



## **An Estimation Method for the Semiparametric Mixed Effects Model**

Huageng Tao; Mari Palta; Brian S. Yandell; Michael A. Newton

*Biometrics*, Vol. 55, No. 1. (Mar., 1999), pp. 102-110.

Stable URL:

<http://links.jstor.org/sici?sici=0006-341X%28199903%2955%3A1%3C102%3AAEMFTS%3E2.0.CO%3B2-3>

*Biometrics* is currently published by International Biometric Society.

---

Your use of the JSTOR archive indicates your acceptance of JSTOR's Terms and Conditions of Use, available at <http://www.jstor.org/about/terms.html>. JSTOR's Terms and Conditions of Use provides, in part, that unless you have obtained prior permission, you may not download an entire issue of a journal or multiple copies of articles, and you may use content in the JSTOR archive only for your personal, non-commercial use.

Please contact the publisher regarding any further use of this work. Publisher contact information may be obtained at <http://www.jstor.org/journals/ibs.html>.

Each copy of any part of a JSTOR transmission must contain the same copyright notice that appears on the screen or printed page of such transmission.

---

JSTOR is an independent not-for-profit organization dedicated to creating and preserving a digital archive of scholarly journals. For more information regarding JSTOR, please contact [support@jstor.org](mailto:support@jstor.org).

# An Estimation Method for the Semiparametric Mixed Effects Model

Huangeng Tao,<sup>1,2</sup> Mari Palta,<sup>2,3,\*</sup> Brian S. Yandell,<sup>1</sup> and Michael A. Newton<sup>1,3</sup>

<sup>1</sup>Department of Statistics, University of Wisconsin,  
1210 W. Dayton Street, Madison, Wisconsin 53706, U.S.A.

<sup>2</sup>Department of Preventive Medicine, University of Wisconsin,  
504 N. Walnut Street, Madison, Wisconsin 53705, U.S.A.

<sup>3</sup>Department of Biostatistics, University of Wisconsin,  
600 Highland Avenue, Madison, Wisconsin 53792, U.S.A.

\**email:* palta@cornfield.epi.wisc.edu

**SUMMARY.** A semiparametric mixed effects regression model is proposed for the analysis of clustered or longitudinal data with continuous, ordinal, or binary outcome. The common assumption of Gaussian random effects is relaxed by using a predictive recursion method (Newton and Zhang, 1999) to provide a nonparametric smooth density estimate. A new strategy is introduced to accelerate the algorithm. Parameter estimates are obtained by maximizing the marginal profile likelihood by Powell's conjugate direction search method. Monte Carlo results are presented to show that the method can improve the mean squared error of the fixed effects estimators when the random effects distribution is not Gaussian. The usefulness of visualizing the random effects density itself is illustrated in the analysis of data from the Wisconsin Sleep Survey. The proposed estimation procedure is computationally feasible for quite large data sets.

**KEY WORDS:** Generalized linear models; Longitudinal data; Mixture model; Recursion method; Random effects; Semiparametric mixed effects model.

## 1. Introduction

Linear and generalized linear mixed effects models are useful in the analysis of longitudinal or clustered data. Several estimation methods have been developed for such models with normally distributed random effects (e.g., Laird and Ware, 1982; Breslow and Clayton, 1993; Hedeker and Gibbons, 1994). The normality assumption is mathematically convenient and can be robust for estimating the fixed effects (Neuhaus and Hauck, 1992). However, estimation efficiency can be compromised when the random effects distribution is incorrectly specified. In many situations, it is of inherent interest to visualize the distribution or its density, e.g., to ascertain skewness in the outcome profile across individuals. A semiparametric mixed model utilizing nonparametric estimation of the random effects distribution allows the greatest flexibility in capturing its potentially important features.

The random effects distribution is referred to as the mixing distribution in the general mixture model. Nonparametric maximum likelihood estimation (NPMLE) is a traditional method that provides a discrete estimate of this distribution (Laird, 1978). With some identifiability assumptions, NPMLE is consistent (Kiefer and Wolfowitz, 1956). Several algorithms have been suggested for computing the NPMLE of the mixing distribution, such as those of Dempster, Laird, and Rubin (1977), Lindsay (1983), Follmann and Lambert (1989), and Lesperance and Kalbfleisch (1992). In practice, the dis-

crete NPMLE estimate can be difficult to interpret because the true random effects distribution is likely to be continuous. A smooth nonparametric maximum likelihood estimator (SNPMLE) using mixtures of Gaussians was recently proposed by Magder and Zeger (1996).

Carroll and Hall (1988) have pointed out that the optimal rate of asymptotic convergence for estimating the density of the mixing distribution is very slow. Acceptable estimates of the random effect density function are therefore obtainable only in large data sets where traditional NPMLE-related methods are computationally too burdensome to be widely used. Newton and Zhang (1999) introduce a predictive recursion algorithm (PR) to nonparametrically estimate a distribution. The PR estimate of a distribution is always continuous when the starting distribution is similarly chosen. We apply PR to semiparametric mixed effects regression models and develop a procedure for simultaneously estimating regression parameters and a random effects density. A Monte Carlo study is conducted to investigate the performance of this method. It should be noted that the current implementation of PR does not give direct estimates of individual random effects. The proposed approach is feasible for fitting semiparametric mixed models with quite large data sets. We illustrate its practical usefulness by application to data from the Wisconsin Sleep Survey (Young, Palta, and Dempsey, 1993).

## 2. Description of Our Method

### 2.1 Mixed Effects Regression Model for Longitudinal Data

Assume that there are  $N$  individuals in the study and that the  $i$ th individual is measured  $n_i$  times. Denote the response for the  $k$ th measurement of the  $i$ th individual by  $y_{ik}$ . Then the linear mixed effects regression model for the response  $y_{ik}$  is given by

$$y_{ik} = \alpha_i + \mathbf{x}'_{ik}\boldsymbol{\beta} + \epsilon_{ik} \stackrel{\text{def}}{=} u_{ik} + \epsilon_{ik}, \quad (1)$$

where  $\{\alpha_i\}$  are independently and identically distributed (i.i.d.) random effects with distribution  $F$ ,  $\mathbf{x}_{ik}$  is a  $p \times 1$  observed covariate vector, and  $\boldsymbol{\beta}$  is the vector of regression coefficients. For future use, we also let  $\mathbf{x}_i = [\mathbf{x}_{i1}, \mathbf{x}_{i2}, \dots, \mathbf{x}_{i\exists}]'$  be an  $\exists \times p$  covariate matrix of fixed effects. The regression errors  $\epsilon_{ik}$ 's are assumed to be i.i.d. as  $N(0, \sigma^2)$  and also independent of the random effect  $\alpha_i$ . Under such a linear mixed effects model, the conditional likelihood for the  $i$ th individual given  $\alpha_i$  is

$$l_i(\boldsymbol{\theta} \mid \alpha_i, \mathbf{y}_i) = \prod_{k=1}^{n_i} [(2\pi\sigma^2)^{-1/2} \exp(-(y_{ik} - u_{ik})^2/2\sigma^2)], \quad (2)$$

where  $\boldsymbol{\theta} = \{\boldsymbol{\beta}, \sigma\}$  and  $\mathbf{y}_i = (y_{i1}, y_{i2}, \dots, y_{i\exists})'$ .

