the Mean Model, additional random effects for the Dispersion model may be considered. For example, sub-groups within groups might have heterogeneous variances even after controlling for any mean-variance relationship. Then, a model of the from:

$$log(\sqrt{V_{e(ij)}}) = \mu_e + \beta_e(t_i + u_{ij}) + d_{e(i)} + d_{e(ij)}.$$

where the $d_{e(ij)}$ are treated as random variables might be more suitable. Similarly, there might be heterogeneous variances at the level of the observation, which following the previous logic suggests the model:

$$log(\sqrt{V_{e(ijk)}}) = \mu_e + \beta_e(t_i + u_{ij}) + d_{e(i)} + d_{e(ij)} + d_{e(ijk)}.$$

Since there is only one observation per level of the observation-level random effect, $d_{e(ijk)}$, the identifiability of these parameters, and their variance, might be called into question. However, the effects are weakly identifiable since their presence will cause the distribution of residuals within a group (or sub-group if $d_{e(ij)}$ is fitted) to have excess kurtosis with respect to the normal. Since the scaled t-distribution can be viewed as a compound distribution of normals whose variances are drawn from an inverse gamma distribution, a model that assumes the e_{ijk} are from a scaled-t, rather than a normal, is equivalent to fitting $d_{e(ijk)}$ as a random effect in the Dispersion Model but assuming they follow an inverse-gamma distribution rather than a log-normal. While the t-distribution approach may be considered less satisfying in that the random effects in the Dispersion Model are effectively following different distributions, a log-normal and inverse-gamma that are matched for their mean and variance are often very similar. The advantage of the t-distribution approach is that the $d_{e(ijk)}$ are effectively integrated out analytically leaving only a single parameter to be estimated (the degrees of freedom) where as the $d_{e(ijk)}$ under the log-normal approach need to be integrated out using MCMC which may be computationally prohibitive. Options for using the t-distribution approach are commented out in the code below (see Juárez and Steel (2010) for a discussion of prior specifications for the degrees of freedom).

Note that the parameterisations above are for the log standard deviations rather than the log variances given in the main manuscript, hence the slightly different notation. However, on the log-scale, reparameterising from the variances to the standard deviations simply scales location and standard deviation effects by two and variances by four. Hence to obtain parameters under the log-variance parameterisation we can multiply `beta_lsds` and `sigma_lsds` by two to get the fixed effects and standard deviations under a log-variance parameterisation. The correlation `r_lsds` is equivalent for both parameterisations. In addition, the mean-variance slopes (β_u and β_e) in the manuscript were omitted and effectively set to zero.

External priors are required for the 'fixed effects', `beta` and `beta_lsds`, and the dispersion parameters `sgroup`, `r_lsds` and `sigma_lsds`. Elements of `beta` and `beta_lsds` are assigned normal priors with zero

