T E C H N I C A L
R E P O R T

0504

A NOTE ON THE EVALUATION OF TRIPLE-GOAL ESTIMATES IN LINEAR MIXED MODELS

GHIDEY, W., LESAFFRE, E. and G. VERBEKE

I A P S T A T I S T I C S
N E T W O R K

INTERUNIVERSITY ATTRACTION POLE

http://www.stat.ucl.ac.be/IAP

A note on the evaluation of triple-goal estimates in linear mixed models

WENDIMAGEGN GHIDEY, EMMANUEL LESAFFRE and GEERT VERBEKE

Biostatistical center, Catholic University of Leuven, B-3000 Leuven, Belgium

Summary

In this article the method of Shen and Louis (1998) is applied to estimate the random effect parameters in a linear mixed model (Laird and Ware, 1982). Shen and Louis' was originally developed for general two-stage hierarchical models to have a single set of estimates that could satisfy three inferential goals simultaneously: (1) optimal estimates for empirical distribution functions (EDF), (2) optimal estimates for the random effect parameter ranks and, (3) good estimates for the individual parameters. The performance of these so-called *triple-goal estimates* in estimating the EDF and subject-specific parameters of the random effects in a linear mixed model is evaluated here using a simulation study under different distributional assumptions of the random effects and measurement error. For normally distributed random effects, the triple-goal estimates perform better in estimating the EDF than the empirical Bayes with a little trade-off to optimally estimate the individual random effect parameters. However, under misspecified distribution of the random effects combined with a large measurement error variability this property is lost. We exemplify this using a mixture of normals as distribution for the random effects. The triple-goal estimates and empirical Bayes estimates can also be determined taking into account the true (non-normal) distribution of the random effects. We show this for a mixture of normals. In that case, the triple-goal estimates once again perform better than the EB estimates in estimating the mixture EDF of the random effects.

1 Introduction

Shen and Louis (1998) developed a new method of parameter estimation in two-stage hierarchical models. These so-called triple-goal estimates are based on optimization of estimation of the empirical distribution and ranks of the parameters with a little trade-off in the optimality of individual parameter estimates. Mathematical and simulation-based analyses showed that for a known prior, the method produces more optimal estimates for the empirical distribution function or histogram than other methods including the posterior means estimates.

We adapt this technique to estimate random effect parameters in a linear mixed model (Laird and Ware, 1982) and compare these estimates with the empirical Bayes (EB) estimates. Special emphasis is given here to the evaluation of their performance when the normality assumption about the distribution of the random effects in the linear mixed model fails. Under violation of the normality assumptions, such as a finite mixture of normals, Verbeke and Lesaffre (1996) have shown that the EB estimates may show a considerable shrinkage towards a unimodal distribution. It is our purpose to evaluate the behaviour of the triple-goal estimates under a similar violation of the normality assumption.

In Section 2, we define the triple-goal also called GR estimates in the context of a linear mixed model. Sections 3 and 4 determine and compare the GR and EB estimates when the underlying random effects distribution is normal or a mixture of two normal distributions, respectively. Section 5 extends the GR and EB estimates to the heterogeneity models of Verbeke and Lesaffre(1996). In Section 6 we draw some conclusions.

2 The linear mixed-effects model

The Laird and Ware (1982) random effects model for longitudinal (repeated) data is

$$\mathbf{Y}_k = X_k\beta + Z_k\mathbf{b}_k + \epsilon_k, \qquad k = 1, \ldots, K \tag{1}$$

where $\mathbf{Y}_k$ is the n_k dimensional response vector, β is a p-dimensional vector of fixed effects, the d-dimensional random effect vector $\mathbf{b}_k$ is distributed as $N(0, D)$, and the n_k-dimensional error terms ϵ_k are distributed as $N(0, \sigma^2 I_{n_k})$. All the $\mathbf{b}_k$ and ϵ_k are independent. X_k is a known fixed effects regressor matrix and Z_k is the random effects regressor matrix.

Hence, marginally $\mathbf{Y}_k \sim N(X_k\beta, V_k), k = 1 \ldots K$ with $V_k = Z_k^T D Z_k + \sigma^2 I_{n_k}$. Further, for a well formulated model it is assumed that $rank(X_k, Z_k) = rank(X_k)$(Morrell $et.al.$, 1997).

The population mean parameters β and the covariance matrix V_k are usually estimated using maximum likelihood estimation. On the other hand, empirical Bayes estimation results in estimates of $\mathbf{b}_k, k = 1 \ldots K$.

The empirical Bayes estimate

Consider the mixed effect model (1) where $\mathbf{b}_k \sim G_0 = N(0, D)$, and $\epsilon_k \sim N(0, \sigma^2 I_{n_k})$.

