Statistica Sinica **19** (2009), 997-1011

ASYMPTOTIC PROPERTIES AND EMPIRICAL EVALUATION OF THE NPMLE IN THE PROPORTIONAL HAZARDS MIXED-EFFECTS MODEL

Anthony Gamst, Michael Donohue and Ronghui Xu

University of California, San Diego

Abstract: A Proportional Hazards Mixed-effects Model (PHMM) was recently proposed, which associates general random effects with arbitrary covariates and includes the univariate frailty model as a special case. In this paper, we establish the asymptotic properties of the nonparametric maximum likelihood estimator (*NPMLE*) of the parameters of the PHMM. This estimator is computed using a Monte Carlo Expectation-Maximization algorithm. The finite sample performance of the *NPMLE* is examined in a series of simulations and compared with the performance of a penalized partial likelihood estimator and an approximate Laplace *EM* estimator. The model and *NPMLE* are applied to an analysis of twin data.

Key words and phrases: Asymptotic efficiency, clustered survival data, consistency, identifiability.

1. Introduction

Clustered survival data arise from various areas of application: genetic and familial studies, multi-center clinical trials, group-randomized trials, studies of unemployment duration, etc. To analyze this type of data, frailty models were introduced to model the correlation among the survival times from the same clusters (cf. Oakes (1992) and Hougaard (2000)). More recently, Ripatti and Palmgren (2000), Vaida and Xu (2000), and Ripatti, Larsen and Palmgren (2002) proposed the proportional hazards model with mixed effects. It includes the frailty model as a special case, but is more general in that it allows random effects on arbitrary covariates. It is therefore able to model covariate by cluster interactions in a way similar to the linear, generalized linear, and non-linear mixed-effects models. For example, in a multi-center clinical trial, a treatment effect that varies from center to center can be modelled as a random treatment effect, where center here serves as the cluster for patients.

Assume that the data consist of possibly right-censored event time observations from n clusters, with n_i observations in the ith cluster, $i = 1, \ldots, n$. Within a cluster the observations are dependent, but conditional on the cluster-specific

$d \times 1$ vector of random effects $\mathbf{b}_i$, the survival times T_{ij} are independent and their hazard functions follow the proportional hazards mixed-effects model (PHMM):

$$\lambda_{ij}(t) = \lambda_0(t) \exp(\boldsymbol{\beta}'\mathbf{Z}_{ij} + \mathbf{b}_i'\mathbf{W}_{ij}), \tag{1.1}$$

where $\lambda_{ij}(t)$ is the hazard function of the jth observation from the ith cluster, $\mathbf{b}_i$ is a vector of random effects for the ith cluster, and $\mathbf{Z}_{ij}$, $\mathbf{W}_{ij}$ are the covariate vectors for the fixed and random effects. In (1.1) $\mathbf{W}_{ij}$ is usually a subset of $\mathbf{Z}_{ij}$, apart from possibly a '1' which represents the cluster effect on the baseline hazard. To insure identifiability, we assume that $E(\mathbf{b}_i) = \mathbf{0}$. For distribution of the random effects we also assume that

$$\mathbf{b}_i \overset{\text{i.i.d.}}{\sim} N(\mathbf{0}, \boldsymbol{\Sigma}). \tag{1.2}$$

The immediate interpretation of the random effects or the variance components is on the log hazards. In some practical applications, it is also helpful to consider the mathematically equivalent linear transformation model formulation of PHMM:

$$g(T_{ij}) = -\boldsymbol{\beta}'\mathbf{Z}_{ij} - \mathbf{b}_i'\mathbf{W}_{ij} + e_{ij}, \tag{1.3}$$

where T is the time to event of interest, $g(\cdot)$ is a monotone transformation corresponding to the cumulative baseline hazard function under (1.1), and the error e has a fixed extreme value distribution with $\text{Var}\,(e) = 1.645$. Under (1.3) the total variance of the transformed survival time is decomposed into different attributes, just as it is under the linear mixed models.

Under PHMM (1.1), Ripatti and Palmgren (2000) developed a penalized partial likelihood estimate of the parameters, and Vaida and Xu (2000) developed the nonparametric maximum likelihood estimator (*NPMLE*), computed using an *EM* algorithm and Markov Chain Monte Carlo (*MCMC*) methods. Ripatti, Larsen and Palmgren (2002) further improved the *MCEM* algorithm with an automated stopping rule. Cortiñas-Abrahantes and Burzykowski (2005) compared the penalized partial likelihood estimate with an approximate Laplace *EM* estimate through simulations. Earlier analytic results include Murphy (1994, 1995) and Parner (1998) on the asymptotics for gamma frailty models, and Kosorok, Lee, and Fine (2001) on the identifiability of frailty models for independent identically distributed (i.i.d.) data. Kosorok, Lee and Fine (2004) studied robust inference under the univariate frailty models. The asymptotic properties of the *NPMLE*, however, remain unproven under the PHMM. The finite sample performance of the *NPMLE* also has not been studied in the literature.

In the next section we show the consistency and asymptotic normality of the *NPMLE* under the PHMM. In Section 3 we study the finite sample property of the estimator using simulation, and compare it to the penalized partial likelihood estimator and the approximate Laplace *EM* estimator mentioned above. Section 4 provides an example of the application of PHMM to the analysis of twin data. The last section contains discussion.

2. Asymptotic Theory under PHMM

In this section we state and sketch the proof of each of our main results. Detailed proofs are contained in an appendix available through `http://www.stat.sinica.edu.tw/statistica`. The main idea follows Murphy (1994, 1995), and we make use of the identifiability argument as in Zeng, Lin and Yin (2005).

The data from subject j in cluster i can be written $\mathbf{y}_{ij} = (X_{ij}, \delta_{ij}, \mathbf{Z}_{ij}, \mathbf{W}_{ij})$, where X_{ij} is the possibly right-censored failure time and δ_{ij} is the failure-event indicator. Let $Y_{ij}(t) = I\{X_{ij} \geq t\}$ and $\mathbf{y}_i = (\mathbf{y}_{i1}, \ldots, \mathbf{y}_{in_i})$ be the data for cluster i. Conditional on the random effect $\mathbf{b}_i$, the log-likelihood for the i-th cluster is

$$l_i(\beta, \lambda; \mathbf{y}_i|\mathbf{b}_i) = \sum_{j=1}^{n_i} \left\{ \delta_{ij} \Big[\log \lambda(X_{ij}) + \beta' \mathbf{Z}_{ij} + \mathbf{b}_i' \mathbf{W}_{ij} \Big] - \Lambda(X_{ij}) e^{\beta' \mathbf{Z}_{ij} + \mathbf{b}_i' \mathbf{W}_{ij}} \right\}, \tag{2.1}$$

where $\Lambda(t) = \int_0^t \lambda(s) ds$. The log-likelihood of the observed data is then

$$L_n(\theta) = \sum_{i=1}^{n} \log \left\{ \int \exp[l_i(\beta, \lambda; \mathbf{y}_i|\mathbf{b}_i)] \phi(\mathbf{b}_i; \mathbf{\Sigma}) d\mathbf{b}_i \right\}, \tag{2.2}$$

where $\theta = (\beta, \mathbf{\Sigma}, \lambda)$ and ϕ is a multivariate normal density.

