Journal of Statistics Applications & Probability
An International Journal

http://dx.doi.org/10.18576/jsap/090301

Bayesian and Frequentist Approaches Robustness in Variance Components Estimation

Hamdy F. F. Mahmoud[1,2]

[1]Department of Statistics, Virginia Polytechnic Institute and State University, Blacksburg, VA 24061, USA
[2]Department of Statistics, Mathematics and Insurance, Faculty of Commerce, Assiut University, Assiut, Egypt

Received: 6 Feb. 2020, Revised: 22 Aug. 2020, Accepted: 6 Sep. 2020
Published online: 1 Nov. 2020

Abstract: Variance components and functions thereof are important in many fields such as industry, agriculture, quantitative genetics and animal breeding. This paper contributes to evaluating the variance components estimation approaches and assessing their robustness to outliers. Using an intensive simulation study and under different settings, it was found that researchers can decide which method of estimation is appropriate to their study to estimate the variance components based on the source of outliers (error term, ε_{ij} or random effects, α_i), the variance components ratio, and the sample size.

Keywords: ANOVA, EM algorithm, Gibbs sampler, MLE, outliers, Variance components

1 Introduction

Variance components and functions thereof are important in many fields such as industry, quantitative genetics, agriculture and animal breeding. There are two main alternative approaches for estimating variance components: frequentist and Bayesian-based approaches. Frequentist approaches include such as analysis of variance [1], maximum likelihood estimate [2], expectation maximization algorithm [3], restricted maximum likelihood estimate [4] and minimum variance quadratic unbiased estimator [5–7]. These frequentist approaches (ANOVA, MLE, EM, RMLE and MVQUE) consider variance components as fixed but with unknown real values or vectors. On the other hand, Bayesian approach estimators, such as Gibbs sampler [8], consider that variance components are random variables with some prior knowledge of their distributions. This prior knowledge is given by a prior distribution or by data from a previous experiment. The prior distribution and the likelihood function of the sample are combined resulting in a posterior distribution. To the best of my knowledge, there are no known criteria for a researcher to choose a variance components estimation approach from these alternatives.

Many articles have studied the frequentist methods: [9] studied the properties of ANOVA, MLE, RMLE and MINQUE, such as unbiasedness property, [10] studied EM algorithm for estimating the generalized linear mixed model by relaxing the normality assumption of random effects, and [11] applied ANOVA, MINQUE, MLE and Gibbs sampling on a numerical example in animal breeding. Other articles studied the performance of prior distributions within Gibbs sampler: [12] studied sampling-based approaches such as Gibbs sampling with an inverted gamma prior for estimating the variance components in terms of how much the method is straightforward and easy to be implemented, [13] studied the noninformative prior properties for variance components, [14] studied the effect of using a flat prior on the posterior distribution, [15] created a flexible software for Bayesian inference, [16] and [17] introduced a half-t family prior and compared it to a noninformative and inverted gamma prior in hierarchical models, [18] argued that the half-Cauchy prior should be replaced with an inverse Gamma prior in Bayesian hierarchical models, [19] studied Cauchy prior in logistic regression, and [20] studied Bayesian estimation and likelihood-based estimators (MLE, RMLE, marginal and penalized quasi-likelihood) in terms of bias and convergence. Despite the extensive literature available, there are still some gaps, particularly, on these points: (1) There are no metrics for evaluating the performance of the two approaches in estimating variance components when outliers are present; (2) comparisons of the frequentist approach and Bayesian approach are

* Corresponding author e-mail: ehamdy@vt.edu

very limited; and (3) in these papers, a case study is considered to compare the methods within each approach and no intensive simulation study has been considered under different scenarios of settings, so intensive simulation studies under different scenarios are needed. In this paper, the two main approaches are evaluated and compared using an intensive simulation study under different settings in two cases: when data have outliers and when there are no outliers. The mean square error (MSE), absolute bias and other statistical measures will be used to evaluate the different methods of estimating variance components. In addition, this paper studies the sensitivity of these approaches for outliers. This paper contributes to evaluating the variance components estimation approaches and assessing their robustness for outliers. The motivation behind this study is to: (1) compare the Bayesian and frequentist approaches using an intensive simulation study at different scenarios, and metrics are presented which can be used to decide when to use these approaches, in a wide range of the variance components ratio ($\frac{\sigma_\alpha}{\sigma_\varepsilon}$); (2) evaluate the two approaches in terms of robustness to outliers as well as when there are no outliers; and (3) finding a criteria based on which a researcher can choose which method should be used. To compare between the frequentist approach and the Bayesian approach in robustness, the one-way balanced random effect model is used.

1.1 The One-Way Statistical Model

In the one-way random effect model with balanced data, there are groups or levels with n observations, $y_i j$ within each level. This model takes the following form:

$$y_{ij} = \mu + \alpha_i + \varepsilon_{ij},\ i = 1,2,\ldots,a,\ j = 1,2,\ldots,n, \tag{1}$$

where μ is the overall mean, α_i and ε_{ij} are unobservable independent random variables with distributions $N(0,\sigma_\alpha^2)$ and $N(0,\sigma_\varepsilon^2)$, respectively. It follows that y_{ij}'s have a joint normal distribution with mean μ and covariance matrix given, by

$$Cov(y_{ij}, y_{kl}) = \begin{cases} \sigma_\alpha^2 + \sigma_\varepsilon^2, i = k, j = l \\ \sigma_\alpha^2, i = k, j \neq l \\ 0, i \neq k. \end{cases}$$

The parameters of this model are μ, σ_α^2 and σ_ε^2, the last two are called variance components.

This model is extensively analyzed in Chapter 3 of [9]. In this paper, it is used to compare between THE frequentist and Bayesian approaches in terms of robustness for outliers. The rest of the paper is outlined as follows. In Section 2, the variance components estimation approached are presented. In Section 3, a Monte Carlo study under different settings is conducted to evaluate the performance of the estimators. Real data application is studied in Section 4. Section 5 contains conclusion and recommendations.

2 Variance Components Estimation Approaches

The Bayesian approach is subjective and uses prior beliefs to define a prior probability distribution on possible values of the population parameters. The frequentist approach uses the likelihood function to do a parametric inference.

2.1 Bayesian Approach

Gibbs sampling or Gibbs sampler estimator is one of the most common estimation approaches in the Bayesian literature. It is a Markov Chain Monte Carlo (MCMC) algorithm for obtaining a sequence of observations which are approximated from a specified probability distribution, when direct sampling is difficult. For model (1.1),

$$y|\mu, \alpha_1, \alpha_2, \ldots, \alpha_a, \sigma_\varepsilon^2 \sim N(\mu + \alpha_i, \sigma_\varepsilon^2) \text{ and} \tag{2}$$

$$\alpha_i|\sigma_\alpha^2 \sim N(0, \sigma_\alpha^2). \tag{3}$$

The joint posterior distribution of all parameters in the model, given the data, can be expressed by Bayes theorem as follows:

$$f(\mu, \boldsymbol{\alpha}, \sigma_\alpha^2, \sigma_\varepsilon^2|y) \propto f(y|\mu, \boldsymbol{\alpha}, \sigma_\alpha^2, \sigma_\varepsilon^2) f(\boldsymbol{\alpha}|\sigma_\alpha^2) f(\mu) f(\sigma_\alpha^2) f(\sigma_\varepsilon^2), \tag{4}$$

Table 1: Common priors for the model parameters

Paramter	Prior
μ	$\mathrm{N}(0,c)$
	$\mu \propto 1$
σ^2	$\sigma \propto 1$
	Jeffery's prior
	inverse-gamma(δ,δ)
	Unif(0, c)
	Half-Cauchy

c is a suitable big value, and δ is a small value.

where $f(\mu)$, $f(\sigma_\alpha^2)$ and $f(\sigma_\varepsilon^2)$ are the prior distributions for μ, σ_α^2 and σ_ε^2, respectively. There are many recommended priors for these model parameters. Table 1 shows the common priors for these parameters. [21] introduced data-level variance prior distributions for the variance components. [17] suggested a uniform prior for σ_α, U(0, c), where $c \to \infty$, [15] suggested an inverse-gamma(δ,δ) prior distribution with δ=0.001, and [16] recommended using Cauchy prior distribution as a default weakly informative choice for the variance components. [18] and [19] studied Cauchy prior distribution in more details. In this paper, the last three priors are considered. Given the posterior distribution function (2.3), one can obtain the conditional distributions of the parameters as follows:

$$f(\sigma_\alpha^2|\mu,\alpha_i,\sigma_\varepsilon^2,y), \tag{5}$$

$$f(\sigma_\varepsilon^2|\mu,\alpha_i,\sigma_\alpha^2,y), \tag{6}$$

$$f(\alpha_i|\mu,\sigma_\varepsilon^2,\sigma_\alpha^2,y), i=1,2,\ldots,a \tag{7}$$

and

$$f(\mu|\alpha_i,\sigma_\alpha^2,\sigma_\varepsilon^2,y). \tag{8}$$

Although we are interested in σ_ε^2 and σ_α^2, all the full conditional distributions are needed to run the Gibbs sampler. The steps are as follows:

1. Set arbitrary initial values for μ, $\boldsymbol{\alpha} = (\alpha_1,\ldots,\alpha_a^T)$, σ_ε^2 and σ_α^2;
2. sample σ_α^2 from (2.4) and update σ_α^2;
3. sample σ_ε^2 from (2.5) and update σ_ε^2;
4. sample α_i from (2.6) and update α_i;
5. sample μ from (2.7) and update μ;
6. repeat [2]-[5] big number of times, N, till convergence is achieved.

