

ARTICLE IN PRESS



Available at
www.ElsevierMathematics.com

POWERED BY SCIENCE @ DIRECT®

Computational Statistics & Data Analysis III (IIII) III-III

COMPUTATIONAL
STATISTICS
& DATA ANALYSIS

www.elsevier.com/locate/csda

1

Examples in which misspecification of a random effects distribution reduces efficiency, and possible remedies

3

5

Alan Agresti^{a,*}, Brian Caffo^b, Pamela Ohman-Strickland^c

7

^a*Department of Statistics, University of Florida, 204 Griffin-Floyd Hall, Box 118545, Gainesville, FL 32611-8545, USA*

9

^b*Department of Biostatistics, Johns Hopkins University, Baltimore, MD 21205, USA*

^c*Division of Biometrics, University of Medicine and Dentistry of New Jersey, New Brunswick, NJ 08903-2688, USA*

11

Received 1 April 2002; accepted 1 December 2003

Abstract

13

This note shows three cases in which a considerable loss of efficiency can result from assuming a parametric distribution for a random effect that is substantially different from the true distribution. For two simple models for binary response data, we studied the effects of assuming normality or of using a nonparametric fitting procedure for random effects, when the true distribution is potentially far from normal. Although usually the choice of random effects distribution has little effect on efficiency of predicting outcome probabilities, the normal approach suffered when the true distribution was a two-point mixture with a large variance component. Likewise, for a simple survival model, assuming a gamma distribution for the frailty distribution when the true one was a two-point mixture resulted in considerable loss of efficiency in predicting the frailties. The paper concludes with a discussion of possible ways of addressing the problem of potential efficiency loss, and makes suggestions for future research.

17

© 2003 Published by Elsevier B.V.

25

Keywords: Binomial; Frailty model; Gamma distribution; Logit model; Nonparametric; Odds ratio

1. Introduction

27

Recently there has been increasing use of random effects in modeling. Much of this has been in the context of the generalized linear mixed model (GLMM) for repeated

* Corresponding author. Tel.: +1-352-392-1941x234; fax: +1-352-392-5175.

E-mail address: aa@stat.ufl.edu (A. Agresti).

1 measurement and other forms of clustered data and in the modeling of clustered sur-
 2 vival times. The distribution of the random effect is usually chosen for computational
 3 convenience. In linear models with normal errors the normal distribution simplifies
 4 calculations. Due partly to the relationship with these linear models and its ease of
 5 generalization to multivariate random effects, the random effects in GLMMs are also
 6 usually assumed to be normal random variables. In survival analysis, if the random
 7 effects, called frailties, are assumed to be gamma random variables, then the predicted
 8 frailties as well as the likelihood itself have closed-form expressions. While alternative
 9 random effect distributions have been proposed and implemented in some cases, little
 10 research has investigated the consequences of misspecifying that distribution.

11 First, consider the GLMM. Let y_{ij} denote observation j in cluster i , $i = 1, \dots, I$,
 12 $j = 1, \dots, n_i$. Let \mathbf{x}_{ij} denote a column vector of values of explanatory variables for that
 13 response, which serve as coefficients of fixed effects in the model, and let \mathbf{z}_{ij} denote a
 14 corresponding vector of coefficients of random effects. Let \mathbf{u}_i denote a vector of random
 15 effect values for cluster i . Let $\mu_{ij} = E(y_{ij}|\mathbf{u}_i)$. The linear predictor for a GLMM has
 the form

$$g(\mu_{ij}) = \mathbf{x}'_{ij}\boldsymbol{\beta} + \mathbf{z}'_{ij}\mathbf{u}_i, \quad (1)$$

17 where $g(\cdot)$ is a link function. Conditional on \mathbf{u}_i , the model assumes that $\{y_{ij}, j =$
 18 $1, \dots, n_i\}$ are independent.

19 The random effect vector \mathbf{u}_i in a GLMM is assumed to have a multivariate normal
 20 distribution $N(\mathbf{0}, \boldsymbol{\Sigma})$, with covariance matrix $\boldsymbol{\Sigma}$ depending on unknown *variance com-*
 21 *ponents*. See, for instance, [Breslow and Clayton \(1993\)](#) and [Wolfinger and O'Connell \(1993\)](#). There is also some literature on modeling using non-normal random effects.
 22 One approach uses conjugate mixture models ([Lee and Nelder, 1996](#)). Another ap-
 23 proach is nonparametric, with a mixture distribution concentrated on a set of mass
 24 points of unspecified number and location (e.g., [Heckman and Singer, 1984](#); [Aitkin,](#)
 25 [1999](#)).

27 Despite its popularity and attractive features, the normality assumption can rarely
 28 be checked very closely. For instance, [Verbeke and Molenberghs \(2000, Section 7.8\)](#)
 29 noted that under a normality assumption for random effects, the predicted random ef-
 30 fects tend to look normally distributed even when the true random effects are generated
 31 from a highly non-normal distribution. An obvious concern of this or any parametric
 32 assumption for the random effects is whether there are any harmful effects of misspec-
 33 ification.

35 In an alternative use of random effects, the frailty model starts with the Cox pro-
 36 portional hazards model and assumes that the random effect has a multiplicative effect
 on the hazards. Let t_{ij} denote the j th failure in cluster i . The hazard is modeled as

$$\lambda(t_{ij}) = \lambda_0(t_{ij})u_i \exp\{x'_{ij}\boldsymbol{\beta}\}, \quad (2)$$

37 where u_i , the random effect or frailty, is generally assumed to have a continuous
 38 unimodal distribution with mean one. For computational ease, the frailty distribution is
 39 usually assumed to be gamma (e.g., [Clayton, 1978](#); [Oakes, 1982](#); [Nielsen et al., 1992](#)),

1 although other possibilities have been proposed such as a positive stable distribution
(Hougaard, 1986) and the inverse Gaussian (Whitmore and Lee, 1991).