Note that the log-likelihood (2.2) has no maximum over the space of absolutely continuous Λ when the sample size is finite. To define the nonparametric maximum likelihood estimator, we extend the parameter space to include all Λ on $[0, \tau]$ that are continuous on the right with left-hand limits, and modify the log-likelihood so that

$$l_i(\beta, \Lambda; \mathbf{y}_i|\mathbf{b}_i) = \sum_{j=1}^{n_i} \left\{ \delta_{ij} \Big[\log \Lambda\{X_{ij}\} + \beta' \mathbf{Z}_{ij} + \mathbf{b}_i' \mathbf{W}_{ij} \Big] - \Lambda(X_{ij}) e^{\beta' \mathbf{Z}_{ij} + \mathbf{b}_i' \mathbf{W}_{ij}} \right\}, \tag{2.3}$$

where $\Lambda\{t\}$ is the size of the jump in Λ at time t.

Let $\hat{\theta}_n = (\hat{\beta}_n, \hat{\mathbf{\Sigma}}_n, \hat{\Lambda}_n)$ be the NPMLE of θ. Taking derivatives with respect to β, $\mathbf{\Sigma}$, and the jumps in Λ gives the score equations

$$0 = \sum_{ij} \Big(\delta_{ij} - \hat{\Lambda}_n(X_{ij}) e^{\hat{\beta}_n' \mathbf{Z}_{ij}} E_{\hat{\theta}_n}(e^{\mathbf{b}_i' \mathbf{W}_{ij}} | \mathbf{y}_i) \Big) \mathbf{Z}_{ij} \tag{2.4}$$

$$n \hat{\mathbf{\Sigma}}_n = \sum_i E_{\hat{\theta}_n}(\mathbf{b}_i \mathbf{b}_i' | \mathbf{y}_i) \tag{2.5}$$

$$\hat{\Lambda}_n(t) = \sum_{ij} \frac{\delta_{ij}(1 - Y_{ij}(t))}{\sum_{kl} Y_{kl}(X_{ij}) e^{\hat{\beta}_n' \mathbf{Z}_{kl}} E_{\hat{\theta}_n}(e^{\mathbf{b}_k' \mathbf{W}_{kl}} | \mathbf{y}_k)}. \tag{2.6}$$

We assume the following fairly standard conditions on the model.

C1. Conditional on the covariates, $\mathbf{Z}_{ij}$ and $\mathbf{W}_{ij}$, the latent censoring times C_{ij} are independent of the event times T_{ij} and the random effects $\mathbf{b}_i$.

C2. There is some $\epsilon > 0$ such that $P(C_{ij} \geq \tau | \mathbf{Z}_{ij}, \mathbf{W}_{ij}) \geq \epsilon$, almost surely.

C3. The baseline hazard function $\lambda_0(t) > 0$ and is continuous on the finite time interval $[0, \tau]$.

C4. The covariates, $\mathbf{Z}_{ij}$ and $\mathbf{W}_{ij}$, are bounded.

C5. The true parameters β and $\mathbf{\Sigma}$ are elements of the interior of a known compact set $\mathcal{K} = \{(\beta, \mathbf{\Sigma}) : |\beta| \leq B$, for some constant B, and $\mathbf{\Sigma}$ is symmetric and positive definite, with eigenvalues bounded away from 0 and ∞.$\}$

C6. The clusters $\{(\mathbf{y}_i, n_i)\}_{i=1}^n$ are i.i.d., and $P(n_i \geq 2) > 0$.

C7. If there is a vector $\mathbf{c}$ and a symmetric matrix $\mathbf{S}$, such that, for $k \neq j = 1, \ldots, n_i$, $\mathbf{c}'[1, \mathbf{Z}_{ij}']' + \mathbf{W}_{ij}' \mathbf{S} \mathbf{W}_{ij} = 0$ and $\mathbf{W}_{ij}' \mathbf{S} \mathbf{W}_{ik} = 0$ almost surely, then $\mathbf{c} = \mathbf{0}$ and $\mathbf{S} = \mathbf{0}$.

Condition C1 is standard and required for identifiability in the presence of censoring. Condition C2 ensures that we observe failures on the interval $[0, \tau]$, so that we can estimate Λ_0 on that interval. Conditions C2 and C3 imply that $\inf_{u \in [0,\tau]} EY_{ij}(u) > 0$, and conditions C4 and C5 imply that $e^{\beta' \mathbf{Z}_{kl}} E_\theta(e^{\mathbf{b}_k' \mathbf{W}_{kl}} | \mathbf{y}_k)$ > 0, so the *NPMLE* $\hat{\Lambda}_n$ is almost surely, eventually, finite on $[0, \tau]$. The addition of condition C6, implies that the score equations can be solved, almost surely, eventually, for $(\hat{\beta}_n, \hat{\mathbf{\Sigma}}_n)$ in $\mathcal{K}$. Intuitively, this is because once enough of the n_i's are ≥ 2, we can tell the difference between $\mathbf{b}_i$ and β, which allows us to construct estimates of Σ. These then imply that $\hat{\theta}_n$ is almost surely, eventually, in a compact set Θ and we can apply Helly's Selection Theorem to conclude that some subsequence of $\hat{\theta}_n$ converges to some θ^* in Θ, as $n \to \infty$. All we have to do now is to show that, for every such convergent subsequence, we must have $\theta^* = \theta_0$.