mean and standard deviations of 10, and `sgroup` a half-Cauchy prior with location 0 and scale 5. `r_lsds` is assigned a uniform prior from -1 to 1 (although parameterised through a Lewandowski-Kurowicka-Joe (LKJ) prior). `GIG_lpdf` is a function (provided by Enrico Fabrizi) for calculating the log-density of the Generalised Inverse Gaussian (GIG) distribution, although only integer values of γ are permitted. `sqrtGIG_lpdf` is a function for calculating the density of a standard deviation had the variance come from a GIG distribution. The elements of `sigma_lsds` squared (i.e. the variances) are assigned a GIG prior with $\lambda = 1$, $\delta = 0.01$ and $\gamma = \sqrt{3 + 9/N_g}$ (see Gardini, Trivisano, and Fabrizi (2021) for notation and details). However, commented out code provides the option for using half-Cauchy priors on `sigma_lsds` instead. The following stan code object is named `DHGLM_stan`.

```
functions{

  // GIG prior: Enrico Fabrizi https://link.springer.com/article/10.1007/s11336-021-09769-y
  //            only integer lambda allowed

  real GIG_lpdf(real y, int lambda, real delta, real gamma){
    real log_p;
    log_p=1.0*lambda*log(gamma/delta)-log(2.0)-log(modified_bessel_second_kind(lambda, delta*gamma))
    +(1.0*lambda-1.0)*log(y)-0.5*(delta*delta/y+gamma*gamma*y);
    return(log_p);
  }

  // GIG_lpdf calculates the log density of y given a GIG distribution.
  // If y are variances, but we are working on the standard deviation scale, sqrt_y,
  // we can calculate the same density as J*GIG(sqrt_y^2) where J is the Jacobian
  // (the partial derivative of y with respect to sqrt_y (i.e 2 * sqrt_y)).

  real sqrtGIG_lpdf(real sqrt_y, int lambda, real delta, real gamma){
    real log_p;
    log_p = log(sqrt_y)+log(2.0);  // Jacobian
    log_p += 1.0*lambda*log(gamma/delta)-log(2.0)-log(modified_bessel_second_kind(lambda, delta*gamma))
    +(1.0*lambda-1.0)*log(sqrt_y^2)-0.5*(delta*delta/sqrt_y^2+gamma*gamma*sqrt_y^2);
    return(log_p);
  }

}

data{
  int<lower=0> Nt;    // total number of observations (Ng*Nt*c if balanced)
  int<lower=0> Ng;    // number of groups
  int<lower=0> c;     // number of subgroups
  real y[Nt];         // observations
  int group[Nt];      // group identifier
  int subgroup[Nt];   // subgroup identifier
  int muvar;          // should the relationship between the mean and variance be modelled
}

parameters{

  // MEAN MODEL

  real beta;                              // intercept for the mean model

  // standard-deviation standardised random effects for mean part of the model:
```

```
  matrix[c,Ng] egroup_by_subgroup_star; // matrix of subgroup random effects within groups
  row_vector[Ng] egroup_star;           // vector of group random effects

  real<lower=0> sgroup;                 // standard-deviations of the group effects

  // VARIANCE MODEL (parameterised in terms of log-standard deviations)

  row_vector[2+2*muvar] beta_lsds; // fixed effects for the variance part of the model
                                   // [1] subgroup log-standard-deviation intercept
                                   // [2] residual log-standard-deviation intercept
                                   // if muvar==1
                                   // [3] slope of subgroup log-standard-deviation on mean