In a Bayesian context and from the theorem of Lindley and Smiths (1972), it follows that the posterior distribution for $\mathbf{b}_k \mid \mathbf{Y}_k$ is normal with (Bryk and Raudenbush, 1992)

$$E(\mathbf{b}_k \mid \mathbf{Y}_k) = (Z_k^T Z_k + \sigma^2 D^{-1})^{-1} Z_k^T (\mathbf{Y}_k - X_k\beta), \tag{2}$$

$$var(\mathbf{b}_k \mid \mathbf{Y}_k) = \sigma^2 (Z_k^T Z_k + \sigma^2 D^{-1})^{-1}. \tag{3}$$

The EB estimate for $\mathbf{b}_k$ equals the mean of the posterior distribution after all the unknown parameters are replaced by their maximum likelihood estimates. It can be seen that the EB estimates $\hat{\mathbf{b}}_k$ are weighted combinations of the overall fixed effect estimate $\hat{\beta}$ and least square estimate based on the linear model $\mathbf{Y}_k = X_k\beta_k + \epsilon_k$ with weights that reflect the precisions of the estimators (Strenio *et.al.*, 1983; Morris, 1983; Dempster *et.al.*, 1981).

The triple-goal or GR estimate

The GR estimate for a univariate random effects vector $\mathbf{b} = (b_1, \ldots, b_K)^T$ is obtained as follows (see Shen and Louis, 1998 for more details): Suppose the true empirical distribution function is $G_K(t) = \frac{1}{K}\sum I_{\{b_k \leq t\}}$, where $I(.)$ is the indicator function and $-\infty < t < \infty$, then

$$\bar{G_K}(t) = E[G_K(t) \mid Y] = \frac{1}{K}\sum P(b_k \leq t \mid Y_k),$$

minimizes the integrated squared error loss given by

$$ISEL(A, G_K) = \int (A(t) - G_K(t))^2 dt,$$

among all the candidate estimates $A(.)$ of $G_K(.)$.

Further, for any candidate vector of ranks $\mathbf{Q}$, $\bar{R_k} = E[R_k \mid Y] = \sum_{j=1}^{K} P(b_k \geq b_j \mid Y_k)$ is the estimate of the true rank R_k of b_k such that it minimizes the squared error loss

$$(1/K)\sum(Q_k - R_k)^2.$$

Thence, the $(\hat{R}_k = rank(\bar{R}_k))^{th}$ order GR estimate of $\mathbf{b}$ equals,

$$\hat{b}_{\hat{R}_k} = \bar{G}_K^{-1}\left(\frac{2\hat{R}_k - 1}{2K}\right) \qquad \hat{R}_k = 1, \ldots, K. \tag{4}$$

The distribution parameters of $b_k \mid \mathbf{Y}_k$ in equations (2) and (3) characterize the distribution function $\bar{G_K}$ and hence $\hat{b}_{\hat{R}_k}$ can be solved from the equality,

$$\begin{aligned}
\bar{G_K}(\hat{b}_{\hat{R}_k}) &= \frac{1}{K}\sum_{k=1}^{K} P(b_k \leq \hat{b}_{\hat{R}_k} \mid \mathbf{Y}_k) \\
&= \frac{1}{K}\sum_{k=1}^{K} \Phi\left(\frac{\hat{b}_{\hat{R}_k} - E(b_k \mid \mathbf{Y}_k)}{sd(b_k \mid \mathbf{Y}_k)}\right) \\
&= \frac{2\hat{R}_k - 1}{2K} \quad ,
\end{aligned}$$

where Φ is the standard normal distribution function.

This is a problem to find the ordinate of a random variable whose value of CDF, which can be considered as a mixture of K normal distribution functions each with mixing weight $\frac{1}{K}$, is given by the last term. Mathematical solution is difficult and hence it has to be solved with numerically. Bi-section method of solving functions has been implemented here.

In practice, $\hat{R}_k$ can be replaced with the respective rank of the EB estimate. For a wide class of models (including the linear mixed effect models) the ranks of both EB and GR estimates are identical to the minimum error loss rank estimates. See Shen and Louis (1998) for detail as to when such equality holds.

3 The GR estimates under a normal random effects distribution

We applied the technique of Shen and Louis to estimate the random intercept parameter after fitting a random intercept model to simulated data. The simulation is based on an observed growth curve data from Ethiopian babies (Lesaffre, *et.al.* 1999). The weight profile of a random sample of 54 infants with at most 7 time point measurements was considered. Further, the MLE estimates of the fixed effects and covariance parameters for the observed sample data are shown in Table 1. These estimates represent the true parameter values for the subsequent simulations in the following manner. First, a random sample of size $K = 54$ random intercept parameters b_k was generated from $b_k \sim N(0, D)$, $k = 1, \ldots, 54$. Then the response vector Y_k given b_k was generated from $Y_k \mid b_k \sim N(X_k\beta + \mathbf{1}b_k, \sigma^2)$ for $\mathbf{1}$ an n_k-dimension column vector of ones. This simulation was carried out 100 times and the EB and GR estimates were determined for each simulated data.