We would like to compare $\hat{\theta}_n$ to θ_0 directly, but $L_n(\hat{\theta}_n) - L_n(\theta_0)$ diverges as $n \to \infty$, due to the absolute continuity of Λ_0. Therefore we have to replace the absolutely continuous Λ_0 with the function

$$\bar{\Lambda}_n(t) = \sum_{ij} \frac{\delta_{ij}(1 - Y_{ij}(t))}{\sum_{kl} Y_{kl}(X_{ij}) e^{\beta_0' \mathbf{Z}_{kl}} E_{\hat{\theta}_0}(e^{\mathbf{b}_k' \mathbf{W}_{kl}} | \mathbf{y}_k)},$$

which is close to Λ_0 but of the same form as $\hat{\Lambda}_n$. Indeed, it is easy to see that $\sup_{u \in [0,\tau]} |\bar{\Lambda}_n(u) - \Lambda(u)| \to 0$ and $\bar{\theta}_n = (\beta_0, \mathbf{\Sigma}_0, \bar{\Lambda}_n) \to \theta_0$, as $n \to \infty$. Now, a version of the Glivenko-Cantelli Theorem implies that $\sup_{\theta\in\Theta}|L_n(\theta)-L(\theta)|\to 0$ as $n\to\infty$, where $L(\theta)=E_{\theta_0}L_n(\theta)$, and therefore that $L_n(\hat{\theta}_n)\to L(\theta^*)$ and $L_n(\bar{\theta}_n)\to L(\theta_0)$ as $n\to\infty$. But we know that $L_n(\hat{\theta}_n)\geq L_n(\bar{\theta}_n)$, by definition, and $L(\theta_0)\geq L(\theta^*)$, by the properties of Kullback-Leibler Information. This means that we must have $L(\theta^*)=L(\theta_0)$.

Finally, Condition C7 is an extension of Condition 2(g) in Parner (1998), which is to avoid collinearity among the covariates; see Parner (1998) for more discussion. This ensures that the semiparametric Fisher information operator is one-to-one, which implies that L has a unique maximum and therefore that $\theta^*=\theta_0$. The above argument proves the following.

Theorem 1. *Under conditions* C1–7, $\|\hat{\beta}_n-\beta_n\|\to 0$, $\|\hat{\mathbf{\Sigma}}_n-\mathbf{\Sigma}_n\|\to 0$, *and* $\sup_{t\in[0,\tau]}|\hat{\Lambda}_n(t)-\Lambda_0(t)|\to 0$, *almost surely, as* $n\to\infty$, *where* $\|\cdot\|$ *is the Euclidean norm.*

We could argue, as above, taking derivatives at the jumps in Λ and working with the finite sample scores in studying the asymptotic distribution of the *NPMLE*, but it is equivalent and more convenient to work with one-dimensional sub-models through the distribution identified by θ. To that end let $h=(h_1,h_2,h_3)$, where h_1 is a d_1-vector, h_2 is a $d_2(d_2-1)$-vector corresponding to the upper-triangle of the symmetric matrix H_2, and h_3 is a function of bounded variation on the interval $[0,\tau]$. Set $\beta_s=\beta+sh_1$, $\Sigma_s=\Sigma+sH_2$, and $\Lambda_s(t)=\int_0^t(1+sh_3(u))d\Lambda(u)$. Take the derivative of $L_n(\theta_s)$ in s at 0 to compute the score operator

$$
\begin{aligned}
S_n(\theta)[h]=&\frac{1}{2}\mathrm{tr}\left(\left\{\mathbf{\Sigma}^{-1}\frac{1}{n}\sum_i E_\theta(\mathbf{b}_i\mathbf{b}_i'|\mathbf{y}_i)-I\right\}\mathbf{\Sigma}^{-1}H_2\right)\\
&+\frac{1}{n}\sum_{ij}\left(\int_0^\tau(h_1'\mathbf{Z}_{ij}+h_3(u))\left\{dN_{ij}(u)-Y_{ij}(u)e^{\beta'\mathbf{Z}_{ij}}E_\theta(e^{\mathbf{b}_i'\mathbf{W}_{ij}}|\mathbf{y}_i)d\Lambda(u)\right\}\right).
\end{aligned}
$$

Note that the random measure in the second term of the equation above is a martingale and that $S_n(\hat{\theta}_n)[h]=0$ for all $h\in\mathcal{H}$. The Martingale Central Limit Theorem from Pollard (1984) implies that $\sqrt{n}S_n(\theta_0)$ converges in distribution to a mean-zero Gaussian process $\mathcal{G}$ on $\ell_\infty(\mathcal{H})$, the set of bounded real-valued functions on $\mathcal{H}$.

Now, let $S(\theta)=E_{\theta_0}S_n(\theta)$. Note that $S(\theta_0)[h]=0$ for all $h\in\mathcal{H}$. It can be seen that $S(\theta)$ is Fréchet differentiable on $\ell_\infty(\mathcal{H})$ at θ_0; that is, there is a bilinear operator $\dot{S}(\theta_0)$ such that $S(\theta)-S(\theta_0)=\dot{S}(\theta_0)[\theta-\theta_0,\cdot]+o_P(n^{-1/2}\vee\|\theta-\theta_0\|)$. Furthermore, under condition C7, $\dot{S}(\theta_0)$ is continuously invertible on its range. The approximation condition $\|(S_n-S)(\theta)-(S_n-S)(\theta_0)\|=o_P(n^{-1/2}\vee\|\theta-\theta_0\|)$ can be shown directly and we have

$$
0=S_n(\hat{\theta}_n)=S_n(\theta_0)+\dot{S}(\theta_0)[\hat{\theta}_n-\theta_0]+o_P\left(n^{-1/2}\right). \tag{2.7}
$$

Note the similarity between this expression and the usual Taylor expansion of the score with finite dimensional parameters. As usual, the trick is making sure that the remainder term is $o_P(n^{-1/2})$. From (2.7) and the fact that $\dot{S}(\theta_0)$ is continuously invertible on its range, we have

$$\sqrt{n}(\hat{\theta}_n - \theta_0) = -\sqrt{n}\dot{S}(\theta_0)^{-1}S_n(\theta_0) + o_P(1). \tag{2.8}$$

Finally, the Continuous Mapping Theorem implies $\sqrt{n}(\hat{\theta}_n - \theta_0)$ tends in distribution to a mean-zero Gaussian process $-\dot{S}(\theta_0)\mathcal{G}$. Now $\hat{\theta}_n$ is efficient by Theorem 5.1 of Bickel, Klaasen, Ritov and Wellner (1993) and the fact that $-\sqrt{n}\dot{S}(\theta_0)^{-1}S_n(\theta_0)$ is the efficient influence function. This gives us the following.