                                   // [4] slope of residual log-standard-deviation on mean

  // standard-deviation standardised random effects for variance part of the model:

  matrix[2,Ng] lsds_star;  // matrix of group-specific random effects for the log standard-deviations
                           // Rows are subgroup (Vu) and residual (Ve)

  vector<lower=0>[2] sigma_lsds;  // standard deviations of the group-specific log standard-deviations

  cholesky_factor_corr[2] Lr_lsds; // Cholesky factor of the correlation matrix
                                   // of group-specific log standard-deviations

}

transformed parameters{

 // MEAN MODEL

  vector[Nt] mu;   // linear predictor for mean part of the model

 // unstandardised random effects for mean part of the model:

  row_vector[Ng] egroup;

  matrix[c,Ng] egroup_by_subgroup;

// VARIANCE MODEL

  vector<lower=0>[Nt] SD;   // residual standard deviation for each observation

// unstandardised random effects for the variance part of the model:

  matrix[2,Ng] sds;

  egroup = egroup_star*sgroup;

  sds = diag_pre_multiply(sigma_lsds, Lr_lsds)*lsds_star;

  sds[1,] += beta_lsds[1];
  sds[2,] += beta_lsds[2];
```

```
  // adding the intercept to the log-standard-deviations

  if(muvar==1){
    sds[1,] += beta_lsds[3]*egroup;
    sds[2,] += beta_lsds[4]*egroup;
  }
  // adding a slope (mean-variance relationship) to the log-standard-deviations

  sds = exp(sds);
  // exponentiate log-standard-deviations to get standard-deviations

  // unstandardised random effects in Mean Model whose variance varies over groups:

  for(i in 1:c){
    egroup_by_subgroup[i,] = sds[1,].*egroup_by_subgroup_star[i,];
  }

  for(i in 1:Nt){
    mu[i] = beta + egroup[group[i]] + egroup_by_subgroup[subgroup[i], group[i]];
    SD[i] = sds[2,group[i]];
  }
  // mean and random parts of the model

}
model{

  // MEAN MODEL

  beta ~ normal(0, 10);            // prior distributions for the fixed effects for the mean model

  egroup_star ~ std_normal();
  to_vector(egroup_by_subgroup_star) ~ std_normal();
  to_vector(lsds_star) ~ std_normal();
  // unit-normal prior distributions for the standardised random effects

  sgroup ~ cauchy(0, 5);
  // prior distributions for the standard-deviations of the group effects

  // VARIANCE MODEL

  beta_lsds ~ normal(0, 10);      // prior for the fixed effects for the variance model

  // priors for the variance of the subgroup/residual log standard-deviations
  sigma_lsds[1]  ~ sqrtGIG(1,0.01,sqrt(3.0+9.0/Ng));
  sigma_lsds[2]  ~ sqrtGIG(1,0.01,sqrt(3.0+9.0/Ng));

  // sigma_lsds ~ cauchy(0, 5);  // replaces the GIG prior if half-Cauchy used

  Lr_lsds ~ lkj_corr_cholesky(1); // prior for the correlation matrix
                                  // of group-specific log standard-deviations.

  y ~ normal(mu, SD);
```

```
  // nu ~ gamma(2,0.1);
  // y ~ student_t(nu, mu, SD)
  // An alternative model to y ~ normal(mu, SD) that deals with observation-level heterogeneity.
  // The residuals are assumed to be t-distributed rather than normal.
}

generated quantities{
  matrix [2,2] r_lsds = multiply_lower_tri_self_transpose(Lr_lsds);
  // returning correlation matrices in the model output from the Cholesky factors
}
```