Table 1: *Fixed effect and covariance parameter values estimated from an observed growth curve data, and used for the subsequent simulations.*

Effect	Parameter	Estimate
Mean structure		
intercept	β_0	2.98
slope	β_1	0.90
Covariance structure		
$var(b_k)$	D	0.59
$var(\epsilon_k)$	σ^2	0.33

Finally, the estimates by the two approaches (GR and EB) were evaluated using Integrated

Squared Error Loss (ISEL). The ISEL measures the discrepancy between the empirical distribution functions of the estimate and the true random effects.

$$\begin{aligned} ISEL(\hat{G_K}, G_K) &= \int (\hat{G_K}(t) - G_K(t))^2 dt \\ &\approx \sum_{t=-\infty}^{\infty} (\hat{G_K}(t) - G_K(t))^2 \Delta t \end{aligned}$$

where $\hat{G}_K(t)$ is the empirical distribution of the estimates, $G_K(t)$ is the true EDF and Δt is very small for a better approximation of the integral. In addition, the squared error loss (SEL) was determined (see below) in order to see the increase of error loss in estimating b_k by GR relative to the EB approach. The EB estimate is optimal for SEL. Let $\hat{b}^j$ be the j^{th} simulation estimated vector. The SEL for the j^{th} simulation is defined as,

$$SEL_{\hat{b}}^j = \frac{1}{54} \sum_{k=1}^{54} (\hat{b}_k^j - b_k^j)^2, \qquad j = 1, \ldots, 100,$$

with $\hat{b}^j = (\hat{b}_1^j, \ldots, \hat{b}_K^j)^T$ and $b^j = (b_1^j, \ldots, b_K^j)^T$.

For simulated data based on parameters in Table 1, the average $ISEL$ for the EB and GR-method is 0.0078, 0.0063, respectively. Thus there is more than 23% increase in the error loss by the EB approach compared to GR in estimating the normal empirical distribution function of the random effects. On the other hand, the average SEL value for the EB and GR-method is 0.06747 and 0.0709, respectively yielding a GR/EB ratio of 1.051. Thus the GR increases the SEL on average by only around 5%.

4 The GR estimates under a mixture of normals

In the previous section the better performance of the GR method is shown when the normality assumption of the $b_k's$ holds. i.e. the simulated random effects were in line with the underlying assumption in the mixed model effect model (1).

Among other things the EB estimates, $\hat{b}_k$, are often used to verify the normality assumption of the random effects (Pinheiro and Bates, 2000, §4.3.2). However, it has been shown by Verbeke and Lesaffre (1996) that when the distribution of the b_k is a mixture of normals, this is not necessarily reflected in the histogram of $\hat{b}_k$; especially when the measurement error variance σ^2 is relatively large.

Here, we will repeat the simulation study of Verbeke and Lesaffre (1996) but examine also the performance of the GR estimates. Indeed, given the GR estimates reproducing the EDF we wanted to see how much they maintained their optimality under a mixture of normals. The random effects are simulated from a mixture of two normal distributions,

$$b_i \sim pN(\mu_1, D) + (1-p)N(\mu_2, D).$$

We assume here that the components have equal mixing weights ($p = 0.5$). Further we fixed the respective population means at -2 and 2 and their variance D as in Table 1. The EB and GR estimates are determined for 100 simulated data sets from the mixed density for two values of σ^2 i.e. 0.33 and 15.

Table 2 presents the performance of the GR approach relative to the EB approach based on ISEL, SEL and two-samples Kolmogorov-Smirnov (KS) test to compare the empirical distributions of the estimates and the true random effects.

Both the SEL and ISEL summary statistics increase with σ^2. Again, the ISEL for GR is smaller than for EB. However, the KS goodness of fit test indicates that both approaches do not perform well ($p < 0.05$) to estimate the EDF. The extent of discrepancy is much higher when σ^2 is relatively larger.