Theorem 2. *Under conditions* C1–7, $\sqrt{n}(\hat{\theta}_n - \theta_0)$ *converges in distribution to a mean-zero Gaussian process on* $\mathcal{H}$*. Furthermore,* $\hat{\theta}_n$ *is asymptotically efficient.*

The proof of Theorem 2 also demonstrates Theorem 3.

Theorem 3. *Let* $\mathbf{s}$ *be the* $d_2(d_2-1)/2$*-vector corresponding to the upper-triangle of the symmetric matrix* $\mathbf{\Sigma}$, $V(h)$ *be the asymptotic variance of*

$$\sqrt{n}\left[h_1'(\hat{\beta}_n - \beta_0) + h_2'(\hat{\mathbf{s}}_n - \mathbf{s}_0) + \int_0^\tau h_3(u)d\{\hat{\Lambda}_n - \Lambda_0\}(u)\right],$$

and $\mathbf{J}_n$ *be the negative Hessian matrix of the log-likelihood* $L_n(\hat{\theta}_n)$ *with respect to* β, $\mathbf{\Sigma}$, *and the jumps in* Λ *at those* X_{ij} *with* $\delta_{ij} = 1$*. Then, under* C1–7, *the variance estimator*

$$n\mathbf{h}_n'\mathbf{J}_n^{-1}\mathbf{h}_n \to V(h)$$

uniformly in probability, where $\mathbf{h}_n$ *is the vector with elements* h_1, h_2, *and* $h_3(X_{ij})$ *at those* X_{ij} *for which* $\delta_{ij} = 1$.

Remark. At the first submission of this manuscript, an Associate Editor pointed out the forthcoming publication of Zeng and Lin (2007), that contains PHMM as a special case. Our work has been done independently, with proof of the asymptotics posted at `http://www.bepress.com/harvardbiostat/paper43`, under Harvard University Biostatistics Working Paper Series dated May, 2006.

3. Simulations

In this section we carry out simulations to compare the finite sample properties of the *NPMLE* obtained using the *MCEM* algorithm (Vaida and Xu (2000)) to the penalized partial likelihood (*PPL*) estimator (Ripatti and Palmgren (2000)). In addition, as mentioned earlier, Cortiñas-Abrahantes and Burzykowski (2005) considered an approximate Laplace *EM* estimator. The Laplace estimator requires large cluster sizes, and we also compare the *MCEM* estimator with the two other estimators in this setting.

Table 1. Parameter estimates and standard errors for *NPMLE* (first row for each parameter value) and *PPL* estimate (second row) ($n = 100$, $n_i = 2$, no covariates)

Parameter	True Value	Mean	Estimated SE	Empirical SE
$\hat{\sigma}^2$	0.2	0.181	0.192	0.148
		0.223	0.091	0.185
$\hat{\sigma}^2$	0.5	0.449	0.252	0.232
		0.532	0.151	0.275
$\hat{\sigma}^2$	1	0.914	0.359	0.363
		1.073	0.232	0.408

We first consider the simulation settings of Ripatti and Palmgren (2000) in Tables 1−4. We carried out 500 simulations for each of the tables with $\lambda_0(t) = 0.1$. In Tables 1 and 2 there are $n = 100$ clusters, and each cluster has $n_i = 2$ observations. The data in Table 1 were generated with no covariates, $\lambda_{ij}(t) = \lambda_0(t)\exp(b_i)$, and with three values of $\sigma^2 = \text{Var}\,(b_i) : 0.2, 0.5$, and 1. It can be seen from the table that the *NPMLE* and the *PPL* estimator are comparable in terms of bias. Although their biases appear to be in different directions, that is not always the case as will be seen in the other tables. The main problem with the *PPL* estimator, as pointed out in Ripatti and Palmgren (2000), is the underestimation of the standard errors. We showed in Theorem 3 that the variance of the *NPMLE* is consistently estimated using the observed information matrix. We see from simulations that, in finite samples, the estimate of standard error associated with the *NPMLE* is closer to the observed standard error than the estimator associated with the *PPL* estimator.

The data of Table 2 were generated with three covariates $(Z_{ij1}, Z_{i2}, Z_{i3})'$, one of which is on the individual observation level, and the other two on the cluster level. We took $Z_{ij1}, Z_{i2} \sim N(0, 1)$, and Z_{i3} binary. Censoring was at 20%. Both the *NPMLE* and the *PPL* estimate of the β's had a slight bias towards zero, but in comparison the *NPMLE* had less bias in almost all cases (except for β_1 when $\sigma^2 = 0.5$). Ripatti and Palmgren (2000) also noted that the bias increases with the variance of the random effects for the *PPL* estimator, but this does not seem to be the case for the *NPMLE*. The *NPMLE* however, consistently had larger variances for $\hat{\beta}_3$ than the *PPL* estimator. It is once again noted that the standard errors of the *PPL* estimator were underestimated, while the standard errors of the *NPMLE* appear to be estimated with reasonable accuracy.

The data of Table 3 were generated similarly to those of Table 2, but with 50 clusters of 4 observations each. Although the $\hat{\beta}$'s of both methods become generally less biased with increased cluster sizes (despite fewer clusters), $\hat{\beta}_2$ and $\hat{\beta}_3$, which correspond to the cluster level covariates, had larger standard errors for both methods. The estimate of the variance component, as compared to Table

Table 2. Parameter estimates and standard errors for *NPMLE* (first row for each parameter value) and *PPL* estimate (second row) ($n = 100$, $n_i = 2$)

Parameter	True Value	Mean	Estimated SE	Empirical SE
$\sigma^2 = 0.5$				
$\hat{\beta}_1$	1	0.984	0.132	0.131
		0.987	0.118	0.144
$\hat{\beta}_2$	-0.7	-0.691	0.126	0.125
		-0.689	0.123	0.133
$\hat{\beta}_3$	0.5	0.484	0.223	0.220
		0.481	0.178	0.176
$\hat{\sigma}^2$	0.5	0.448	0.255	0.250
		0.538	0.176	0.267
$\sigma^2 = 1$				
$\hat{\beta}_1$	1	0.971	0.139	0.140
		0.965	0.122	0.139
$\hat{\beta}_2$	-0.7	-0.689	0.146	0.145
		-0.668	0.140	0.153
$\hat{\beta}_3$	0.5	0.475	0.264	0.272
		0.471	0.181	0.186
$\hat{\sigma}^2$	1	0.896	0.364	0.359
		0.962	0.243	0.371
$\sigma^2 = 2$				
$\hat{\beta}_1$	1	0.980	0.146	0.138
		0.921	0.126	0.149
$\hat{\beta}_2$	-0.7	-0.688	0.179	0.188
		-0.643	0.167	0.182
$\hat{\beta}_3$	0.5	0.485	0.332	0.328
		0.469	0.185	0.188
$\hat{\sigma}^2$	2	1.822	0.593	0.599
		1.766	0.355	0.598

2, was improved under the *PPL*, but became less accurate for the *NPMLE*. It is understood that for the *NPMLE* the number of clusters is the effective 'sample size' for estimating $\mathrm{Var}\,(b_i)$, while the improved performance of the *PPL* with the cluster sizes might be explained by the Laplace approximation that was used in the *PPL*.