Function for simulating observations from a DHGLM

A function is implemented for simulating data under the DHGLM described above assuming a balanced design. `n` observations are simulated for each of `c` subgroups for each of `Ng` groups. `beta` specifies the overall intercept (mean in this case) of the observations and `sgroup` the standard deviation of the group effects (the Mean Model). `beta_lsds` can be of length two, in which case it specifies the intercept (mean in this case) of the log standard deviations of subgroup effects followed by residual effects. If `beta_lsds` is of length four, then the third and fourth elements specify the slope of the log standard deviations (subgroup and residual respectively) on the group effects from the Mean Model. `C_lsds` is the 2x2 covariance matrix for the two log standard deviations, with subgroup in row/column one, and residual in row/column two. `beta_lsds` and `C_lsds` define the Dispersion Model.

```
sim_DHGLM<-function(Ng, c, n, beta, sgroup, beta_lsds,  C_lsds){

  ########################################################
  # Function for simulating data from a multi-way DHGLM #
  ########################################################

  # Data Structure

  # Ng:         number of groups
  # c:          number of subgroups
  # n:          number of observations within subgroups

  # Mean Model

  # beta:       intercept of the Mean Model
  # sgroup:     standard deviation of group effects

  # Dispersion Model

  # beta_lsds: fixed effects for log(sd) part of the model
  #            [intercepts followed by mean-log(sd) slopes]
  # C_lsds:    covariance matrix of log(sd)'s
  # in beta_lsds and C_lsds, Vu is followed by Ve

  ############################################
  # set up a data-frame for balanced design #
  ############################################

  data <- as.data.frame(matrix(NA,Ng*c*n,3))
  names(data) <- c("group","subgroup", "group_by_subgroup")
```

```
data$group <- rep(1:Ng,c*n)
data$subgroup <- rep(1:c, each = Ng * n)
data$group_by_subgroup <- match(paste(data$group,data$subgroup),
  c(t(outer(unique(data$group), unique(data$subgroup), paste))))
# gets group_by_subgroup indicies with subgroup varying the fastest
# (i.e. with groups 1,2 and subgroups a,b and c,
# group_by_subgroup indicies 1-6 index 1-a, 1-b, 1-c, 2-a, 2-b,2-c)

if(!length(beta)==1){
  stop("beta (intercept) should be of length 1")
}
# check whether beta has the right number of fixed effects [1]]

if(!length(beta_lsds)%in%c(2,4)){
  stop("beta_lsds should be of length 2 (intercepts only)
     or of length 4 (intercepts + slopes on group effects)")
}
# check whether beta_lsds has the right number of fixed effects [2 or 4]]

if(any(!diag(C_lsds)>0)){
  stop("C_lsds should be positive definite")
}
if(abs(C_lsds[2,1]/prod(sqrt(diag(C_lsds))))>1){
  stop("C_lsds should be positive definite")
}

# check whether C_lsds is positive definite

ny<-Ng*n*c

egroup<-rnorm(Ng, 0, sgroup)
# simulate group random effects for mean part of the model

beta_muvar<-matrix(0,1,2)
if(length(beta_lsds)==4){
  beta_muvar[c(1,2)]<-beta_lsds[c(3,4)]
}
# Slopes for the mean-log(sd) relationship organised into matrix form (1 x 2)
# When premutiplied by the group effects (Ng x 2) it gives the predicted log(sd) given the mean

mu_lsds<-t(beta_lsds[1:2]+t(egroup%*%beta_muvar))
# matrix (Ng x 2) of predicted log(sd)s form fixed effects

sds<-exp(mu_lsds+mvrnorm(Ng, rep(0,2), C_lsds))
# matrix (Ng x 2) of sds

egroup_by_subgroup = matrix(rnorm(Ng*c, 0, sds[,1]),Ng,c)
# simulate matrix of group by subgroup effects

mu<-beta+egroup[data$group]+t(egroup_by_subgroup)[data$group_by_subgroup]
# combine fixed and random effects

data$y<-rnorm(ny, mu, sds[data$group,2])
```

```
  # simulate observations conditional on linear predictors (fixed+random) and group specific Ve.

  return(data)
}
```

Simulate data for a DHGLM and fit Stan model

Below, we simulate data, fit the DHGLM in Stan and plot the MCMC chains. Running multiple chains for longer would be advisable.

```
Ng <- 1000 # number of groups
c <- 4     # number of subgroups
n <- 5     # number of observations within subgroups

beta <-10                 # intercept of  Mean Model
sgroup<-1                 # between-group standard-deviation
beta_lsds<-c(-1, 0)       # intercepts, no mean-variance relationship.

sigma_lsds <- c(0.8,0.9)  # standard deviations of subgroup and residual log standard-deviations
r_lsds<-0.3               # correlation between subgroup and residual log standard-deviations

C_lsds<-matrix(r_lsds*prod(sigma_lsds), 2, 2)
diag(C_lsds)<-sigma_lsds^2 # (co)varinaces for subgroup and residual log standard-deviations

sim_data<-sim_DHGLM(Ng=Ng, c=c, n=n, beta=beta, sgroup=sgroup, beta_lsds=beta_lsds,  C_lsds=C_lsds)
# simulate data

sim_stan<-list(
            Nt=Ng*c*n,
            Ng=Ng,
            c=c,
            muvar=0,
            y=sim_data$y,
            group=sim_data$group,
            subgroup=sim_data$subgroup
          )
# stan list

model_output<-sampling(DHGLM_stan, data = sim_stan, chains = 1, refresh=-1)
# fit model

pars<-c("beta","beta_lsds[1]","beta_lsds[2]","r_lsds[1,2]","sgroup","sigma_lsds[1]","sigma_lsds[2]")
# parameters to plot

post<-mcmc(as.data.frame(model_output)[pars])
plot(post) # plot MCMC trace and density plot
```