3 Thus far, there has been limited study about the effect of misspecification of the ran-
5 dom effects distribution. In survival analysis, Klein et al. (1992) surveyed the proposed
7 frailty distributions, applying each one to the Framingham heart study data. Although
9 the different frailty distributions had different theoretical implications for the patterns
11 of association between failure times, the alternative specifications did not have a large
13 effect on estimated covariate effects. However, these models all assumed a continuous
random effect with a unimodal frailty distribution. On the other hand, Heckman and
Singer (1984) discussed a case in survival analysis in which bias does occur. Ex-
amining models for censored longitudinal economic data, they showed that estimates
of fixed parameters in a particular Weibull regression model were highly sensitive to
misspecification.

Less dramatic evidence has occurred for other types of models. Neuhaus et al.
(1992) investigated this for a logit model with a random intercept. They argued
that there is little bias in the estimation of the fixed regression effects but some
bias in the mean of the random intercepts when the random effects distribution is
nonsymmetric. They also suggested that standard error estimates are reasonably well
behaved under misspecification. See Chen et al. (2002) for mention of other, more
recent, papers that made the same conclusion for other models. However, Heagerty
and Zeger (2000) argued that regression parameters in random effects models have
bias that is more sensitive to random effects assumptions than their counterparts in
the corresponding marginal models. To illustrate this, they considered a violation
of the usual form of model in which the variance of the random effects depends
on values of covariates. They concluded that between-cluster effects may be more
sensitive than within-cluster effects to correct specification of the random effects
distribution.

Despite some conflicting evidence, the conventional wisdom among data analysts
seems to be that the choice of random effects distribution is not crucial to quality
of inference about regression effects. The purpose of this note is to show, however,
that this may not be so when there is a severe polarization of subjects in the form
of a binary latent class model. We observed that misspecification of this form in the
random effects distribution has the potential for a serious drop in efficiency in the
prediction of random effects and the estimation of other parameters. This is illus-
trated with simulations based on two simple logit models (Sections 2 and 3) and one
failure time hazards model (Section 4). Although usually the choice of random ef-
fects distribution had little effect on efficiency, the parametric approach suffered when
the true distribution was a two-point mixture with a large variance. In the exam-
ple presented in Section 3, even a within-cluster fixed effect is poorly estimated in
this case.

Since the random effects distribution cannot be simply checked, this brings up the
important issue of how to guard against the potential loss of efficiency if the true
random effects distribution is quite far from the assumed one. We describe some pro-
posals for addressing this. This is an important but apparently difficult issue to address
in future research.

2. Random effects model for proportions

The first example is a simple one-way random effects model for binary data. In cluster i , conditional on u_i , y_{ij} is a Bernoulli random variable with expectation μ_i satisfying

$$\text{logit}(\mu_i) = \alpha + u_i, \quad i = 1, \dots, I, \quad j = 1, \dots, n, \quad (3)$$

where $E(u_i) = 0$ and $\text{Var}(u_i) = \sigma^2$. Conditional on u_i , $\sum_j y_{ij}$ has a binomial distribution with n trials and parameter $\mu_i = \exp(\alpha + u_i) / [1 + \exp(\alpha + u_i)]$. We simulated samples from this model for all combinations of $I = 10$ and 30 , $n = 10$ and 30 , $\alpha = 0$ and 1.0 , and $\sigma = 0, 0.5$, and 1.0 , and for various distributions for u_i , including normal, uniform, exponential (shifted so $E(u_i) = 0$), binary with probability 0.5 at each point, and degenerate at a single point (i.e., no random effect and thus I identically distributed binomials). We focused primarily on how well one could estimate α and predict $\{\mu_i\}$ (given u_i) under the usual normal random effects assumption and with a nonparametric approach (Aitkin, 1999).

For simulation k , $k = 1, \dots, 10,000$, let $\hat{\alpha}_k$, $\hat{\sigma}_k$, and $\{\hat{\mu}_{ik}, i = 1, \dots, I\}$ denote the ML estimates of α and σ and the predictions of $\{\mu_{ik}\}$ based on the estimated posterior means of $\{\mu_{ik}\}$ using the posterior distributions of $\{u_{ik}\}$ given the data. We computed the mean, standard error of the mean, and median of $\{|\hat{\alpha}_k - \alpha|\}$, $\{|\hat{\sigma}_k - \sigma|\}$, and $\{\sum_i |\hat{\mu}_{ik} - \mu_{ik}| / I\}$. The relative sizes of the medians were similar to the means for normal and nonparametric fitting schemes, and are not reported here. However, there is a caveat regarding what these sample mean distances estimate. The nonparametric fitting has a positive probability of infinite mass points; thus $E|\hat{\sigma}_k - \sigma|$ does not exist. Likewise, $\hat{\alpha}_k$ can be infinite even for the model without a random effect. The probability of such behavior is very small for the values of n and I used. Thus, we report these sample means as a measure of estimation quality, keeping in mind that they actually estimate expected values conditional on estimates being finite. The standard errors of the means were nearly all less than 0.003 .

Table 1 shows some results when $\alpha = 0$ and $\sigma = 1.0$, representing moderate heterogeneity. The overwhelming impression Table 1 conveys is that the random effects assumption has little influence. Assuming normality does not hurt when the true distribution is far from normality, and using a nonparametric approach when the true distribution is normal does not result in much efficiency loss. In the latter case closer analysis reveals that the nonparametric predictions have absolute deviations from true values averaging about 5–10% higher than ones based on the normal approach. The exceptions where results differ considerably (by more than 20%) for the two approaches are the cases highlighted with *. In two cases for which the true distribution is a two-point mixture and $n = 30$, the normal approach lost considerable efficiency in predicting $\{\mu_{ik}\}$.