When convergence is achieved, the points from Nth iteration are sample points from the appropriate marginal distributions. Because our interest is in making inferences about σ_ε^2 and σ_α^2, no attention will be paid hereafter to α_i and μ.

2.2 Frequentist Approach

There are many methods under frequent approach, in this paper three estimators are considered and compared to the Bayesian approach using Gibbs sampler under various prior distributions. The three methods are: analysis of variance (ANOVA), maximum likelihood estimator (MLE), and expectation maximization algorithm (EM).

2.2.1 Analysis of Variance (ANOVA)

ANOVA method is based on the law of total variance, where the total variability in a response variable is partitioned into components attributable to different sources of variation. The method of moments is used to estimate the variance components by setting the mean square values to the expected mean square. These estimators are unbiased estimators for the variance components. Consider the sums of squares of the ANOVA in Table 2 for model (1.1):
$SSA = n\sum_{i=1}^{a}(\bar{y}_{i.}-\bar{y}_{..})^2$, $SSE = \sum_{i=1}^{a}\sum_{j=1}^{n}(\bar{y}_{ij}-\bar{y}_{i.})^2$ and $SST = \sum_{i=1}^{a}\sum_{j=1}^{n}(\bar{y}_{ij}-\bar{y}_{..})^2$,
where $\bar{y}_{i.}$ is the mean of the observations in group i, and $\bar{y}_{..}$ is the overall sample mean. The expectation of the first two

Table 2: ANOVA table for the one-way random model

Source	df	Sum of squares (SS)	Mean sum of squares (MSS)
Model	$a-1$	SSA	MSA=SSA/$(a-1)$
Error	$a(n-1)$	SSE	MSE=SSE/$a(n-1)$
Total	$an-1$	SST	

sums of squares are: $E(SSA) = n(a-1)\sigma_\alpha^2 + (a-1)\sigma_\varepsilon^2$, and $E(SSE) = a(n-1)\sigma_\varepsilon^2$. So ANOVA estimates of σ_α^2 and σ_ε^2 are obtained by matching the sums of the squares to their expectations.

$$\hat{\sigma}_\varepsilon^2 = \frac{SSE}{a(n-1)} \text{ and } \hat{\sigma}_\alpha^2 = \frac{1}{n}[\frac{SSA}{a-1} - \hat{\sigma}_\varepsilon^2]. \tag{9}$$

2.2.2 Maximum Likelihood Estimate (MLE)

The maximum likelihood method of estimation was developed by [1], and first time applied to variance components estimation was by [22] and [23]. Under normality assumption of the error and random effects terms, the variance components can be obtained by maximizing twice the negative log-likelihood function. In terms of ANOVA table sum of squares, the MLEs for this model are simply the solutions of the maximum likelihood estimating equations as follows:

$$\hat{\sigma}_\varepsilon^2 = \frac{SSE}{a(n-1)} \text{ and } \hat{\sigma}_\alpha^2 = \frac{1}{n}[(1-\frac{1}{a})\frac{SSA}{a-1} - \hat{\sigma}_\varepsilon^2]. \tag{10}$$

From the formulas of estimating the variance components using ANOVA and MLE, one can see that they are equal when the number of random effect levels $a \to \infty$, otherwise, the MLE estimate of the variance component of random effect, σ_α^2, is smaller than ANOVA estimate for the variance of the random effects.

2.2.3 Expectation Maximization (EM)

In the EM algorithm, α_i $(i = 1, 2, \ldots, a)$ are treated as missing (or latent) data and $\mathbf{y}$ is treated as observed (or incomplete) data. Hence, we can have "complete" data, which is $\{\mathbf{y}, \boldsymbol{\alpha}\}$. EM algorithm has steps: expectation and maximization. Let $\tau = \sigma_\alpha^2/\sigma_\varepsilon^2$, EM algorithm steps for estimating the variance components for the model (1.1) are as follows:

0-Step: Obtain starting values $\sigma_\alpha^{2(0)}$, $\sigma_\varepsilon^{2(0)}$ and $\mu^{(0)}$,
E-Step: Calculate
$E(\alpha_i^{'}\alpha_i|\mathbf{y}) = \tau^2 \sum_{i=1}^a \frac{(\bar{y}_{i.}-\mu)^2}{(\tau+1/n)^2} + a\sigma_\alpha^2 - \sigma_\alpha^2 \tau \sum_{i=1}^a \frac{1}{(\tau+\frac{1}{n})}$ and
$E(\varepsilon_i^{'}\varepsilon_i|\mathbf{y}) = \sum_{i=1}^a \sum_{j=1}^n (y_{ij}-\mu)^2 - \tau \sum_{i=1}^a \frac{(\bar{y}_{i.}-\mu)^2}{(\tau+\frac{1}{n})^2}(2+n\tau) + \sigma_\varepsilon^2 \sum_{i=1}^a \frac{\tau}{(\tau+\frac{1}{n})}$.
M-Step: Calculate
$t_1 = a\sigma_\alpha^{2(m+1)} = \tau^{2(m)} \sum_{i=1}^a \frac{(\bar{y}_{i.}-\mu^{(m)})^2}{(\tau^{(m)}+1/n)^2} + a\sigma_\alpha^{2(m)} - \frac{a\sigma_\alpha^{2(m)}\tau^{(m)}}{(\tau^{(m)}+\frac{1}{n})}$,
$t_0 = an\sigma_\varepsilon^{2(m+1)} = \sum_{i=1}^a \sum_{j=1}^n (y_{ij}-\mu^{(m)})^2 + \frac{a\tau^{(m)}\sigma_\varepsilon^{2(m)}}{(\tau^{(m)}+\frac{1}{n})} - \frac{\tau^{(m)}(2+n\tau^{(m)})}{(\tau^{(m)}+\frac{1}{n})^2} \sum_{i=1}^a (\bar{y}_{i.}-\mu^{(m)})^2$
and $\mu^{m+1} = \frac{\bar{y}_{..}}{n\tau^{(m)}+1} + (1 - \frac{1}{n\tau^{(m)}+1})\mu^{(m)}$.
I-Step: Iterate between E-Step and M-Step till convergence happens, and when convergence is achieved, declare $\hat{\sigma}_\alpha^{2(m+t)}$ and $\hat{\sigma}_\varepsilon^{2(m+t)}$ as EM algorithm estimates.

3 Monte Carlo Study

In order to study the performance of the estimates of frequentist approach estimators and Gibbs sampler under different prior distributions, an intensive simulation study is performed. The data have been generated from the model (1.1) under different number of levels of a and n ($a = 20$, and $n = 10$, $a = 10$ and $n = 20$ and $a = 10$ and $n = 4$), and different values of σ_α within interval (0.1, 10) with a grid search of 0.2 and different values of σ_α within interval (10, 0.1) with a grid search