Sampling designs

Given that DHGLM estimates of the (co)variance of variances cannot be obtained analytically, it seems unlikely that exact expressions for how power changes with sampling design and effort could be found. Instead, we simulate data sets of 3200 observations under the set of parameters defined above. We simulate 10 data

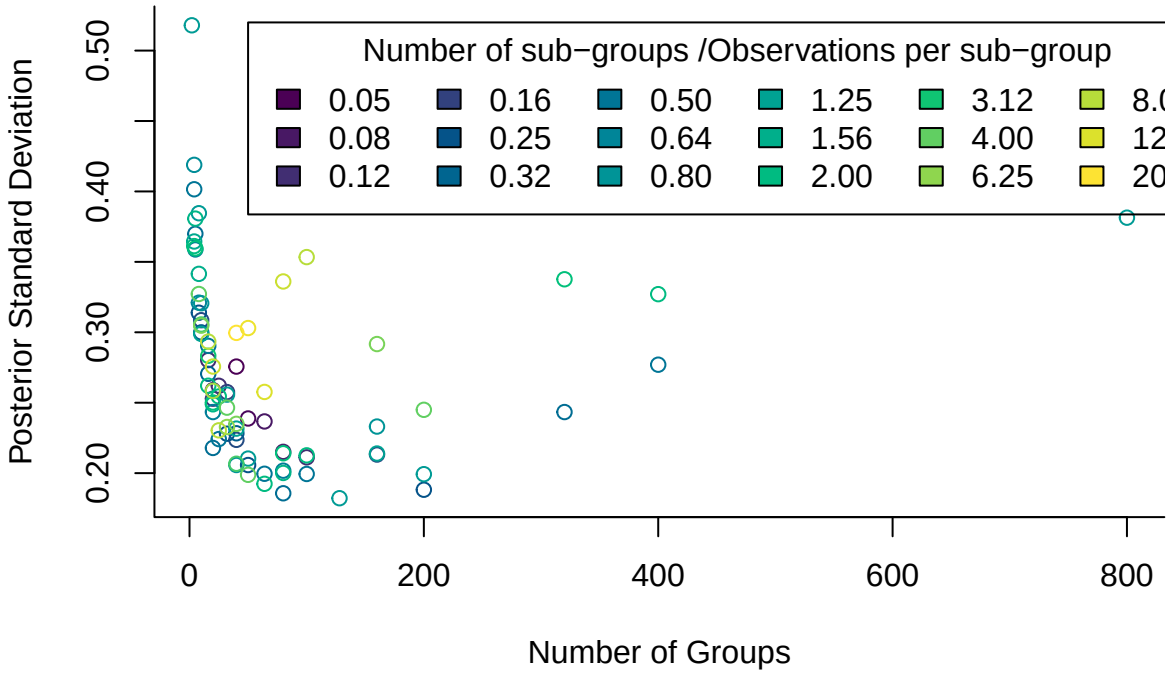
sets for each of the 83 possible designs where `n` and `c` range between 2 and 40, analyse them using `DHGLM_stan` and store the posterior standard deviation of the correlation of the log-scale variances.

```
design_obs<-expand.grid(2:40,2:40)
# generate all combinations of n and c for values ranging from 2 to 40.

design_obs<-design_obs[which(3200%%apply(design_obs,1, prod)==0),]
# save combinations where 3200/(nc)=Ng is integer

design_obs<-cbind(3200/apply(design_obs, 1, prod), design_obs, NA, NA)
# add Ng to design_obs and columns for storing the posterior mean and sd.

colnames(design_obs)<-c("Ng", "n", "c", "post.mean", "post.sd")


design_obs<-design_obs[rep(1:nrow(design_obs),10),]
# duplicate design_obs 10X.

for(i in 1:nrow(design_obs)){
# iterate through designs

  Ng<-design_obs[i,"Ng"]
  c<-design_obs[i,"c"]
  n<-design_obs[i,"n"]
  N<-n*c

  sim_data<-sim_DHGLM(Ng=Ng, c=c, n=n, beta=beta, sgroup=sgroup, beta_lsds=beta_lsds, C_lsds=C_lsds)
  # simulate data

  sim_stan<-list(
            Nt=Ng*c*n,
            Ng=Ng,
            c=c,
            muvar=0,
            y=sim_data$y,
            group=sim_data$group,
            subgroup=sim_data$subgroup)
  # format data for stan

  model_output<-sampling(DHGLM_stan, data = sim_stan, chains = 1, iter = 5000, refresh=-1)
  # fit model

  design_obs$post.mean[i]<-mean(model_output@sim$samples[[1]]["r_lsds[1,2]"][[1]])
  design_obs$post.sd[i]<-sd(model_output@sim$samples[[1]]["r_lsds[1,2]"][[1]])
  # store posterior mean and standard deviation of the log-scale correlation between variances
  print(i)
}
```