Table 2: *ISEL and SEL performances of GR and EB estimates and results of the KS-statistics for two measurement errors variances.*

	$\sigma^2 = 0.33$		$\sigma^2 = 15$	
Measures	EB	GR	EB	GR
ISEL	0.0217	0.0203	0.1267	0.0869
SEL	0.1505	0.1593	1.848	1.839
P-value of KS-goodness of fit statistic	0.0019	0.0006	0	0

Figures 1 and 2 also depict this lack of fit to the true mixture distribution. For smaller measurement error, $\sigma^2 = 0.33$, Figure 1 shows that the two approaches almost equally approximate the bimodal mixture distribution but both fits are not adequate (see Table 2 for the KS statistic) at 0.05 level of significance.

On the other hand, when $\sigma^2 = 15$, the histograms and the super-imposed EDF curves in Figure 2 clearly show that the estimates do not approximate well the distribution of the random effects. The distributions of the estimates rather show a homogeneous normal distribution than a bimodal mixed distribution.

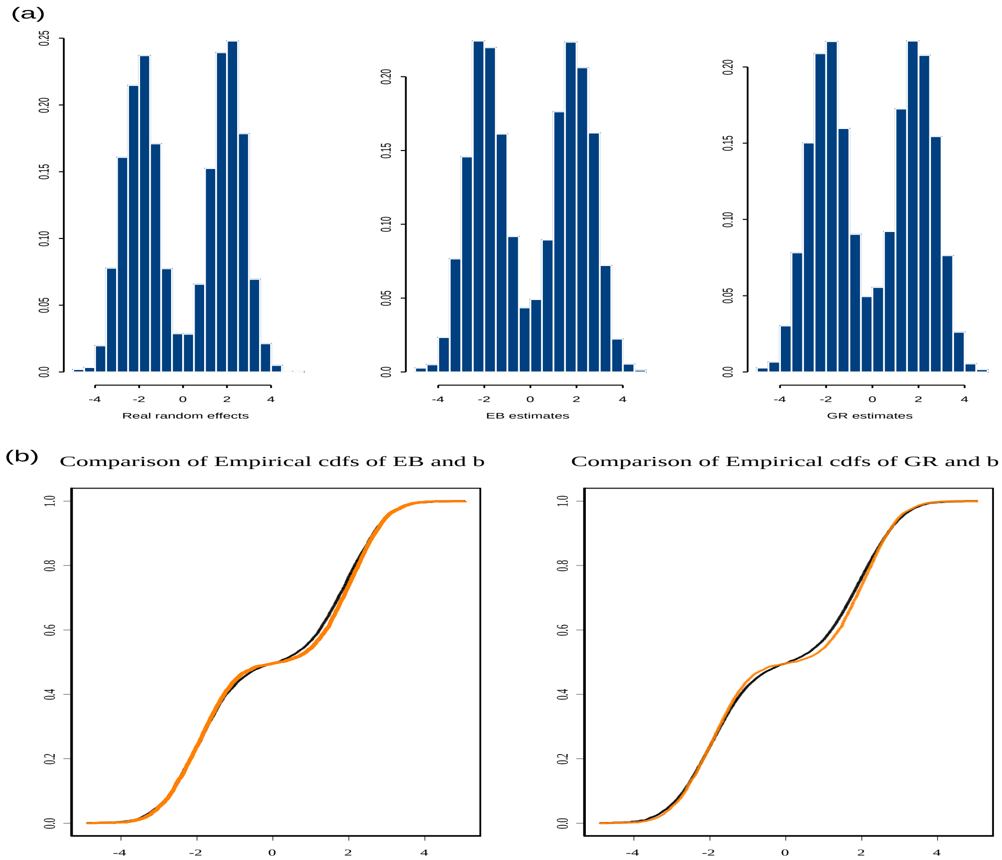

Figure 1: (a) Histograms, and (b) EDF's of the EB and GR estimates (dark lines) superimposed over the true random effects distribution (gray line), for $\sigma^2 = 0.33$.

As a conclusion, it is evident from these numerical results that the GR method is equally sensitive to violations of the distributional assumptions of the random effects as the EB estimates, and they also hardly show the heterogeneity in the b's. The problem is worsened when the measurements errors are large. In retrospect this result could have been expected. Indeed, it can be seen from equations (2) and (4) that both the EB and GR estimates are based on the posterior distribution of the random effects. Therefore, when the random effect b_k is wrongly assumed as normally distributed then the posterior $b_k \mid Y_k$ will consequently and wrongly be assumed as normal with the mean and variance parameters in equations (2) and (3), respectively.