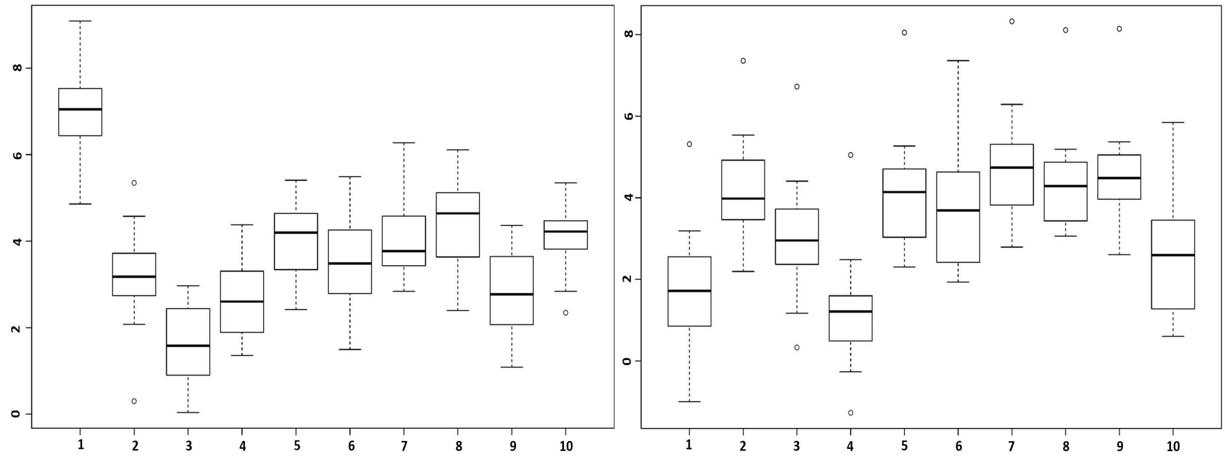

Fig. 1: Boxplots of two random data sets: one has 5% outliers in α_i (left) and one has 5% outliers in ε_{ij} (right) at $\sigma_\alpha = 1$, $\sigma_\varepsilon = 1$, $\mu = 3$, $n = 20$ and $a = 10$

of 0.2. So, the ratio $(\frac{\sigma_\alpha}{\sigma_\varepsilon})$ will be between 0.01 and 100. The grand mean μ is set to be 3. In uniform prior distribution, the range is set to be (0, 100), inverse-gamma (δ, δ) prior distribution is set with $\delta = 0.001$ as recommended in [15], and for half-Cauchy prior, the scale is set to be 25, a weakly informative prior distribution as recommended in [16]. Each setting is replicated 1000 times at each value of the ratio $\frac{\sigma_\alpha}{\sigma_\varepsilon}$ and the variance components are estimated. In Gibbs sampler, number of iterations is set to be 30000, and burn in is 1000. Two types of outlier contamination were considered: factor contamination and error contamination. In the first case, 5 % of the random effects α_i $(i = 1, 2, \ldots, a)$ were replaced by a constant outlier and in the second case, 5 % of the error term ε_{ij} (corresponding to different groups) were replaced by a constant outlier. Therefore, we analyzed three different situations:

Case 1: uncontaminated data; no outliers in the simulated data

Case 2: there are 5 % outliers in α_i (α_1 is replaced by 4)

Case 3: there are 5 % outliers in ε_{ij} (ε_{i1} is replaced by 4)

Figure 1 (left) shows a simulated data having 5 % outliers in α_i, and Figure 1 (right) shows a simulated data having 5 % outliers in ε_{ij}.

3.1 Case 1: No Outliers

Figure 2 (a, b), shows that the absolute bias has the same form of MSE that is a function of the ratio of $(\frac{\sigma_\alpha}{\sigma_\varepsilon})$ as well as a function of the sample size. A close look at the form when the ratio is between 0 and 4, Figure 2 (c) shows that when the ratio is small, ANOVA estimates are better than MLE and EM estimates in terms of bias for σ_α. Also, it shows that the difference between the estimates gets smaller when the ratio gets bigger, and the relationship depends on the sample size. The ratio at which, the three methods have the same performance depends on the sample size, the smaller the sample size, the larger the ratio; when n = 20 and a = 10, the ratio is about 0.5, and when n = 4 and a = 10, the ratio is about 1. Figure 2 (d) reveals that the absolute bias for σ_ε monotone decreases when the ratio increases.

For the three frequentist estimators, Figure 3 (left) shows that MSE of σ_α is a dependent on the ratio of $(\frac{\sigma_\alpha}{\sigma_\varepsilon})$ as well as a function of the sample size. When the ratio is small, MSE is small, and when the ratio gets higher, MSE gets higher till the ratio value reaches a high value, approximately 20, it gets stable. The same pattern can be seen for different sample sizes, but the higher sample size, the smaller the MSE. Figure 3 (right) shows that MSE of σ_ε is a monotone decreasing by the ratio, and it reaches around zero value when the ratio reaches 10.

For Gibbs sampler under different prior distributions, at n = 20 and a = 10, Figure 4 (left) shows MSE of σ_α as a function of the ratio. One can see that when the ratio is small, using inverse gamma prior is not a good choice compared to uniform and half-Cauchy priors, but when the ratio is big, inverse gamma is a good choice compared to the other two prior distributions. The change point is around ratio of 1. Figure 4 (right) of σ_ε shows that inverse-gamma prior is better in case of a smaller ratio, but when the ratio reaches about 1, the performance of all priors is comparable.

Figure 5 (left) and 5 (right) show results of the absolute bias of σ_α and σ_ε as a function of the ratio, respectively. They reveal the same form of relationships, inverse gamma as a prior is not a good choice compared to uniform and half-Cauchy

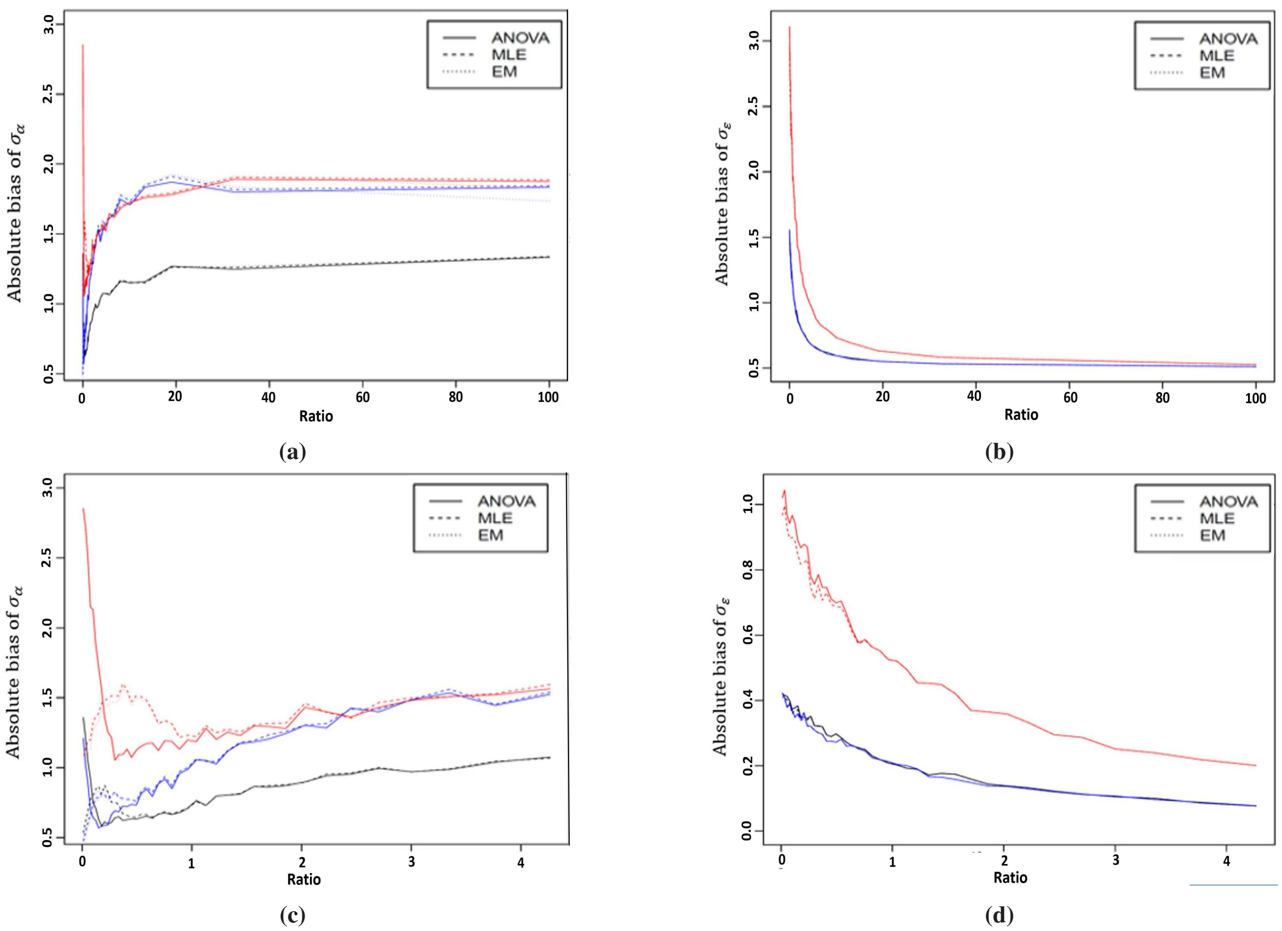

Fig. 2: Absolute bias for σ_α (left panel) and σ_ε (right panel) at different ratios of variance components ($\frac{\sigma_\alpha}{\sigma_\varepsilon}$): black is for $n = 20$ and $a = 10$, blue is for $n = 10$ and $a = 20$, and red is for $n = 10$ and $a = 4$