To model ordinal response data, we use the threshold concept as described by Hedeker and Gibbons (1994). Assume that the continuous response  $y_{ik}$  in equation (1) is latent and classified into  $J$  categories by the underlying threshold values  $\boldsymbol{\tau} = \{\tau_0 = -\infty, \tau_1, \tau_2, \dots, \tau_{J-1}, \tau_J = \infty\}$ . Then the conditional probability that  $y_{ik}$  is in the  $j$ th category given the random effect  $\alpha_i$  is

$$P(\tau_{j-1} \leq y_{ik} < \tau_j \mid \alpha_i, \boldsymbol{\beta}, \boldsymbol{\tau}, \sigma) = \Phi[(\tau_j - u_{ik})/\sigma] - \Phi[(\tau_{j-1} - u_{ik})/\sigma], \quad (3)$$

where  $\Phi(\cdot)$  represents the cumulative standard normal distribution function. As

$$P(y_{ik} < \tau_j \mid \alpha_i, \boldsymbol{\beta}, \boldsymbol{\tau}, \sigma) = \Phi[(\tau_j - u_{ik})/\sigma],$$

this model is a generalized linear model with probit link for the observed response dichotomized by each threshold or a cumulative probit model for the ordinal response (see, e.g., McCullagh and Nelder, 1989). All the parameters in model (3) are not simultaneously identifiable because the origin and unit of  $u_{ik}$  can be chosen arbitrarily with corresponding adjustment of  $\tau_1$  and  $\sigma$ . For convenience, we usually apply the convention  $\tau_1 = 0$  and  $\sigma = 1$ . Then the conditional likelihood function for the  $i$ th individual given  $\alpha_i$  in this ordinal response regression model is

$$l_i(\boldsymbol{\theta} \mid \alpha_i, \mathbf{y}_i) = \prod_{k=1}^{n_i} \prod_{j=1}^J [\Phi(\tau_j - u_{ik}) - \Phi(\tau_{j-1} - u_{ik})]^{\delta_{ikj}}, \quad (4)$$

where  $\boldsymbol{\theta} = \{\boldsymbol{\beta}, \boldsymbol{\tau}\}$  and

$$\delta_{ikj} = \begin{cases} 1 & \text{if } y_{ik} \in j\text{th category} \\ 0 & \text{otherwise.} \end{cases}$$

Alternatively, when the logistic distribution is assumed for the regression error  $\epsilon_{ik}$ , the standard normal distribution function  $\Phi(\cdot)$  in equation (3) can be replaced by the logistic distribu-

tion function

$$\Psi(\tau_j - u_{ik}) = (1 + \exp[-(\tau_j - u_{ik})])^{-1}.$$

In a parallel way, we then obtain a binary logit model for observed responses dichotomized at any threshold or a cumulative logit model for the ordinal responses (McCullagh and Nelder, 1989).

Some remarks are in order regarding the identifiability of the mixing, or random effects, distribution. Teicher (1961) showed that the mixing distribution in a binomial response model without any covariate  $x_{ik}$  is not identifiable. Follmann and Lambert (1989) indicated that the maximum number of identifiable points of a discrete mixing distribution in a binary response model is positively related to the number of distinct  $x_{ik}$ . In practice, estimating the random effects distribution is more meaningful when the fixed covariates have many distinct values.

### 2.2 Estimating the Distribution of $\alpha_i$ 's

In this section, we use the likelihood functions  $l_i(\boldsymbol{\theta} \mid \alpha_i, \mathbf{y}_i)$  given by equation (2) for the linear mixed effects model (or equation (4) for the ordinal mixed effects model) to estimate the density of the random effects  $\alpha_i$ 's by the PR method (Newton and Zhang, 1999).

For a given  $\boldsymbol{\theta}$ , the PR estimate of the density function of the random effects  $f_{\boldsymbol{\theta}}(\cdot)$  is calculated by the following recursive equation from  $i = 1$  to  $N$  (randomly ordered):

$$f_{\boldsymbol{\theta}}^i(\alpha) = (1 - w_i)f_{\boldsymbol{\theta}}^{i-1}(\alpha) + w_i \frac{f_{\boldsymbol{\theta}}^{i-1}(\alpha) \cdot l_i(\boldsymbol{\theta} \mid \alpha, \mathbf{y}_i)}{c_i(\boldsymbol{\theta})}. \quad (5)$$

The user-supplied weights  $w_i$  must decrease as  $i$  increases (see Appendix). The final PR step gives  $f_{\boldsymbol{\theta}}^N(\cdot)$ , which estimates the unknown density of the random effects  $\alpha_i$ . Note that the second term on the right-hand side of (5) is the Bayesian posterior density function of the  $\alpha_i$ 's given the  $i$ th observation with the prior density  $f_{\boldsymbol{\theta}}^{i-1}(\alpha)$ . The denominator  $c_i(\boldsymbol{\theta})$  is the normalizing constant of the posterior density. Hence, the PR algorithm averages the density estimate from the previous step and the posterior density given the current observation. Although  $f_{\boldsymbol{\theta}}^N$  depends on the order by which observations are processed, this dependence may be weak and often can be ignored.

As discussed in the Appendix, we allow the weights to be functions of a parameter  $\rho$ , defined on the interval  $(0, 1]$ , and fit this parameter by our algorithm. The amount of smoothing of the posterior is influenced by  $\rho$ . A choice or estimate close to 1 leads to a posterior distribution less influenced by individual observation vectors and hence closer to the prior. The prior distribution can be chosen as, e.g., uniform or Gaussian. We show in our Monte Carlo studies in Section 3.0 that estimation properties are quite insensitive to this choice. A key feature of (5) is that an approximation to a Dirichlet process posterior is produced in  $O(N)$  steps (Newton, Quintana, and Zhang, 1998).