For $\sigma = 0.5$ (with $\alpha = 0$), a weaker degree of heterogeneity, for all cases the results for normal and nonparametric fitting were similar. The efficiency gain for the nonparametric approach was then minor when the true distribution was two-point. When $\sigma = 0$, the true model is the fixed effects one in which all In trials are identical with probability 0.5 . In that case, not reported in these tables, results were also similar. For the runs with

Table 1
Mean distances of estimates from parameters in model (3) with $\alpha = 0$, $\sigma = 1.0$

I	n	True	Assumed	$ \hat{\alpha} - \alpha $	$ \hat{\mu}_i - \mu_i $	$ \hat{\sigma} - \sigma $
10	10	Normal	Normal	0.31	0.11	0.36
			Nonparametric	0.31	0.12	0.36
		Uniform	Normal	0.32	0.10	0.32
			Nonparametric	0.32	0.11	0.31
	Exponential	Normal	0.30	0.10	0.40	
		Nonparametric	0.29	0.11	0.44	
	Two-point	Normal	0.34	0.10	0.29	
		Nonparametric	0.32	0.09	0.25	
10	30	Normal	Normal	0.28	0.06	0.26
			Nonparametric	0.27	0.07	0.23
		Uniform	Normal	0.28	0.06	0.20
			Nonparametric	0.28	0.07	0.18
	Exponential	Normal	0.26	0.06	0.34	
		Nonparametric	0.25	0.07	0.31	
	Two-point	Normal	0.29	0.062 ^a	0.14	
		Nonparametric	0.28	0.037 ^a	0.12	
30	10	Normal	Normal	0.18	0.10	0.19
			Nonparametric	0.18	0.11	0.20
		Uniform	Normal	0.19	0.10	0.18
			Nonparametric	0.18	0.10	0.18
	Exponential	Normal	0.18	0.10	0.22	
		Nonparametric	0.17	0.10	0.26	
	Two-point	Normal	0.17	0.06	0.08 ^a	
		Nonparametric	0.18	0.07	0.14 ^a	
30	30	Normal	Normal	0.16	0.06	0.13
			Nonparametric	0.16	0.07	0.13
		Uniform	Normal	0.16	0.06	0.11
			Nonparametric	0.16	0.07	0.10
	Exponential	Normal	0.15	0.06	0.19	
		Nonparametric	0.15	0.06	0.19	
	Two-point	Normal	0.17	0.061 ^a	0.08 ^a	
		Nonparametric	0.16	0.023 ^a	0.06 ^a	

^aCases with a difference of 20% or more.

- 1 $\alpha = 1.0$, the distribution of probabilities has mean above 0.5 and is skewed. Similar
 2 results occurred. The only case with a major difference was predicting probabilities
 3 when the distribution was two-point and $\sigma = 1.0$ with $n = 30$, in which the estimated
 4 average distance between $\hat{\mu}_i$ and μ_i was less than half as large for the nonparametric
 5 approach. For some cases, however, the nonparametric approach gave poorer estimates
 6 of the variance component.
- 7 To investigate the indication that for predicting probabilities the two-point distribution
 8 tended to favor the nonparametric approach more as σ and n increase, we also simulated
 9 other combinations with larger values of σ and n . Table 2 shows results for $n = 30$ and
 10 100 and for $\sigma = 1.0$ and 2.0, when $I = 30$ and $\alpha = 0$. In all these cases with two-point

Table 2

Mean distances of estimates from parameters in model (3) for $n = 30$ and 100 and $\sigma = 1.0$ and 2.0 , when $\alpha = 0$ and $I = 30$

σ	n	True	Assumed	$ \hat{\alpha} - \alpha $	$ \hat{\mu}_i - \mu_i $	$ \hat{\sigma} - \sigma $
1.0	30	Normal	Normal	0.16	0.06	0.13
			Nonparametric	0.16	0.07	0.13
		Two-point	Normal	0.17	0.061 ^a	0.08 ^a
			Nonparametric	0.16	0.023 ^a	0.06 ^a
1.0	100	Normal	Normal	0.16	0.04	0.12
			Nonparametric	0.15	0.04	0.11
		Two-point	Normal	0.12 ^a	0.032 ^a	0.12 ^a
			Nonparametric	0.15 ^a	0.010 ^a	0.04 ^a
2.0	30	Normal	Normal	0.31	0.06	0.29
			Nonparametric	0.29	0.07	0.27
		Two-point	Normal	0.50 ^a	0.045 ^a	0.49 ^a
			Nonparametric	0.30 ^a	0.013 ^a	0.09 ^a
20.0	100	Normal	Normal	0.27	0.04	0.27
			Nonparametric	0.28	0.04	0.24
		Two-point	Normal	0.30	0.014 ^a	0.75 ^a
			Nonparametric	0.30	0.007 ^a	0.06 ^a

^aCases with a difference of 20% or more.

1 distributions, the nonparametric approach performed substantially better, while losing
 2 relatively little efficiency if the random effects distribution is truly normal. In some
 3 cases the nonparametric approach also gave much better estimates of the variance
 component.

5 For this model, ML fitting of the nonparametric random effects approach usually
 6 converged with relatively few mass points. In fact, fitting a model having only two
 7 mass points often gave results quite similar to the full nonparametric approach.