We take the average posterior standard deviation (averaged over the 10 data sets for each design) and show how it varies according to the number of groups (N_g) and how replication within a group is partitioned within subgroup and between subgroups (n/c).

```
design_obs_means<-design_obs %>%  group_by(Ng, c, n) %>%
  summarise(
    post.sd = mean(post.sd), post.mean = mean(post.mean)
```

```
)

plot(design_obs_means$post.sd~design_obs_means$Ng, type="n", ylab="Posterior Standard Deviation", xlab="

design_obs_means$cn.ratio<-design_obs_means$c/design_obs_means$n

col_fac<-sort(unique(design_obs_means$cn.ratio))
design_obs_means$col_fac<-match(design_obs_means$cn.ratio, col_fac)

for(i in 1:length(col_fac)){
  points(design_obs_means$post.sd[which(design_obs_means$col_fac==i)]~design_obs_means$Ng[which(design_

}
legend(50, 0.52, legend=formatC(round(col_fac[seq(1, length(col_fac), 2)],2),2, format="f"), fill=hcl.co
```



The optimal design has a modest number of observations within each group ($n = c = 5$) but the number of groups is large ($N_g = 128$). Although many designs have comparable precision, ensuring the number of groups is at least as large as the number of observations per group seems warranted. When deciding how observations are partitioned within a group it seems best to keep n and c roughly comparable, or to slightly favour n over c. The leading designs are:

```
head(design_obs_means[order(design_obs_means$post.sd),1:4])
```

```
## # A tibble: 6 x 4
## # Groups:   Ng, c [6]
##      Ng     c     n post.sd
##   <dbl> <int> <int>   <dbl>
```

```
## 1   128     5     5   0.182
## 2    80     4    10   0.186
## 3   200     2     8   0.188
## 4    64    10     5   0.192
## 5    50    16     4   0.199
## 6   200     4     4   0.199
```

The best design is likely to depend on the true underlying parameter values, and we advocate rerunning these simulations before designing the experiment if it is believed the true underlying parameter values are likely to deviate from those used.

Gardini, Aldo, Carlo Trivisano, and Enrico Fabrizi. 2021. "Bayesian Analysis of ANOVA and Mixed Models on the Log-Transformed Response Variable." *Psychometrika* 86 (2): 619–41. https://doi.org/10.1007/s11336-021-09769-y.

Juárez, Miguel A, and Mark FJ Steel. 2010. "Model-Based Clustering of Non-Gaussian Panel Data Based on Skew-t Distributions." *Journal of Business & Economic Statistics* 28 (1): 52–66.

Wagner, Günter P. 2023. "Models of contingent evolvability suggest dynamical instabilities in body shape evolution." In *Evolvability: A Unifying Concept in Evolutionary Biology?*, edited by Thomas F Hansen, David Houle, Mihaela Pavlicev, and Christophe Pélabon, 199–219. MIT Press.