Newton and Zhang (1999) prove the consistency of the PR algorithm under certain conditions for estimating a survival function under interval censoring. Our Monte Carlo results (see Section 3.1) support a similar result for mixed effects models.

2.3 Estimating  $\theta$

The marginal likelihood given the observed data,  $\mathbf{y}_i$ ,  $i = 1, \dots, N$ , is

$$L(\theta, f | \mathbf{y}_i, i = 1, \dots, N) = \prod_{i=1}^N \int_a^b l_i(\theta | \alpha, \mathbf{y}_i) f(\alpha) d\alpha. \quad (6)$$

To estimate the parameter vector  $\theta$ , we can construct the log-profile likelihood by replacing the true  $f(\alpha)$  in (6) with the PR estimate. Details are given in the Appendix.

2.4 Estimating the Standard Error of Parameter Estimates

The standard way to estimate the asymptotic variance of maximum likelihood estimators is by the inverse of the observed information matrix. For parametric models, the profile likelihood is known to have a biased information matrix and a biased score test statistic. Some adjustments can be made to reduce such bias to  $O(1/N)$  (McCullagh and Tibshirani, 1990; Diccio and Martin, 1996). However, there are very few theoretical results regarding the profile likelihood in the semiparametric model. We use the bootstrap method to estimate the standard error of parameter estimates (Efron and Tibshirani, 1993). The original data contain  $N$  observations,  $\{(\mathbf{y}_i, \mathbf{x}_i), i = 1, \dots, N\}$ , which are treated as the population. Then  $B$  bootstrap data sets of size  $N$  pairs are randomly sampled with replacement from this population. The sample standard deviation of the parameter estimates for the  $B$  bootstrap sets estimates the true standard error.

3. Simulation Studies

3.0 Description of the Simulation Experiment

We conducted a simulation study to compare the parameter estimates of the semiparametric with those of the Gaussian mixed model. This simulation experiment is similar to that in Magder and Zeger (1996), which allowed the sample size and the mixing distribution to vary. Specifically, data were simulated in a two-stage process. First,  $N$  different independent scalar random variables,  $\alpha_i$ , were simulated from a mixing distribution  $F$  as described below. Then the  $i$ th cluster of data was computed as in model (1) with  $\beta = (\beta_1, \beta_2) = (2, 5)$  and  $\epsilon_i$  a vector generated by independent Gaussians with mean 0 and variance 4. The cluster size was fixed at 4. The terms  $\mathbf{x}_i$  were  $(4 \times 2)$  matrices with the first column drawn independently from the standard Gaussian distribution for each  $i$  and the second column set to 0 for  $i = 1, \dots, N/2$  and set to 1 for  $i = N/2 + 1, \dots, N$ . Thus,  $\beta_1$  represents the effect of a within cluster covariate and  $\beta_2$  the effect of a cluster specific covariate.

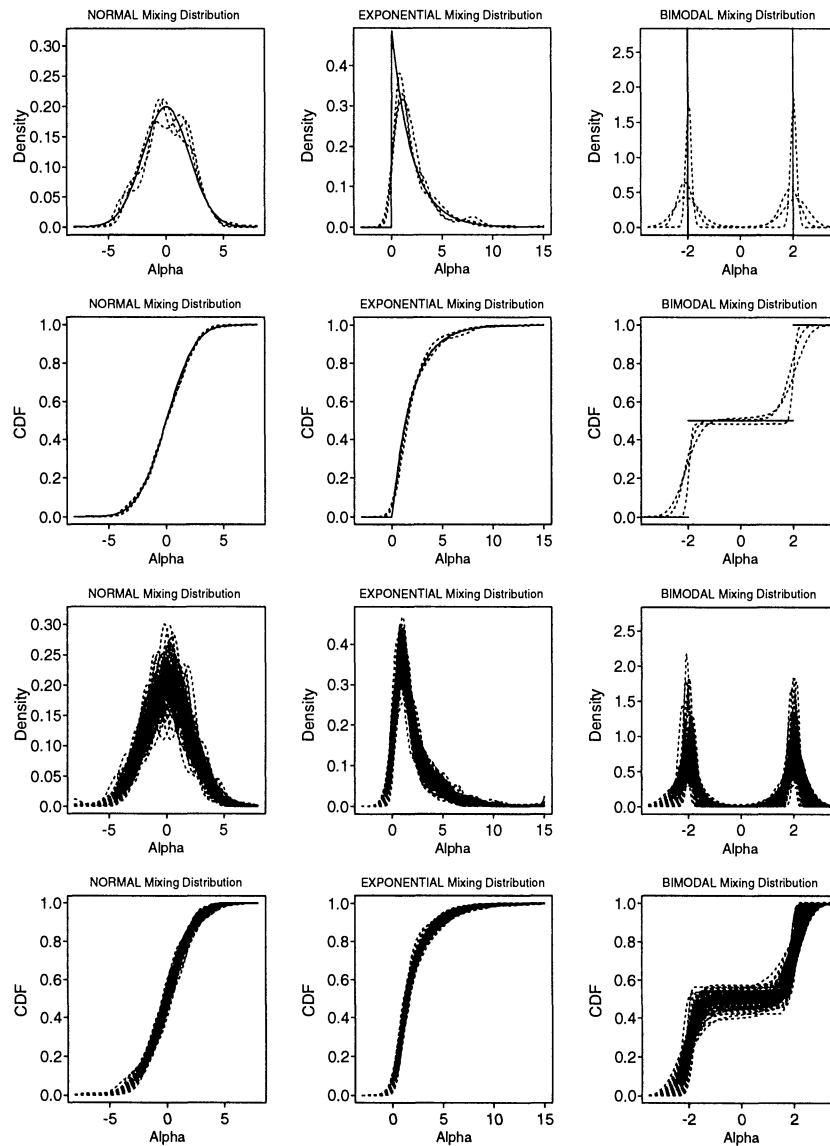
The sample size and mixing distribution were varied to result in six different types of simulated data sets. We considered small samples consisting of 50 independent clusters and large samples consisting of 600 independent clusters. Three different distributions with variance 4 were considered for the random effect  $\alpha_i$ : a Gaussian distribution with mean zero, a skewed exponential distribution with mean 2, and a discrete distribution with equal point masses placed at  $-2$  and 2. One hundred data sets of each type were simulated.