3. Random effects model for log odds ratio

9 The second example refers to estimating the mean log odds ratio for several $2 \times$
 10 2 contingency tables. Here, (y_{i1}, y_{i2}) are each based on n trials in partial table i .
 11 Conditional on a random effect u_i , they are independent binomials with log odds ratio
 $\beta + u_i$. Specifically, conditional on u_i , y_{ij} is $\text{bin}(n, \mu_{ij})$ where

$$\text{logit}(\mu_{i1}) = \alpha + (\beta + u_i)/2, \quad \text{logit}(\mu_{i2}) = \alpha - (\beta + u_i)/2 \quad (4)$$

13 and where $E(u_i) = 0$ and $\text{Var}(u_i) = \sigma^2$.

14 This model for binary responses is useful when heterogeneity occurs among odds
 15 ratios in different studies or for different clusters of observations. For instance, in
 16 comparing two treatments on a binary response with data from several centers, it is
 17 unrealistic to assume exactly the same odds ratio in each center (i.e., $\sigma = 0$). Allow-
 18 ing $\sigma > 0$ provides a more sensible model allowing heterogeneity (e.g., [Beitler and](#)
 19 [Landis, 1985](#); [Agresti and Hartzel, 2000](#); [Hartzel et al., 2001](#)). Interest here focuses

Table 3

Mean distances of estimates from parameters in model (4) for various β and σ , when $I = 10$ and $n = 30$

β	σ	True	Assumed	$ \hat{\beta} - \beta $	$ \hat{\mu}_{ij} - \mu_{ij} $	$ \hat{\sigma} - \sigma $
0.0	1.0	Normal	Normal	0.30 ^a	0.043	0.14
			Nonparametric	0.39 ^a	0.048	0.13
	Two-point	Normal	0.30 ^a	0.046 ^a	0.10	
		Nonparametric	0.24 ^a	0.038 ^a	0.09	
0.0	2.0	Normal	Normal	0.54	0.044	0.24
			Nonparametric	0.56	0.052	0.26
	Two-point	Normal	0.56 ^a	0.045 ^a	0.12	
		Nonparametric	0.21 ^a	0.024 ^a	0.14	
2.0	1.0	Normal	Normal	0.30 ^a	0.041	0.16
			Nonparametric	0.42 ^a	0.045	0.15
	Two-point	Normal	0.31	0.042	0.11	
		Nonparametric	0.27	0.036	0.10	
2.0	2.0	Normal	Normal	0.56 ^a	0.042	0.24 ^a
			Nonparametric	0.69 ^a	0.048	0.33 ^a
	Two-point	Normal	0.57 ^a	0.041 ^a	0.15	
		Nonparametric	0.26 ^a	0.024 ^a	0.15	

^aCases with a difference of 20% or more.

1 on estimating the expected log odds ratio β . This provides an overall treatment effect
 2 measure, allowing for heterogeneity in the odds ratios. In practice, α would also vary
 3 somewhat in i , but we focus on quality of estimation of β .

4 Simulations required substantially more time for this model. Also, as many as 100
 5 quadrature points were sometimes needed to adequately approximate the log likelihood
 6 when σ was large. For 1000 simulations, Table 3 compares ML estimates of β and
 7 σ and predictions of μ_{ij} , when $\alpha = 0$ and when true random effects distributions are
 8 normal or two-point, assuming normality or nonparametric fitting. (We also considered
 9 $\alpha = 1.0$ and obtained similar results.) Standard errors of estimates of β and σ are on
 10 the order of 0.01 or less. Again, when the random effects distribution is truly two-point
 11 but one assumes normality, considerable loss of efficiency can result when σ is large.
 12 Results for the expected log odds ratio β are somewhat more dramatic than for $\{\mu_{ij}\}$.

13 We also considered bias in estimating β for this model. Both approaches performed
 14 well, and results are not shown here. When $\beta = 0$ neither estimate is biased (conditional
 15 on finite estimates) by the symmetry of the model, and when $\beta = 2.0$ the normal
 estimates performed well even when the true distribution was a two-point mixture.

17 4. Frailty model for survival

18 The final example entails estimation of a simple version of the hazard function (2)
 19 for survival data. Using the notation given previously, in cluster i , conditional on u_i ,
 t_{ij} is assumed to be a random failure time with constant hazard function

$$\lambda_i(t) = \alpha u_i, \quad i = 1, \dots, I, \quad j = 1, \dots, n. \quad (5)$$

Table 4
Mean distances of estimates from parameters in model (5) with $\alpha = 2.0$ and $\sigma = 1.0$

I	n	True	Assumed	$ \hat{\alpha} - \alpha $	$ \hat{\lambda}_i - \lambda_i $	$ \hat{\sigma} - \sigma $
10	10	Gamma	Gamma	0.32	0.46	0.34
			Nonparametric	0.32	0.53	0.38
		Uniform	Gamma	0.33	0.45	0.31
			Nonparametric	0.33	0.49	0.29
		Two-point	Gamma	0.34	0.47 ^a	0.28
			Nonparametric	0.33	0.38 ^a	0.24
10	30	Gamma	Gamma	0.27	0.28 ^a	0.26
			Nonparametric	0.26	0.37 ^a	0.29
		Uniform	Gamma	0.28	0.28	0.23
			Nonparametric	0.27	0.33	0.20
		Two-point	Gamma	0.29	0.28 ^a	0.17 ^a
			Nonparametric	0.28	0.15 ^a	0.13 ^a
30	10	Gamma	Gamma	0.18	0.44	0.18 ^a
			Nonparametric	0.18	0.49	0.24 ^a
		Uniform	Gamma	0.19	0.43	0.22
			Nonparametric	0.19	0.44	0.18
		Two-point	Gamma	0.20	0.43 ^a	0.16 ^a
			Nonparametric	0.18	0.28 ^a	0.13 ^a
30	30	Gamma	Gamma	0.16	0.28 ^a	0.14 ^a
			Nonparametric	0.15	0.34 ^a	0.17 ^a
		Uniform	Gamma	0.16	0.27	0.17 ^a
			Nonparametric	0.16	0.30	0.11 ^a
		Two-point	Gamma	0.16	0.27 ^a	0.10
			Nonparametric	0.16	0.06 ^a	0.09

^aCases with a difference of 20% or more.