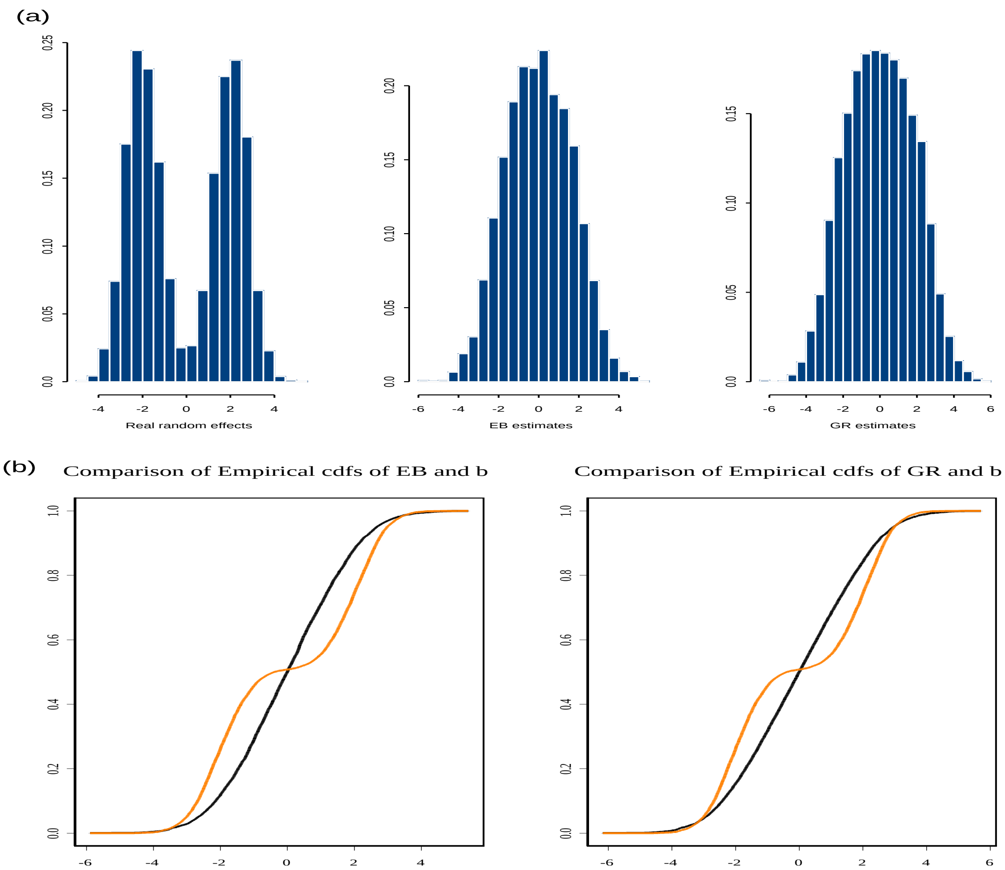

Figure 2: (a) Histograms, and (b) EDF's of the EB and GR estimates (dark lines) superimposed over the true random effects distribution (gray line), for $\sigma^2 = 15$.

5 Estimates for the heterogeneity model

Verbeke and Lesaffre (1996) defined the heterogeneity model as an extension of the linear mixed model (1). This model accommodates the clustered b_k's by further assuming that the random effects are sampled from a mixture of g normals with cluster specific means μ_j, common covariance matrix D and mixing weight p_j, i.e. $\mathbf{b}_k \sim \sum p_j N(\mu_j, D)$ such that $E(\mathbf{b}_k) = \sum p_j \mu_j = 0$ for $j = 1, \ldots, g$. It then follows that marginally $\mathbf{Y}_k \sim \sum p_j N(X_k\beta + Z_k\mu_j, V_k)$ and conditionally $\mathbf{Y}_k \mid \mathbf{b}_k \sim N(X_k\beta + Z_k\mathbf{b}_k, \sigma^2 I_{n_k})$. Observe that the mixture model can approximate arbitrarily close any deviation from normality (see Dalal *et.al.* 1983).

For a fixed component with mean μ_j, it can be derived from the Lindley and Smiths theorem

(1972) that

$$\mathbf{b}_k \mid \mathbf{Y}_k, \mu_j \sim N(\mu_j + DZ_k^T W_k(\mathbf{Y}_k - X_k\beta - Z_k\mu_k), \sigma^2 (Z_k^T Z_k + \sigma^2 D^{-1})^{-1}).$$

The posterior distribution of $\mathbf{b}_k$ given $\mathbf{Y}_k$ is a mixture of the individual component densities given by,

$$\mathbf{b}_k \mid \mathbf{Y}_k \sim \sum_{j=1}^{g} p_{kj} N(\mu_j + DZ_k^T W_k(\mathbf{Y}_k - X_k\beta - Z_k\mu_k), \sigma^2 (Z_k^T Z_k + \sigma^2 D^{-1})^{-1}), \quad (5)$$

with p_{kj} the posterior probability for the kth subject to belong to the jth component, as mixing weights. The EB, the mean of the posterior distribution, equals,

$$\hat{\mathbf{b}}_k = E(\mathbf{b}_k \mid \mathbf{Y}_k, \hat{\beta}, \hat{D}, \hat{\sigma}^2) = \hat{D} Z_k^T \hat{W}_k(\mathbf{Y}_k - X_k\hat{\beta}) + (I - \hat{D} Z_k^T \hat{W}_k Z_k) \sum_{j=1}^{g} p_{kj}\hat{\mu}_j, \quad (6)$$

where the maximum likelihood estimates of the parameters are fitted using the mixture model.