Table 1  
Simulation results

Mixing distribution	Model	Mean(ISE)	MSE( $\hat{\beta}_1$ )	MSE( $\hat{\beta}_2$ )	MSE( $\hat{\sigma}^2$ )	$\bar{\rho}$	$-\log(L)$
Small data sets ( $N = 50$ )							
Gaussian	Gaussian	.0354	.0209	.3523	.1986	—	460.06
	Semipara1	.0513	.0221	.4141	.2055	.8266	458.62
	Semipara2	.0532	.0222	.4346	.2030	.7581	458.37
Exponential	Gaussian	.0825	.0211	.4263	.2211	—	458.45
	Semipara1	.0425	.0209	.2407	.2291	.7883	452.62
	Semipara2	.0434	.0214	.2551	.2349	.7427	452.48
Bimodal	Gaussian	.2360	.0188	.4449	.1525	—	459.79
	Semipara1	.1200	.0176	.1541	.2035	.5380	451.09
	Semipara2	.1217	.0178	.1568	.2060	.5199	451.15
Large data sets ( $N = 600$ )							
Gaussian	Gaussian	.0027	.0023	.0293	.0156	—	5551.8
	Semipara1	.0050	.0024	.0307	.0160	.8511	5550.2
	Semipara2	.0049	.0024	.0312	.0159	.8385	5549.5
Exponential	Gaussian	.0543	.0019	.0301	.0178	—	5548.0
	Semipara1	.0072	.0018	.0213	.0191	.8145	5476.9
	Semipara2	.0071	.0019	.0207	.0191	.8046	5475.9
Bimodal	Gaussian	.2070	.0017	.0445	.0184	—	5552.7
	Semipara1	.0478	.0016	.0109	.0216	.5076	5450.5
	Semipara2	.0486	.0016	.0110	.0218	.5034	5450.7

Semipara1: with a normal starting distribution in PR.

Semipara2: with a uniform starting distribution in PR.



**Figure 1.** Each plot in the first row contains three randomly selected density estimates (dashed lines) as well as the true density (solid lines) for one type of mixing distribution. The plots in the second row are the corresponding distribution functions. The plots in the two bottom rows present the estimated density curves and CDFs from 100 data sets ( $N = 600$ ) for each mixing distribution.

These simulated data sets were fit by three approaches: (1) SAS PROC MIXED (Littell et al., 1996) for the Gaussian mixed effects model, (2) the semiparametric mixed model using a Gaussian form (1) for the starting density  $f^0(\alpha)$  in PR, and (3) the semiparametric mixed model using a uniform starting distribution on the interval  $(a, b)$  with  $a$  and  $b$  chosen as  $\text{mean} \pm 3 \text{ SD}$  from (1).

### 3.1 Results for Estimating the Mixing Distribution

To evaluate the quality of the estimates of the mixing distribution, we computed the integrated squared error (ISE) between the estimated distribution and the true distribution,  $\int [\hat{F}(\alpha) - F(\alpha)]^2 d\alpha$ . Table 1 summarizes the results for each type of data set. The average ISE of the estimated mixing

distributions of the Gaussian mixed model were smaller than those of the semiparametric mixed model when the true distribution was Gaussian. For the other two types of mixing distribution, the semiparametric mixed model gave much better estimates (smaller ISEs) than did the Gaussian mixed model. The ISE values from the semiparametric mixed model were smaller in the large sample cases, lending support to the conjecture that PR is asymptotically consistent for estimating the mixing distribution. The semiparametric estimation procedure performs very similarly with the two different starting distributions. Figure 1 gives the plots of the density estimates and the distribution function estimates (CDF estimates) for the larger sample data sets using the uniform starting distribution. Both the density estimates and

the CDF estimates capture the basic features of the true distribution in each case. The variation of the CDF estimates tends to be small when the true distribution is continuous.

3.2 *Estimates of Fixed Effects*

The mean squared errors (MSEs) in estimation of  $\beta_1$  (the time-dependent effect) in Table 1 are very similar in the Gaussian and semiparametric mixed models. There were greater differences in the MSEs of  $\beta_2$  (the time-invariant effect) between the two models. With the discrete bimodal mixing distribution, the MSEs of estimates of  $\beta_2$  from the Gaussian mixed model are almost three times those from the semiparametric mixed model. This is consistent with findings of Magder and Zeger (1996). Note that the weight parameter  $\rho$  is smaller on average for the discrete bimodal mixing distribution case. This is consistent with  $\rho$  playing a role in adjusting the smoothness of the estimated distribution. Additional simulation results (not shown here) obtained by setting  $\rho = 1$  resulted in larger MSEs. Simulations for ordinal mixed effects models (also not shown) led to conclusions similar to those for the mixed effects models with continuous outcome. The average loglikelihood from the semiparametric mixed model is always greater than that from the Gaussian mixed model, which reflects a better fit with the former.

3.3 *Other Simulation Results*

Magder and Zeger (1996) compared several estimation methods for the semiparametric mixed model by a Monte Carlo study with a model similar to that in equation (1). However, they allowed the cluster size to vary from 1 to 6 and used the skewed mixing distribution  $[.25 \cdot N(\mu = 14, \sigma^2 = 10) + .75 \cdot \chi^2(4)] / 4 / \sqrt{109}$  instead of the exponential distribution in Table 1. For each mixing distribution, they simulated 400 data sets with the sample size equal to 180 clusters. We compared the results of using our algorithm on data sets simulated as in their study with their results using NPMLE and SNPMLE. Table 2 indicates that, for all three mixing distributions, the PR-based estimators are either best or second best for recovering the regression parameters and the random effects distribution.