1 We take $E(u_i) = 1$ to ensure that the average hazard rate for the population of clusters
 2 is the baseline hazard $\lambda_0(t)$. For simplicity, the baseline hazard α is assumed constant
 3 for all t . In frailty models, u_i is usually assumed to be a gamma random variable. As
 4 in Section 2, we use simulation to study how well α is estimated and λ_i is predicted
 5 assuming both a nonparametric and parametric form for the mixing distribution under
 6 different true mixing distributions. No censoring was included in the simulations. For
 7 simulation k , $k = 1, \dots, 10,000$, $\hat{\alpha}_k$ and $\hat{\sigma}_k$ denote the estimated population hazard and
 8 frailty standard deviation, and $\hat{\lambda}_{ik}$ represents the predicted frailty for individual i .
 9 Table 4 summarizes results for $\alpha = 2.0$ and $\sigma = 1.0$, showing effects of I and n
 10 when they equal 10 and 30. Standard errors for sample means were estimated via
 11 Monte Carlo to be less than 0.004 for $|\hat{\alpha} - \alpha|$ and $|\hat{\sigma} - \sigma|$ and 0.002 for $|\hat{\lambda}_i - \lambda_i|$.
 12 Similar to results for the simple logit model of Section 2, misspecification of the
 13 frailty distribution does not result in efficiency loss when estimating the average hazard.
 14 However, misspecification of the frailty distribution did result in efficiency loss in
 15 estimating the predicted values and standard deviation. This is especially true when
 16 the fitted frailty distribution is gamma and the true distribution is a two-point mixture,
 17 and somewhat less so when the frailty distribution is fitted nonparametrically when the

Table 5
Mean distances of estimates from parameters in model (5) for various values of σ , when $\alpha=2.0$ and $n=I=30$

σ	True	Assumed	$ \hat{\alpha} - \alpha $	$ \hat{\lambda}_i - \lambda_i $	$ \hat{\sigma} - \sigma $
0.5	Gamma	Gamma	0.09	0.24	0.09
		Nonparametric	0.09	0.27	0.10
	Two-point	Gamma	0.09	0.25 ^a	0.07
		Nonparametric	0.09	0.19 ^a	0.07
0.75	Gamma	Gamma	0.12	0.27	0.11
		Nonparametric	0.12	0.32	0.13
	Two-point	Gamma	0.13	0.26 ^a	0.08 ^a
		Nonparametric	0.12	0.12 ^a	0.06 ^a
1.0	Gamma	Gamma	0.16	0.28 ^a	0.14 ^a
		Nonparametric	0.15	0.34 ^a	0.17 ^a
	Two-point	Gamma	0.16	0.27 ^a	0.10
		Nonparametric	0.16	0.06 ^a	0.09
1.5	Gamma	Gamma	0.26	0.32 ^a	0.26
		Nonparametric	0.27	0.39 ^a	0.31
	Two-point	Gamma	0.24	0.28 ^a	0.25 ^a
		Nonparametric	0.23	0.08 ^a	0.07 ^a
1.75	Gamma	Gamma	0.45	0.46	0.47
		Nonparametric	0.47	0.51	0.52
	Two-point	Gamma	0.28	0.28 ^a	0.48 ^a
		Nonparametric	0.27	0.09 ^a	0.08 ^a

^aCases with a difference of 20% or more.

1 true distribution is gamma. The loss in efficiency increases when the cluster size n
 2 increases. When the true frailty distribution was uniform, nonparametric fitting resulted
 3 in better estimation of the standard deviation of the mixing distribution when both the
 4 number of clusters and the cluster size were large. Yet, there was a slight tendency
 5 for predicted values from the fitted gamma distribution to be closer to the true random
 6 effects than predicted values from the nonparametric model.

7 Table 5 presents additional results, examining the effect of the size of σ when the
 8 number of clusters and the size of each was fixed at 30. When the true frailty distri-
 9 bution is two-point, as σ increases fitting a gamma distribution results in increasingly
 10 poorer efficiency. This efficiency loss occurs in predicting the random effects and even
 11 more so in estimating σ . Nonparametric fitting of the frailty distribution when the true
 12 distribution is gamma also results in some efficiency loss in estimating σ and pre-
 13 dicting hazards. However, the efficiency loss is much less severe than in the reverse
 14 setting.

15 These results are consistent with those of Heckman and Singer (1984), who found
 16 that complete characterization of the mixing distribution was difficult, even when the
 17 mean and standard deviation could be accurately estimated. However, when modeling
 18 the mixture distribution nonparametrically, poor estimation of the mixing distribution
 19 did not translate into poor estimation of the corresponding marginal distribution of
 failure times.

1 5. Proposals for addressing misspecification issues

3 In the three examples of models in this article, we have seen that harmful effects
4 can result from assuming a continuous random effects distribution such as the normal
5 or gamma when actually the subjects fall into one of two quite distinct classes. In such
6 cases, a more appropriate model is a type of latent class model, assuming that subjects
7 are a mixture of two types. This suggests that in posing a model, it is wise to give
8 some careful thought to whether such a binary mixture distribution is plausible. In some
9 applications it may arise naturally, when subjects vary according to an unmeasured
10 binary factor such as gender or genetic type.