Finally, for Table 4 we had nested random effects. There were 50 clusters, each cluster had a left and a right sub-cluster, and each sub-cluster had two observations. There were three independent random effects, one at the cluster level, and one each for the two sub-clusters. Their variances were σ_1^2, σ_2^2 and σ_3^2. The covariates were the same as in Tables 2 and 3. The estimates of the β's by the two methods were comparable, although again, the standard errors

Table 3. Parameter estimates and standard errors for *NPMLE* (first row for each parameter value) and *PPL* estimate (second row) ($n = 50$, $n_i = 4$)

Parameter	True Value	Mean	Estimated SE	Empirical SE
$\sigma^2 = 0.5$				
$\hat{\beta}_1$	1	0.993	0.118	0.117
		0.976	0.113	0.127
$\hat{\beta}_2$	-0.7	-0.703	0.140	0.140
		-0.673	0.139	0.145
$\hat{\beta}_3$	0.5	0.504	0.258	0.260
		0.476	0.174	0.171
$\hat{\sigma}^2$	0.5	0.433	0.190	0.195
		0.493	0.168	0.237
$\sigma^2 = 1$				
$\hat{\beta}_1$	1	0.977	0.120	0.120
		0.978	0.116	0.118
$\hat{\beta}_2$	-0.7	-0.688	0.172	0.181
		-0.690	0.173	0.179
$\hat{\beta}_3$	0.5	0.503	0.325	0.333
		0.491	0.177	0.188
$\hat{\sigma}^2$	1	0.893	0.308	0.319
		0.978	0.268	0.346
$\sigma^2 = 2$				
$\hat{\beta}_1$	1	0.980	0.123	0.124
		0.964	0.117	0.128
$\hat{\beta}_2$	-0.7	-0.695	0.225	0.241
		-0.666	0.224	0.232
$\hat{\beta}_3$	0.5	0.501	0.428	0.409
		0.483	0.178	0.186
$\hat{\sigma}^2$	2	1.778	0.533	0.548
		1.928	0.462	0.600

were underestimated for the *PPL*. The *NPMLE* gave much better estimates of the variance components, especially for σ_1^2.

Note that the approximate Laplace estimator considered in Cortiñas-Abrahantes and Burzykowski (2005) is not suitable for small cluster sizes as in the above settings. In the simulations of Cortiñas-Abrahantes and Burzykowski (2005), the cluster sizes were $n_i = 20$ or 100. In the following we take their settings and compare the *MCEM* to the *PPL* and the Laplace estimators in Tables 5 and 6. The results for the latter two estimators are copied from Cortiñas-Abrahantes and Burzykowski (2005).

The data were generated for a pair of bivariate failure times for each subject

Table 4. Parameter estimates and standard errors for *NPMLE* (first row for each parameter value) and *PPL* estimate (second row) (nested random effects, $n = 50$, $n_{i1} = n_{i2} = 2$)

Parameter	True Value	Mean	Estimated SE	Empirical SE
$\hat{\beta}_1$	1	0.978	0.136	0.136
		0.921	0.116	0.139
$\hat{\beta}_2$	-0.7	-0.692	0.150	0.154
		-0.691	0.131	0.157
$\hat{\beta}_3$	0.5	0.507	0.272	0.281
		0.511	0.235	0.249
$\hat{\sigma}_1$	0.2	0.173	0.198	0.181
		0.093	0.082	0.156
$\hat{\sigma}_2$	0.5	0.428	0.378	0.361
		0.401	0.209	0.352
$\hat{\sigma}_3$	1	0.932	0.506	0.490
		0.865	0.312	0.457

j in cluster i, $i = 1, \ldots, n$, $j = 1, \ldots, n_i'$ ($n_i = 2n_i'$ in our notation), according to

$$\lambda_{ij1}(t) = \lambda_{01}(t) \exp(\beta Z_{ij} + b_{i1}),$$
$$\lambda_{ij2}(t) = \lambda_{02}(t) \exp(\beta Z_{ij} + b_{i2}).$$

$\lambda_{01}(t) = 0.5$, $\lambda_{01}(t) = 1$, and

$$\begin{pmatrix} b_{i1} \\ b_{i2} \end{pmatrix} \sim N \left\{ \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{pmatrix} \right\}.$$

Note that the above model allows two different baseline hazard functions, and it is straightforward to modify all three methods considered here to fit such a model. The single covariate Z_{ij} was binary, $\beta = 1$, and censoring was again at 20%. We generated 250 datasets for each simulation. In Table 5, $\sigma_1^2 = \sigma_2^2 = 0.2$, and in Table 6, $\sigma_1^2 = \sigma_2^2 = 1$. In both tables the values of the correlation coefficient considered were $\rho = 0.5$ and 0.9.

The results are reported here in terms of bias and mean squared error (MSE). From the tables we see that the *MCEM* estimator had the smallest MSE in the majority of cases, while between the other two methods the Laplace estimator tended to have smaller MSE when $\sigma_1^2 = \sigma_2^2 = 0.2$, and the opposite held when $\sigma_1^2 = \sigma_2^2 = 1$. The main difficulty of the Laplace and the *PPL* estimators lies in the estimation of the variance components (Cortiñas-Abrahantes and Burzykowski (2005)) when the number of clusters is small ($n = 10$) and the variance components are small ($\sigma_1^2 = \sigma_2^2 = 0.2$); here the relative bias can be over 60%. The *MCEM* estimator performed much better in comparison, especially for