4. **Analysis of the Wisconsin Sleep Survey Data**

We illustrate the application of the semiparametric ordinal logistic regression model in the analysis of longitudinal data from the Wisconsin Sleep Survey (Young et al., 1993). Questions on the frequency and loudness of snoring were included as indicators of possible sleep-disordered breathing. We examined responses to the question, "How often do you snore according to what others have told you?" with possible answers: 0 = Never snore, 1 = Only snore once or a few times ever, 2 = Snore a few nights per month, 3 = Snore at least once a week, 4 = Snore several (3-5) nights per week, and 5 = Snore every night or almost every night. This or similar questions on snoring have been extensively used in sleep research. It is generally believed by researchers in the field that many people are not themselves aware of snoring and that relying on the subject having been told by others that they snore yields a more accurate assessment. A cohort of 3360 state employees, originally of age 30-60, responded twice, on average 4 years apart. The frequency of each response to the question in the two survey waves is given in the second and third columns of Table 3. In the first survey, 3096 individuals answered the question as well as questions pertaining to relevant covariates;

**Table 2**  
*Comparison of methods*

Mixing distribution	Method	Mean (ISE)	MSE ( $\hat{\beta}_1$ )	MSE ( $\hat{\beta}_2$ )
Moderate sample size ( $N = 180$ )				
Gaussian	Gaussian <sup>a</sup>	.0099	.0107	.1281
	NPMLE <sup>a</sup>	.0776	.0107	.1759
	SNPMLE <sup>a</sup>	.0166	.0107	.1421
	PR	.0157	.0107	.1397
Skewed	Gaussian <sup>a</sup>	.0580	.0109	.1356
	NPMLE <sup>a</sup>	.0759	.0107	.1142
	SNPMLE <sup>a</sup>	.0184	.0106	.0988
	PR	.0195	.0105	.0880
Bimodal	Gaussian <sup>a</sup>	.2145	.0083	.1211
	NPMLE <sup>a</sup>	.0877	.0076	.0439
	SNPMLE <sup>a</sup>	.1326	.0079	.0557
	PR	.0814	.0076	.0505

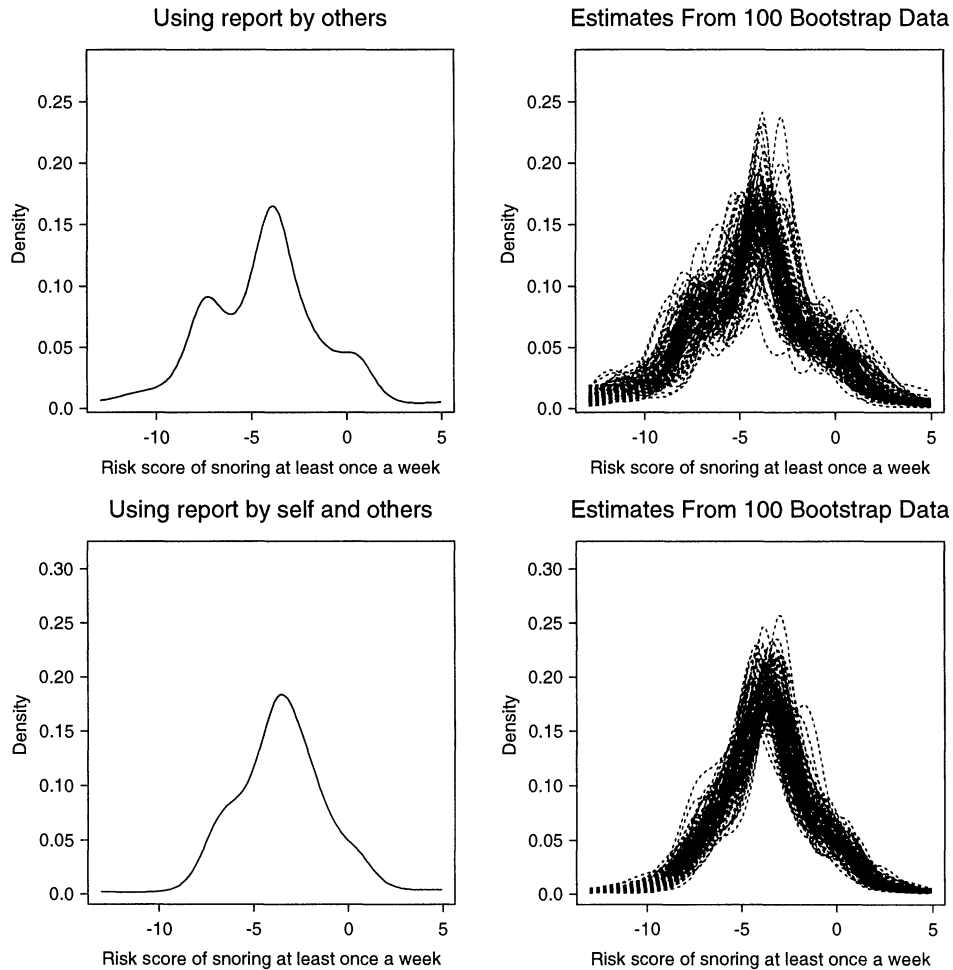
<sup>a</sup> These are from Table 3 of the paper by Magder and Zeger (1996).

in the second survey, 3034 individuals provided complete answers. About 30% of respondents went up by at least one snoring frequency category between surveys, and about 20% went down by at least one step. The large percentage changing in both directions points to the possible presence of random reporting error.

We considered the three covariates found to be the most predictive of sleep-disordered breathing in previous analyses (Young et al., 1993): sex (1 = female; 0 = male), age (in years), and body mass index (BMI, defined as body weight in kilograms divided by the square of height in meters). We also included the interaction between sex and age (calculated by  $\text{sex} \cdot (\text{age} - 40)$ ). We fit an ordinal logistic model using our algorithm with the following results. The estimated threshold parameters and  $\rho$  (with standard errors) were  $\{\hat{\tau}_2 = 1.700(0.069), \hat{\tau}_3 = 4.224(0.122), \hat{\tau}_4 = 5.627(0.136), \hat{\tau}_5 = 7.362(0.170), \rho = 0.759(0.080)\}$ . The estimated regression coefficients of sex, age, BMI and  $\text{sex} \cdot (\text{age} - 40)$  were, respectively,  $-1.636(0.134)$ ,  $0.100(0.010)$ ,  $0.264(0.012)$ , and  $-0.057(0.013)$ . Thus, snoring level increases with increasing

**Table 3**  
*Frequency (%) of individuals at each snoring level in repeat waves of Wisconsin Sleep Survey*

Category	As told by others		As told or known by self	
	Survey 1	Survey 2	Survey 1	Survey 2
0	754 (24.4)	613 (20.2)	502 (15.2)	405 (12.3)
1	401 (13.0)	396 (13.1)	735 (22.3)	756 (22.9)
2	717 (23.2)	663 (21.9)	815 (24.7)	754 (22.8)
3	356 (11.5)	368 (12.1)	376 (11.4)	387 (11.7)
4	358 (11.6)	400 (13.2)	360 (10.9)	404 (12.2)
5	510 (16.5)	594 (19.6)	510 (15.5)	594 (18.0)
Total	3096	3034	3298	3300



**Figure 2.** Density functions of individual risk scores for snoring for females at age = 40 and BMI = 27. The two top plots are based on report of snoring by others. The plots in the bottom panels are based on reports by both self and others. Each of the two left panels gives the density estimate based on the data, and the corresponding right panels give the density estimates for 100 bootstrap samples.