11 For instance, Follman and Lambert (1989) analyzed data on the effect of the dosage
12 of a poison on the death rate of a protozoan of a particular genus. They assumed that
13 there were two varieties (unmeasured) of that genus. Thus, they modeled the probability
14 of death at dosage level x as equal to $\rho\pi_1(x) + (1 - \rho)\pi_2(x)$, where $\text{logit}[\pi_i(x)] = \alpha_i + \beta x$
15 and ρ is unknown. The fit of this model to 426 binary observations at 8 dosage levels
16 (summarized by a deviance of 3.4 with $df = 4$) was much better than that of a single
17 logistic regression model (deviance of 24.7 with $df = 6$), which is the special case with
18 $\rho = 1$. Their example illustrates the potential discrepancy of results and efficiency loss
19 with a normal random effects assumption when a two-point mixture model fits better.
20 The two-point mixture model has $\hat{\beta} = 124.8$ with $SE = 25.2$, for which $\hat{\beta}/SE = 4.9$. The
21 normal mixture model has $\hat{\beta} = 65.5$ with $SE = 19.5$, for which $\hat{\beta}/SE = 3.4$.

22 In the absence of a theoretical framework suggesting whether the normal or a bi-
23 nary approach may be more valid, the obvious question arises about what to do to
24 diagnose and to protect oneself against potential effects of misspecification. In particu-
25 lar, the fact that the two-point distribution presented problems for the ordinary normal
26 approach with the three simple models discussed in this paper suggests that it is also
27 likely to be problematic for a wide variety of other models. In this section, we sum-
28 marize some proposals for addressing this issue that may be worth studying in future
29 research.

29 5.1. Always use a nonparametric approach

30 The safest approach might seem to be always to use a nonparametric rather than
31 a parametric approach for the random effects distribution. Although the nonparametric
32 approach is discrete, it can well approximate a normal distribution by using several
33 mass points, yet it can also accommodate the binary mixture with large variance as a
34 special case.

35 The nonparametric random effects approach does seem promising when the true ran-
36 dom effects distribution is plausibly binary. However, it has its own disadvantages. It
37 can lose some efficiency when a parametric assumption would not be badly violated.
38 Our simulation results also showed that its variance component estimate may be poor.
39 In addition, in most applications the number of mixture mass points is unknown. Thus,
40 standard asymptotic theory does not apply, and model comparison is awkward. Also,
41 identifiability problems can arise (e.g., Follman and Lambert, 1991). Finally, this ap-
42 proach is not as readily adapted to multivariate random effects modeling as the normal

1 distribution, for instance to provide a simple multivariate mixture model that has a
common variance and common correlation parameter.

3 5.2. Use a mixture of normals

5 Some authors have suggested replacing a normal random effects distribution by a
finite mixture of normals (e.g., [Everitt and Hand, 1981](#); [Magder and Zeger, 1996](#);
7 [Verbeke and Molenberghs, 2000](#)). For instance, a model with a random intercept that
might normally be assumed to have a $N(\mu, \sigma^2)$ distribution might instead be assumed
9 to have a $\rho N(\mu_1, \sigma^2) + (1 - \rho)N(\mu_2, \sigma^2)$ distribution for some mixture parameter ρ .
An appealing aspect of this proposal is that it can accommodate in a simple manner a
11 wide variety of shapes. In particular, it includes the extreme deviation from normality
of a two-point mixture distribution as the special case of the mixture of two normals
with $\sigma=0$. Thus, this would seem to protect against the occurrence of this problematic
13 distribution, for which this paper observed possible efficiency loss for all three models
with the normal assumption.

15 We tried this approach with the [Follman and Lambert \(1989\)](#) example. This is a
case where the two point-mixture model has a large estimated variance, and one would
17 hope that the mixture of normals approach would fit much better than a single normal
random intercept. The Follman and Lambert two-point mixture model gave fit

$$\hat{\pi}(x) = 0.34\hat{\pi}_1(x) + 0.66\hat{\pi}_2(x)$$

19 with

$$\text{logit}[\hat{\pi}_1(x)] = -196.2 + 124.8x, \quad \text{logit}[\hat{\pi}_2(x)] = -205.7 + 124.8x.$$

By comparison, a logistic random intercept model for which the random intercept
21 was assumed to follow a mixture of normals had estimated mixture probability of
0.34 for a $N(-196.4, 0.3)$ component and estimated mixture probability of 0.66
23 for a $N(-205.7, 0.3)$ component, with estimated slope of 124.8. With the very small es-
timated variance component, the fit is essentially the same as the two-point mixture
25 model, which fits the data much better than an ordinary logistic-normal GLMM. The
maximized log-likelihoods were -177.4 for the two-point mixture model and the mix-
27 ture of normals model, compared to -187.1 for the single normal mixture. We fitted
the mixture-of-normals random intercept model using EM with Gauss–Hermite quadra-
29 ture. Fitting was sensitive to starting values, and we checked results using marginal
maximization via simulated annealing.

31 These results are very promising for the mixture of normals approach. To get further
confirmation, we tried this approach for the model for the log odds ratio of Section
33 3, which [Table 3](#) summarized. When the true distribution was two-point, we hoped
this approach would give results similar to those with nonparametric fitting. However,
35 in the simulation study the mean absolute distance of $\hat{\beta}$ from β was very similar to
that obtained by assuming normality. Moreover, the estimate of σ tended not to be
37 as good as with the normal approach, both in this case and when the true random
effects distribution was normal. We found this disappointing, and because of it, we
39 do not feel we can give an unqualified endorsement of this approach. Further research

1 is needed to analyze why this approach failed in this case and to see if one can
2 characterize situations for which it could be expected to work as well as it seems to
3 for the Follman and Lambert (1989) example.