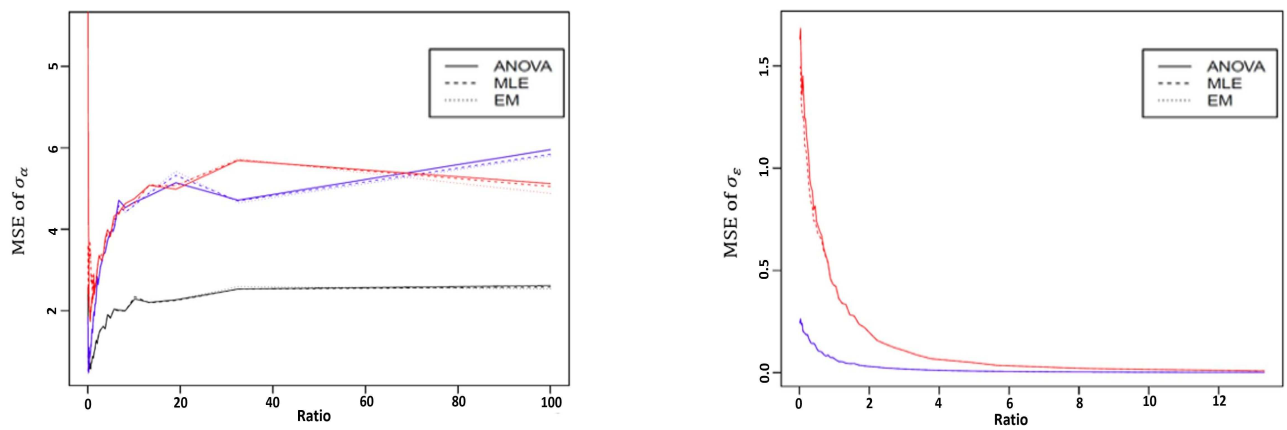

Fig. 3: Mean square error (MSE) for σ_α (left) and σ_ε (right) at different ratios of variance components ($\frac{\sigma_\alpha}{\sigma_\varepsilon}$): black is for $n = 20$ and $a = 10$, blue is for $n = 10$ and $a = 20$, and red is for $n = 10$ and $a = 4$

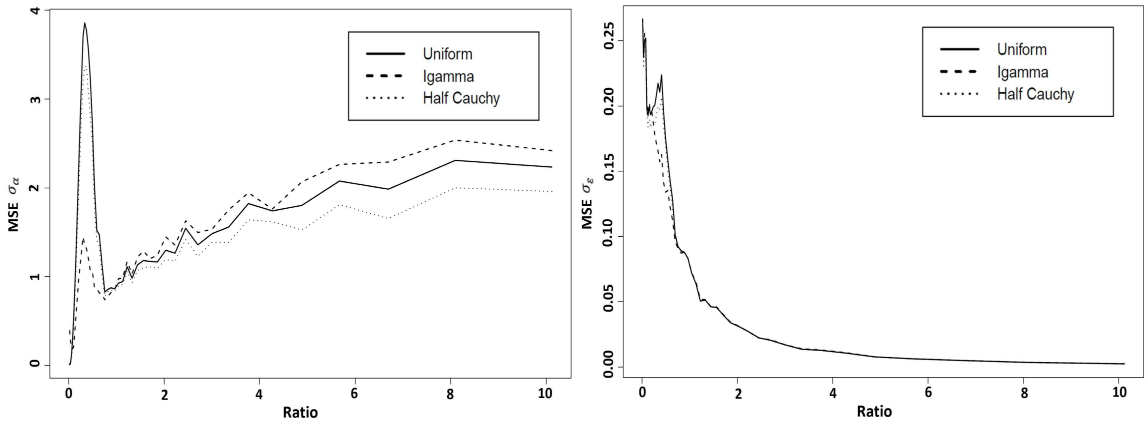

Fig. 4: Mean square error (MSE) for σ_α (left) and σ_ε (right) at different ratios of variance components ($\frac{\sigma_\alpha}{\sigma_\varepsilon}$), for Gibbs sampler using different priors at $n = 20$ and $a = 10$

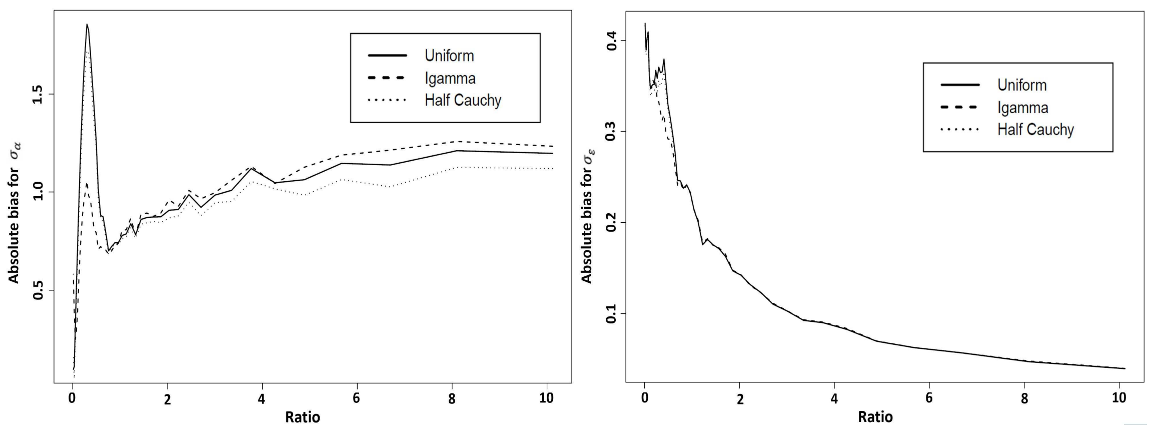

Fig. 5: Absolute bias for σ_α (left) and σ_ε (right) at different ratios of variance components ($\frac{\sigma_\alpha}{\sigma_\varepsilon}$), for Gibbs sampler using different priors at $n = 20$ and $a = 10$

priors when the ratio is small, but when the ratio is big, inverse gamma is a good choice compared to the other two prior distributions.

For a smaller sample size, n = 10 and a = 4, Figure 6 (left) and 7 (left) show the form of MSE and absolute bias of σ_α with the ratio ($\frac{\sigma_\alpha}{\sigma_\varepsilon}$). They show the same forms in Figure 4 (left) and 5 (left), except that the gap between estimates is higher when the ratio is smaller than 2 ($\frac{\sigma_\alpha}{\sigma_\varepsilon} \approx 2$) because the total sample size is smaller (n = 10 and a = 4) compared to n = 20 and a = 10. In addition, MSE and absolute bias of σ_ε also have the same forms in Figure 4 (right) and 5 (right), they reach the same performance when the ratio reaches 2.

Table 3 displays summary results of a selected setting, when the ratio is equal to 1 ($\frac{\sigma_\alpha=1}{\sigma_\varepsilon=1} = 1$), n=10 and a=20. The results confirm the previous findings where ANOVA is better among frequentist approaches and inverse gamma is a good prior choice that is based on MSE and bias criteria. Table 4 shows the results at a smaller ratio ($\frac{\sigma_\alpha=0.5}{\sigma_\varepsilon=1} = 0.5$) and the same sample size. It reveals the same findings with a bigger gap between the estimates. In sum, when there are no outliers in a real data set, and the ratio of the variance components is small, researchers should use inverse gamma prior in Gibbs sampler or ANOVA, and if the ratio is big, they should use half-Cauchy within Gibbs sampler.