Similarly, the GR estimates in (4) can be extended to the heterogeneity model by modifying the posterior distribution that characterizes $\bar{G}_K(t)$. Now, for a given estimate of the unknown parameters, the $\hat{R_k}^{th}$ order GR estimate is the solution of the equality,

$$\begin{aligned} \bar{G_K}(\hat{\mathbf{b}}_{\hat{R}_k}) &= \frac{1}{K}\sum_{k=1}^{K} P(\mathbf{b}_k \le \hat{\mathbf{b}}_{\hat{R}_k} \mid \mathbf{Y}_k) \\ &= \frac{1}{K}\sum_{k=1}^{K}\sum_{j=1}^{g} p_{kj}\Phi\left(\frac{\hat{\mathbf{b}}_{\hat{R}_k} - E(\mathbf{b}_k \mid \mathbf{Y}_k, \mu = \mu_j)}{sd(\mathbf{b}_k \mid \mathbf{Y}_k, \mu = \mu_j)}\right) \quad (7) \end{aligned}$$

with $E(\mathbf{b}_k \mid \mathbf{Y}_k, \mu = \mu_j) = \mu_j + DZ_k^T W_k(\mathbf{Y}_k - X_k\beta - Z_k\mu_k)$ and $sd(\mathbf{b}_k \mid \mathbf{Y}_k, \mu = \mu_j) = \sigma(Z_k^T Z_k + \sigma^2 D^{-1})^{-1/2}$ are the component specific means and standard deviation of $\mathbf{b}_k \mid \mathbf{Y}_k$, respectively. In practice the unknown parameters are estimated by the maximum likelihood estimation using an EM algorithm (Titterington *et.al.* 1985).

We calculated the heterogeneous EB and GR estimates from simulated data based on a mixture of two normally distributed random effects as in the previous section. We took three different levels of measurement error variability. The homogeneous model (based on linear mixed model (1)) and heterogeneous model EB estimates are compared using ISEL measure.

Table 3 summarizes the simulation results on the error loss performances of the EB estimates for the homogeneity model and the heterogeneity model.

For the smallest measurement error variance, 0.33, there was no relevant error loss reduction by the heterogeneity model estimate (EB.2) relative to the homogeneity model estimate (EB.1). Further, as the measurement error increases, the error loss measures increase for both models. The

Table 3: *SEL and ISEL performances of homogeneous and heterogeneous models EB estimates for varying measurement errors. EB.1 is based on normality assumption of the distribution of the random effects while EB.2 is based on the mixture of two normally distributed components*

	$\sigma^2 = 0.33$		$\sigma^2 = 9$		$\sigma^2 = 16$	
Measures	EB.1	EB.2	EB.1	EB.2	EB.1	EB.2
ISEL	0.0061	0.0060	0.0644	0.0536	0.1199	0.1128
SEL	0.0736	0.0685	1.261	1.173	1.945	1.986

ISEL for the homogeneous model was always larger than for the heterogenous model. This contrast is clearly seen for $\sigma^2 = 9$. However, at $\sigma^2 = 16$, the ISEL ratio is 1.009 indicating both models performed equivalently worst in estimating the empirical distribution function.

Figure 3 compares the EDF's of the estimates against the EDF of the true random effects parameters. For the smaller σ^2 the EDF's of the two models estimates fit very well to the underlying mixture distribution of b. This is in line with the results in Table 3. However, as σ^2 increases the homogeneity model estimate (EB.1) cannot show the mixture pattern appropriately in contrast to the heterogeneity model estimates (EB.2).

But, despite the better performance of the heterogeneity estimators to reflect the mixture its ISEL value is as high as for the homogeneity model estimates. This is due to its behaviour at the edges of the distribution. The EDF of heterogeneity model estimators (EB.2) shows higher discrepancies at the tails of the distribution in comparison to the homogeneity model estimators. On the other hand, the latter estimators are worse in the middles of the distributions.

We also compared the ISEL performances of the heterogeneity model GR estimates (equation (7)) against the EB estimates (equation (6)). This is intended to see whether the GR estimates will perform better when the underlying assumption is in line with the mixture distribution of the random effects. Recall that in Section 3 a similar analysis has been carried out for homogeneous normally distributed random effects.