5.3. Comparing predicted residuals

5 For the case of the survival model, we investigated the possibility of comparing the
6 predicted frailties with what we will call *residual frailties*. The residual frailties were
7 calculated separately for each individual, assuming the mean hazard from the model to
8 be the true mean but assuming nothing further about the form of the random effects
9 distribution. In the current context, the residual frailty for an individual is the ratio of
10 the constant hazard one would calculate for that individual if one ignored the infor-
11 mation from the rest of the population relative to the population mean hazard. These
12 residual frailties are not shrunk towards 1.0 as the predicted frailties would be. We
13 considered two statistics based on comparing these residuals: (1) the average squared
14 distance between these predicted and residual frailties and (2) the average of abso-
15 lute values of the differences between the estimated variance under the assumed model
16 and the squared distance between the two residuals. The first statistic compares the
17 predicted effects under the hypothesized model with nonparametric estimates of the
18 random effects. The focus of this statistic is whether the predicted means of the hazard
19 given the observed data are accurate. On the other hand, the second statistic compares
20 the estimated conditional variances. The estimated variances for the predicted values
21 are compared with nonparametric estimates of those variances. The calculation of this
22 statistic assumes that the predicted values are accurate estimates of the conditional
23 means for each subject.

24 These two statistics were used to compare the two models fitted for the frailties in
25 this paper. They correctly selected the model with gamma frailty distribution over 80%
26 of the time when the gamma distribution was in fact the correct distribution. However,
27 these statistics had difficulty when the true distribution of the random effects was a
28 two-point distribution. For the latter case, the technique was effective only with cluster
29 sizes of at least about 45. This is too large to be useful for most applications, since
30 more commonly the number of clusters is large relative to the size of the individual
31 clusters.

32 At this stage, it is unclear how promising this approach might be with further de-
33 velopment, and additional research is needed. Alternatively, one can try to generalize
34 methods that have been proposed for using residuals in linear random effects mod-
35 els, such as those discussed by Lange and Ryan (1989) based on weighted normal
36 quantile plots of standardized linear combinations of the random effect predictions (see
37 also Houseman et al., 2002). However, as noted by Verbeke and Molenberghs (2000,
38 p. 89), since these depend on the random effects as well as the error terms, such
39 plots cannot differentiate a wrong distributional assumption for the random effects or
40 the error terms from a wrong choice of covariates. See Glidden (1999) and Shih and
41 Louis (1995) for numerical and graphical techniques for checking the adequacy of a
gamma assumption in a semiparametric gamma frailty model that allows unspecified
marginal distributions.

1 5.4. Model selection criteria

3 The choice of a random effects distribution is simply one element of the choice of a
4 model. Various criteria have been proposed for choosing among models, whether the
5 models be nested such as random effects models assuming normality or assuming a
6 mixture of normals, or nonnested such as random effects models assuming normality or
7 assuming an unspecified distribution on a finite number of mass points. One possible
8 approach would be to use one of these criteria, such as AIC, to select among a set of
9 candidate models. However, the maximized likelihood refers to the marginal distribu-
10 tion, integrating over the random effects distribution, and quite different random effects
11 distributions can generate similar marginal distributions. So, for instance, it is not clear
12 that AIC would detect cases in which the normal random effects assumption is much
13 poorer than the assumption of a binary random effect, unless the marginal fits were
quite different. It may be a more promising research problem to develop an AIC-type
measure in terms of the conditional distribution at the random effects level.

15 5.5. Other approaches

17 Alternative approaches are undergoing development currently that have promise and
18 that deserve attention in future research of effects of misspecification. For instance,
19 Chen et al. (1992) have adapted for the GLMM the semi-nonparametric approach of
20 Gallant and Nychka (1987) for which the random effects density belongs to a class of
21 smooth densities that contains a wide variety of shapes including the normal as a special
22 case. In a simulation using a logit model with a mixture of normals for a random effect,
23 this approach seemed effective in detecting the non-normality. However, its results were
24 very similar to those assuming normal random effects for that particular model. It would
25 be interesting to see if this approach does as well as the fully nonparametric approach
when the true random effects distribution is binary with large variance.

6. Summary

27 In summary, the conventional wisdom seems to be that the choice of random effects
28 distribution is not crucial. This is mainly due to studies such as Neuhaus et al. (1992)
29 and Chen et al. (2002) and articles quoted therein. It is also because for some simple
30 models, different distributions yield the same ML estimate. For instance, this happens
31 when the model with an arbitrary mixture distribution is saturated and fits perfectly
32 for data that are consistent with the model. An example is the simple logit model
33 for binary matched pairs that is a special case of (4) when $n = 1$ in each row of
34 I tables corresponding to I subjects. For it, Neuhaus et al. (1994) showed that for
35 an arbitrary mixture distribution, the same ML estimate occurs as with conditional
36 maximum likelihood (treating $\{u_i\}$ as fixed effects and eliminating them by conditioning
37 on their sufficient statistics) when the sample correlation is nonnegative in the 2×2
table cross-classifying the two observations from each pair.

1 We have seen in this paper that although the conventional wisdom is apparently often
 2 true, it does not always hold. It was not the purpose of this study to conduct a detailed
 3 investigation in terms of a great variety of models. However, we have highlighted cases
 4 for some very simple models in which misspecification can be a problem. Specifically,
 5 this can happen when the mixture distribution departs dramatically from the usual
 6 parametric choice, in the form of a two-point distribution with large variance. Although
 7 it is not surprising that severe misspecification of a random effects distribution can affect
 8 quality of prediction of characteristics involving the random effects (such as predictions
 9 of probabilities), we have seen that it can also affect fixed effects (e.g., β in Table 3).