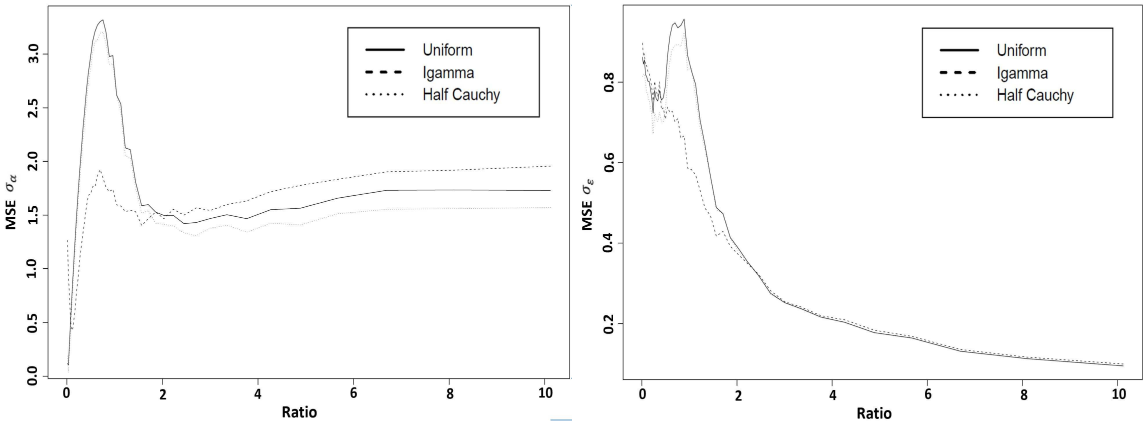

Fig. 6: Mean square error (MSE) for σ_α (left) and σ_ε (right) at different ratios of variance components ($\frac{\sigma_\alpha}{\sigma_\varepsilon}$), for Gibbs sampler using different priors at $n = 10$ and $a = 4$

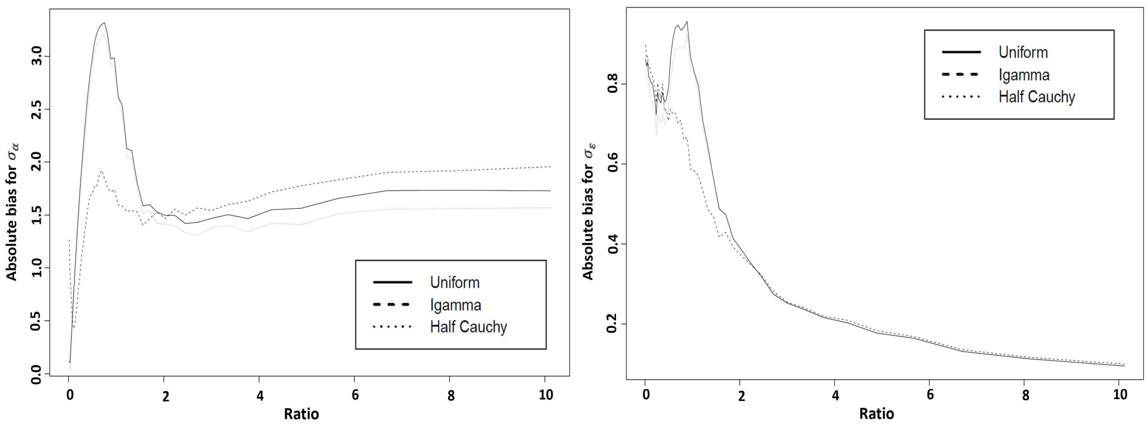

Fig. 7: Absolute bias for σ_α (left) and σ_ε (right) at different ratios of variance components ($\frac{\sigma_\alpha}{\sigma_\varepsilon}$), for Gibbs sampler using different priors at $n = 10$ and $a = 4$

Table 3: Summary results of 1000 simulated data sets without outliers for the frequentist approaches and Gibbs sampler with different prior distributions. $\sigma_\alpha = 1$, $\sigma_\varepsilon = 1$, so the ratio =1 at $n = 20$ and $a = 10$.

Approach			Mean ± SE	Median	75 percentile	Bias	MSE
	ANOVA	σ_ε	0.999± 0.001	0.997	1.036	0.0003	0.003
		σ_α	0.983 ± 0.006	0.976	1.103	-0.0165	0.032
Friquentist	EM	σ_ε	0.999± 0.001	0.997	1.036	-0.0003	0.003
		σ_α	0.955± 0.006	0.949	1.073	-0.0441	0.032
	MLE	σ_ε	0.999± 0.001	0.997	1.036	-0.0003	0.003
		σ_α	0.955± 0.006	0.949	1.073	-0.0441	0.032
	Uniform	σ_ε	0.998±0.002	0.997	1.032	-0.0016	0.003
		σ_α	0.949±0.008	0.961	1.062	-0.0501	0.034
Bayesian	IGamma	σ_ε	1.003± 0.002	1.002	1.036	0.0030	0.003
		σ_α	1.015± 0.008	1.025	1.131	0.0156	0.034
	H-Cauchy	σ_ε	0.998± 0.002	0.997	1.031	-0.0015	0.003
		σ_α	0.950±0.002	0.960	1.060	-0.0496	0.034

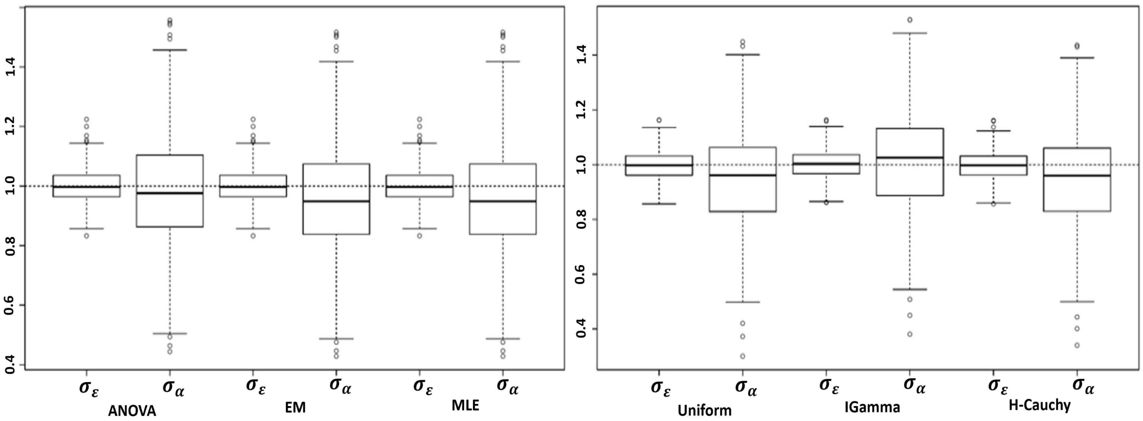

Fig. 8: Boxplot of 1000 estimates of σ_α and σ_ε at both values equal to 1 ($\frac{\sigma_\alpha=1}{\sigma_\varepsilon=1}=1$) for frequentist approaches (left) and Gibbs sampler (right) at $n=20$ and $a=10$

Table 4: Summary results of 1000 simulated data sets without outliers for the friquentist approaches and Gibbs sampler with different prior distributions with the setting of $\sigma_\alpha=0.5$ and $\sigma_\varepsilon=1$, so the ratio = 0.5 at $n=20$ and $a=10$

Approach			Mean ± SE	Median	75 percentile	Bias	MSE
	ANOVA	σ_ε	0.997±0.002	0.997	1.032	-0.002	0.003
		σ_α	0.488± 0.004	0.487	0.563	-0.011	0.013
Friquentist	EM	σ_ε	0.997± 0.002	0.997	1.032	-0.002	0.003
		σ_α	0.470± 0.004	0.469	0.545	-0.029	0.013
	MLE	σ_ε	0.997± 0.002	0.997	1.032	-0.002	0.003
		σ_α	0.470± 0.004	0.469	0.545	-0.029	0.013
	Uniform	σ_ε	1.007±0.002	1.003	1.039	0.007	0.002
		σ_α	0.426±0.006	0.434	0.526	-0.073	0.006
Bayesian	IGamma	σ_ε	1.007±0.002	1.004	1.039	0.007	0.003
		σ_α	0.487±0.006	0.495	0.578	0.008	0.003
	H-Cauchy	σ_ε	1.004± 0.002	1.001	1.037	0.004	0.003
		σ_α	0.444±0.005	0.445	0.528	-0.055	0.018

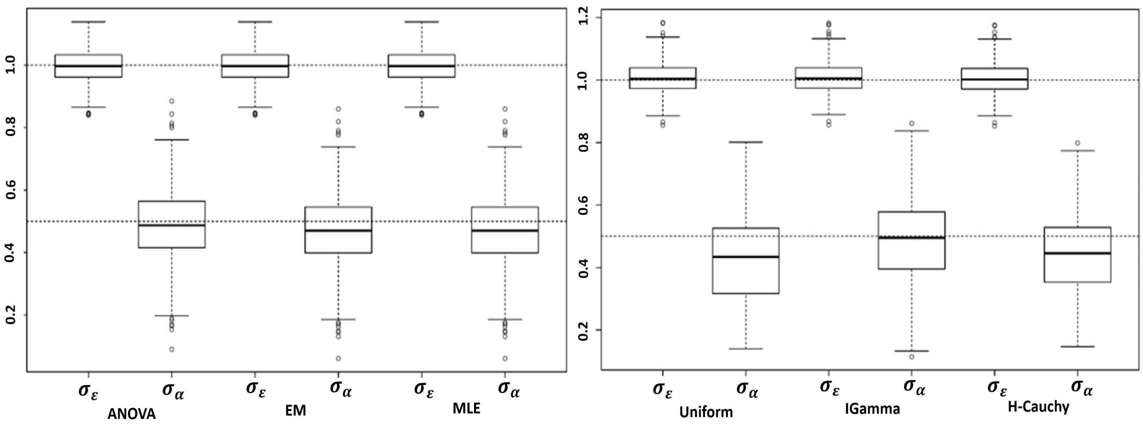

Fig. 9: Boxplot of 1000 estimates of σ_α and σ_ε, at both values equal to 0.5 ($\frac{\sigma_\alpha=0.5}{\sigma_\varepsilon=1}=0.5$), for frequentist approaches (left) and Gibbs sampler (right) at $n=20$ and $a=10$

Table 5: Summary results of 1000 simulated data sets with 5% outliers for α_i for the frequestist approaches and Gibbs sampler with different prior distributions. $\sigma_\alpha = 1$, $\sigma_\varepsilon = 1$, so the ratio is equal to 1 at $n = 20$ and $a = 10$.