For the 100 simulated values from the mixture distribution of two normal components, the average ISEL's for the EB and GR estimates are 0.0062 and 0.0051, respectively with a relative ratio of 1.21. Similarly, the average SEL value was 0.071 for the EB estimator while for the GR estimate 0.077 was obtained. This result again conforms to the previous conclusion that under the correct model assumptions the GR estimates perform better in estimating the EDF with a small trade off in estimating the individual random effects.

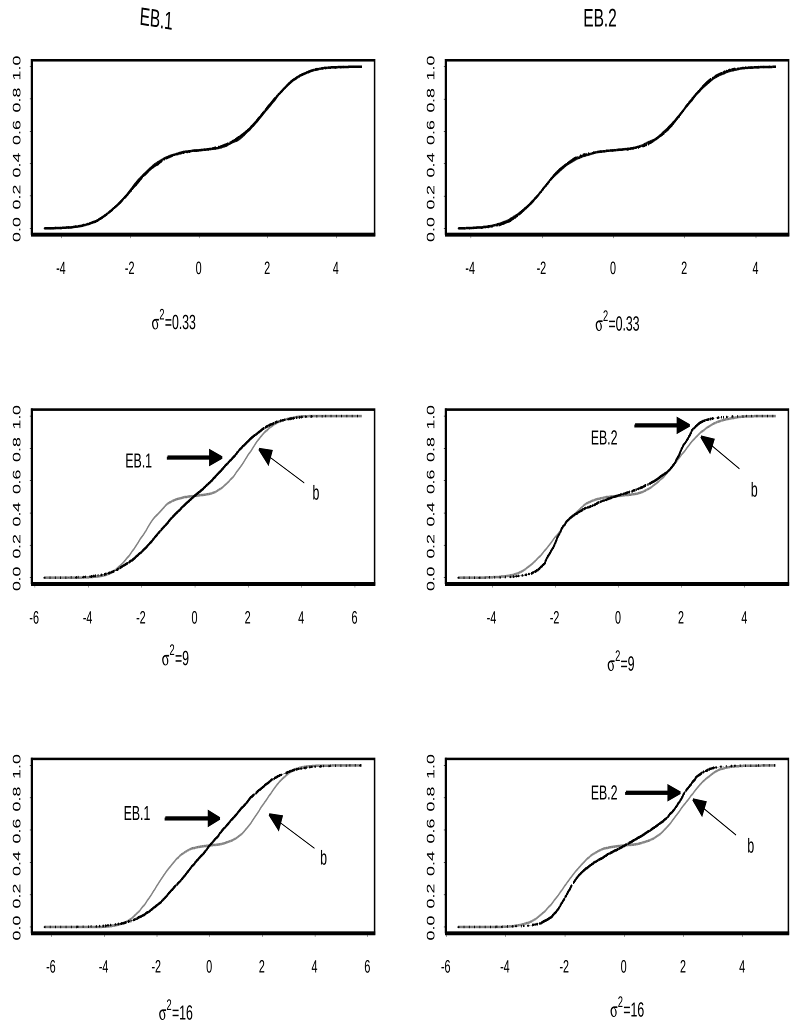

Figure 3: EDF estimates (dark lines) from the homogeneous model empirical Bayes (EB.1) and heterogenous model empirical Bayes (EB.2) estimates superimposed over the true random effects distribution (gray line), for $\sigma^2 = 0.33, 9, \&16$.

Finally, we checked the behaviour of the EB and GR estimators when a mixture model is assumed (using equations (6) and (7)) while in fact the homogeneous normal model applies to the random effects.

From Figure 4 we can observe that all random effect estimators perform similar. Hence, it can be concluded that assuming heterogeneity doesn't give a distorted picture of the estimated random