age (especially in males) and BMI. Overall, males tend to snore more frequently than females. Based on the ordinal regression model as described in Section 2.1, the risk score of Ever snoring for the  $i$ th individual is  $\alpha_i + \mathbf{x}_i\beta$ . Risk scores for snoring with other frequencies are obtained by adding the appropriate  $\tau_j$ . As the semiparametric model provides estimates of  $\tau_j$ ,  $\beta$ , and the density of  $\alpha_i$ , we can obtain the risk score density at any snoring frequency and fixed covariate values  $\mathbf{x}$ . The top left panel in Figure 2 gives the plot of the estimated risk score density function with sex = female, age = 40 and BMI = 26.94 (the average BMI values for females with age = 40). We note that this plot displays two peaks. The top right panel shows the results of 100 bootstrap samples of the data. Most of these curves display the bimodality obtained from the original sample, indicating that this feature is unlikely to have arisen by chance. Further examination indicated that subjects never being told they snore contributed heavily to the lower mode of the distribution. This pointed to the possibility that

some subjects who snore might not have been told of this by others.

A secondary question on the survey inquired, “Aside from what others have told you, how often, if ever, have you had the feeling or awareness that you have been snoring?” Responses to this question were scored in the same categories as those described previously. As we suspected that individuals are likely to snore with at least the frequency reported by either themselves or others, we reran the analysis using the maximum of the two responses. The frequency distribution of this new outcome variable at the two survey occasions is given in the last two columns of Table 3. The response distribution was quite similar to that of the original variable, with the exception of more responses of Few times ever as compared to Never snore. Further analysis showed the shift to be stronger for women. Combining the questions also eliminated some missing data. Estimated threshold parameters and  $\rho$  for the new outcome were (with standard errors)  $\{\hat{\tau}_2 = 2.67(0.080), \hat{\tau}_3 = 4.82(0.112), \hat{\tau}_4 = 5.99(0.140),$

$\hat{\tau}_5 = 7.45(0.172)$ ,  $\rho = 0.781(0.065)$ . The estimated regression coefficients of sex, age, BMI, and the interaction of (age - 40) by sex were, respectively,  $-1.45(0.132)$ ,  $0.072(0.0076)$ ,  $0.20(0.011)$ , and  $-0.042(0.011)$ . Thus, sex and BMI had somewhat weaker coefficients. With the new outcome variable, we find that the left peak of the risk score distribution is no longer present (see the bottom left panel in Figure 2). We conclude that relying on a report of snoring only by others tends to differentially misclassify individuals who snore relatively rarely.

The analysis of the Wisconsin Sleep Survey data demonstrates the use of the random effects density estimate to identify misclassification in an ordinal response. Thus, in this example, the plot is interpreted as the distribution of the sum of between-individual variation in risk and individual-averaged measurement error. In other applications, a similar bimodality might reflect the omission from the regression model of an important binary covariate, which creates the appearance of two risk groups. On the basis of the knowledge of the development and risk factors for sleep-disordered breathing, such an interpretation is not believable here.

## 5. Discussion

We have proposed a semiparametric mixed effects model for the analysis of longitudinal continuous or ordinal data. Traditional nonparametric methods for estimating the random effects distribution are computationally burdensome. The PR method is a new alternative for providing a smooth nonparametric estimate of a distribution. Although the general asymptotic theory for the PR method is still under development, the existing work (Newton and Zhang, 1999) lends support for expecting the method to have desirable properties. Our Monte Carlo results provide empirical evidence of consistency for the particular models examined here. As anticipated, the semiparametric method improves the mean squared error of the fixed effects estimators in situations with nonnormally distributed random effects. Our simulation studies indicate that there is little loss compared to estimators assuming normality even when the normality assumption is true.

The primary advantage of our method is in being able to visualize the distribution of the random effects. If the link function and form of relationship between covariates and outcome are correct or predefined, the random effects distribution shows the distribution of risk scores in the population. This is often of interest from a public health point of view, as one might want to know whether, even conditioning on the known covariates, a fraction of the population has very high risk scores while most cluster at low risk or whether risk scores tend to be more or less symmetrically distributed across the population. A bimodal distribution might indicate the presence of separate, hitherto unknown risk groups. On the other hand, the distribution can be used as a check of model fit, as it is akin to the distribution of residuals in ordinary regression analysis. Unusual or unexpected risk score patterns might point to erroneous functional relationships (which can then be examined by, e.g., transformation of covariates) or differential measurement errors in the response (as in our example and in Verbeke and Lesaffre, 1996).

We were able to analyze quite large data sets with both ordinal and continuous response with reasonable computational speed. For example, it took only about 30 seconds on a Sun

Sparc 10 station to run a data set of 180 clusters in the Monte Carlo study in Section 3.3.

Although computational efficiency of our methods is quite high for the estimation aspect, the bootstrap evaluation of the standard errors of coefficients is somewhat time consuming, especially for ordinal response. It is possible that standard errors can be approximated by utilizing the information matrix from the partial likelihood. We opted not to pursue this approach, as it lacks theoretical basis. For practical application, we recommend at this point to use methods based on either GEE (Liang and Zeger, 1986) or normal random effects for approximate model development. Note that, for the linear model, marginal GEE-based estimators are consistent also for regression coefficients based on a mixed effects model (Zeger, Liang, and Albert, 1988). Candidates for final models can be checked and fully described in terms of random effects distribution by the proposed method.

Monte Carlo results indicate that the fit of the PR distribution estimate is better than or comparable to that of some other NPMLE-related methods, including the SNPMLE of Madger and Zeger (1996). The latter method was not yet extended to the case of ordinal outcome. SNPMLE has the advantage of providing a parameterization of the mixing distribution, facilitating statistical inference. Our approach has the potential advantage of greater flexibility in the mixing distribution, and this might be preferable when graphical assessment of the distribution and of model fit are the primary concern. We extended PR by estimating a smoothing parameter  $\rho$  from the data. Our Monte Carlo results indicate that this improves the fit for truly discrete or multimodal mixing distributions. SNPMLE requires prespecification of a smoothing parameter  $h$ . Madger and Zeger give an overview of other related methods that can be used for smooth nonparametric fitting of mixing distributions, including the work of Laird and Louis (1989), Stefanski and Carroll (1990), Zhang (1990), and Davidian and Gallant (1992). A comprehensive comparison of these methods, some of which apply in situations slightly different from ours, is beyond the scope of this paper.