Approach			Mean ± SE	Median	75 percentile	Bias	MSE
	ANOVA	σ_ε	0.997± 0.002	0.997	1.034	-0.003	0.003
		σ_α	1.315± 0.005	1.313	1.405	0.315	0.118
Friquentist	EM	σ_ε	0.997± 0.002	0.997	1.034	-0.003	0.003
		σ_α	1.280± 0.005	1.278	1.367	0.280	0.096
	MLE	σ_ε	0.997± 0.002	0.997	1.034	-0.003	0.003
		σ_α	1.280± 0.005	1.278	1.367	0.280	0.096
	Uniform	σ_ε	0.996± 0.006	0.998	1.033	-0.003	0.003
		σ_α	1.280±0.006	1.272	1.372	0.280	0.099
Bayesian	IGamma	σ_ε	1.001± 0.006	1.003	1.039	0.002	0.003
		σ_α	1.360± 0.002	1.351	1.455	0.360	0.152
	H-Cauchy	σ_ε	0.996± 0.006	0.998	1.033	-0.003	0.003
		σ_α	1.280± 0.006	1.274	1.369	0.280	0.099

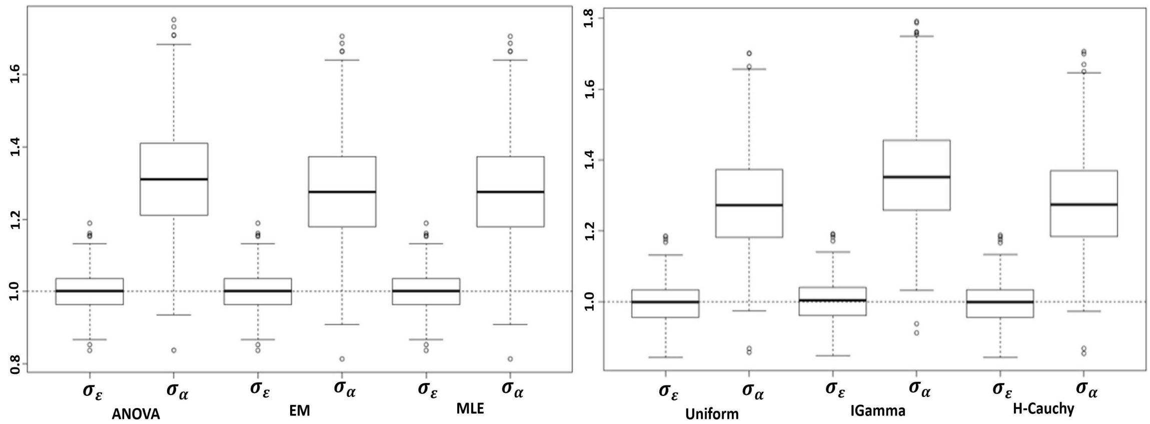

Fig. 10: Boxplot of 1000 estimates of σ_α and σ_ε, at both values equal to 1 ($\frac{\sigma_\alpha=1}{\sigma_\varepsilon=1} = 1$), for frequentist approaches (left) and Gibbs sampler (right) at $n = 20$ and $a = 10$. There are 1000 replicates with 5 % outliers in α_i

3.2 Case 2: Factor Contamination

Tables 5-6, and Figure 6-7 show the results for factor contamination. In case there are outliers in α_i at n = 20 and a = 10, Table 5 shows that among frequentist approaches, EM and MLE estimates are better than ANOVA estimates of variance components. Among Gibbs sampling priors, half-Cauchy and uniform priors are good compared to inverse gamma prior. The same finding one can see from Table 6 with bigger difference between estimates because of smaller sample size. As a result, ANOVA and Gibbs with inverse gamma prior are more sensitive to outliers compared to the other estimators when there are outliers in α_i.

3.3 Case 3: Error Contamination

Tables 7-8 and Figure 8-9 show the results for error contamination. In case there are outliers in ε_{ij}, among frequentist approaches, ANOVA estimate is better than MLE and EM estimate of variance components. Among Gibbs sampling priors, inverse gamma prior is better than half-Cauchy and uniform priors. The same results one can see from Table 8 (compared to Table 7), but there is a big gap between the estimates because of smaller sample size. In sum, when there are outliers in σ_ε, ANOVA and inverse gamma are less sensitive to outliers compared to the other estimates.

From these findings, researchers should not use ANOVA or inverse gamma if there are outliers in α_i, and they should use them when there are outliers in ε_{ij}.

Table 6: Summary results of 1000 simulated data sets with 5% outliers in α_i for the frequestist approaches and Gibbs sampler with different prior distributions with setting of $\sigma_\alpha = 1$, $\sigma_\varepsilon = 1$, so the ratio is equal to 1 at $n = 4$ and $a = 10$

Approach			Mean ± SE	Median	75 percentile	Bias	MSE
Friquentist	ANOVA	σ_ε	0.984± 0.004	0.984	1.069	-0.0153	0.015
		σ_α	1.556± 0.008	1.545	1.713	0.5559	0.362
	EM	σ_ε	0.984±0.004	0.984	1.069	-0.0153	0.015
		σ_α	1.467±0.007	1.458	1.618	0.4670	0.267
	MLE	σ_ε	0.984±0.004	0.984	1.069	-0.0153	0.015
		σ_α	1.467±0.007	1.458	1.618	0.4675	0.267
Bayesian	Uniform	σ_ε	0.994±0.0064	0.993	1.0768	-0.0057	0.020
		σ_α	1.433±0.0121	1.458	1.617	0.4325	0.261
	IGamma	σ_ε	1.0155±0.006	1.0159	1.0994	0.0155	0.018
		σ_α	1.667±0.0119	1.6772	1.8459	0.6668	0.516
	H-Cauchy	σ_ε	0.9909±0.006	0.9903	1.0746	-0.0090	0.019
		σ_α	1.4395±0.0115	1.4582	1.6143	0.4394	0.259

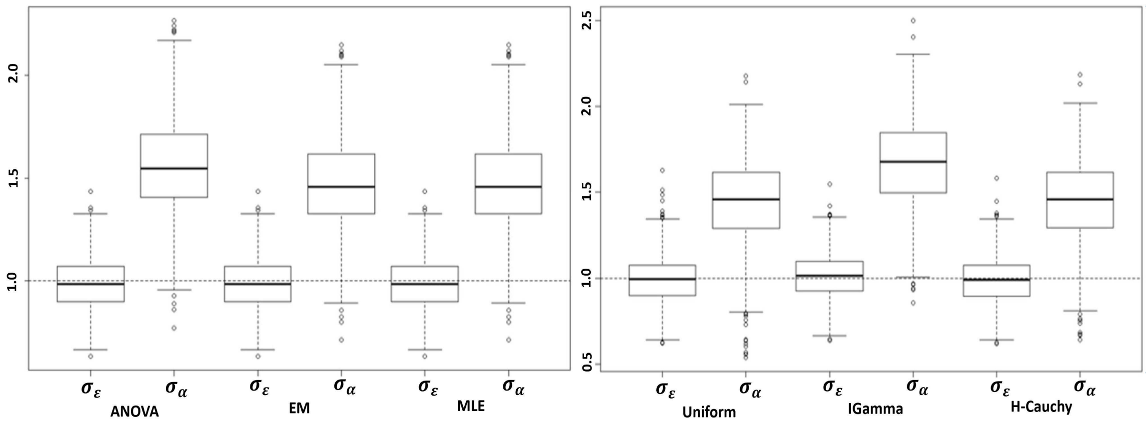

Fig. 11: Boxplot of 1000 estimates of σ_α and σ_ε at both values equal to 1 ($\frac{\sigma_\alpha=1}{\sigma_\varepsilon=1} = 1$) for frequentist approaches (left) and Gibbs sampler (right) at $n = 4$ and $a = 10$. 1000 replicates with 5% outliers in α_i

Table 7: Summary results of 1000 simulated data sets with 5% outliers for ε_{ij} for the frequentist approaches and Gibbs sampler with different prior distributions. $\sigma_\alpha = 1$, $\sigma_\varepsilon = 1$, so the ratio =1, $n = 20$ and $a = 10$.