Table 5. Bias and mean squared error (in parentheses) for the *NPMLE* (first row for each n), the approximate Laplace estimator (second row), and the *PPL* estimator (third row) ($\sigma_1^2 = \sigma_2^2 = 0.2$)

n	n_i	σ_{12}	$\hat{\beta}$	$\hat{\sigma}_1^2$	$\hat{\sigma}_2^2$	$\hat{\sigma}_{12}$	$\hat{\rho}$
10	20	0.1	-0.015 (0.028)	-0.024 (0.019)	-0.029 (0.022)	-0.011 (0.011)	0.012
			0.069 (0.097)	0.069 (0.020)	0.063 (0.019)	0.022 (0.014)	-0.042
			0.007 (0.077)	0.058 (0.030)	0.035 (0.026)	0.034 (0.029)	0.046
50			0.006 (0.005)	-0.007 (0.005)	-0.009 (0.005)	-0.004 (0.003)	-0.001
			0.069 (0.019)	0.010 (0.003)	0.011 (0.002)	0.001 (0.003)	-0.019
			-0.010 (0.011)	0.000 (0.005)	-0.008 (0.004)	-0.003 (0.003)	-0.003
100			0.004 (0.003)	0.003 (0.003)	0.001 (0.002)	-0.001 (0.001)	-0.010
			0.074 (0.013)	0.006 (0.002)	0.008 (0.001)	0.000 (0.002)	-0.014
			-0.008 (0.005)	-0.004 (0.003)	-0.006 (0.002)	-0.003 (0.001)	-0.002
10	100	0.1	-0.007 (0.006)	-0.015 (0.011)	-0.008 (0.010)	-0.011 (0.005)	-0.027
			0.015 (0.012)	-0.020 (0.010)	-0.018 (0.009)	-0.013 (0.006)	-0.018
			0.004 (0.013)	-0.019 (0.011)	-0.020 (0.010)	-0.006 (0.007)	0.018
50			0.000 (0.001)	-0.006 (0.002)	0.000 (0.002)	-0.003 (0.001)	-0.007
			0.010 (0.002)	-0.005 (0.002)	-0.005 (0.002)	-0.003 (0.001)	-0.005
			0.002 (0.002)	-0.003 (0.002)	-0.005 (0.002)	-0.001 (0.001)	0.004
100			0.001 (0.000)	-0.005 (0.001)	-0.002 (0.001)	-0.002 (0.001)	-0.002
			0.012 (0.001)	-0.001 (0.001)	-0.002 (0.001)	-0.001 (0.001)	-0.001
			-0.001 (0.001)	-0.001 (0.001)	0.000 (0.001)	0.000 (0.000)	0.001
10	20	0.18	0.010 (0.035)	0.007 (0.027)	-0.033 (0.015)	-0.013 (0.014)	0.001
			0.080 (0.076)	0.122 (0.044)	0.101 (0.035)	0.083 (0.016)	-0.055
			0.020 (0.073)	0.118 (0.046)	0.090 (0.038)	0.083 (0.019)	-0.033
50			-0.004 (0.006)	-0.004 (0.005)	-0.002 (0.004)	-0.003 (0.003)	-0.002
			0.071 (0.019)	0.062 (0.007)	0.055 (0.006)	0.046 (0.004)	-0.023
			0.000 (0.011)	0.056 (0.016)	0.041 (0.012)	0.041 (0.009)	-0.009
100			0.000 (0.003)	-0.002 (0.002)	-0.003 (0.002)	-0.003 (0.001)	-0.004
			0.076 (0.013)	0.029 (0.002)	0.022 (0.002)	0.019 (0.001)	-0.015
			-0.003 (0.006)	0.024 (0.007)	0.018 (0.006)	0.018 (0.003)	-0.005
10	100	0.18	0.001 (0.008)	-0.015 (0.011)	-0.006 (0.011)	-0.014 (0.009)	-0.024
			0.015 (0.012)	-0.009 (0.011)	-0.009 (0.010)	-0.010 (0.009)	-0.011
			0.007 (0.013)	-0.006 (0.008)	-0.005 (0.007)	-0.004 (0.007)	0.003
50			-0.003 (0.001)	-0.003 (0.002)	-0.005 (0.002)	-0.004 (0.002)	-0.003
			0.010 (0.002)	-0.004 (0.002)	-0.005 (0.002)	-0.005 (0.002)	-0.004
			0.002 (0.002)	0.000 (0.004)	-0.003 (0.003)	-0.001 (0.003)	0.002
100			0.002 (0.001)	-0.003 (0.001)	-0.003 (0.001)	-0.003 (0.001)	0.000
			0.012 (0.001)	0.000 (0.001)	-0.001 (0.001)	-0.001 (0.001)	-0.001
			-0.001 (0.001)	-0.001 (0.001)	-0.001 (0.001)	0.000 (0.001)	0.003

$n_i = 20$ where the biases for the other two methods are severe. The same effects were also seen in estimating the correlation coeffcient ρ, where the *MCEM*

Table 6. Bias and mean squared error (in parentheses) for the *NPMLE* (first row for each n), the approximate Laplace estimator (second row), and the *PPL* estimator (third row) ($\sigma_1^2 = \sigma_2^2 = 1$)