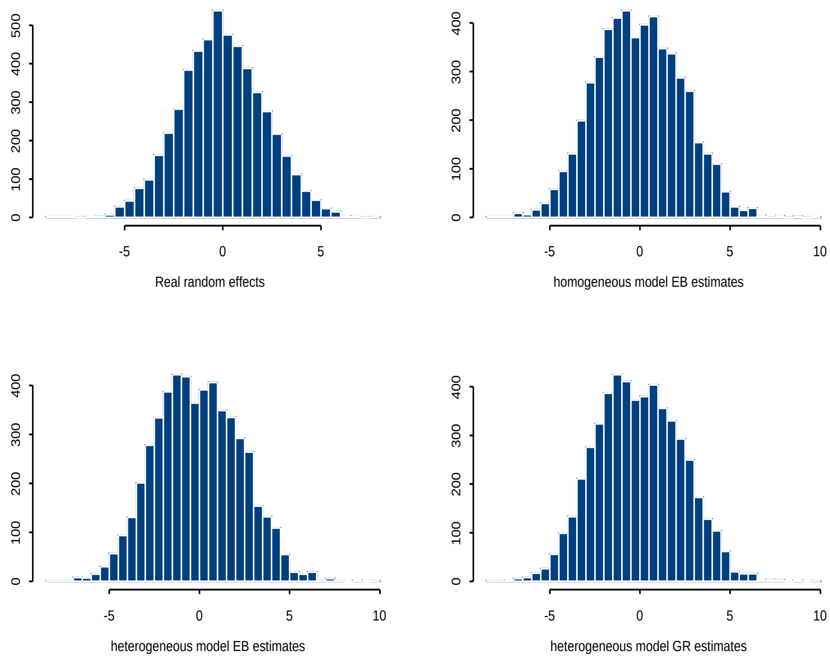

Figure 4: Histograms for the random effect estimates of homogeneous and heterogenous models.

effects distribution.

6 Conclusion

The triple-goal (GR) estimate originally developed for two-stage hierarchical models has been introduced in the context of a linear mixed model. This method, which applies only to a univariate random effect parameter, has been compared with the empirical Bayes estimates of random intercept parameter under different conditions of the distributional assumptions. We showed that the GR method induces a better estimate for the EDF than the empirical Bayes estimate when the normality assumption about the random effects in the linear mixed model holds. When this distribution is misspecified it suffers from the same limitation as the empirical Bayes estimates, i.e. non-robustness. Thus the method of Shen and Louis cannot be used as diagnostic tool for inspecting the appropriateness of normality assumptions in a fitted linear mixed model. On the other hand, when the misspecification of the normality assumption is because of presence of finite mixture of normally distributed components, the better performance of estimating the EDF can still be maintained by incorporating these components. In general, the triple-goal estimates perform well only when based on the correct underlying distribution of the random effects. But in practice, it is not guaranteed to have prior knowledge about the random effects distribution. Therefore, there is still the need for explorative tools, which are independent of the prior knowledge of the underlying distribution, to check distribution of the random effects.

Acknowledgements

Financial support from the IAP research network nr P5/24 of the Belgian State (Federal Office for Scientific, Technical and Cultural Affairs) is gratefully acknowledged.

References

1. Bryk, A.S. and Raudenbush, S.W., *Hierarchical Linear Models:Applications and Data Analysis Methods*, Newbury Park, CA:Sage, 1992.

2. Dalal, S. and Hall, W.J., 'Approximating priors by mixtures of natural conjugate priors', *Journal of the Royal Statistical society, Series B*, 45, 278-286 (1983).

3. Dempster, A.P., Rubin, D.B. and Tsutakawa, R.K., 'Estimation in covariance components models', *Journal of the American Statistical Association*, 76, 341-353 (1981).

4. Laird, N.M. and Ware, J. H. 'Random effects models for longitudinal data', *Biometrics*, 38, 963-974 (1982).

5. Lesaffre, E., Asefa, M., and Verbeke, G., 'Assessing the goodness-of-fit of the Laird and Ware model - an example: the Jimma Infant Survival Differential Longitudinal Study', *Statistics in Medicine*, 18, 835-854 (1999).

6. Lindley, D.V., and Smiths, A.F.M., 'Bayes estimates for the linear model', *Journal of the Royal Statistical society, Series B*, 34, 1-41 (1972).

7. Morell, C.H., Pearson, J.D., and Brant, L.J., 'Linear transformations of linear mixed-effects models', *The American Statistician*, 51, 338-343 (1997).

8. Morris, C.N.,'Parametric empirical Bayes inference: theory and applications', *Journal of the American Statistical Association*, 78, 47-65 (1983).

9. Pinheiro, J.C., and Bates, D.M., *Mixed-effects models in S and S-plus*, New York: Springer-Verlag, 2000.

10. Shen, W. and Louis, T.A, 'Triple-Goal estimates in two-Stage, hierarchical models', *Journal of the Royal Statistics Society, Series B*, 60, 455-471 (1998).

11. Strenio, J.F., Weisberg, H.J., and Bryk, A.S., 'Empirical Bayes estimation of individual growth-curve parameters and their relationship to covariates', *Biometrics*, 39, 71-86 (1983).

12. Titterington, D.M., Smith, A.F.M, Makov, U.E., *Statistical Analysis of Finite Mixture Distributions*, Chichester:John Wiley, 1985.

13. Verbeke, G. and Lesaffre, E., 'A linear mixed-effects model with heterogeneity in the random-effects population', *Journal of the American Statistical Association*, 91, 217-221 (1996).