Approach			Mean ± SE	Median	75 percentile	Bias	MSE
Friquentist	ANOVA	σ_ε	1.322±0.001	1.321	1.035	0.321	0.105
		σ_α	0.959± 0.007	0.947	1.078	-0.041	0.036
	EM	σ_ε	1.322±0.001	1.321	1.035	0.321	0.105
		σ_α	0.930± 0.007	0.919	1.047	-0.069	0.038
	MLE	σ_ε	1.322± 0.001	1.321	1.349	0.321	0.105
		σ_α	0.930±0.007	0.919	1.047	-0.069	0.038
Bayesian	Uniform	σ_ε	1.325±0.001	1.324	1.353	0.324	0.107
		σ_α	0.922±0.007	0.931	1.062	-0.077	0.047
	IGamma	σ_ε	1.330± 0.001	1.329	1.359	0.329	0.110
		σ_α	1.000± 0.007	1.006	1.140	0.001	0.041
	H-Cauchy	σ_ε	1.325± 0.001	1.324	1.353	0.324	0.107
		σ_α	0.925± 0.007	0.931	1.062	-0.074	0.045

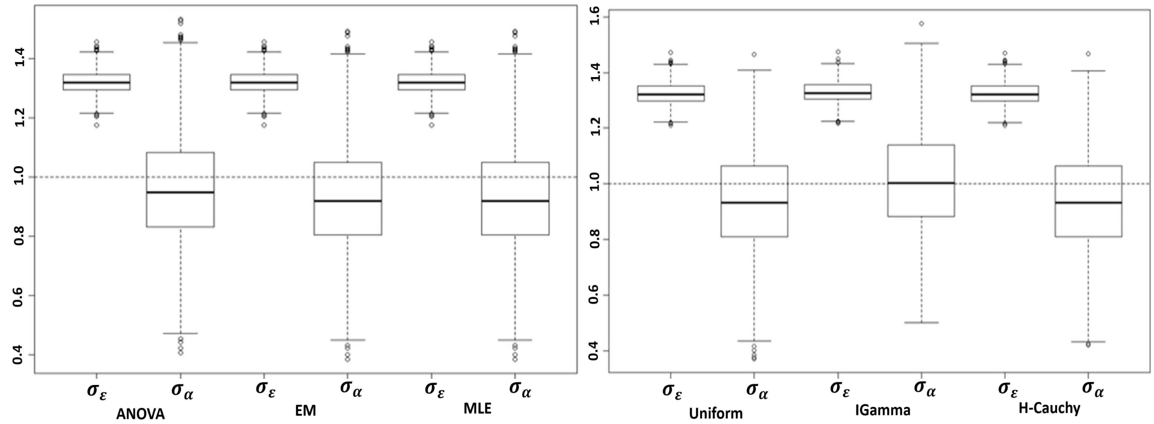

Fig. 12: Boxplot of 1000 estimates of σ_α and σ_ε, at both values equal to 1 ($\frac{\sigma_\alpha=1}{\sigma_\varepsilon=1}=1$), for frequentist approaches (left) and Gibbs sampler (right) at $n=20$ and $a=10$. There are 1000 replicates with 5 % outliers in ε_{ij}

Table 8: Summary results of 1000 simulated data sets with 5 % outliers for ε_{ij} for the frequentist approaches and Gibbs sampler with different prior distributions. $\sigma_\alpha=1$ and $\sigma_\varepsilon=1$, so the ratio =1, $n=4$ and $a=10$.

Approach			Mean ± SE	Median	75 percentile	Bias	MSE
	ANOVA	σ_ε	1.316±0.003	1.314	1.386	0.3167	0.111
		σ_α	0.954±0.011	0.943	1.183	-0.0450	0.116
Friquentist	EM	σ_ε	1.315±0.003	1.313	1.385	0.3154	0.110
		σ_α	0.854±0.012	0.861	1.097	-0.1457	0.150
	MLE	σ_ε	1.315±0.003	1.313	1.385	0.3154	0.110
		σ_α	0.853±0.012	0.861	1.097	-0.1465	0.151
	Uniform	σ_ε	1.420±0.006	1.423	1.515	0.4203	0.197
		σ_α	0.514±0.016	0.352	0.710	-0.4856	0.378
Bayesian	IGamma	σ_ε	1.385±0.005	1.388	1.469	0.3846	0.163
		σ_α	0.846±0.019	0.804	1.173	-0.1531	0.205
	H-Cauchy	σ_ε	1.389±0.005	1.388	1.480	0.3885	0.168
		σ_α	0.609±0.015	0.501	0.820	-0.3909	0.278

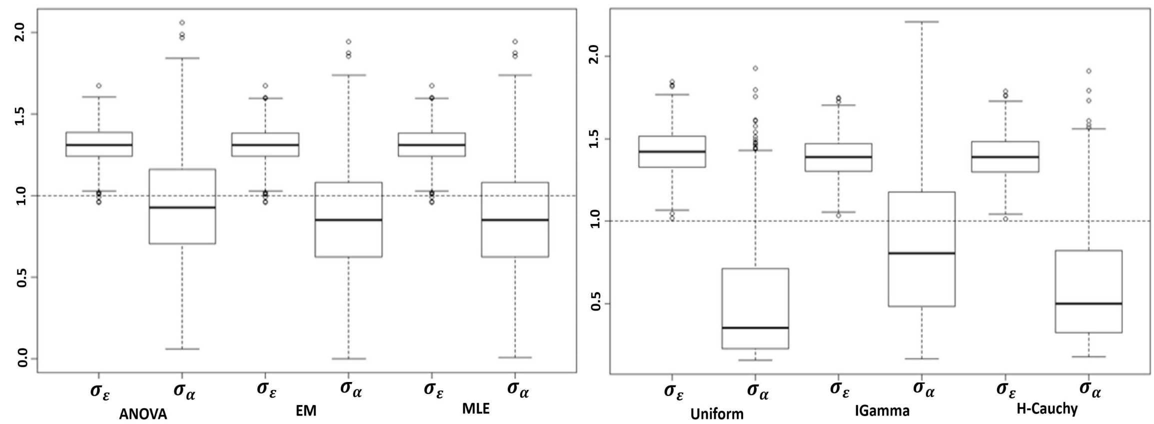

Fig. 13: Boxplot of 1000 estimates of σ_α and σ_ε, at both values equal to 1 ($\frac{\sigma_\alpha=1}{\sigma_\varepsilon=1}=1$), for frequentist approaches (left) and Gibbs sampler (right) at $n=4$ and $a=10$. There are 1000 replicates with 5% outliers in ε_{ij}

4 Real Data Application

The real data that is described in [24] is used to demonstrate our approach. The data is collected to study the association between exposure to radon and its progeny, and lung cancer. Five detectors are used from each of the six laboratories. Each detector gives one measurement of α-energy that is emitted from radioactive radon gas in a house. In total, the data contains 25 measurements. Data is displayed in Fig. 14. On average, laboratories should give close measurements of α-energy because they are applied to the same houses. Figure 14 shows that the laboratory number three gives a different mean compared to the other five laboratories.

The laboratories represents the random effect, so the one-way random effect model can be used to describe the data. [25] shows that laboratory number three is an outlier compared to the other five laboratories, so the data has factor contamination.

The methods of estimating variance components described in this article are applied to the data, and the results are displayed in Table 9 assuming there is factor contamination. It shows that ANOVA and Gibbs sampler with inverted-gamma prior estimates of the variance of random effects, σ^2_α, are higher than the other methods estimates. This means that these methods are sensitive to outliers in random effects. The sample size of this data is small with $a=5$ and $n=5$, and the ratio $\frac{\sigma_\alpha}{\sigma_\varepsilon}=1.2$. As this study is recommended under this setting, MLE, EM algorithm, Gibbs sample with Uniform or Half-Cauchy give better estimate than other methods.