n	n_i	σ_{12}	$\hat{\beta}$	$\hat{\sigma}_1^2$	$\hat{\sigma}_2^2$	$\hat{\sigma}_{12}$	$\hat{\rho}$
10	20	0.5	-0.028 (0.035)	-0.099 (0.276)	-0.104 (0.276)	-0.084 (0.140)	-0.037
			0.078 (0.082)	-0.100 (0.482)	-0.105 (0.392)	-0.086 (0.157)	-0.039
			-0.006 (0.081)	-0.075 (0.252)	-0.114 (0.242)	-0.056 (0.136)	-0.009
50			0.002 (0.007)	-0.032 (0.051)	-0.032 (0.061)	-0.021 (0.029)	-0.005
			0.068 (0.020)	-0.026 (0.190)	-0.026 (0.082)	-0.028 (0.038)	-0.015
			-0.005 (0.013)	-0.018 (0.033)	-0.029 (0.035)	-0.010 (0.013)	0.002
100			0.000 (0.004)	0.011 (0.029)	0.000 (0.035)	0.004 (0.019)	0.001
			0.075 (0.013)	-0.018 (0.080)	-0.016 (0.043)	-0.019 (0.022)	-0.011
			-0.005 (0.006)	-0.017 (0.033)	-0.021 (0.027)	-0.007 (0.018)	0.003
10	100	0.5	0.005 (0.006)	-0.036 (0.210)	-0.076 (0.237)	-0.042 (0.134)	-0.014
			0.011 (0.012)	-0.107 (0.249)	-0.099 (0.216)	-0.061 (0.128)	-0.011
			0.007 (0.013)	-0.101 (0.241)	-0.089 (0.204)	-0.034 (0.119)	-0.004
50			0.000 (0.001)	-0.005 (0.052)	-0.008 (0.044)	-0.007 (0.027)	-0.004
			0.009 (0.002)	-0.024 (0.081)	-0.021 (0.041)	-0.014 (0.029)	-0.002
			0.003 (0.002)	-0.021 (0.051)	-0.024 (0.039)	-0.009 (0.027)	-0.001
100			0.002 (0.000)	0.020 (0.027)	0.001 (0.022)	0.003 (0.013)	-0.002
			0.011 (0.001)	-0.009 (0.026)	0.003 (0.021)	-0.003 (0.015)	-0.002
			-0.001 (0.001)	-0.010 (0.022)	0.004 (0.019)	0.004 (0.014)	0.004
10	20	0.9	0.005 (0.036)	-0.084 (0.343)	-0.072 (0.343)	-0.085 (0.255)	-0.016
			0.081 (0.084)	-0.119 (0.269)	-0.124 (0.264)	-0.120 (0.168)	-0.013
			-0.005 (0.076)	-0.114 (0.182)	-0.140 (0.180)	-0.121 (0.075)	-0.007
50			-0.002 (0.006)	-0.042 (0.052)	-0.036 (0.053)	-0.039 (0.041)	-0.004
			0.069 (0.020)	-0.023 (0.073)	-0.031 (0.060)	-0.031 (0.044)	-0.007
			-0.008 (0.012)	-0.021 (0.051)	-0.029 (0.052)	-0.022 (0.016)	0.000
100			0.000 (0.004)	-0.005 (0.027)	-0.015 (0.033)	-0.012 (0.022)	-0.003
			0.076 (0.013)	-0.020 (0.035)	-0.029 (0.027)	-0.027 (0.021)	-0.005
			-0.007 (0.006)	-0.015 (0.031)	-0.024 (0.029)	-0.016 (0.025)	0.001
10	100	0.9	0.008 (0.007)	-0.097 (0.181)	-0.111 (0.179)	-0.098 (0.157)	-0.005
			0.013 (0.012)	-0.107 (0.266)	-0.098 (0.236)	-0.102 (0.194)	-0.011
			0.005 (0.012)	-0.099 (0.237)	-0.096 (0.210)	-0.085 (0.197)	0.003
50			-0.004 (0.001)	-0.012 (0.042)	-0.009 (0.038)	-0.010 (0.035)	-0.001
			0.010 (0.002)	-0.020 (0.053)	-0.026 (0.047)	-0.024 (0.043)	-0.004
			-0.001 (0.002)	-0.020 (0.054)	-0.024 (0.046)	-0.025 (0.050)	-0.005
100			0.000 (0.001)	0.001 (0.018)	-0.008 (0.018)	-0.003 (0.016)	0.000
			0.012 (0.001)	-0.007 (0.030)	-0.003 (0.023)	-0.006 (0.023)	-0.002
			-0.002 (0.001)	-0.006 (0.025)	-0.001 (0.021)	-0.002 (0.021)	0.001

estimator appeared much more accurate for the small sample sizes. With an increasing number of clusters and an increasing number of subjects per cluster, both the bias and MSE decreased for all three methods.

4. An Example

We consider the application of model (1.1) to twin data. Ripatti, Gatz, Pedersen and Palmgren (2003) applied (1.1) to twin data, where the twins share a common unobserved random effect. In the Vietnam Era Twin (VET) Registry data, zygosity was known for 3,372 complete twin pairs (1,874 monozygotic pairs and 1,498 dizygotic pairs). The monozygotic (*MZ*) twins are often assumed to have the same genes, and the dizygotic (*DZ*) twins share half of their genes. Using the two different types of zygosity we can examine the relative contributions of genetic and environmental factors to age at onset of disease, in this case, age at onset of alcohol dependence (DSM-III-R, Robins, Helzer, Cottler and Goldring (1989)). Prior to the development of survival analysis tools, this type of research in psychiatric epidemiology had been done using linear mixed effects models (when there was no censoring) and structural equation models.

To model the dependence structure of the *MZ* and *DZ* twins using PHMM, for an *MZ* twin pair i, let b_{1i} denote the contribution from the common genetic G and the common environmental C factors, and b_{2i} and b_{3i} denote the unique environmental E factors for twin 1 and twin 2, respectively, so that

$$\lambda_{i1}(t) = \lambda_0(t)\exp(b_{1i} + b_{2i}),$$
$$\lambda_{i2}(t) = \lambda_0(t)\exp(b_{1i} + b_{3i}).$$

We can write $\mathrm{Var}\,(b_{1i}) = \sigma_G^2 + \sigma_C^2$, and $\mathrm{Var}\,(b_{2i}) = \mathrm{Var}\,(b_{3i}) = \sigma_E^2$. For a *DZ* twin pair i, let b_{4i} denote the common genetic (1/2) and the common environmental factors, and b_{5i} and b_{6i} denote the unique genetic (1/2) and the unique environmental factors for twin 1 and twin 2, respectively, so that

$$\lambda_{i1}(t) = \lambda_0(t)\exp(b_{4i} + b_{5i}),$$
$$\lambda_{i2}(t) = \lambda_0(t)\exp(b_{4i} + b_{6i}).$$

We can then write $\mathrm{Var}\,(b_{4i}) = \sigma_G^2/2 + \sigma_C^2$ and $\mathrm{Var}\,(b_{5i}) = \mathrm{Var}\,(b_{6i}) = \sigma_G^2/2 + \sigma_E^2$. Due to the construction of these six random effects, to decompose the total variance it is necessary to assume that they are independent of each other. Therefore the variance matrix for $\mathbf{b}$ is $\mathbf{\Sigma} = \mathrm{diag}(\sigma_G^2 + \sigma_C^2, \sigma_E^2, \sigma_E^2, \sigma_G^2/2 + \sigma_C^2, \sigma_G^2/2 + \sigma_E^2, \sigma_G^2/2 + \sigma_E^2)$.