## ACKNOWLEDGEMENTS

This work was supported by grant PO1 HL42242 from the National Heart Lung and Blood Institute and grant R01 CA53786 from the National Cancer Institute.

## RÉSUMÉ

On propose un modèle de régression semi-paramétrique à effets mixtes pour l'analyse de données longitudinales ou en cluster, avec une variable d'intérêt continue, ordinale, ou binaire. On s'affranchit de l'hypothèse habituelle d'effets aléatoires gaussiens en utilisant la méthode de prédiction récursive partielle (Newton et Zhang, 1996) pour obtenir un estimateur de densité non paramétrique lissé. Une nouvelle stratégie est introduite pour accélérer l'algorithme. Les estimations des paramètres sont obtenues en maximisant la vraisemblance profilée marginale par la méthode de recherche de la direction conjuguée due à Powell. Des résultats de Monte Carlo sont présentés pour montrer que la méthode améliore l'erreur quadratique moyenne des estimations des effets fixes, lorsque la distribution des effets aléatoires n'est pas gaussienne. L'utilité de la visualisation de la densité des effets aléatoires est illustrée par l'analyse des données provenant du Wisconsin Sleep Survey. La procédure d'estimation proposée est aisément



mise en œuvre du point de vue des calculs, pour de larges ensembles de données.

## REFERENCES

- Breslow, N. E. and Clayton, D. G. (1993). Approximate inference in generalized linear mixed models. *Journal of the American Statistical Association* **88**, 9–25.
- Carroll, R. J. and Hall, P. (1988). Optimal rates of convergence for deconvolving a density. *Journal of the American Statistical Association* **83**, 1184–1186.
- Davidian, M. and Gallant, A. R. (1992). Smooth nonparametric maximum likelihood estimation for population pharmacokinetics, with application to quinidine. *Journal of Pharmacokinetics and Biopharmaceutics* **20**, 529–556.
- Dempster, A. P., Laird, N. M., and Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society, Series B* **39**, 1–38.
- Diciccio, T. J. and Martin, M. A. (1996). Information bias and adjusted profile likelihoods. *Journal of the Royal Statistical Society, Series B* **58**, 189–203.
- Efron, B. and Tibshirani, R. (1993). *An Introduction to the Bootstrap*. London: Chapman and Hall.
- Follmann, D. A. and Lambert, D. (1989). Generalizing logistics regression by nonparametric mixing. *Journal of the American Statistical Association* **84**, 295–300.
- Hedeker, D. and Gibbons, R. D. (1994). A random-effects ordinal regression model for multilevel analysis. *Biometrics* **50**, 933–953.
- Kiefer, J. and Wolfowitz, J. (1956). Consistency of the maximum likelihood estimator in the presence of infinitely many incidental parameters. *The Annals of Mathematical Statistics* **27**, 887–906.
- Laird, N. (1978). Nonparametric maximum likelihood estimation of a mixing distribution. *Journal of the American Statistical Association* **73**, 805–811.
- Laird, N. M. and Louis, T. A. (1989). Smoothing the nonparametric estimate of a prior distribution by roughening: A computational study. In *ASA Proceedings of Business and Economic Statistics Section*, 613–620.
- Laird, N. M. and Ware, J. H. (1982). Random effects models for longitudinal data. *Biometrics* **38**, 963–974.
- Lesperance, M. L. and Kalbfleisch, J. D. (1992). An algorithm for computing the nonparametric MLE of a mixing distribution. *Journal of the American Statistical Association* **87**, 120–126.
- Liang, K. Y. and Zeger, S. L. (1986). Longitudinal data analysis using generalized linear models. *Biometrika* **73**, 13–22.
- Lindsay, B. G. (1983). The geometry of mixture likelihoods: A general theory. *The Annals of Statistics* **11**, 86–94.
- Littell, R. C., Milliken, G. A., Stroup, W. W., and Wolfinger, R. D. (1996). *SAS System for Mixed Models*. Cary, North Carolina: SAS Institute.
- McCullagh, P. and Nelder, J. A. (1989). *Generalized Linear Models*. London: Chapman and Hall.
- McCullagh, P. and Tibshirani, R. (1990). A simple method for the adjustment of profile likelihoods. *Journal of the Royal Statistical Society, Series B* **52**, 325–344.
- Magder, L. S. and Zeger, S. L. (1996). A smooth nonparametric estimate of a mixing distribution using mixtures of Gaussians. *Journal of the American Statistical Association* **91**, 1141–1152.
- Neuhauser, J. M. and Hauck, W. W. (1992). The effects of mixture distribution misspecification when fitting mixed-effects logistic models. *Biometrika* **79**, 755–762.
- Newton, M. A. and Zhang, Y. (1999). A recursive algorithm for non-parametric analysis with missing data. *Biometrika*, in press.
- Newton, M. A., Quintana, F. A., and Zhang, Y. (1998). Nonparametric Bayes methods using predictive updating. In *Practical Nonparametric and Semiparametric Bayesian Statistics*, P. Muller, D. Sinha, and D. Dey (eds), 45–61. New York: Springer-Verlag.
- Press, W. H., Teukolsky, S. A., Vetterling, W. T., and Flannery, B. P. (1992). *Numerical Recipes: The Art of Scientific Computing*. Cambridge: Cambridge University Press.
- Stefanski, L. and Carroll, R. J. (1990). Deconvoluting kernel density estimators. *Statistics* **21**, 169–184.
- Teicher, H. (1961). Identifiability of mixtures. *The Annals of Mathematical Statistics* **32**, 244–248.
- Verbeke, G. and Lesaffre, E. (1996). A linear mixed-effect model with heterogeneity in the random-effects population. *Journal of the American Statistical Association* **91**, 217–221.
- Young, T., Palta, M., Dempsey, J., Skatrud, J., Weber, S., and Badr, S. (1993). The occurrence of sleep-disordered breathing among middle-aged adults. *New England Journal of Medicine* **328**, 1230–1235.
- Zeger, S. L., Liang, K. Y., and Albert, P. S. (1988). Models for longitudinal data: A generalized estimating equation approach. *Biometrics* **44**, 1049–1060.
- Zhang, C. (1990). Fourier methods for estimating mixing densities and distributions. *The Annals of Statistics* **18**, 806–831.

Received January 1997. Revised November 1997.

Accepted January 1998.