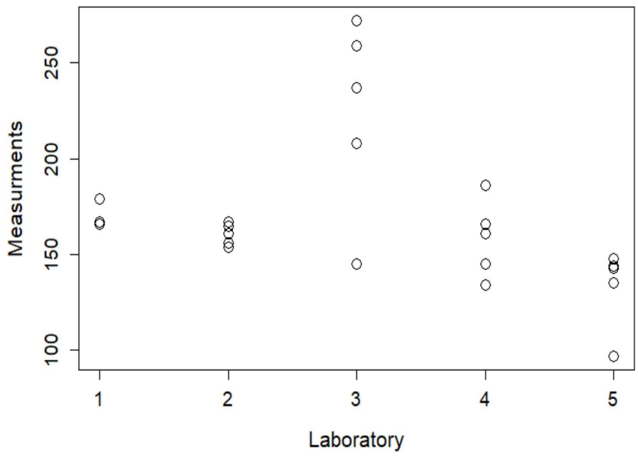

Fig. 14: Scatter plot of laboratories and their detector measurements

Table 9: Variance components estimates (σ^2_α, σ^2_ε) using frequentist approaches and Gibbs sampler with different priors

		σ^2_α	σ^2_ε
Frequentest	ANOVA	986.4	689.6
	MLE	761.5	689.6
	EM	761.4	689.6
Gibbs	Uniform	758.8	689.3
	Half-Cauchy	759.7	690.4
	IGamma	970.6	691.2

5 Conclusion and Recommendation

The performance of the frequentist approaches and Gibbs sampler under different prior distributions for variance components are studied using an intensive simulation study at different situations (no outliers exist, factor contamination,

Table 10: Summary results of 1000 simulated data sets with 5 % outliers for ε_{ij} for the frequentist approaches and Gibbs sampler with different prior distributions, $\sigma_\alpha = 1$, $\sigma_\varepsilon = 1$, so the ratio =1, $n = 4$ and $a = 10$.

	No outliers	Outliers in α_i	Outliers in ε_{ij}
Small sample ($n = 4$ and $a = 10$)	ANOVA or Gibbs with inverse gamma	MLE, EM, or Gibbs with uniform or Cauchy prior	ANOVA or Gibbs with inverse gamma
Small sample ($n = 20$ and $a = 10$)	ANOVA or Gibbs with inverse gamma	MLE, EM, or Gibbs with uniform or Cauchy prior	ANOVA or Gibbs with inverse gamma

and error contamination). The evaluation is based on MSE and absolute bias criteria. The findings advise researchers to select a method of variance components estimation based on sample variances along with the study goal. They need to calculate the sample variance components and the variance components ratios, and based on the ratio, they can select a method of estimation. When there are no outliers in a real data set, and the ratio of the variance components is small, researchers should use inverse gamma prior within Gibbs sampler or ANOVA. However, if the ratio is big, they should use Half-Cauchy within Gibbs sampler or any frequentist approch, they have the same performance. How much the ratio, $\frac{\sigma_\alpha}{\sigma_\varepsilon}$, is it small or big? it depends on the sample size. Based on these results, when the sample size is about 40, the change point ratio is about 2, and when the sample is about 200, the change point ratio is about 1. Table 9 displays the summary of our findings for small ratio, $\frac{\sigma_\alpha}{\sigma_\varepsilon}$, at the three cases: no outliers, factor contamination, and error contamination.

As a result, Gibbs sampler with uniform or Cauchy prior, MLE and EM are less sensitive if we have outliers in α_i, and ANOVA and Gibbs sampler with inverse gamma prior are less sensitive when outliers exist in ε_{ij}. However, in this study, the sensitivity of variance components for the one-way random effects model is studied, but it can be easily generalized to more complicated models.

Acknowledgment

The authors are grateful to the anonymous referee for a careful checking of the details and for the helpful comments that improved this paper.

Conflicts of Interest

The authors declare that there is no conflict of interest regarding the publication of this article

References

[1] Fisher, R. A. *Statistical Methods for Research Workers*. Oliver and Boyd: London, (1925).
[2] Herbach L. H. Properties of model II type analysis of variance tests A: optimum nature of the F-test for model II in balanced case. *Annals of Mathematical Statistics*, **30**, 939-959, (1959).
[3] Dempster, A. P., Laird, N. M. and Rubin, D. B. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society*, Series B. **39** (1), 1-38, (1977).
[4] Anderson R. L. and Bancroft T. A. *Statistical Theory in Research*. McGraw-Hill: New York, (1952).
[5] Rao C. R. Estimation of variance and covariance components: MINQUE theory. *Journal of Multivariate Analysis*, **1**, 257-275, (1971a).
[6] Rao C. R. Minimum variance quadratic unbiased estimation of variance components. *Journal of Multivariate Analysis*, **1**, 445-456, (1971b).
[7] Rao C. R. Estimation of variance and covariance components in linear models. *Journal of American Statistical Association*, **67**, 112-115, (1972).
[8] German, S. and German, D. Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **6** (6), 721-741, (1984).
[9] Searle, S. R. An overview of variance component estimation. *Metrika*, **42** (1), 215-230, (1995).
[10] Chen, J., Zhang, D. and Davidian, M. A Monte Carlo EM algorithm for generalized linear mixed models with flexible random effects distribution. *Biometrics*, **3** (3), 347-360, (2002).
[11] Rasch, D. and Masata, O. Methods of variance component estimation. *Czech Journal of Animal Science*, **51** (6), 227-235, (2006).

[12] Gelfand, A. E. and Smith, A. F. M. Sampling-based approaches to calculating marginal densities. *Journal of the American Statistical Association*, **85**, 398-409, (1990).
[13] Daniels, M. A prior for the variance in hierarchical models. *Canadian Journal of Statistics*, **27**, 569-580, (1999).
[14] Lee, C. and Wang, C. D. Bayesian inferences on variance components using Gibbs sampling with various priors. *Asian-Australasian Journal of Animal Sciences*, **14** (8), 1051-1056, (2001).
[15] Spiegelhalter, D. J., Thomas, A., Best, N. G., Gilks, W. R. and Lunn, D. BUGS: Bayesian inference using Gibbs sampling. MRC Biostatistics Unit, Cambridge, England, http://www.mrc-bsu.cam.ac.uk/bugs/, (1994, 2003).
[16] Gelman, A. Prior distributions for variance parameters in hierarchical models. *Bayesian Analysis*, **3** (3), 515-534, (2006).
[17] Gelman, A., Carlin, J. B., Stern, H. S. and Rubin, D. B. *Bayesian Data Analysis*. Chapman and Hall: London, (2003).
[18] Polson, N. G. and Scott, J. G. On the half-Cauchy prior for a global scale parameter. *Bayesian Analysis*, **7** (4), 887-902, (2012).
[19] Ghosh, J., Li, Y. and Mitra, R. On the use of Cauchy prior distributions for Bayesian logistic regression. *Bayesian Analysis*, **13** (2), 359-383, (2018).
[20] Browne, W. J. and Draper, D. A comparison of Bayesian and likelihood-based methods for fitting multilevel models. *Bayesian Analysis*, **3**, 473-514, (2005).
[21] Box, G. E. P. and Tiao, G. C. Bayesian inference in statistical analysis. Reading, Mass: Addison-Wesley, (1973).
[22] Crump, S. L. *The Estimation of Variance in Multiple Classifications*. Ph.D. thesis, Iowa State University, Ames, Iowa, USA, (1947).
[23] Crump, S. L. The present status of variance components analysis. *Biometrics*, **1**, 1-16, (1951).
[24] Kreienbrock, L., Poffijn, A., Tirmache, M., Feider, M., Kies, A. and Darby, S.C. Intercomparisons of passive radon-detectors under field conditions in epidemiological studies. *Health Physics*, **76**, 558-563, (1999).
[25] Wellmann, J. and Gather. U. Identification of outliers in a one-way random effects model. *Statistical Papers*, **44**, 335-348, (2003).

Hamdy F. F. Mahmoud is an assistant professor at Assiut University in Egypt and a collegiate assistant professor at Virginia Tech, USA. He earned bachelor's and master's degrees from the Faculty of Economics and Political Sciences, Cairo University, Egypt, and master's and doctoral degrees from Virginia Tech in 2010 and 2014, respectively, all in statistics. During his graduate studies at Virginia Tech, Mahmoud received the Jesse C. Arnold Award for Teaching Excellence for the 2012/13 academic year, and was nominated for Graduate Teaching Instructor of Record Award in 2014. He is a reviewer for several international journals in statistics. Mahmoud focuses his research on developing statistical models, especially in semi/nonparametric regression models and applying different types of data, such as epidemiology, spatial, spatially-temporally data and environmental data.