After fitting (1.1) to the data, the *NPMLE* gives $\hat{\sigma}_G^2 = 1.08(0.15)$, $\hat{\sigma}_C^2 = 0.27(0.05)$, and $\hat{\sigma}_E^2 = 0.02(0.01)$, where the standard errors are given in parentheses. According to the convention in genetic epidemiology, these are to be reported as percentages of the total variation in the age at onset of alcohol dependence. Direct calculation gives 79%, 20% and 1%, respectively, of genetic, common environmental, and unique environmental contributions. This is very different from what has been reported in the past in studies of alcohol dependence; in particular, the genetic contribution is about twice as high as reported in the literature. Note that what we have here is in fact a decomposition of the total variance on the log hazard scale in model (1.1), not on the variable age at onset itself. The equivalent formulation (1.3) gives the decomposition for the transformed age at onset, i.e. with the additional error variance $\text{Var}\,(e) = 1.645$, given the hazard function of an individual. With $\hat{\sigma}_G^2 + \hat{\sigma}_C^2 + \hat{\sigma}_E^2 + 1.645$ as the estimated total variation, the percent contributions are $\hat{G} = 36\%$, $\hat{C} = 9\%$, and $\hat{E} = 55\%$. The corresponding standard errors are 5%, 2%, and 4%, respectively, obtained using the Delta method. The 36% genetic contribution is very comparable to what has been reported in the literature on alcohol dependence, and agrees quite well with the results obtained by our collaborators using structural equation modelling (not shown here) for the same dataset.

5. Discussion

In this paper we studied the asymptotic as well as finite sample properties of the nonparametric maximum likelihood estimator under the proportional hazards mixed-effects model. We have established that the *NPMLE* is consistent, asymptotically Gaussian, and efficient. In contrast the theoretical properties of the *PPL* has not been rigorously studied.

Since the initial proposal of the PHMM in 2000, there has not been widely agreed-upon finite sample implementation of estimation under the model. As having been known in the literature, as well as confirmed in our simulation, inference based on the *PPL* can be problematic at least due to its difficulty in estimating the variance. Bootstrap has been proposed to estimate the variance of the *PPL* estimator; however, it is not clear that the bootstrap will be consistent if the consistency of the *PPL* estimator is not established. For the implementation of the *NPMLE*, currently the only alternative to the *MCEM* algorithm is based on Laplace approximation. The Laplace approximation requires reasonably large cluster sizes, and is not suitable for certain data structures such as twins. We have shown in our simulation that the *MCEM* algorithm is numerically stable and that the inference procedure is accurate.

Nonethless the *MCEM* algorithm is computationally intensive, and its convergence can sometimes be slow, especially when the random components are relatively small. Even under today's rapidly growing computational power, there is incentive to develop more efficient algorithms for practical use of the PHMM. Various faster *EM* algorithms have been proposed; for the linear mixed models, for example, Meng and van Dyk (1997, 1998) proposed to speed up the common *EM* algorithms by the addition of a 'working parameter' which transfers the random effect variance into a regression slope. A separate project is currectly under way to implement such faster algorithms under the PHMM.

References

Bickel, P. J., Klaasen, C. A., Ritov, Y. and Wellner, J. A. (1993). *Efficient and Adaptive Estimation for Semiparametric Models.* Springer, New York.

Cortiñas-Abrahantes, J. and Burzykowski, T. (2005). A version of the EM algorithm for proportional hazard model with random effects. *Biometrical J.* **6**, 847-862.

Hougaard, P. (2000). *Analysis of Multivariate Survival Data.* Springer, New York.

Kosorok, M. R., Lee, B. L. and Fine, J. P. (2001). Semiparametric inference for proportional hazards frailty regresssion models. Tech. rep., Department of Biostatistics, University of Wisconsin.

Kosorok, M. R., Lee, B. L. and Fine, J. P. (2004). Robust inference for univariate proportional hazards frailty regression models. *Ann. Statist.* **32**, 1448-1491.

Meng, X.-L. and van Dyk, D. (1997). The EM algorithm — an old folk-song sung to a fast new tune. *J. Roy. Statist. Soc. Ser. B* **59**, 511-567.

Meng, X.-L. and van Dyk, D. (1998). Fast EM-type implementations for mixed effects models. *J. Roy. Statist. Soc. Ser. B* **60**, 559-578.

Murphy, S. (1994). Consistency in a proportional hazards model incorporating a random effect. *Ann. Statist.* **22**, 712-731.

Murphy, S. (1995). Asymptotic theory for the frailty model. *Ann. Statist.* **23**, 182–198.

Oakes, D. (1992). Frailty models for multiple event times. In *Survival Analysis: State of the Art* (*Edited by J.P. Klein and P.K. Goel.*), 371-379. Kluwer Academic (Dordrecht; Norwell, MA).

Parner, E. (1998). Asymptotic theory for the correlated Gamma-frailty model. *Ann. Statist.* **26**, 183-214.

Ripatti, S., Gatz, M., Pedersen, N. L. and Palmgren, J. (2003). Three-state frailty model for age at onset of dementia and death in swedish twins. *Genetic Epidemiology* **24**, 139-149.

Ripatti, S., Larsen, K., and Palmgren, J. (2002). Maximum likelihood inference for multivariate frailty models using an automated Monte Carlo EM algorithm. *Lifetime Data Anal.* **8**, 349-360.

Ripatti, S. and Palmgren, J. (2000). Estimation of multivariate frailty models using penalized partial likelihood. *Biometrics* **56**, 1016-1022.

Robins, L., Helzer, J., Cottler, L. and Goldring, E. (1989). *Diagnostic Interview Schedule Version III Revised.* National Institute of Mental Health, Rockville, MD.

Vaida, F. and Xu, R. (2000). Proportional hazards model with random effects. *Statist. Medicine* **19**, 3309-3324.

Zeng, D. and Lin, D. Y. (2007). Maximum likelihood estimation in semiparametric regression models with censored data. *J. Roy. Statist. Soc. Ser. B* **69**, 507-564.

Zeng, D., Lin, D. Y. and Yin, G. (2005). Maximum likelihood estimation for proportional odds model with random effects. *J. Amer. Statist. Assoc.* **100**, 470-483.

Division of Biostatistics and Bioinformatics, University of California, San Diego, La Jolla, CA92093 U.S.A.

E-mail: acgamst@ucsd.edu

Division of Biostatistics and Bioinformatics, University of California, San Diego, La Jolla, CA92093 U.S.A.

E-mail: mdonohue@ucsd.edu

Division of Biostatistics and Bioinformatics, University of California, San Diego, La Jolla, CA92093 U.S.A.

Department of Mathematics, University of California, San Diego, U.S.A.

E-mail: rxu@ucsd.edu

(Received April 2007; accepted July 2008)