10 Finally, the issue of modeling the random effects distribution in such a way to protect
 11 against possible poor consequences of misspecification is an important but possibly
 12 difficult one for future research. It will be interesting to see results of research directed
 13 toward some of the issues discussed in the previous section. In the absence of much
 14 guidance yet about these issues, what should a methodologist do? Lacking information
 15 about the random effects distribution, a sensible strategy seems to be to use both a
 16 parametric and a nonparametric approach. When the results from the two approaches
 17 differ substantially, caution is suggested. In some cases, such as in the [Follman and
 Lambert \(1989\)](#) example, the nature of a suspected unmeasured variable may suggest
 18 whether a normal or binary random effects distribution seems more plausible.

Acknowledgements

The research of Agresti and Caffo was partially supported by grants from NSF and NIH.

References

- 21 [Agresti, A., Hartzel, J., 2000. Strategies for comparing treatments on a binary response with multi-center
 22 data. *Statist. Med.* 19, 1115–1139.](#)
- 23 [Aitkin, M., 1999. A general maximum likelihood analysis of variance components in generalized linear
 24 models. *Biometrics* 55, 117–128.](#)
- 25 [Beitler, P.J., Landis, J.R., 1985. A mixed-effects model for categorical data. *Biometrics* 41, 991–1000.](#)
- 26 [Breslow, N.E., Clayton, D.G., 1993. Approximate inference in generalized linear mixed models. *J. Amer.
 27 Statist. Assoc.* 88, 9–25.](#)
- 28 [Chen, J., Zhang, D., Davidian, M., 2002. A Monte Carlo EM algorithm for generalized linear models with
 29 flexible random effects distribution. *Biostatistics* 3, 347–360.](#)
- 30 [Clayton, D.G., 1978. A model for association in bivariate life tables and its application in epidemiological
 31 studies of familial tendency toward choric disease. *Biometrika* 65, 141–151.](#)
- 32 [Everitt, B.S., Hand, D.J., 1981. *Finite Mixture Distributions*. Chapman & Hall, London.](#)
- 33 [Follman, D.A., Lambert, D., 1989. Generalizing logistic regression by nonparametric mixing. *J. Amer. Statist.
 34 Assoc.* 84, 295–300.](#)
- 35 [Follman, D.A., Lambert, D., 1991. Identifiability of finite mixtures of logistic regression models. *J. Statist.
 36 Plann. Inference* 27, 375–381.](#)
- 37 [Gallant, A.R., Nychka, D.W., 1987. Semi-nonparametric maximum likelihood estimation. *Econometrica* 55,
 38 363–390.](#)
- 39 [Glidden, D.V., 1999. Checking the adequacy of the gamma frailty model for multivariate failure times.
 40 *Biometrika* 86, 381–393.](#)

- 1 Hartzel, J., Liu, I., Agresti, A., 2001. Describing heterogeneous effects in stratified ordinal contingency tables,
with application to multi-center clinical trials. *Comput. Statist. Data Anal.* 35, 429–449.
- 3 Heagerty, P.J., Zeger, S.J., 2000. Marginalized multilevel models and likelihood inference. *Statist. Sci.* 15,
1–19.
- 5 Heckman, J., Singer, B., 1984. A method for minimizing the impact of distributional assumptions in
econometric models for duration data. *Econometrica* 52, 271–320.
- 7 Hougaard, P., 1986. A class of multivariate failure time distributions. *Biometrika* 73, 571–678.
- Houseman, E.A., Ryan, L.M., Coull, B.A., 2002. Rotated residuals for assessing the adequacy of a linear
9 model with correlated outcomes. Unpublished paper, Harvard School of Public Health.
- Klein, J.P., Moeschberger, M.L., Li, Y.H., Wang, S.T., 1992. Estimating random effects in the Framingham
11 heart study. In: Klein, J.P., Goel, P. (Eds.), *Survival Analysis: State of the Art*. Kluwer Academic, Boston,
MA, pp. 99–120.
- 13 Lange, N., Ryan, L., 1989. Assessing normality in random effects models. *Ann. Statist.* 17, 624–642.
- Lee, Y., Nelder, J.A., 1996. Hierarchical generalized linear models. *J. Roy. Statist. Soc. Ser. B* 58, 619–
15 678.
- Magder, L.S., Zeger, S.L., 1996. A smooth nonparametric estimate of a mixing distribution using mixtures
17 of Gaussians. *J. Amer. Statist. Assoc.* 91, 1141–1151.
- Nielsen, G.G., Gill, R.D., Andersen, P.K., Sørensen, T.I.A., 1992. A counting process approach to maximum
19 likelihood estimation in frailty models. *Scand. J. Statist.* 19, 25–43.
- Neuhaus, J.M., Hauck, W.W., Kalbfleisch, J.D., 1992. The effects of mixture distribution misspecification
21 when fitting mixed-effects logistic models. *Biometrika* 79, 755–762.
- Neuhaus, J.M., Kalbfleisch, J.D., Hauck, W.W., 1994. Conditions for consistent estimation in mixed-effects
23 models for binary matched-pairs data. *Canad. J. Statist.* 22, 139–148.
- Oakes, D., 1982. A model for association in bivariate survival data. *J. Roy. Statist. Soc. Ser. B* 44,
25 414–422.
- Shih, J.H., Louis, T.A., 1995. Assessing gamma frailty models for clustered failure time data. *Lifetime Data*
27 *Anal.* 1, 255–273.
- Verbeke, G., Molenberghs, G., 2000. *Linear Mixed Models for Longitudinal Data*. Springer, Berlin.
- 29 Whitmore, G.A., Lee, M.L.T., 1991. A multivariate survival distribution generated by an inverse Gaussian
mixture of exponentials. *Technometrics* 33, 39–50.
- 31 Wolfinger, R., O’Connell, M., 1993. Generalized linear mixed models: a pseudo-likelihood approach.
J. Statist. Comput. Simulation 48, 233–243.