## APPENDIX

The sequence of weights  $w_i$  in (5) should decrease to zero as the sample size  $N$  goes to infinity. Following principles of Newton and Zhang (1999), the sequence can be chosen as  $w_i = 1/(i+1)^\rho$ , for  $i = 1, \dots, N$ , where  $\rho$  is a constant between 0 and 1. The convergence rate of the PR algorithm, as  $N \rightarrow \infty$  depends on the weights  $w_i$ , as does the amount of smoothing in finite samples. A choice of  $\rho$  close to 1 gives less weight to individual observations and thus produces an estimate closer to the smooth prior. Consequently, the choice of  $\rho$  influences the estimate of the model parameter  $\theta$  as well as the estimate of the random effects distribution. However, a uniformly best choice of the sequence  $w_i$  does not appear to exist. To address this problem, we extend the PR algorithm by incorporating  $\rho$  as a parameter into (5) and choose the

best  $\rho$  based on the data. Letting  $\theta' = \{\theta, \rho\}$ , (5) is modified as follows:

$$f_{\theta'}^i(\alpha) = (1 - w_i(\rho)) f_{\theta'}^{i-1}(\alpha) + w_i(\rho) \frac{f_{\theta'}^{i-1}(\alpha) \cdot l_i(\theta | \alpha, \mathbf{y}_i)}{c_i(\theta')}, \quad (7)$$

where  $w_i(\rho) = \frac{1}{(i+1)^\rho}$  and  $c_i(\theta')$  is defined as

$$c_i(\theta') = \int_a^b f_{\theta'}^{i-1}(\alpha) \cdot l_i(\theta | \alpha, \mathbf{y}_i) d\alpha. \quad (8)$$

The starting density  $f_{\theta'}^0(\alpha)$  is a prior guess of the true random effect distribution. In our simulation studies, it was chosen as a uniform distribution on an interval (a, b), which approximately supports all the probability mass of the random effects distribution  $F$  or as a Gaussian distribution obtained from the Gaussian mixed effects model. Reasonable lower and upper bounds for the interval of integration can be obtained by estimating the mean and standard deviation of the distribution by the Gaussian mixed effects model, e.g., the lower (upper) bound = mean  $\pm$  3 SD. Note that in consideration of the Gauss–Legendre quadrature integration method (see below), we confine the support of  $f$  to a finite interval. The weight parameter  $\rho$  is estimated together with the model parameter  $\theta$  by maximizing the log-profile marginal likelihood function:

$$M(\theta' | \mathbf{y}_i, i = 1, \dots, N) = \sum_{i=1}^N \log \int_a^b l_i(\theta | \alpha, \mathbf{y}_i) f_{\theta'}^N(\alpha) d\alpha. \quad (9)$$

Although the log-profile likelihood (9) is a continuous function of  $\theta'$ , it involves recursive functions. Thus, it is very difficult to obtain an explicit form for its derivatives. We use Powell's conjugate direction search method (Press et al., 1992), which requires only evaluations of the log-profile likelihood itself to maximize this function. In our experience, Powell's method usually takes only three or four iterations to converge adequately in a model with five or six parameters. The log-profile likelihood of our model is not guaranteed to be convex. Thus, Powell's method might obtain only a local maximum. We used different starting values to examine whether the algorithm converges to the same value, and stability was observed in the simulations.

In obtaining the PR estimate  $f_{\theta'}^N(\alpha)$  at a given  $\theta'$  in equation (7), the sequence of normalizing constants  $c_i(\theta')$ ,  $i = 1, 2, \dots, N$ , must be calculated. Gauss–Legendre quadrature summation is used for the numerical integration in computing these normalizing constants, i.e.,

$$c_i(\theta') = \int_a^b f_{\theta'}^{i-1}(\alpha) \cdot l_i(\theta | \alpha, \mathbf{y}_i) d\alpha$$

$$\approx \sum_{j=1}^K v_j f_{\theta'}^{i-1}(a_j) l_i(\theta' | a_j, \mathbf{y}_i), \quad (10)$$

where the  $a_j$  and  $v_j$ , for  $j = 1, 2, \dots, K$ , are the abscissas and weights given by Gauss–Legendre quadrature (Press et al., 1992). In our simulation studies and data analysis, usually 20 abscissas ( $K = 20$ ) were adequate to get quite an accurate numerical integration (The accuracy of the numerical integration can be assessed by checking whether the area under the final estimated density curve is approximately equal to 1). In fact, Gauss–Legendre quadrature approximation to an integration is exact if the integrand is a polynomial of order up to  $2K - 1$ .

Another computational problem in calculating equation (10) is the evaluation of the recursive function  $f_{\theta'}^{i-1}(\alpha)$  when  $i$  is large. We memorized the last-step density estimate  $f_{\theta'}^{i-1}(\alpha)$  at the Gauss–Legendre abscissas  $a_j$ , for  $j = 1, \dots, K$ . We used these for the next PR step to calculate the normalizing constant  $c_i(\theta')$  by equation (10) and the density estimate  $f_{\theta'}^i(\alpha)$  at  $\alpha = \{a_1, \dots, a_K\}$  by equation (7). At the end of PR recursion, the memorized values  $f_{\theta'}^N(a_j)$ ,  $j = 1, \dots, N$ , are sufficient for the numerical integration in the log-profile likelihood function, i.e.,

$$\begin{aligned} M(\theta' | \mathbf{y}_i, i = 1, \dots, N) &= \sum_{i=1}^N \log \int_a^b l_i(\theta | \alpha, \mathbf{y}_i) f_{\theta'}^N(\alpha) d\alpha \\ &\approx \sum_{i=1}^N \log \sum_{j=1}^K v_j l_i(\theta' | a_j, \mathbf{y}_i) f_{\theta'}^N(a_j) \end{aligned} \quad (11)$$

After the algorithm converges to  $\hat{\theta}'$ , we run PR once more to record the normalizing constants  $c_i(\hat{\theta}')$  so that the density estimate of the random effect distribution can be evaluated by equation (7).

In this paper, we concentrated on the semiparametric mixed model with only one random individual effect. The method can be extended to multiple random effects by replacing  $\alpha_i$  in model (1) with  $\mathbf{z}'_{ik} \alpha_i$ , where  $\alpha_i$  is a vector of random effects and  $\mathbf{z}_{ik}$  is the observed covariate vector at the  $k$ th measurement for the  $i$ th cluster. For two random effects, the Gauss–Legendre summation for integration involves a  $20 \times 20$  grid of points during iterations, as compared to only 20 points for one random effect. The resulting decrease in speed of the algorithm occurs for other methods as well.